A Note on the Nonparametric Estimation of the Conditional Mode by Wavelet Methods

Salim Bouzebda and Christophe Chesneau

Received: 12 September 2020; Accepted: 30 October 2020; Published: 31 October 2020

Abstract: The purpose of this note is to introduce and investigate the nonparametric estimation of the conditional mode using wavelet methods. We propose a new linear wavelet estimator for this problem. The estimator is constructed by combining a specific ratio technique and an established wavelet estimation method. We obtain rates of almost sure convergence over compact subsets of $\mathbb{R}^d$.

A general estimator beyond the wavelet methodology is also proposed, discussing adaptivity within this statistical framework.

Keywords: conditional mode estimation; strong consistency; nonparametric estimation; wavelet estimation

MSC: 60F10; 62G07; 62F05; 62H15

1. Introduction

Let us start by describing the mathematical context of the study. Consider a $\mathbb{R}^d \times \mathbb{R}$ valued independent vectors $(X_i, Y_i)_{i \geq 1}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The distribution and probability density functions of $(X_1, Y_1)$ are denoted by $F(\cdot)$ and $f(\cdot)$, respectively. The conditional density function of $Y_1$ given $X_1 = x$ is denoted by $f(\cdot \mid x)$. By assuming that $f(\cdot \mid x)$ is unimodal, the conditional mode of $Y_1$ given $X_1 = x$, denoted by $\Theta(x)$, is defined by the following equation:

$$f(\Theta(x) \mid x) = \max_{y \in \mathbb{R}} f(y \mid x).$$ (1)

A kernel estimator of the conditional mode $\Theta(x)$ is defined as the random variable $\hat{\Theta}(x)$ that maximizes the kernel estimator $\hat{f}(\cdot \mid x)$ of $f(\cdot \mid x)$, which is

$$\hat{f}(\hat{\Theta}(x) \mid x) = \max_{y \in \mathbb{R}} \hat{f}(y \mid x).$$ (2)

Kernel-type estimators have been studied in depth in different dependency contexts. See [1–7], to name a few. [4] motivated the use of conditional mode by pointing out that the prediction of the $Y$-values given $X$-values is done by regression function estimation. However, such an approach is not reasonable in certain circumstances, including the case where the conditional density of $Y$ given $X$ is far from being a symmetric function, or even worse when it is not unimodal. Prediction via conditional mode estimation is in these cases much more appropriate. In this regard, we may refer to [7]. Most kernels used are symmetrical and, when chosen, are fixed. This can be effective for estimating curves with unbounded supports, but not for those with compact support or a subset of the entire real line. The same for those having discontinuities at boundary points.
An interesting feature of wavelet methods in statistics is to provide efficient procedures under mild assumptions on the regularity of the unknown function. In addition, procedures that adapt to the regularity of the object to be estimated can be developed. These procedures are called adaptive. The complete background on wavelets can be found in [8–11]. The use of wavelets in various curve estimation problems is studied in [12]. We can refer also to the articles of [13–19].

In the present work, we propose a wavelet based estimator of the conditional mode constructed from the methodology of [4] and an efficient linear wavelet estimator introduced by [20]. We establish its rates of almost sure convergence over compact subsets of \( \mathbb{R}^d \) under mild assumptions on the involved unknown functions; the Besov spaces are considered. A general method is also discussed, beyond the wavelet estimation setting. To the best of our knowledge, the developed methodology and result respond to a problem that has not been studied systematically up to the present, which is the basic motivation of this work.

The presentation of the note is as follows. In Section 2, we first present the estimator that we consider throughout this note. In Section 3, we declare the main assumptions and give the main theoretical result, with discussions. Section 4 is about a general method allowing more flexibility on the choice of the estimation method. Section 5 concludes the note.

2. Estimation Procedures

The mathematical context of the multiresolution analysis as well as the considered estimator are presented in this section.

2.1. Multiresolution Analysis

We start by giving some notation and definitions that are needed for later use. For an accessible introduction to the theory of wavelets, we refer to [8–11]. According to [8], a multiresolution analysis on the Euclidean space \( \mathbb{R}^d \) can be seen as a decomposition of the space \( L_2(\mathbb{R}^d) \) into an increasing sequence of closed subspaces \( \{ V_j : j \in \mathbb{Z} \} \) in a such a way that

(i) \( V_j \subset V_{j+1}, \ j \in \mathbb{Z}, \)
(ii) \( \cap_j V_j = \{0\}, \ \cup_j V_j = L_2(\mathbb{R}^d), \)
(iii) \( f(x) \in V_j \Leftrightarrow f(2x) \in V_{j-1}, \ f(x) \in V_j \Rightarrow f(x+k) \in V_j, \ k \in \mathbb{Z}^d, \)

where \( V_0 \) is closed subspace under integer translation. Assume that there exists a scale function \( \phi(\cdot) \in L_2(\mathbb{R}^d) \) fulfilling

\[
\int_{\mathbb{R}^d} \phi(x)dx = 1,
\]

such that \( \{ \phi_k(x) = \phi(x-k) : k \in \mathbb{Z}^d \} \) constitute an orthonormal basis for the subspace \( V_0. \) Notice that

\[
\{ \phi_{j,k}(x) = 2^{jd/2}\phi(2^jx-k) : k \in \mathbb{Z}^d \}
\]

is an orthonormal basis for the subspace \( V_j. \) We call the multiresolution analysis \( r \)-regular if \( \phi(\cdot) \in C^r \) and all its partial derivatives up to total order \( r \) are rapidly decreasing, i.e., for every integer \( i > 0, \) there exists a constant \( A_i > 0 \) such that, for all \( |\beta| \leq r, \)

\[
\left| \left( D^\beta \phi \right)(x) \right| \leq \frac{A_i}{(1 + |x|)^r},
\]

(3)

where \( (D^\beta \phi)(x) = \partial^\beta \phi(x)/(\partial^{\beta_1}x_1 \cdots \partial^{\beta_d}x_d), \) \( \beta = (\beta_1, \ldots, \beta_d) \) and \( |\beta| = \sum_{i=1}^{d} \beta_i. \) The multiresolution is assumed to be \( r \)-regular in our framework. Let us denote by \( W_j \) the orthogonal complement of \( V_j \) in \( V_{j+1}, \) in the sense that \( V_j \oplus W_j = V_{j+1}. \) This implies that the space \( L_2(\mathbb{R}^d) \) can be decomposed as
\( L_2(\mathbb{R}^d) = \bigoplus_{j \in \mathbb{Z}} W_j \), or, equivalent way, as \( L_2(\mathbb{R}^d) = V_0 \bigoplus \bigoplus_{j \geq 0} W_j \). Then, we can see that there exist \( N = 2^d - 1 \) wavelet functions \( \{\psi_i(x), i = 1, \ldots, N\} \) satisfying the following:

(W.1) an orthonormal basis for \( W_0 \) is given by \( \{\psi_i(x - k) : k \in \mathbb{Z}^d, i = 1, \ldots, N\} \),

(W.2) an orthonormal basis for \( L_2(\mathbb{R}^d) \) is given by \( \{\psi_{i,j,k}(x) : i = 1, \ldots, N, k \in \mathbb{Z}^d, j \in \mathbb{Z}\} \) where \( \psi_{i,j,k}(x) = 2^{jd/2} \psi_i(2^jx - k) \),

(W.3) \( \psi_i(\cdot) \) has the same regularity as \( \phi(\cdot) \) and both functions are compactly supported on \([-L, L]^d\) for some \( L > 0 \).

For any function \( f(\cdot) \) assumed to be in \( L_2(\mathbb{R}^d) \), for any integer \( m \), we have the orthonormal representation,

\[
f(x) = \sum_{k \in \mathbb{Z}^d} a_{m,k} \phi_{m,k}(x) + \sum_{j \geq m} \sum_{i=1}^N b_{i,j,k} \psi_{i,j,k}(x),
\]

where

\[
a_{m,k} = \int_{\mathbb{R}^d} f(u) \phi_{m,k}(u) du, \quad b_{i,j,k} = \int_{\mathbb{R}^d} f(u) \psi_{i,j,k}(u) du.
\]

One can remark that the orthogonal projection of \( f(\cdot) \) on \( V_\ell \) may be written in two equivalent ways, for any \( m \leq \ell \),

\[
(P_{V_\ell} f)(x) := \sum_{k \in \mathbb{Z}^d} a_{\ell,k} \phi_{\ell,k}(x)
= \sum_{k \in \mathbb{Z}^d} a_{m,k} \phi_{m,k}(x) + \sum_{j=m}^{\ell} \sum_{i=1}^N b_{i,j,k} \psi_{i,j,k}(x).
\]

Recall that a function \( f(\cdot) \in L_2(\mathbb{R}^d) \) pertains to the Besov balls \( B^d_{s,p,q} \) if and only if there exists a constant \( C > 0 \) such that

\[
I_{s,p,q} (g) := \|P_{V_0} g\|_L^p + \left( \sum_{j \geq 0} \|2^j\|P_{W_j} g\|_L^p \right)^{\frac{1}{p}} \leq C,
\]

with the usual modifications for the cases \( p = \infty \) or \( q = \infty \). For a more general definition of Besov spaces, we refer to Appendix A.

2.2. Main Estimator

We consider the estimator \( \hat{\sigma}_{y}(y \mid x) \) defined by the ratio:

\[
\hat{\sigma}_{y}(y \mid x) = \frac{\hat{\sigma}_{y}(y, x)}{f(x)},
\]

where \( \hat{\sigma}_{y}(y, x) \) is an estimator for \( \partial_y f(y, x) \) defined by

\[
\hat{\sigma}_{y}(y, x) = \sum_{k \in \mathbb{Z}^d+1} \hat{a}_{\ell,k} \phi_{\ell,k}(y, x), \quad \hat{a}_{\ell,k} = -\frac{1}{n} \sum_{i=1}^n \partial_y \phi_{\ell,k}(Y_i, X_i),
\]

and \( \hat{f}(x) \) is an estimator for \( f(x) \) defined by

\[
\hat{f}(x) = \sum_{k \in \mathbb{Z}^d} \hat{a}_{m,k} \phi_{m,k}(x), \quad \hat{a}_{m,k} = \frac{1}{n} \sum_{i=1}^n \phi_{m,k}(X_i).
\]
Here, the resolution levels $\ell = \ell(n)$ and $m = m(n) \to \infty$ at a rates specified below. Since we assume that $\phi(\cdot)$ and $\psi(\cdot)$ have a compact support so that the summations above are finite for each fixed $x$ (note that in this case the support of $\phi(\cdot)$ and $\psi(\cdot)$ is a monotonically increasing function of their degree of differentiability [9]). We focus our attention on multivariate linear estimators which will be shown to have uniform almost sure convergence rates over compact sets, in a similar way as in [21,22] in the setting of the strongly mixing processes. The considered wavelet estimators $\hat{\omega}_y f(y, x)$ and $\hat{f}(x)$ are derived to those proposed in [20]. See also [23,24] for the use of this estimator in other statistical contexts.

Now, assume that there is some compact subset $C_x$ such that $f(y \mid x)$ has a unique mode $\Theta(x)$ on $C_x$. Then, our estimator for the mode is defined by assuming that there is some unique $\hat{\Theta}(x)$ in $C_x$ such that

$$\hat{\omega}_y f(\hat{\Theta}(x) \mid x) = 0. \quad (6)$$

The idea of such construction was proposed by [4] for kernel estimators. We extend it to wavelet methodology, which is a nontrivial extension.

3. Assumptions and Main Results

This section is devoted to our main result.

3.1. Assumptions

We formulate the following assumptions. Let $D$ be a compact set in $\mathbb{R}^d$ and $x \in D$.

(A.1) $f(y \mid x)$ is two times continuous differentiable;

(A.2) There is some compact subset $C_x$ such that $f(y \mid x)$ has an unique mode $\Theta(x)$ on $C_x$ and there exists a constant $c > 0$ such that $\inf_{y \in C_x} |\partial_y f(y \mid x)| \geq c > 0$;

(A.3) We have $\lim_{|y| \to \infty} f(y, x) = 0$;

(A.4) $\partial_y f(\cdot, \cdot) \in B^{d+1}_{\infty,p}$ and $f(\cdot) \in B^{d+1}_{p',q'}$;

(A.5) There exists a constant $c' > 0$ such that $f(x) \geq c' > 0$.

These assumptions are only technical and necessary to prove our main theorem.

3.2. Main Result

Theorem 1 below presents uniform rates of almost sure convergence over compact subsets of $\mathbb{R}^d$.

**Theorem 1.** Suppose that (A.1)–(A.5) hold. Let $x \in D$. Consider the estimator $\hat{\Theta}(x)$ characterized by (6) with $\ell$ and $m$ such that

$$2^\ell = O\left(\left(\frac{\log n}{n}\right)^{\frac{1}{3+2(p-1)/p}}\right), \quad 2^m = O\left(\left(\frac{\log n}{n}\right)^{\frac{1}{3+2(p-1)/p'}}\right).$$

Then, we have, as $n \to \infty$,

$$|\hat{\Theta}(x) - \Theta(x)| = O\left(\max\left(\frac{\log n}{n}, \left(\frac{\log n}{n}\right)^{\frac{1}{3+2(p-1)/p}}, \left(\frac{\log n}{n}\right)^{\frac{1}{3+2(p-1)/p'}}\right)\right), \text{ a.s.}$$

**Proof.** We follow the lines of [4], Proof of Theorem 2. By (A.1), the Taylor expansion of the function $\partial_y f(y \mid x)$ yields

$$\partial_y f\left(\Theta(x) \mid x\right) = \partial_y f\left(\Theta(x) \mid x\right) + \left(\Theta(x) - \Theta(x)\right) \partial_y^2 f\left(\Theta^*(x) \mid x\right), \quad (7)$$
with $\Theta^*(x)$ being a point between $\Theta(x)$ and $\hat{\Theta}(x)$. Making use of the condition (A.2) and the definition of $C_{x}$, we obtain

$$\partial_{y}f(\Theta(x) \mid x) = \partial_{y}f(\hat{\Theta}(x) \mid x) = 0.$$ 

Moreover, always by (A.2), we readily infer that

$$|\hat{\Theta}(x) - \Theta(x)| = \left|\frac{\partial_{y}f(\hat{\Theta}(x) \mid x) - \partial_{y}f(\Theta(x) \mid x)}{\partial_{y}f(\Theta^*(x) \mid x)}\right| \leq \frac{1}{c} \left|\partial_{y}f(\hat{\Theta}(x) \mid x) - \partial_{y}f(\Theta(x) \mid x)\right| \leq \frac{1}{c} \sup_{y \in C_{x}} \left|\partial_{y}f(y \mid x) - \partial_{y}f(y \mid x)\right|. \tag{8}$$

An appropriate decomposition and the triangular inequality yield

$$\left|\partial_{y}f(y \mid x) - \partial_{y}f(y \mid x)\right| = \left|\frac{\partial_{y}f(y, x)}{f(y)} - \frac{\partial_{y}f(y, x)}{f(x)}\right| = \left|\frac{f(y)\partial_{y}f(y, x) - f(x)\partial_{y}f(y, x)}{f(x)f(y)}\right| \leq \frac{1}{f(x)} \left(\sup_{y \in C_{x}} \left|\partial_{y}f(y, x) - \partial_{y}f(y, x)\right| + \sup_{y \in C_{x}} \left|\partial_{y}f(y, x) - \partial_{y}f(y, x)\right|\right). \tag{9}$$

By (A.3) and (A.4), it follows from [20], Theorem and Remark (1) and suitable choices of the parameters that

$$\left|\hat{f}(x) - f(x)\right| = O\left(\left(\frac{\log n}{n}\right)^{\frac{d}{d' - d'}}\right), \text{ a.s.}$$

and

$$\sup_{y \in C_{x}} \left|\partial_{y}f(y, x) - \partial_{y}f(y, x)\right| = O\left(\left(\frac{\log n}{n}\right)^{\frac{s -(d+1)p}{2(d + 1 - (d+1)p)}}\right), \text{ a.s.}$$

Combining these inequalities with condition (A.5) and, for large enough $n$, $|\hat{f}(x)| \geq |\hat{f}(x) - f(x)| - f(x)$, we get

$$\left|\hat{\Theta}(x) - \Theta(x)\right| \leq O\left(\max\left(\left(\frac{\log n}{n}\right)^{\frac{s -(d+1)p}{2(d + 1 - (d+1)p)}}, \left(\frac{\log n}{n}\right)^{\frac{1}{2(d - d')}}\right)\right), \text{ a.s.}$$

This completes the proof of Theorem 1. \qed

**Remark 1.** Theorem 1 implies that the strong consistency of $\hat{\Theta}(x)$ under the considered assumptions.

**Remark 2.** From Theorem 1, if $p = p' = +\infty$, we have, as $n \to \infty$,

$$\left|\hat{\Theta}(x) - \Theta(x)\right| = O\left(\max\left(\left(\frac{\log n}{n}\right)^{\frac{1}{2(d + 1 - (d+1)p)}}, \left(\frac{\log n}{n}\right)^{\frac{1}{2d'}}\right)\right), \text{ a.s.}$$

and, if $ds \geq (d + 3)s'$, it is reduced to

$$\left|\hat{\Theta}(x) - \Theta(x)\right| = O\left(\left(\frac{\log n}{n}\right)^{\frac{1}{2d'}}\right), \text{ a.s.}$$
4. On a General Estimation Method: One Step to the Adaptivity

In this section, a more general estimator of the conditional mode is considered; we adopt the same notations and assumptions as above but beyond the methodology of wavelets. Specifically, we now consider an estimator \( \hat{\partial_y f}(y \mid x) \) of \( \partial_y f(y \mid x) \) defined under the following ratio form:

\[
\hat{\partial}_y f(y \mid x) = \frac{\hat{\partial}_y f(y, x)}{f(x)} I(|\hat{f}(x)| \geq c'/2),
\]

(10)

where \( I(.) \) is the indicator function, \( c' \) refers to (A.5), \( \hat{\partial}_y f(y, x) \) is a generic estimator for \( \partial_y f(y, x) \) (wavelets, kernel, splines, etc.) and \( \hat{f}(x) \) is a general estimator for \( f(x) \) (wavelets, kernel, splines, etc.). Both may be adaptive or not.

Under this general setting, the result below presents a general upper bound for \( |\hat{\Theta}(x) - \Theta(x)| \) with \( \hat{\Theta}(x) \) characterized by (6). This result has the features to hold for any \( n \in \mathbb{N}^* \), and to involve \( \sup_{y \in C_x} |\hat{\partial}_y f(y, x) - \partial_y f(y, x)| \) and \( |\hat{f}(x) - f(x)| \).

**Proposition 1.** Suppose that (A.1)–(A.5) hold. Let \( x \in D \). Consider the estimator \( \hat{\Theta}(x) \) characterized by (6) with the general estimator given in (10). Then, for any \( n \in \mathbb{N}^* \), we have

\[
|\hat{\Theta}(x) - \Theta(x)| \leq \Omega_x \left( \sup_{y \in C_x} |\hat{\partial}_y f(y, x) - \partial_y f(y, x)| + |\hat{f}(x) - f(x)| \right),
\]

where

\[
\Omega_x = \frac{2}{cc'} \max \left[ \sup_{y \in C_x} |\partial_y f(y, x)| \left( 1 + \frac{1}{c'} \right), 1 \right].
\]

**Proof.** By putting (7) and (8) together, we have

\[
|\hat{\Theta}(x) - \Theta(x)| \leq \frac{1}{c} \sup_{y \in C_x} |\hat{\partial}_y f(y \mid x) - \partial_y f(y \mid x)|.
\]

Now, based on (10), we have

\[
|\hat{\partial}_y f(y \mid x) - \partial_y f(y \mid x)| \leq |\hat{\partial}_y f(y \mid x) - \partial_y f(y \mid x)| I(|\hat{f}(x)| \geq c'/2) + \sup_{y \in C_x} |\partial_y f(y \mid x)| I(|\hat{f}(x)| < c'/2).
\]

Owing to (A.5), the last term can be bounded as

\[
I(|\hat{f}(x)| < c'/2) \leq I(|\hat{f}(x) - f(x)| > c'/2) \leq \frac{2}{c'} |\hat{f}(x) - f(x)|.
\]

By virtue of (9) and \( (1/|\hat{f}(x)|) I(|\hat{f}(x)| \geq c'/2) \leq 2/c' \), we immediately get

\[
|\hat{\partial}_y f(y \mid x) - \partial_y f(y \mid x)| I(|\hat{f}(x)| \geq c'/2) \leq \frac{2}{c'} \left( \sup_{y \in C_x} |\hat{\partial}_y f(y, x) - \partial_y f(y, x)| + \frac{\sup_{y \in C_x} |\partial_y f(y, x)|}{c'} |\hat{f}(x) - f(x)| \right).
\]

The desired result follows by combining the above inequalities together. \( \square \)
Remark 3. Based on Proposition 1, the following inequalities hold. There exists a constant $C > 0$ such that the uniform risk of $\hat{\Theta}(x)$ satisfies

$$
\mathbb{E} \left( \sup_{x \in D} \left| \hat{\Theta}(x) - \Theta(x) \right| \right) \leq C \left( \mathbb{E} \left( \sup_{(x,y) \in D \times C_x} \left| \partial_y f(y,x) - \partial_y \hat{f}(y,x) \right| \right) + \mathbb{E} \left( \sup_{x \in D} \left| \hat{f}(x) - f(x) \right| \right) \right).
$$

In particular, the choices of adaptive estimators for $\hat{\partial}_y f(y,x)$ and $\hat{f}(x)$ in (10) which are efficient in the uniform risk yields an adaptive and efficient estimator $\hat{\Theta}(x)$ for $\Theta(x)$ in the uniform risk.

An adaptive wavelet estimator based on hard thresholding for $\hat{f}(x)$ efficient in the uniform risk has been developed by [25]. However, to our knowledge, an adaptive wavelet estimator for $\hat{\partial}_y f(y,x)$ that proved to be efficient in the uniform risk has not received particular attention, and need further non-trivial developments beyond the scope of this note. Consequently, one can say that such a ratio estimator is semi-adaptive. In this regard, the “full” adaptive nonparametric estimation of the conditional mode needs further investigations that are left for the future.

5. Conclusions

This note proposed some results on the estimation of the conditional mode by wavelet methods. A ratio wavelet estimator is proposed and we showed that it achieves fast rates of almost sure convergence over compact subsets of $\mathbb{R}^d$. A general approach is also discussed, taking a step in the direction of adaptivity. The possible perspectives of this work are to study the asymptotic normality of the proposed estimator. In addition, a full adaptive extension through wavelet thresholding or others adaptive techniques remains aground of free research. These two main branches need deeper developments, which we leave for future investigations.

Author Contributions: S.B. and C.C. have contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This project has no funding.

Acknowledgments: The authors thank the two reviewers for their detailed and constructive comments.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Besov Spaces

Notice that there are several equivalent definitions for the Besov spaces $B_{s,p,q}$, for $s > 0$, $1 \leq p \leq \infty$, and $1 \leq q \leq \infty$. Let $(S_\tau f)(x) = f(x - \tau)$. For $0 < s < 1$, set

$$
\gamma_{s,p,q}(f) = \left( \int_{\mathbb{R}^d} \left( \frac{\|S_\tau f - f\|_{L^p}}{\|\tau\|^s} \right)^q d\tau \right)^{1/q}, \quad \gamma_{s,p,\infty}(f) = \sup_{\tau \in \mathbb{R}^d} \frac{\|S_\tau f - f\|_{L^p}}{\|\tau\|^s}.
$$

For $s = 1$, let us define

$$
\gamma_{1,p,q}(f) = \left( \int_{\mathbb{R}^d} \left( \frac{\|S_\tau f - f\|_{L^p}}{\|\tau\|^d} \right)^q d\tau \right)^{1/q}, \quad \gamma_{1,p,\infty}(f) = \sup_{\tau \in \mathbb{R}^d} \frac{\|S_\tau f - f\|_{L^p}}{\|\tau\|^d}.
$$

For $0 < s < 1$ and $1 \leq p, q \leq \infty$, put

$$
B_{s,p,q} = \{ f \in L^p : \gamma_{s,p,q}(f) < \infty \}.
$$
For $s > 1$, set $s = [s]^- + \{s\}^+$ with $[s]^-$ an integer and $0 < \{s\}^+ \leq 1$. Set $B_{s,p,q}$ the space of functions in $L_p(\mathbb{R}^d)$ such that $D^j f \in B_{s^-,-p,q}$ for all $|j| \leq [s]^-$.

Define the norm

$$
\|f\|_{B_{s,p,q}} = \|f\|_{L_p} + \sum_{|j| \leq [s]^-(D^j f)}.
$$

We refer to [26,27] and the Appendix of [22] for more details. Notice that any function $f \in B_{s,p,q}$ is in $L_p(\mathbb{R}^d)$, where $s > 0$ denotes the real valued smoothness parameter of the function $f(\cdot)$. An important characterization of $B_{s,p,q}$ based on wavelets coefficients is provided by [8]. The multiresolution analysis is assumed to be $r$-regular and $s < r$. Hence, we have $f \in B_{s,p,q}$ if and only if

$$
J_{s,p,q}(f) = \|P_{0,0} f\|_{L_p} + \left( \sum_{j > 0} \left( 2^{jq} \|P_{j,0} f\|_{L_p} \right)^q \right)^{1/q} < \infty,
$$

with the classical modification for the sup-norm where $q = \infty$. We have $f \in B_{s,p,q}$ being equivalent to the following:

$$
J'_{s,p,q}(f) = \|a_0 \cdot\|_{L_p} + \left( \sum_{j > 0} \left( 2^{j(s+d(1/2-1/p))} \|b_j \cdot\|_{L_p} \right)^q \right)^{1/q} < \infty,
$$

where

$$
\|b_j \cdot\|_{L_p} = \left( \sum_{l=1}^d \sum_{k \in \mathbb{Z}^d} |b_{j,l,k}|^p \right)^{1/p}.
$$

References

1. Ezzahrioui, M.; Ould-Saïd, E. Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *J. Nonparametr. Stat.* 2008, 20, 3–18.

2. Ferraty, F.; Laksaci, A.; Vieu, P. Functional time series prediction via conditional mode estimation. *C. R. Math. Acad. Sci. Paris* 2005, 340, 389–392.

3. Berlinet, A.; Gannoun, A.; Matzner-Løber, E. Normalité asymptotique d’estimateurs convergents du mode conditionnel. *Canad. J. Stat.* 1998, 26, 365–380.

4. Quintela-Del-Río, A.; Vieu, P. A nonparametric conditional mode estimate. *J. Nonparametr. Stat.* 1997, 8, 253–266.

5. Ould-Saïd, E. A note on ergodic processes prediction via estimation of the conditional mode function. *Scand. J. Stat.* 1997, 24, 231–239.

6. Samanta, M.; Thavanesswaran, A. Nonparametric estimation of the conditional mode. *Comm. Stat. Theory Methods* 1990, 19, 4515–4524.

7. Collomb, G.; Härdle, W.; Hassani, S. A note on prediction via estimation of the conditional mode function. *J. Stat. Plann. Inference* 1987, 15, 227–236.

8. Meyer, Y. Wavelets and operators. In *Cambridge Studies in Advanced Mathematics*; Translated from the 1990 French original by D. H. Salinger; Cambridge University Press: Cambridge, UK, 1992; Volume 37.

9. Daubechies, I. Ten lectures on wavelets. In *CBMS-NSF Regional Conference Series in Applied Mathematics*; Society for Industrial and Applied Mathematics (SIAM): Philadelphia, PA, USA, 1992; Volume 61.

10. Vidakovic, B. Statistical modeling by wavelets. In *Wiley Series in Probability and Statistics: Applied Probability and Statistics*; A Wiley-Interscience Publication; John Wiley & Sons Inc.: New York, NY, USA, 1999.

11. Mallat, S. *A Wavelet Tour of Signal Processing*, 3rd ed.; The Sparse Way, with Contributions from Gabriel Peyré; Elsevier/Academic Press: Amsterdam, The Netherlands, 2009.

12. Härdle, W.; Kerkhaycharian, G.; Picard, D.; Tsybakov, A. Wavelets, approximation, and statistical applications. In *Lecture Notes in Statistics*; Springer: New York, NY, USA, 1998; Volume 129.

13. Hall, P.; Patil, P. Formulas for mean integrated squared error of nonlinear wavelet-based density estimators. *Ann. Stat.* 1995, 23, 905–928.

14. Hall, P.; Patil, P. On wavelet methods for estimating smooth functions. *Bernoulli* 1995, 1, 41–58.
15. Donoho, D.L.; Johnstone, I.M.; Kerkyacharian, G.; Picard, D. Density estimation by wavelet thresholding. *Ann. Stat.* **1996**, *24*, 508–539.
16. Hall, P.; Penev, S. Cross-validation for choosing resolution level for nonlinear wavelet curve estimators. *Bernoulli* **2001**, *7*, 317–341.
17. Doosti H.; Afshari M.; Nouriand H.A. Wavelets for nonparametric stochastic regression with mixing stochastic process. *Comm. Stat. Theory Methods* **2008**, *37*, 373–385.
18. Chesneau, C. On the adaptive wavelet estimation of a multidimensional regression function under $\alpha$-mixing dependence: Beyond the standard assumptions on the noise. *Comment. Math. Univ. Carol.* **2013**, *4*, 527–556.
19. Chaubey, Y.P.; Chesneau, C.; Doosti, H. Adaptive wavelet estimation of a density from mixtures under multiplicative censoring. *Statistics* **2015**, *49*, 638–659.
20. Prakasa Rao, B.L.S. Nonparametric estimation of partial derivatives of a multivariate probability density by the method of wavelets. In *Asymptotics in Statistics and Probability, Festschrift for George G. Roussas*; Puri, M.L., Ed.; VSP: Leiden, The Netherlands, 2000; pp. 321–330.
21. Masry, E. Multivariate probability density estimation by wavelet methods: strong consistency and rates for stationary time series. *Stoch. Process. Appl.* **1997**, *67*, 177–193.
22. Masry, E. Wavelet-based estimation of multivariate regression functions in Besov spaces. *J. Nonparametr. Stat.* **2000**, *12*, 283–308.
23. Chaubey, Y.P.; Doosti, H. Wavelet based estimation of the derivatives of a density for $m$-dependent random variables. *J. Iran. Stat. Soc.* **2005**, *4*, 97–105.
24. Chaubey, Y.P.; Doosti, H.; Prakasa Rao, B.L.S. Wavelet based estimation of the derivatives of a density with associated variables. *Int. J. Pure Appl. Math.* **2006**, *27*, 97–106.
25. Giné, E.; Nickl, R. Uniform Limit theorem for wavelet density estimators. *Ann. Probab.* **2009**, *37*, 1605–1646.
26. Bergh, J.; Lofström, J. Interpolation spaces. An introduction. In *Grundlehren der Mathematischen Wissenschaften, No. 223*; Springer: Berlin, Germany, 1976.
27. Triebel, H. Theory of function spaces. In *Monographs in Mathematics*; Birkhäuser Verlag: Basel, Switzerland, 1983; Volume 78.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).