Adaptive Simultaneous Confidence Intervals in Nonparametric Estimation

Karine Tribouley
Université Paris X, Equipe Modal'X, Bâtiment G, 92001 Nanterre Cedex, France.
karine.tribouley@u-paris10 .fr

Abstract

We present non linear wavelet methods to compute simultaneous confidence intervals for $f(x)$ when f is a functional parameter issued from a non parametric model. The levels of the intervals are at least γ, and we prove that they achieve the minimum diameter up to a logarithmic term. The procedure is data-driven and the adaptation is made via the Lepski's algorithm.

1 Introduction

To study the problem of computing confidence intervals, the statistician has to deal with the following two choices:

- Either he searches for a confidence interval with an exact level γ which minimizes the coverage error,
- or he searches for the shortest interval of level at least γ.

In the statistical literature, both points of view have been studied in the non parametric framework. For the first point, in the case of the problem of pointwise confidence intervals, see (among many others) Hall (1992a, 1992b), Neumann, (1995), Picard and Tribouley (2000), Tribouley (1999). These methods require good approximations of the empirical repartition function of the considered estimators and, consequently, are difficult to generalize to more complex problems, such as simultaneous confidence intervals.

In this paper, we deal with the construction of simultaneous confidence intervals. More precisely, if $[A, B]$ is the support of interest and γ is a confidence level, we search a collection $I(x)$ satisfying

$$P\left(f(x) \in I(x), \forall x \in [A, B]\right) \geq \gamma, \tag{1}$$

where f is a function issued from a regression model or a density model. The simultaneity is an interesting property, for instance when some information on the shape of f is desired (for example, monotonicity, maxima or minima). Moreover, it provides some knowledge on the variance of the local estimation (see for example, Härdle and Marron, (1990)). Let us remark that our aim is not to control in probability the quantity $\|\hat{f} - f\|_\infty$ ($\hat{f}$ is an estimator of f) because our simultaneous confidence intervals may have a diameter varying with the position x. For confidence L_∞balls, you may see Loader and Sun (1994) for the linear regression model, and for confidence L_2 balls, Iouditski and Lacroix-Lambert (2000), and Hoffmann and Lepskii (2002).

Adopting the second point of view, Hall and Titterington (1988) build simultaneous confidence intervals through a linear interpolation of pointwise confidence intervals. The discretisation scheme is such that it is adapted to the regularity of the unknown function f. Moreover, Hall and Titterington (1988) prove that their confidence intervals are optimal for the diameter criterion. More precisely, among the family of considered functions (by instance C^1 functions with bounded derivatives), no other procedure can produce confidence intervals of level at least γ with a smaller diameter. The results are satisfying in practice, but, as it is usual for non parametric estimation, the method requires to work with regular functions (at least C^1) and with the a priori knowledge of the regularity index. This index allows them to choose the adequate discretisation path. In addition, Hall and Titterington (1988) asks for the knowledge of a bound for the first derivative of f (or of the second derivative if we are interested by second order methods). Somehow it is the more restrictive constraint on the procedure because, in practice, the diameter of the intervals depends strongly on this bound. Härdle and Marron (1990) adopt a similar point of view, using bootstrap methods and assuming that f is twice differentiable.

In this paper, we propose to follow Hall and Titterington (1988), considering a wavelet basis instead of a kind of Shauder basis. Using wavelets, for a given x, only a small number of coefficients has to be estimated. Hence, the variance term does not explode, and this gives intervals $I(x)$ well located in x. Moreover, it is possible to give intervals for functions less regular than C^1 functions.

The resolution level used to build the confidence intervals is fixed by a standard trade-off between the bias term and the variance term. More precisely, we choose the multiresolution index j such that the bias term becomes negligible with respect to the variance term. Therefore, no correction is required in order to take the bias term into account. We minimize the variance term, which is equivalent to minimize the diameter of the interval. An optimal confidence interval (in the sense that the order of the diameter is equal to the optimal order given in Hall and Titterington (1988)) is then obtained. This procedure is not yet adaptive because the smoothing index j depends on the regularity of f. Thus, we provide an interval for which the construction does not (or only slightly) depends on the unknown function f. The adaptation is obtained choosing the multiresolution index by Lepski's algorithm (Lepski, (1991)). This algorithm provides a statistic which under-estimates the optimal index. This phenomenon may have no importance for the estimation problem, but, for some specific situations, leads to a slower estimation rate than the optimal one. This is the case here: the diameter of the simultaneous almost adaptive confidence intervals loses (in order) a log factor with respect to the optimal diameter given in Hall and Titterington (1988).

In order to prove results concerning coverage and diameter, we need assumptions similar to Picard and Tribouley (2000), and to Tribouley (1999) for the problem of the pointwise confidence interval built according to the first point of view (minimization of the coverage error). We assess that the wavelet coefficients of the function correctly indicate the regularity index. We quantify this adaptation capacity thanks to a parameter ρ_n and our procedure depends on this parameter. Hence, our procedure is almost adaptive since this parameter is a priori unknown.

The paper is organized as follows. In Section 2, we introduce the models and assumptions. In Section 3, we describe the construction of the simultaneous confidence intervals and we give the results about the coverage and the diameter. In Section 4, some preliminary results are stated and the proofs of the theorems of Section 3 are postponed to Section 5.

2 Models and Assumptions

2.1 Models and Notation

We consider the usual nonparametrical models. The Gaussian regression model is defined by

$$Y_i = f\left(\frac{i}{n}\right) + \sigma\,\epsilon_i, \quad i = 1, \ldots, n$$

where $\sigma > 0$, and the ϵ's are independent standard Gaussian variables. In the density model, $X_1, \ldots, X_n$ denote the n independent variables with common density f. For each model, the parameter of interest is f. We suppose that f is compactly supported and, without loss of generality, we denote $[0, M]$ the support (with $M = 1$ in the regression model). We are interested by finding confidence intervals $I(x)$ for any $x \in [A, B]$ strictly included in $[0, M]$. These confidence intervals have to be simultaneous in the sense of (1).

Comments

- In the regression model, the normality hypothesis is not necessary and may be replaced by an assumption on the moments of the errors (see Picard and Tribouley, (2000)). However, this simplifies the proofs (no Edgeworth expansion is required). Also, for the sake of simplicity, we suppose that the variance σ^2 of the errors is known. But, as usual for the regression model, it can be estimated, with no consequence on the others results.

- The compactness and the knowledge of the support of f in the density case is a strong assumption. However, it is not a restriction in practice since the number of data is finite and we always restrict ourselves to the interval $[\min X_i, \max X_i]$.

Let N be a fixed positive constant. Let ϕ and ψ be a scaling function and an associated wavelet function. We assume that these functions are compactly supported on $[0, 2N-1]$ and that the q-th moment of the wavelet ψ vanishes for $q = 0, \dots N$. See for example the Daubechies's wavelets (Daubechies, (1992)). For any function h, we denote by $h_{j,k}(x)$ the function $2^{j/2} h(2^j x - k)$.

In the sequel, we use the following notations: $O(u_n)$ is a quantity t_n such that $\lim_{n\to\infty} |t_n|/u_n \le C$ for a positive constant C and $o(u_n)$ is a quantity t_n such that $\lim_{n\to\infty} |t_n|/u_n = 0$.

2.2 Estimation of f

For any j, k, the scaling and wavelet coefficients of the function f are defined respectively by

$$\alpha_{j,k} = \int \phi_{j,k} f \quad \text{and} \quad \beta_{j,k} = \int \psi_{j,k} f.$$

which are estimated by their empirical counterparts

$$\hat{\alpha}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \phi_{j,k}(X_i) \quad \text{and} \quad \hat{\beta}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \psi_{j,k}(X_i)$$

in the density model and by

$$\hat{\alpha}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \phi_{j,k}(\frac{i}{n}) Y_i \quad \text{and} \quad \hat{\beta}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \psi_{j,k}(\frac{i}{n}) Y_i$$

in the regression model. Motivated by the following expansion of f on the wavelet basis

$$\forall j \ge 0, \qquad \forall x, \ f(x) = \sum_k \alpha_{j,k} \phi_{j,k}(x) + \sum_{l=j}^{\infty} \sum_k \beta_{l,k} \psi_{l,k}(x), \tag{2}$$

we consider the family

$$j \ge 0, \quad \hat{f}_j(x) = \sum_k \hat{\alpha}_{j,k} \phi_{j,k}(x)$$

of estimators of $f(x)$.

2.3 Functional assumptions

Let us describe the class of functions we consider. Let s, M_1, M_2, m_f be some positive real and (ρ_n) a sequence of positive numbers.

- **Assumption** $A_1(M_1, s)$**.** f belongs to the space defined thanks to the wavelet coefficients by

$$\Big\{ f, \quad \forall j,k, \quad |\beta_{j,k}| \leq M_1 \, 2^{-j(s+1/2)} \Big\}.$$

- **Assumption** $A_2(M_2, s, \rho_n)$**.** Under the assumption $A_1(M_1, s)$, for any $n \geq 2$, there exists j between $j_s - \rho_n$ and j_s such that

$$\exists k, \quad |\beta_{j,k}| \geq M_2 \, 2^{-j_s(s+1/2)},$$

where j_s is (up to a logarithmic term) the optimal parameter for the estimation problem (the order of 2^{j_s} is $(n/\log n)^{1/1+2s}$).

- **Assumption** $A_3(m_f)$**.** For any x in the interval of interest $[A, B]$, $f(x) \geq m_f$.

Comments

- If s is not an integer, A_1 holds for Lipschitz functions. More precisely, the functions in

$$L_s(M_1) \quad = \quad \Big\{ f : R \to R; \quad \forall (x,y) \in R^2, |f^{([s])}(x) - f^{([s])}(y)| \leq c M_1 |x-y|^{\alpha}, \\ \text{where } s = [s] + \alpha, \ 0 < \alpha \leq 1 \}$$

satisfy A_1. The Besov norm and the Lipschitz norm are linked thanks to the constant $c > 0$.

- A_2 is obviously satisfied for large ρ_n, but becomes more restricting when ρ_n is decreasing. The sequence ρ_n gives a measure of the capacity for f to be adaptively estimated. Let us remark that we link this capacity to the number of data n: the more data are available, the stronger the condition can be. In the sequel, we consider functions f such that A_2 holds for $\rho_n \leq O(\log_2 \log n)$.

- A_3 is required for the density model: it is a standard assumption in order to bound from below the variance term. If A_3 does not hold, we restrict the interval $[A, B]$ of the study.

For more details about the assumptions A_1, A_2, we refer to Picard and Tribouley, (2000).

3 Construction of the confidence interval

3.1 Simultaneous confidence interval for the coefficients

Let us recall that N is the number of vanishing moments of the wavelet ψ and M is the length of the support of f. We consider some index $j \geq \log_2((2N+1)/M)$. We restrict the study to the coefficients such that k is varying in

$$\mathcal{K}_j = \{0, \ldots, 2^j M - (2N+1)\}. \tag{3}$$

At some given level j, the number of coefficients is denoted by $K_j = 2^j M - 2N$. The expectation and the variance of the empirical coefficients are respectively

$$\mu_{j,k} = \alpha_{j,k} \quad \text{and} \quad \sigma^2_{j,k} = \frac{1}{n}\left[\left(\int \phi^2_{j,k} f \right) - \left(\int \phi_{j,k} f \right)^2 \right],$$

for the density model, and

$$\mu_{j,k} = \alpha_{j,k} + r_n \quad \text{and} \quad \sigma^2_{j,k} = \frac{\sigma^2}{n^2} \sum_{i=1}^{n} \phi^2_{j,k}(\frac{i}{n})$$

for the regression model. The Riemann approximation term is given by

$$r_n \quad = \quad \left[\frac{1}{n} \sum_{i=1}^{n} \phi_{j,k}(\frac{i}{n}) f(\frac{i}{n}) \; - \; \int \phi_{j,k} f \right] .$$

For the case of the density model, we estimate the variance by

$$\hat{\sigma}^2_{j,k} = \frac{1}{n^2} \sum_{i=1}^{n} \phi^2_{j,k}(X_i) - \frac{1}{n^2(n-1)} \sum_{i \neq l = 1}^{n} \phi_{j,k}(X_i)\phi_{j,k}(X_l).$$

Let $\gamma \in [0,1]$ be the minimal confidence level and d_j be the sequence satisfying $\mathcal{N}(d_j) = 1 - \frac{(1-\gamma)}{2K_j}$, $\mathcal{N}$ being the repartition function of the standard Gaussian random variable. By the standard large deviation theory,

$$d_j = \sqrt{\log(\frac{2}{(1-\gamma)} K_j)} = 0(\sqrt{\log 2^j}). \tag{4}$$

We propose the family of confidence intervals $I_j(k) = [\tilde{\alpha}^1_{j,k}, \tilde{\alpha}^2_{j,k}]$ with extremities given by

$$\begin{aligned} \tilde{\alpha}^1_{j,k} &= \hat{\alpha}_{j,k} - d_j v_{j,k} \\ \tilde{\alpha}^2_{j,k} &= \hat{\alpha}_{j,k} + d_j v_{j,k} \end{aligned} \tag{5}$$

where the quantity $v^2_{j,k}$ denotes either $\sigma^2_{j,k}$ (for the regression case) or $\hat{\sigma}^2_{j,k}$ (for the density case). These intervals are simultaneous confidence intervals for the coefficients $\{(\alpha_{j,k}), k \in \mathcal{K}_j\}$ in the sense of the following proposition.

Proposition 1 *Let us define j_0 and j_∞ by*

$$2^{j_0} = \log n, \quad 2^{j_\infty} = \sqrt{\frac{n}{\log n}}$$

for the density model and,

$$2^{j_0} = n^{\frac{1}{1+2N}}, \quad 2^{j_\infty} = \sqrt{\frac{n}{\log n}}$$

for the regression model. Let j be in $\{j_0, \dots j_\infty\}$.

For the density case, we assume that f is bounded from below on the interval $[(2N-1)2^{-j}, M-(2N-1)2^{-j}]$. For the regression case, we assume that there exist $s \geq \frac{N}{1+2N}$ and $M_1 > 0$ such that $A_1(M_1, s)$ holds. We have

$$P\left(\alpha_{j,k} \in I_j(k), \forall k \in \mathcal{K}_j\right) \geq \gamma - o(1).$$

Comments.

- For the regression case, the choice of j_0 and the condition $s \geq \frac{N}{1+2N}$ are necessary to keep the Riemann approximation r_n negligible. For the density case, the constraint on j_∞ is due to the Gaussian approximation for the law of the empirical coefficients. Similarly to the regression model, it leads us to consider some function f at least $1/2-$ regular (in the sense of A_1).

- We do not consider all the non zero coefficients $\alpha_{j,k}$. In particular, we do not compute the coefficients needed to synthesize f near the extremities of its support. This implies that the simultaneous confidence interval will be built for x belonging in $[A, B]$ strictly nested in the support of f. The distances needed between $[A, B]$ and $[0, M]$ are given in Theorem 1 and Theorem 2.

3.2 Simultaneous confidence interval for the function f.

Let $p \geq 1$ and $a > 0$. Motivated by the expansion (2), we propose a family of confidence intervals

$$I_j(x,p) = [\tilde{f}_j^1(x), \tilde{f}_j^2(x)] \tag{6}$$

with extremities given by

$$\begin{aligned} \tilde{f}_j^1(x) &= \hat{f}_j(x) - (1+a)d_j^p \sum_k v_{j,k}|\phi_{j,k}(x)| \\ \tilde{f}_j^2(x) &= \hat{f}_j(x) + (1+a)d_j^p \sum_k v_{j,k}|\phi_{j,k}(x)| \end{aligned}$$

for j varying between j_0 and j_∞ defined in Proposition 1. Let us set the following constant

$$b(a) = \frac{M_1^2(2N)^3\|\psi\|_\infty^2}{a^2\phi(x_0)^2C_v^2}, \tag{7}$$

where C_v is the constant depending on m_f given in (10) or (11) and x_0 is an integer such that $\phi(x_0) \neq 0$.

We have the following results concerning the coverage probability of the simultaneous intervals.

Theorem 1 *Let $s > 1/2$ and $M_1, m_f > 0$. We suppose $A_1(s, M_1)$ holds. Moreover, for the density model, we suppose that $A_3(m_f)$ is true. Let j_s be defined by*

$$2^{j_s} = \left(b(a)\,\frac{n}{\log n}\right)^{\frac{1}{1+2s}}.$$

Then, for any j varying between j_s and j_∞, $p = 1$ and $a > 0$, we get

$$P\left(f(x) \in I_j(x,p), \forall x \in [A,B]\right) \geq \gamma + o(1)$$

as soon as

$$\frac{(2N-1)}{2^j} \leq A < B \leq M - \frac{(2N-1)}{2^j}.$$

Theorem 2 *Let $M_2 > M_1$. We suppose in addition that $A_2(M_2, s, \rho_n)$ holds for some $\rho_n \leq O(\log\log n)$. Then, for $p \geq 1 + (1+2N)\rho_n(\log_2\log n)^{-1}$ and $a > a_0(M_1, M_2)$, if A, B verify*

$$\frac{2N-1}{2^{j_0}} \leq A < B \leq M - \frac{2N-1}{2^{j_0}},$$

we have

$$P\left(f(x) \in I_{\hat{j}}(x,p), \forall x \in [A,B]\right) \geq \gamma + o(1)$$

where $\hat{j}$ is computed in the following way

$$\hat{j} = \sup\left(j = j_0, \dots j_\infty,\ \exists k \in \mathcal{K}_j, |\hat{\beta}_{j,k}| \geq T\sqrt{\frac{\log n}{n}}\right).$$

The threshold constant T satisfies $T_1 \leq T \leq T_2$ where

$$\begin{aligned} T_1 &= \frac{M_1}{b(a)} + 2\|\psi\|_2\sigma\sqrt{N}, \\ T_2 &= \frac{M_2}{b(a)} - 2\|\psi\|_2\sigma\sqrt{N}, \\ a_0(M_1, M_2) &= \frac{2\sqrt{\sigma}(2N+1)^{7/4}\|\psi\|_\infty^{3/2}}{C_v|\phi(x_0)|}\frac{M_1}{\sqrt{M_2 - M_1}} \end{aligned}$$

for the regression model and

$$T_1 = \frac{M_1}{b(a)} + \frac{\|\psi\|_\infty}{6}\left(1+\sqrt{1+36\frac{\|f\|_\infty}{\|\psi\|_\in fty^2}}\right),$$

$$T_2 = \frac{M_2}{b(a)} - \frac{\|\psi\|_\infty}{6}\left(1+\sqrt{1+36\frac{\|f\|_\infty}{\|\psi\|_\in fty^2}}\right),$$

$$a_0(M_1,M_2) = \left(1+\sqrt{1+36\frac{\|f\|_\infty}{\|\psi\|_\in fty^2}}\right)^{1/2} \frac{(2N+1)^{3/4}\|\psi\|_\infty^{3/2}}{\sqrt{3}C_v|\phi(x_0)|}\frac{M_1}{\sqrt{M_2-M_1}}$$

for the density model.

Comments

- Using Proposition 1, we easily obtain a bound on the coverage probability when I_j is interpreted as a confidence interval for the expectation of $\hat{f}_j$. Theorem 1 explains how to choose the multiresolution level j such that the bias term becomes negligible and then to obtain a simultaneous confidence interval with bounded coverage probability for f. Let us remark that in Theorem 1, the constant a used in the construction of I_j may be chosen arbitrarily small and $p = 1$ is convenient.

- Contrary, in the adaptive procedure, the parameters p and a are useful because we overestimate the bias term. We need to increase the variance term (by a constant thanks a and by a logarithmic term thanks to p) such that the bias term becomes negligible in order.

- We determine a_0 such that the constrains on the threshold constant are compatible i.e. $T_1 < T_2$. But, in practice, we choose the quantity a as small as we wish; Theorem 2 is valid for functions satisfying the assumptions for $M_1 < h(a)M_2$ where h is a function depending on $N, \phi, \psi, \sigma, m_f, s$.

Theorem 1 and Theorem 2 give a collection of simultaneous confidence intervals of level at least γ. Among the collection, we have now to exhibit the interval with the shortest diameter (in order). We denote by $|I|$ the diameter of I.

Theorem 3 *Under the assumptions of Theorem 1, I_{j_s} is the shortest interval among the family of intervals described above and, in the regression case*

$$\sup_{x\in[A,B]} |I_{j_s}(x,1)| \le O\left((\frac{n}{\log n})^{-\frac{s}{1+2s}} \right).$$

In the density case, the diameter is random and

$$\lim_{n,C\longrightarrow+\infty} P\left(\inf_{x\in[A,B]} |I_{j_s}(x,1)| \ge C(\frac{n}{\log n})^{-\frac{s}{1+2s}} \right) = 0.$$

Theorem 4 *Under the assumptions of Theorem 2, we have*

$$\lim_{n,C\longrightarrow+\infty} P\left(\inf_{x\in[A,B]} |I_{\hat{j}}(x,p)| > C\,(\log n)^{\frac{p-1}{2}} (\frac{n}{\log n})^{-\frac{s}{1+2s}} \right) = 0.$$

Comments

- The diameters given in Theorem 3 are of the same order than the optimal diameter of Hall and Titterington (1988). The interval $I_{j_s}(x,1)$ is then optimal among all simultaneous confidence intervals of level at least γ. The (near) adaptive interval $I_{\hat{j}}(x,p)$ is (near) optimal (up to a logarithmic factor).

- The loss of a logarithmic factor in Theorem 3 is due to the property of simultaneity and this is unavoidable (see Hall and Titterington, (1998)). In Theorem 4, another logarithmic factor appears, due to the property of adaptivity.

4 Preliminaries

4.1 Exponential inequalities

First, we recall the Bernstein inequality and next we state large deviation inequalities for the wavelet coefficients and the random smoothing index $\hat{\jmath}$.

Lemma 1 *Let $Z_1, \dots Z_n$ be n independent random variables such that, for $i = 1, \dots n$, $EZ_i = 0$, $V(Z_i) = v^2$ and $|Z_i| < +\infty$. Then*

$$\forall \lambda > 0, \qquad P(\frac{1}{n}\sum_i Z_i \geq \lambda) \leq \exp\left(-\frac{n\lambda^2}{2(v^2 + \lambda/3|Z|_\infty)}\right).$$

Let us define

$$\gamma(x) = x^2(8N\|\psi\|_2^2\sigma^2)^{-1} \tag{8}$$

in the regression case and

$$\gamma(x) = x^2(2\|f\|_\infty + 2x/3\|\psi\|_\infty)^{-1} \tag{9}$$

in the density case. Let us recall that j_s is defined by $2^{j_s} = \left(bn\,(\log n)^{-1}\right)^{\frac{1}{1+2s}}$ for some $b > 0$.

Lemma 2 *In the regression model, we assume that $A_1(M_1, s)$ holds for some $s \geq N/(1+2N)$ and some $M_1 > 0$. Let T be some positive constant. Then there exists $C > 0$ such that, for any $\delta \leq \gamma(T)$,*

$$\forall j \leq \log_2(\frac{n}{\log n}), \quad P\left(|\hat{\beta}_{j,k} - \beta_{j,k}| \geq T\sqrt{\frac{\log n}{n}}\right) \leq Cn^{-\delta}.$$

Lemma 3 *Let $s > 0$ ($s \geq N/(1+2N)$ in the regression model) and $M_1 > 0$. Let us assume that $A_1(M_1, s)$ holds. Then, as soon as the threshold constant T satisfies $T > M_1 b^{-1}$, there exists $C > 0$ such that, for any $\delta \leq \gamma(T - M_1 b^{-1})$,*

$$P(\hat{\jmath} \geq j_s) \leq Cn^{-\delta}$$

Lemma 4 *Let $M_2 > 0$ and $\rho_n > 0$. Under the same assumption as in the previous lemma and if $A_2(M_2, s, \rho_n)$ holds such a way that $M_1 < M_2$, then, as soon as the threshold constant T satisfies $M_1 b^{-1} < T < M_2 b^{-1}$, there exists $C > 0$ such that, for any $\delta \leq \gamma(T - M_1 b^{-1}) \wedge \gamma(M_2 b^{-1} - T)$,*

$$P(\hat{\jmath} + (1+2N)\rho_n \leq j_s) \leq Cn^{-\delta}$$

Proofs. The proofs are very usual in the wavelet framework. In the density model, Lemma 2 is an application of Lemma 1. In the regression model, we first have to bound the bias term and then to apply another kind of Bernstein inequality (see by instance Petrov, (1995), page 57). Lemma 3 and Lemma 4 are consequences of Lemma 2 and of the definition of $\hat{\jmath}$. For more detailed proofs, see Picard and Tribouley (2000) in the regression case and Tribouley, (1999), in the density case.

4.2 Edgeworth expansion (density case)

Lemma 5 *Let us denote $\mu^{(3)}_{j,k} = E(\phi_{j,k}(X) - \int \phi_{j,k} f)^3$ and $\mu^{(2)}_{j,k} = E(\phi_{j,k}(X) - \int \phi_{j,k} f)^2$. For all $n, j > 0$,*

$$P(\frac{\alpha_{j,k} - \hat{\alpha}_{j,k}}{\sigma_{j,k}} < x) \quad = \quad \mathcal{N}(x) - \frac{1}{\sqrt{n}}\frac{\mu^{(3)}_{j,k}}{6(\mu^{(2)}_{j,k})^{3/2}}(x^2 - 1)\mathcal{N}'(x) + O\left(\frac{2^j}{n}\right)$$

uniformly in x.

Proof. We follow the lines of Feller, (1966): we prove a second order Edgeworth expansion and we bound the term of second order with the evaluation of

$$\begin{aligned} |\mu^{(3)}_{j,k}| &\leq 2^{j/2+3}\|f\|_\infty(\|\phi\|_3^3 + 2^{-2j}\|\phi\|_1^3) \leq O(2^{j/2}), \\ \mu^{(4)}_{j,k} &\leq 2^{j+4}\|f\|_\infty(\|\phi\|_4^4 + 2^{-3j}\|\phi\|_1^4) \leq O(2^{j}). \end{aligned}$$

On the other hand, because f is lower bounded on $[(2N-1)2^{-j}, M-(2N-1)2^{-j}]$ (say by m_f) and because we consider only the coefficients such that $k = 0, \ldots, 2^j-(2N-1)$ (hence the support of $\phi_{j,k}1_{[0,M]}$ always contains the interval $[0, 2N-1]$), we get

$$\mu^{(2)}_{j,k} \geq \left(m_f - 2^{-j}\|f\|_\infty\right) \geq O(1). \tag{10}$$

4.3 Accuracy of the variance estimation (density case)

Lemma 6 *For any constant C larger than $(18\|f\|_\infty\|\phi\|_4^4 + C_v\|\phi\|_\infty)^{1/2}C_v^{-1}$, there exists a positive constant c not depending on C and n such that,*

$$\forall j,k, \quad P\left(|\hat\sigma_{j,k} - \sigma_{j,k}| \geq C\frac{2^{j/2}}{n}(\log n)^{1/2}\right) \leq c\,n^{-1}.$$

Proof. There exists some constant C_v (depending on m_f) such that $\sigma^2_{j,k} \geq C_v^2 n^{-1}$ (see the bound (10)). Then, for any $0 < u < \sigma_{j,k}$, we have

$$\begin{aligned} P(|\hat\sigma_{j,k} - \sigma_{j,k}| \geq u) &= P(\frac{\hat\sigma_{j,k}}{\sigma_{j,k}} \geq 1 + \frac{u}{\sigma_{j,k}}) + P(\frac{\hat\sigma_{j,k}}{\sigma_{j,k}} \leq 1 - \frac{u}{\sigma_{j,k}}) \\ &\leq P(|\hat\sigma^2_{j,k} - \sigma^2_{j,k}| \geq 2u\sigma_{j,k} + u^2) + P(|\hat\sigma^2_{j,k} - \sigma^2_{j,k}| \geq 2u\sigma_{j,k} - u^2) \\ &\leq 2P(|\hat\sigma^2_{j,k} - \sigma^2_{j,k}| \geq \frac{C_v\, u}{\sqrt{n}}). \end{aligned}$$

Let us remark that

$$\begin{aligned} \hat\sigma^2_{j,k} - \sigma^2_{j,k} &= \frac{1}{n^2}\sum_i \left(\phi^2_{j,k}(X_i) - \int \phi^2_{j,k} f\right) + \frac{2}{n^2}\sum_i\left(\phi_{j,k}(X_i)\int \phi_{j,k} f - (\int \phi_{j,k} f)^2\right) \\ &+ \frac{1}{n^2(n-1)}\sum_{i\neq l}(\phi_{j,k}(X_i) - \int \phi_{j,k} f)(\phi_{j,k}(X_l) - \int \phi_{j,k} f) \end{aligned}$$

To bound the first and second term, we apply Lemma 1 with

$$Z_i = \phi^2_{j,k}(X_i) - \int \phi^2_{j,k} f, \qquad V(Z_i) \leq 2^j\|f\|_\infty\|\phi\|_4^4, \quad |Z_i| \leq 2^j\|\phi\|_\infty, \quad \lambda = \frac{C_v u\sqrt{n}}{3}$$

and with

$$Z_i = \phi_{j,k}(X_i)\int \phi_{j,k} f - (\int \phi_{j,k} f)^2, \qquad V(Z_i) \leq 2^{-j}\|f\|^2_\infty\|\phi\|_1, \quad |Z_i| \leq \|\phi\|^2_\infty\|f\|_\infty, \quad \lambda = \frac{C_v u\sqrt{n}}{6}.$$

There exist $c > 0$ and $c' > 0$ ($c = C_v^2/9(2\|f\|_\infty\|\phi\ |_4^4 + C_v\|\phi\|_\infty/9)^{-1}$) such that

$$P\left(\frac{1}{n^2}\sum_{i=1}^n(\phi^2_{j,k}(X_i) - \int \phi^2_{j,k} f) \geq \frac{C_v\, u}{3\sqrt{n}}\right) \quad \leq \quad \exp - cn^2u^2 2^{-j}$$

and

$$P\left(\frac{2}{n^2}\sum_i\left(\phi_{j,k}(X_i)\int\phi_{j,k}f-(\int\phi_{j,k}f)^2\right)\geq\frac{C_v\,u}{3\sqrt{n}}\right)\quad\leq\quad\exp-c'n^2u^2.$$

For the last term, let us put

$$Z_i=\phi_{j,k}(X_i)-\int\phi_{j,k}f,\qquad EZ_i=0,\qquad EZ_i^2\leq\int\phi_{j,k}^2f\leq\|f\|_\infty.$$

Observing that

$$E(\sum_{i\neq l}Z_iZ_l)^2\quad=\quad\sum_{i\neq l}EZ_i^2EZ_l^2\leq O(n^2),$$

and using Chebychev inequality, we deduce that,

$$P\left(\frac{1}{n^2(n-1)}\sum_{i\neq l}(\phi_{j,k}(X_i)-\int\phi_{j,k}f)(\phi_{j,k}(X_l)-\int\phi_{j,k}f)\geq\frac{C_v\,u}{3\sqrt{n}}\right)\quad\leq\quad O(u^{-2}n^{-3}).$$

Choosing $u=C2^{j/2}\sqrt{\log n}n^{-1}$ with $C\geq c^{-1/2}$ (which is possible because $j\leq j_\infty$), we get the result.

4.4 Technical lemma

Lemma 7 *Let X_n be a sequence of random variables admitting the Edgeworth expansion*

$$P(X_n<t)=\mathcal{N}(t)+p_n(t)\mathcal{N}'(t)+O(u_n)$$

with some polynomials p_n of bounded order with bounded coefficients. We assume that the sequence Y_n of random variables satisfies

$$P(|Y_n|>v_n)\leq w_n.$$

Then

$$P(X_n+Y_n<t)=P(X_n<t)+O(u_n+v_n+w_n).$$

Proof. The result follows immediately from the inequalities

$$P(X_n<t+v_n)-P(|Y_n|>v_n)\leq P(X_n+Y_n<t)\leq P(X_n<t+v_n)+P(|Y_n|>v_n)$$

and the Lipschitz equicontinuity of the functions $\mathcal{N}(t)+p_n(t)\mathcal{N}'(t)$.

5 Proofs

5.1 Proof of Proposition 1

5.1.1 Regression case

Let us fix j in $\{j_0,\ldots,j_\infty\}$ and k in $\mathcal{K}_j$. Because the number of terms in the sum is

$$\#\{i,\phi_{j,k}(\frac{i}{n})\neq 0\}=(2N-1)n2^{-j},$$

there exists a constant C_v such that

$$v_{j,k}^2 \geq C_v^2 \frac{1}{n}. \tag{11}$$

Thanks to the regularity assumption on f, we get

$$\begin{aligned} |r_n| &\leq \sum_{i=1}^{n} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \left|\phi_{j,k}(\frac{i}{n}) - \phi_{j,k}(x)\right| |f(x)|\, dx + \sum_{i=1}^{n} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \left|f(\frac{i}{n}) - f(x)\right| \left|\phi_{j,k}(\frac{i}{n})\right| dx \\ &\leq O\left(n^{-(s\wedge 1)} 2^{-j/2} + \frac{2^{j/2}}{n}\right) \end{aligned} \tag{12}$$

which is bounded by $O\left(\frac{2^{j/2}}{n}\right)$ as soon as $s \geq \frac{N}{1+2N}$. Combining (12) and (11), we obtain:

$$|r_n| \leq O\left((\frac{2^j}{n})^{1/2} \frac{1}{d_j} \frac{d_j}{\sqrt{n}}\right) = O\left(\frac{2^{j/2}}{n}\right).$$

Since

$$P\left(|\frac{\mu_{j,k} - \hat{\alpha}_{j,k}}{\sigma_{j,k}}| < d_j - \frac{|r_n|}{\sigma_{j,k}}\right) \quad \leq \quad P\left(|\frac{\alpha_{j,k} - \hat{\alpha}_{j,k}}{\sigma_{j,k}}| < d_j\right),$$

and, thanks to the definition of d_j,

$$\begin{aligned} P\left(|\frac{\mu_{j,k} - \hat{\alpha}_{j,k}}{\sigma_{j,k}}| < d_j - \frac{|r_n|}{\sigma_{j,k}}\right) &\geq 1 - \frac{1-\gamma}{K_j} - O\left(\frac{|r_n|}{\sigma_{j,k}} \exp -d_j^2/2\right) - O\left(\frac{j}{n^2}\right) \\ &\geq 1 - \frac{1-\gamma}{K_j} - O\left((2^j n)^{-1/2}\right), \end{aligned}$$

we obtain the bound

$$P(\mathcal{E}_k) \quad \geq \quad 1 - \frac{1-\gamma}{K_j} - O\left((2^j n)^{-1/2}\right)$$

where we set $\mathcal{E}_k = \{\left|\frac{\alpha_{j,k} - \hat{\alpha}_{j,k}}{v_{j,k}}\right| \leq d_j\}$. We finish the proof observing that

$$P(\cap_{k\in\mathcal{K}_j} \mathcal{E}_k) \quad = \quad P(\overline{\cup_{k\in\mathcal{K}_j} \overline{\mathcal{E}_k}}) \geq 1 - \sum_{k\in\mathcal{K}_j} P(\overline{\mathcal{E}_k}) \geq \gamma - o(1).$$

5.1.2 Density case

Let us fix j in $\{j_0, \ldots, j_\infty\}$ and k in $\mathcal{K}_j$. Using Lemma 6 and Lemma 7, we obtain

$$\begin{aligned} P(|\alpha_{j,k} - \hat{\alpha}_{j,k}| \leq d_j \hat{\sigma}_{j,k}) &= P(|\hat{\alpha}_{j,k} - \alpha_{j,k}| \leq d_j \sigma_{j,k} + d_j(\hat{\sigma}_{j,k} - \sigma_{j,k})) \\ &= P(|\frac{\alpha_{j,k} - \hat{\alpha}_{j,k}}{\sigma_{j,k}}| \leq d_j) + O\left(d_j \frac{2^{j/2}}{n} (\log n)^{1/2} + n^{-1} + \frac{2^j}{n}\right). \end{aligned}$$

Applying now Lemma 5 with $|\mu_{j,k}^{(3)}| \leq O(2^{j/2})$ and $\mu_{j,k}^{(2)} \geq O(1)$ (see (10)), it follows

$$P\left(|\frac{\alpha_{j,k} - \hat{\alpha}_{j,k}}{\sigma_{j,k}}| \leq d_j\right) \quad \geq \quad 1 - \frac{1-\gamma}{K_j} - O\left(\frac{2^{j/2}}{\sqrt{n}} \exp -(d_j^2/2)\, d_j^2 + \frac{2^j}{n}\right).$$

and then, we conclude that

$$P(|\alpha_{j,k} - \hat{\alpha}_{j,k}| \leq d_j \hat{\sigma}_{j,k}) \quad \geq \quad 1 - \frac{1-\gamma}{K_j} - \frac{1}{K_j} O(\epsilon_j)$$

for

$$\epsilon_j = K_j \left(\frac{2^j}{n}\right)^{1/2} \exp -(d_j^2/2)\, d_j^2 + K_j \frac{1}{n} + K_j d_j \frac{2^{j/2}}{n} \sqrt{\log n} + K_j \frac{2^j}{n} .$$

Since $d_j = O(\sqrt{\log 2^j})$, $K_j = O(2^j)$ and $j \leq j_\infty$, we have $\epsilon_j \leq o(1)$. We finish the proof as in the regression model.

5.2 Proof of Theorem 1

Let us fix $j \geq j_0$ and let x be in $[A, B]$ for A, B such that $(2N-1)2^{-j} \leq A < B \leq M - (2N-1)2^{-j}$. Using the regularity assumption for f, we bound the bias term:

$$\begin{aligned} |B_j(x)| &\leq |\sum_{l \geq j} \sum_k \beta_{j,k} \psi_{j,k}(x)| \\ &\leq M_1 (2N-1) \|\psi\|_\infty \, 2^{-js}. \end{aligned} \tag{13}$$

On the other hand, thanks to (4) and (10), we have the following bound for the variance term

$$\begin{aligned} V_j(x) &= d_j \sum_k v_{j,k} |\phi_{j,k}(x)| \\ &\geq (\log \frac{2}{(1-\gamma)} K_j)^{1/2} \frac{C_v}{\sqrt{n}} 2^{j/2} \, |\phi(x_0)| \\ &\geq |\phi(x_0)| C_v \left(\frac{2^j \log 2^j}{n} \right)^{1/2}, \end{aligned} \tag{14}$$

where $x_0 = 2^j x - k$ for some k chosen in $\{2^j x - (2N-1), \ldots, 2^j x\}$ such that $\phi(x_0) \neq 0$. Note that x_0 does not vary with x. We deduce from (13) and (14), that, if j_s is given by $2^{j_s} = (b(a) n (\log n)^{-1})^{1/(1+2s)}$ for some constant $a > 0$ and where $b(a)$ is defined in (7), then for any $j \geq j_s$, we have

$$|B_j(x)| \quad \leq \quad a V_j(x).$$

We apply now Proposition 1, combining with Expansion (2),

$$\begin{aligned} \gamma &\leq P\left(- d_j v_{j,k} |\phi_{j,k}(x)| \leq (\alpha_{j,k} - \hat{\alpha}_{j,k}) \phi_{j,k}(x) \leq \ d_j v_{j,k} |\phi_{j,k}(x)|, \ \forall k \in \mathcal{K}_j, \ \forall x \in [A,B]\right) \\ &\leq P\left(- d_j \sum_k v_{j,k} |\phi_{j,k}(x)| \leq \sum_k (\alpha_{j,k} - \hat{\alpha}_{j,k}) \phi_{j,k}(x) \leq \ d_j \sum_k v_{j,k} |\phi_{j,k}(x)|, \ \forall x \in [A,B]\right) \\ &= P\Big(- V_j(x) + B_j(x) + \hat{f}_j(x) \leq P_j f(x) + B_j(x) \leq \ V_j(x) + B_j(x) + \hat{f}_j(x), \ \forall x \in [A,B]\Big) \\ &\leq P\Big(-(1+a)\, V_j(x) + \hat{f}_j(x) \leq f(x) \leq \ (1+a) V_j(x) + \hat{f}_j(x), \ \forall x \in [A,B]\Big). \end{aligned}$$

Since we can choose a arbitrary small, we get the result in the non adaptive case.

5.3 Proof of Theorem 2

We follow the same way as in the non adaptive case. In view to bound the bias term, we use the definition of $\hat{j}$ and the bound (13).

$$|B_{\hat{j}}(x)| \quad = \quad |B_{\hat{j}}(x)| 1_{\{\hat{j} \leq j_s\}} + |B_{\hat{j}}(x)| 1_{\{\hat{j} \geq j_s\}}$$

$$
\begin{aligned}
&\leq \sum_{\hat{j}}^{j_s}\sum_k |\beta_{j,k}\psi_{j,k}(x)| 1_{\{|\hat{\beta}_{j,k}|\leq T\sqrt{\frac{\log n}{n}}\}} + 2|B_{j_s}(x)| \\
&\leq \sum_{\hat{j}}^{j_s}\sum_k |\beta_{j,k}\psi_{j,k}(x)| 1_{\{|\beta_{j,k}|\leq 2T\sqrt{\frac{\log n}{n}}\}} + 2|B_{j_s}(x)| \\
&\quad + \sum_{j_0}^{j_s}\sum_k |\beta_{j,k}\psi_{j,k}(x)| 1_{\{|\hat{\beta}_{j,k}-\beta_{j,k}|\geq T\sqrt{\frac{\log n}{n}}\}} \\
&\leq C\left(\frac{2^{j_s}\log n}{n}\right)^{1/2} + Z.
\end{aligned}
$$

Using Markov Inequality and Lemma 2, we prove that the variable

$$
Z = \sum_{j_0}^{j_s}\sum_k |\beta_{j,k}\psi_{j,k}(x)| 1_{\{|\hat{\beta}_{j,k}-\beta_{j,k}|\geq T\sqrt{\frac{\log n}{n}}\}}
$$

satisfies

$$
\begin{aligned}
P(|Z|\geq n^{-\delta_1}) &\leq n^{\delta_1} E\sum_{j_0}^{j_s}\sum_k |\beta_{j,k}\psi_{j,k}(x)| 1_{\{|\hat{\beta}_{j,k}-\beta_{j,k}|\geq T\sqrt{\frac{\log n}{n}}\}} \\
&\leq C\sum_{j_0}^{j_s} 2^{-js} n^{\delta_1} P\left(|\hat{\beta}_{j,k}-\beta_{j,k}|\geq T\sqrt{\frac{\log n}{n}}\right) \\
&\leq O(n^{-\delta_2})
\end{aligned}
$$

for any $\delta_1 > 0$ and any $\delta_2 \leq \gamma(T) - \delta_1$ where the functional $\gamma(.)$ is defined in (8) or (9). Next, we give a lower bound for the variance term. Let us put $\tau_n = (2N+1)\rho_n$.

$$
\begin{aligned}
V_{\hat{j}}(x) &= d_{\hat{j}}^p \sum_k v_{\hat{j},k}|\phi_{\hat{j},k}(x)| \left(1_{\{\hat{j}+\tau_n\geq j_s\}} + 1_{\{\hat{j}+\tau_n\leq j_s\}}\right) \\
&\geq \left(\frac{2^{(j_s-\tau_n)}(\log 2^{(j_s-\tau_n)})^p}{n}\right)^{1/2} 1_{\{\hat{j}\geq j_s-\tau_n\}} \\
&\geq \left(\frac{2^{j_s}\log n}{n}\right)^{1/2}\left(\frac{2^{-\tau_n}(\log 2^{(j_s-\tau_n)})^p}{\log n}\right)^{1/2} - \tilde{Z}
\end{aligned}
$$

where

$$
\tilde{Z} = \left(\frac{2^{(j_s-\tau_n)}(\log 2^{(j_s-\tau_n)})^p}{n}\right)^{1/2} 1_{\{\hat{j}+\tau_n\leq j_s\}}.
$$

Using Markov Inequality and Lemma 2, we prove that $\tilde{Z}$ satisfies

$$
\begin{aligned}
P(|\tilde{Z}|\geq n^{-\tilde{\delta}_1}) &\leq n^{\tilde{\delta}_1}\left(\frac{2^{j_s}(\log n)^p}{n}\right)^{1/2} P(\hat{j}+\tau_n\leq j_s) \\
&\leq O(n^{-\tilde{\delta}_2})
\end{aligned}
$$

for any $\tilde{\delta}_1 > 0$ and any $\tilde{\delta}_2 \leq \gamma(T - M_1 b^{-1}) \wedge \gamma(M_2 b^{-1} - T) + 1/2 - 1/(2+4s) - \tilde{\delta}_1$. If p is chosen such that $p \geq 1 + \tau_n(\log\log n)^{-1}$ then $|B_{\hat{j}}(x)| \leq aV_{\hat{j}}(x) + \tilde{Z} - Z$ for any constant $0 < a$. Choosing now $\delta_1 = \tilde{\delta}_1 = 1/2$ and $\delta_2 = \tilde{\delta}_2 > 0$ (which is possible because of the constraint on the threshold constant T), we finish the proof as previously.

5.4 Proof of Theorem 3

Let us recall that j_s is defined by $2^{j_s} = \left(b(a)\,\frac{n}{\log n}\right)^{1/1+2s}$ for $b(a)$ given in (7). In the regression case, we have, for any j varying in $\{j_s, \dots j_\infty\}$

$$\forall x \in [A,B], \quad |I_j(x,1)| = 2d_j \sum_k v_{j,k}|\phi_{j,k}(x)| \le 0 \left(\log(2^j)\frac{2^j}{n}\right)^{1/2}$$

which is minimum in order for $j = j_s$ and then

$$\inf_{j\in\{j_s,\dots j_\infty\}} \sup_{x\in[A,B]} |I_j(x,1)| \le O\left(\frac{n}{\log n}\right)^{-\frac{s}{1+2s}}.$$

In the density case, we have, for any $j \ge j_s$ and $p \ge 1$

$$\forall x \in [A,B], \quad |I_j(x,p)| = Z(j) + \tilde{Z}(j)$$

where

$$\tilde{Z}(j) \quad = \quad 2d_j^p \sum_k \sigma_{j,k}|\phi_{j,k}(x)| \;\text{ and }\; Z(j) = 2d_j^p \sum_k (\hat{\sigma}_{j,k} - \sigma_{j,k})|\phi_{j,k}(x)|. \qquad (15)$$

Let us take $p = 1$. We observe that, for some $C > 0$,

$$P\left(\forall x \in [A,B], \quad Z(j) \ge C\left(\frac{2^j \log 2^j}{n}\right)^{1/2}\right) \quad \le \quad P\left(\sup_k |\hat{\sigma}_{j,k} - \sigma_{j,k}| \ge C(2N-1)^{-1}\|\phi\|_\infty^{-1}\frac{1}{\sqrt{n}}\right)$$

Using now Lemma 6 and since $j \le j_\infty$, we obtain for C large enough

$$P\left(\inf_{x\in[A,B]} |I_j(x,1)| \ge C\left(\log 2^j\,\frac{2^j}{n}\right)^{1/2}\right) \quad \le \quad O(n^{-1}),$$

and then, the shortest confidence interval is $I_{j_s}(x,1)$ which diameter is of the order of $(\frac{n}{\log n})^{-\frac{s}{1+2s}}$.

5.5 Proof of Theorem 4

Let $v_n > 0$ and $C > 0$. We recall that j_s is defined by $2^{j_s} = \left(b(a)\,\frac{n}{\log n}\right)^{1/1+2s}$ for $b(a)$ given in (7). Let c be some positive constant which does not depend on n or on C and which may change from line to line. In the regression model, we have

$$\begin{aligned}
P\left(\inf_{x\in[A,B]} |I_{\hat{j}}(x,p)| \ge C\,v_n\right) \quad &\le \quad (C\,v_n)^{-1} \sup_{x\in[A,B]} \left(E|I_{\hat{j}}(x,p)|1_{\{\hat{j}\le j_s\}} + E|I_{\hat{j}}(x,p)|1_{\{\hat{j}>j_s\}}\right) \\
&\le \quad (C\,v_n)^{-1} c\left(\left(\log 2^{j_s}\right)^{p/2}\left(\frac{2^{j_s}}{n}\right)^{1/2} + \left(\log 2^{j_\infty}\right)^{p/2}\left(\frac{2^{j_\infty}}{n}\right)^{1/2} P(\hat{j} > j_s)\right)
\end{aligned}$$

Using Lemma 3, there exists some $\delta \le \gamma(T - M_1/b)$ (the functional $\gamma(.)$ is defined in (8)) such that

$$P\left(\inf_{x\in[A,B]} |I_{\hat{j}}(x,p)| \ge C\,v_n\right) \quad \le \quad \frac{c}{C}\left(1 + \left(\frac{n}{\log n}\right)^{\frac{s}{1+2s}} n^{-\text{elta}}\right).$$

as soon as $v_n = (\log n)^{\frac{p-1}{2}}(\frac{n}{\log n})^{-\frac{s}{1+2s}}$. Because of our constraint on the threshold constant T, it is possible to choose δ such that

$$\frac{s}{1+2s} \leq \delta \leq \gamma(T - \frac{M_1}{b(a)})$$

and we deduce that

$$\lim_{C,n \longrightarrow +\infty} P\left(\inf_{x\in[A,B]} |I_{\hat{j}}(x,p)| \geq C\, v_n\right) = 0.$$

In the density model, we split $|I_{\hat{j}}|$ in the same way as in (15)

$$\forall x \in [A,B], \quad |I_{\hat{j}}(x,p)| = Z(\hat{j})(1_{\{\hat{j}+\tau_n \leq j_s\}} + 1_{\{\hat{j}+\tau_n \geq j_s\}}) + \tilde{Z}(\hat{j})(1_{\{\hat{j}\geq j_s\}} + 1_{\{\hat{j}\leq j_s\}}).$$

Obviously, we bound the last term

$$\tilde{Z}(\hat{j})1_{\{\hat{j}\leq j_s\}} \quad \leq \quad (2N+1)\|\phi\|_\infty \|f\|_\infty^{1/2} \left(\frac{2^{j_s}(\log 2^{j_s})^p}{n}\right)^{1/2} \tag{16}$$

and then, v_n is again of the order of $(\log n)^{\frac{p-1}{2}}(\frac{n}{\log n})^{-\frac{s}{1+2s}}$. Using Markov Inequality and Lemma 3, we get

$$P(\forall x \in [A,B],\ \tilde{Z}(\hat{j})1_{\{\hat{j}\geq j_s\}} \geq Cv_n) \leq (Cv_n)^{-1}(2N+1)\|\phi\|_\infty \|f\|_\infty^{1/2}\left(\frac{2^{j_\infty}(\log 2^{j_\infty})^p}{n}\right)^{1/2} n^{-\tilde{\delta}}, \tag{17}$$

for some $\tilde{\delta} \leq \gamma(T - M_1/b(a))$ (the functional $\gamma(.)$ is defined in (9)). In the same way, using Lemma 4 and bounding $\hat{\sigma}_{j,k}$ by $2^{j/2}n^{-1/2}\|\phi\|_\infty$, we get

$$P(\forall x \in [A,B], Z(\hat{j})1_{\{\hat{j}+\tau_n \leq j_s\}} \geq Cv_n) \leq (Cv_n)^{-1}4(2N+1)\|\phi\|_\infty^{3/2}(\log 2^{j_\infty})^{p/2}\frac{2^{j_\infty}}{n^{1/2}}n^{-\delta}, \tag{18}$$

for some $\delta \leq \gamma(T - M_1/b(a)) \wedge \gamma(M_2/b(a) - T)$. For the first term, we have

$$P\left(\forall x \in [A,B],\ Z(\hat{j})1_{\{\hat{j}+\tau_n > j_s\}} \geq Cv_n\right) \quad \leq \quad \sum_{j=j_s}^{j_\infty} P\left(\sup_k |\hat{\sigma}_{j,k} - \sigma_{j,k}| \geq C\left(\log 2^j \frac{2^j}{n^2}\right)^{1/2} \epsilon_j\right)$$

for $\epsilon_j = v_n \left(\log 2^{j_s-\tau_n}\right)^{-p/2} 2^{-(j_s-\tau_n)/2}\left(\log 2^j \frac{2^j}{n^2}\right)^{-1/2}$. Let us remark that the assumption $s > 1/2$ implies that $\epsilon_j \geq 1$ for any $j \leq j_\infty$. It follows from Lemma 6 that

$$P(\forall x \in [A,B],\ Z(\hat{j})1_{\{\hat{j}+\tau_n \leq j_s\}} \geq Cv_n) \quad \leq \quad O\left(\frac{\log n}{n}\right). \tag{19}$$

Since we have

$$\begin{aligned} P\left(\inf_{x\in[A,B]} |I_{\hat{j}}(x,p)| \geq C\, v_n\right) \quad &\leq \quad P\left(\inf_{x\in[A,B]} Z(\hat{j})1_{\{\hat{j}+\tau_n\leq j_s\}} \geq 3C\, v_n - \tilde{Z}(\hat{j})1_{\{\hat{j}\leq j_s\}}\right) \\ &\quad + P\left(\inf_{x\in[A,B]} \tilde{Z}(\hat{j})1_{\{\hat{j}\geq j_s\}} \geq C\, v_n\right) + P\left(\inf_{x\in[A,B]} Z(\hat{j})1_{\{\hat{j}\geq j_s\}} \geq C\, v_n\right), \end{aligned}$$

we combine (16), (17), (18) and (19) and we obtain the result if we can choose

$$\frac{s}{1+2s} - \frac{1}{4} \leq \tilde{\delta} \leq \gamma(\frac{M_2}{b(a)} - T) \quad \text{and} \quad \frac{s}{1+2s} - \frac{1}{2} \leq \delta \leq \gamma(T - \frac{M_1}{b(a)})$$

which is possible under the constraint on T.

References

[1] Adak, S., Johnstone, I., Marron, J., Neumann, M. and Patil, P. Exact risk analysis of wavelet regression. *J. Comput. Graphical Stat.* **7** (1998) 278-309.

[2] Bowman, A.W. and Härdle, W. Bootstrapping in Nonparametric Regression: Local Adaptive Smoothing and Confidence Bands. *J.A.S.A.* (1988) 102-110.

[3] Daubechies, I. *Ten Lectures on Wavelets. SIAM: Philadelphia.* (1992)

[4] Donoho, D.L. De-noising by soft thresholding. *I.E.E.E. Trans. on Inf. Theory* **41** (1995) 613-627.

[5] Faraway, T. Bootstrap selection of bandwidth and confidence bands for nonparametric regression. *J.Statist. Comput. Simulation.* **37** (1990) 37-44.

[6] Feller, W. *An introduction to Probability Theory and Its Applications.* Wiley. (1966)

[7] Hall, P. On bootstrap confidence intervals in nonparametric regression. *Ann. Statist.* **20** (1992a) 695-711.

[8] Hall, P. Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.* **20** (1992b) 675-694.

[9] Hall, P., Titterington, D., On confidence bands in nonparametric density estimation and regression. *Journal of Multivariate Analysis.* **27** (1988) 228-254.

[10] Härdle, W., Marron, J.S. Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* (1990) 778-796.

[11] Hoffmann, M., Lepskii, O.V. Random rates in anisotropic regression. *Ann. Statist.* (2002) 325-396.

[12] Iouditski, A., Lacroix-Lambert, S. Confidence Intervals for Adaptive Regression Estimation on the Besov. *Preprint RR 3643 INRIA Rhones Alpes.* (2000).

[13] Lepskii, O.V. Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36** (1991) 682-697.

[14] Loader, C., Sun, J. Simultaneous confidence bands for linear regression and smoothing. *Ann. Statist.* **22** (1994) 1328-1345.

[15] Neumann, M. Automatic bandwidth choice and confidence intervals in nonparametric regression. *Ann. Statist.* **23** (1995) 1937-1959.

[16] Petrov, V. V. *Limit Theorems of Probability* Theory, Clarendon Press, Oxford (1995).

[17] Picard, D., Tribouley, K. Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.* **28** (2000) 298-335.

[18] Tribouley, K. Adaptive confidence interval for the density. *SJS.* To appear (2003).