Adaptive confidence bands in the nonparametric fixed design regression model

Pierre-Yves Massé, William Meiniel

École Normale Supérieure de Cachan

July 20, 2012

Abstract

In this note, we consider the problem of existence of adaptive confidence bands in the fixed design regression model, adapting ideas in Hoffmann and Nickl [10] to the present case. In the course of the proof, we show that sup-norm adaptive estimators exist as well in regression.

1 Introduction

We observe random variables Y_i's and assume that, for $n \in \mathbb{N}$,

$$Y_i = f(x_i) + \varepsilon_i, \quad 1 \le i \le n \tag{1}$$

where $x_i = \frac{i}{n}$, $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ are independent and identically distributed random variables defined on $(\Omega, \mathcal{F}, P)$, and f is the unknown regression function. We further assume the variance σ^2 is known. Our aim is to reconstruct f from the sample.

Let us fist define the parameter space. We assume f belongs to some Hölder space. The Hölder space $\mathcal{C}^t$ for $0 < t \le 1$ is the space of continuous functions f on $[0, 1]$ such that

$$\|f\|_{\mathcal{C}^t} = \|f\|_\infty + \sup_{x \ne y} \frac{|f(x) - f(y)|}{|x - y|^t} < \infty \quad \text{if} \ \ 0 < t < 1$$

$$\|f\|_{\mathcal{C}^t} = \sup_{x \ne y} \frac{|f(x+y) + f(x-y) - 2f(x)|}{|y|} < \infty \quad \text{if} \ \ t = 1.$$

If $t > 1$, $\mathcal{C}^t$ is the space of functions such that the $\lfloor t \rfloor$-th derivative of f exists, is continuous and belongs to $\mathcal{C}^{t-\lfloor t \rfloor}$ if t is not an integer, and such that the t-th derivative of f exists, is continuous and is in $\mathcal{C}^1$ otherwise. Define then $\|f\|_{\mathcal{C}^t} = \|f^{\lfloor t \rfloor}\|_{\mathcal{C}^{t-\lfloor t \rfloor}}$. Therefore, for all t, $\|f\|_{\mathcal{C}^t}$ is a norm of f.

Working with this definition may prove difficult, and we thus use the wavelet basis characterisation of $\mathcal{C}^t$: whether or not f belongs to it depends on the size of the coefficients of its decomposition over this basis. We use Daubechies wavelets on the unit interval, as in [10]. Denote, accordingly, by ϕ and ψ the scaling functions, $\phi_m = \phi(\cdot - m)$, $\psi_{jm} = 2^{j/2}\psi(2^j \cdot -m)$ and, for $< \cdot, \cdot >$ the usual inner product in $L^2([0, 1])$,

$$\|f\|_{t,\infty} = \max\left(\sup_m | < \phi_m, f > |, \sup_{j,m} 2^{j(s+1/2)}| < \psi_{jm}, f > | \right).$$

Now, Theorem 4.4 in [5] gives the equivalence : for $t > 0$, $\mathcal{C}^t$ is the set of continuous functions on $[0, 1]$ such that $\|f\|_{t,\infty} < \infty$. Moreover, $\|f\|_{t,\infty}$ is a norm equivalent to $\|f\|_{\mathcal{C}^t}$.

For fixed $B > 0$, define

$$\Sigma(t) = \left\{ f : \mathbb{R} \to \mathbb{R} \;\middle|\; ||f||_{t,\infty} \leq B \right\}.$$

One has typically no knowledge of the regularity of f. But estimators which achieve the optimal risk in sup-norm loss for whatever t have been constructed. They are called adaptive, and exist in the density and the white noise cases, as shown respectively by Giné and Nickl [8] and Goldenshluger and Lepski [9]. Now, Brown and Low [2] have proven that the white noise and regression cases are asymptotically equivalent, and therefore one may legitimately expect adaptive estimators to exist as well in the latter. For the sake of reference we prove it in section 3.2.2, that is, we prove the following theorem, where

$$r_n(t) = \left(\frac{\log n}{n} \right)^{t/(2t+1)} \qquad t > 0.$$

Theorem 1. *Let $Y_1, \ldots, Y_n$ verify $Y_i = f(i/n) + \varepsilon_i$ where the ε_i's are i.i.d. $\mathcal{N}(0, \sigma^2)$.*
Then, for every integer $l > 0$, there exists an estimator $\hat{f}_n(x) := \hat{f}_n(x, Y_1, \ldots, Y_n, l)$ and an integer n_0 such that, for every t, $0 < t \leq l$, some constant $D(B, l)$ and every $n \geq n_0$, we have

$$\sup_{f \in \Sigma(t)} E_f \|\hat{f}_n - f\|_\infty \leq D(B, l) r_n(t).$$

We use the following notations. For any random variable $X : \Omega \to \mathbb{R}^n$ measurable with respect to $\sigma(Y_1, ..., Y_n)$ and every function $f \in \Sigma(r)$, $E_f(X)$ is the expectation of X when the function in Equation 1 is f. For all $f \in \Sigma(r)$, and all $A \in \mathcal{B}(\mathbb{R}^n)$, $P_f\left((Y_1, \ldots, Y_n) \in A \right) = E_f[1_A(Y_1, \ldots, Y_n)]$.

Thanks to these estimators, one may then want to construct confidence bands for f, that is data-driven sets which cover f at all points simultaneously.

Definition 1. *A confidence band is a family of random intervals*

$$C_n = C(Y_1, \ldots, Y_n) = \left\{ [c_n(y), c'_n(y)] \right\}_{y \in [0,1]}.$$

Define the diameter of C_n, $|C_n| = \sup_{y \in [0,1]} |c'_n(y) - c_n(y)|$.

For instance, Claeskens and Van Keilegom [4] construct such bands for the function and its derivatives. However, just as with estimators, one would like the diameter of the band to be optimal for all t, that is :

$$\forall t > 0, \ \sup_{f \in \Sigma(t)} E_f |C_n| \leq L r_n(t)$$

for some constant L independent of t. Here, we restrict ourselves to the case where f belongs to $\Sigma(s)$ or $\Sigma(r)$ for $0 < r < s$ given. Now, results by Low [13], by Genovese and Wasserman [6], and also our findings below, imply that to construct such sets is not possible for $\Sigma(s) \cup \Sigma(r)$. To circumvent this problem, two paths have recently been considered. Both restrict the parameter space, so that such negative results no longer apply, but at the same time the parameter space remains as large as possible. The one studied by Giné and Nickl [7] in the density model consists in removing permanently some functions from the model, according to a self-similarity condition. Bull [3] then extends their results, in particular to the white noise model. Hoffmann and Nickl [10] on the other hand let the parameter space evolve with n. In this article, we follow their approach.

Denote $d(f, \Sigma) = \inf_{g \in \Sigma} \|f - g\|_\infty$ the distance which derives from the sup-norm. For $\rho_n > 0$, define

$$\tilde{\Sigma}(r, \rho_n) = \left\{ f \in \Sigma(r) \ \middle| \ d(f, \Sigma(s)) \geq \rho_n \right\}.$$

Consider the model

$$\mathcal{P}_n = \Sigma(s) \cup \tilde{\Sigma}(r, \rho_n),$$

for $0 < r < s$.

Definition 2 (Honesty)**.** *The confidence band C_n is called asymptotically honest with level $0 < \alpha < 1$ for $\mathcal{P}_n$ if it satisfies the asymptotic coverage inequality*

$$\liminf_n \ \inf_{f \in \mathcal{P}_n} \ P_f(f \in C_n) \geq 1 - \alpha.$$

Definition 3 (Adaptivity)**.** *The confidence band C_n is called adaptive over $\mathcal{P}_n$ if there exists a constant L such that for every $n \in \mathbb{N}$,*

$$\sup_{f \in \Sigma(s)} E_f |C_n| \leq L r_n(s), \qquad \sup_{f \in \tilde{\Sigma}(r, \rho_n)} E_f |C_n| \leq L r_n(r).$$

One is in turn interested in finding the smallest possible seqence ρ_n. Hoffmann and Nickl provide a lower bound for ρ_n which is sharp as they are able to construct confidence bands for ρ_n of the order of this lower bound. The present article adapts their proofs to the regression model.

To compute the lower bound, Hoffmann and Nickl reduce to a testing problem whose minimax rate of testing is ρ_n, and we proceed likewise. To construct confidence bands, we use adaptive estimators, which existence we prove, as well as a concentration inequality for a certain gaussian process.

We first give the main result, and carry on with its proof. It relies on some additional results, which are discussed in a subsequent section.

2 Existence of adaptive and honest confidence bands

2.1 The main result

We may now give a precise statement of the existence of confidence bands.

Theorem 2. *Let $s > r > 0$ and $B > 0$ be given.*

- *Assume $r > 1/2$ and let $\alpha < 1/2$. Suppose C_n is a confidence band that is honest with level α over $\mathcal{P}_n$, and adaptive. Then necessarily,*

$$\liminf_n \frac{\rho_n}{r_n(r)} > 0. \tag{2}$$

- *Let $0 < \alpha < 1$. There exists a sequence ρ_n satisfying*

$$\limsup_n \frac{\rho_n}{r_n(r)} < \infty \tag{3}$$

and a confidence band C_n that is honest with level α over $\mathcal{P}_n$, and adaptive.

The first part of the theorem means that the crown removed around $\Sigma(s)$ has to be large enough. Indeed, there are functions in $\Sigma(r)$ we cannot distinguish (as Hoffmann and Nickl [10] explain) from those of $\Sigma(s)$, and they must thus be excluded from the model, as detailed in the next section. As explained in [10], the sets constructed cannot be easily computed. In the next two sections we prove this theorem, using auxiliary results we discuss later.

2.2 Proof of the lower bound

In all what follows, $(j_n^*)_{n \geq 0}$ is a positive sequence such that

$$2^{-j_n^* r} \simeq \left(\frac{\log n}{n} \right)^{\frac{r}{2r+1}},$$

where $\simeq$ means that the two sequences involved are dominated by each other. Before starting the proof, we define a certain set we use in it, then precise the notations we use for tests.

For j big enough and any m, ψ_{jm} is supported in the interior of $[0, 1]$. Furthermore, since ψ has a compact support, denote $[a, b] \supset supp(\psi)$ and $c_0^{-1} = \lceil b - a \rceil$. Now, the

$$\psi_{j, mc_0^{-1}}, \ m = 1, \ldots, c_0(2^j - 1)$$

have disjoint supports. Note that c_0 depends only on ψ.
Let $(j_n)_{n \geq 0}$ be a positive sequence tending to ∞. For all $1 \leq m \leq c_0(2^{j_n} - 1)$, let

$$f_m = 2^{-j_n(r+1/2)} \psi_{j_n, c_0^{-1} m}.$$

(f_m depends on n but since it is apparent that the set $\mathcal{M}_n$ depends on n, we do not repeat it for f_m so as to simplify notations.) Define

$$\mathcal{M}_n(j_n) = \{f_1, ..., f_{c_0(2^{j_n} - 1)}\}.$$

$\mathcal{M}_n(j_n)$ is a subset of $\Sigma(r)$ and, under suitable choices of j_n and ρ_n precised below, it is in fact a subset of $\tilde{\Sigma}(r, \rho_n)$.

The existence of confidence bands for $\mathcal{P}_n$ is related to the possibility of testing accurately the alternative :

$$H_0 : f = 0 \quad \text{against} \quad H_1 : f \in \mathcal{M}_n(j_n). \tag{4}$$

A test T_n based on a sample of size n is any

$$T_n : \Omega \to \{0, 1\}$$

which is measurable with respect to $\sigma(Y_1, ..., Y_n)$, the σ-algebra generated by $Y_1, ..., Y_n$. We denote by $\mathcal{T}_n$ the set of all tests T_n.

To assess the quality of the test T_n designed to solve the testing problem 4, we use the sum of the errors of first and second type :

$$r(T_n, j_n) = E_{f_0}(T_n) + \sup_{f \in \mathcal{M}_n(j_n)} E_f(1 - T_n),$$

where $f_0 = 0$ identically.

We need the two following results. The first one shows that for $2^{-j_n r} \ll \left(\frac{\log n}{n} \right)^{\frac{r}{2r+1}}$, we cannot test efficiently f_0 againt $\mathcal{M}_n(j_n)$. Now, $2^{-j_n r}$ is intuitively the distance in the sup-norm between $\mathcal{M}_n(j_n)$ and $\Sigma(s)$. As a result, Proposition 1 means that close to $\Sigma(s)$, we find sets which cannot be statistically distinguished from it. The proof is in section 3.1.

Proposition 1. *Let $(j_n)_{n \geq 0}$ be any sequence satisfying $2^{-j_n r} = o(2^{-j_n^*})$ as $n \to \infty$. Then*

$$\liminf_{n \to \infty} \inf_{T_n \in \mathcal{T}_n} r(T_n, j_n) \geq 1.$$

Once we are given a confidence band, we can construct an obvious test to decide the testing problem 4. Let

$$T_n^0 = 1_{C_n \cap \mathcal{M}_n(j_n) \neq \emptyset}.$$

We accept f_0 if no f_m is in C_n and reject otherwise. If j_n satisfies certain conditions, T_n^0 is a relevant test for the testing problem 4. The problem is that it may be possible to find j_n such that both these conditions are satisfied and $2^{-j_n r} \ll \left(\frac{\log n}{n}\right)^{\frac{r}{2r+1}}$, leading to a contradiction. Lemma 1 indeed shows that if ρ_n goes to 0 too quickly, we may find j_n such that both $r(T_n^*, j_n) \to 0$ and the conditions of Proposition 1 are satisfied. To remedy this, the exclusion zone around $\Sigma(s)$ must contain all functions which prevent adaptation, and therefore ρ_n must be big enough.

Lemma 1. *Assume* $\lim_{n\to\infty} \frac{\rho_n}{r_n(r)} = 0$. *Then there exists* j_n *such that*

- $\mathcal{M}_n(j_n) \subset \tilde{\Sigma}(r, \rho_n)$, *for all* n *large enough, depending only on* B, r *and* s.

- $r_n(s)/2^{-j_n r} \to 0$ *as* $n \to \infty$.

- $2^{-j_n r} = o(2^{-j_n^* r})$ *as* $n \to \infty$.

Proof. Let $j_n = \min(\lfloor -\frac{\log 2\rho_n}{r \log 2} \rfloor - 1, \lfloor -\frac{r_n(s) \log n}{r \log 2} \rfloor)$. Then

$$\max(2\rho_n, r_n(s) \log n) \leq 2^{-j_n r}.$$

First, for all $1 \leq m \leq 2^{j_n} - 1$,

$$d(f_m, \Sigma(s)) \geq 2^{-j_n r}(1 - B 2^{j_n(r-s)})$$

(as in Hoffmann and Nickl, in section 2.1, inequality (2.5)). Since ρ_n is less than the right-hand side for all n large enough depending only on B, r and s, $\mathcal{M}_n(j_n)$ is included in $\tilde{\Sigma}(r, \rho_n)$ for all n large enough. Then,

$$\frac{r_n(s)}{2^{-j_n r}} \leq \frac{r_n(s)}{r_n(s) \log n} \to 0$$

as $n \to \infty$.

Finally, write $j_n = \min(\lfloor a_n \rfloor, \lfloor b_n \rfloor)$; $a_n, b_n \to \infty$ as $n \to \infty$. So, $\lfloor a_n \rfloor \sim a_n$, but

$$2^{-a_n r} = o(r_n(r)) \quad \text{and} \quad 2^{-\lfloor a_n \rfloor r} \leq 2^{-a_n r} 2^r \quad \text{so} \quad 2^{-\lfloor a_n \rfloor r} = o(2^{-j_n^* r}).$$

Likewise, $2^{-\lfloor b_n \rfloor r} = o(2^{-j_n^* r})$. So $2^{-j_n r} = o(2^{-j_n^* r})$.

$\square$

We now prove Proposition 2, which is equivalent to the first part of Theorem 2. We proceed by *reductio ad absurdum*. The main thing is to prove T_n^0 is consistent for the testing problem 4.

Proposition 2. *Assume* $\liminf_{n\to\infty} \frac{\rho_n}{r_n(r)} = 0$. *Then we cannot find an honest and adaptive confidence band of level* $\alpha < 1/2$ *for* $\mathcal{P}_n$.

Proof. Assume otherwise. Take a subsequence of $\frac{\rho_n}{r_n(r)}$ which tends towards 0, and still note n the subscript. Take j_n as in Lemma 1. We show first that $r(T_n^*, j_n) \to 0$ as $n \to \infty$.

$$
\begin{aligned}
E_{f_0}(T_n^0) &= P_{f_0}(C_n \cap \mathcal{M}(j_n) \neq \emptyset) \\
&= P_{f_0}\left(C_n \cap \mathcal{M}_n(j_n) \neq \emptyset, f_0 \in C_n\right) + P_{f_0}\left(C_n \cap \mathcal{M}_n(j_n) \neq \emptyset, f_0 \notin C_n\right) \\
&\leq P_{f_0}\left(|C_n| \geq d(\Sigma(s), \mathcal{M}_n(j_n))\right) + P(f_0 \notin C_n) \\
&\leq \frac{E_{f_0}(|C_n|)}{d(\Sigma(s), \mathcal{M}_n(j_n))} + \alpha + o(1)
\end{aligned}
$$

5

thanks to Markov's inequality and honesty of C_n.

So

$$E_{f_0}(T_n^0) \leq \frac{Lr_n(s)}{2^{-j_n r}(1-B)} 2^{j_n(r-s)}) + \alpha + o(1) = \alpha + o(1)$$

thanks to Lemma 1 (second point) and adaptivity of C_n.

For $f_m \in \mathcal{M}_n(j_n)$,

$$
\begin{aligned}
E_{f_m}(1 - T_n^0) &= P_{f_m}(C_n \cap \mathcal{M}_n(j_n) = \emptyset) \\
&\leq P_{f_m}(f_m \notin C_n) \\
&\leq \sup_{f \in \mathcal{P}_n} P_f(f \notin C_n)
\end{aligned}
$$

thanks to Lemma 1 (first point). So

$$\sup_{f_m \in \mathcal{M}_n(j_n)} E_{f_m}(T_n^0) \leq \sup_{f \in \mathcal{P}_n} P_f(f \notin C_n) \leq \alpha + o(1)$$

by honesty of C_n. Therefore we have

$$\limsup_{n \to \infty} r(T_n^0, j_n) \leq 2\alpha < 1$$

but

$$\liminf_{n \to \infty} \inf_{T_n \in \mathcal{T}_n} r(T_n, j_n) \geq 1,$$

thanks to Lemma 1 (third point) and Proposition 1, which is a contradiction. $\qquad \square$

2.3 Proof of upper bound

2.3.1 Construction of C_n

Let $\rho_n = \lambda r_n(r)$, with λ chosen below. We will estimate the function f to construct confidence bands. We need two different estimators. We use the local polynomial estimator of order l with $h = \left(\frac{\log n}{n}\right)^{\frac{1}{2r+1}}$, $f_n(h)$, which we note f_n in this section, defined in Theorem 3. Thanks to Equation 10, we know that for some $b > 0$, for all $f \in \Sigma(r)$,

$$\|E_f f_n - f\|_\infty \leq br_n(r) \tag{5}$$

$\hat{f}_n$ is the estimator given by Theorem 1. $\hat{f}_n$ is adaptive over $\Sigma(s) \cup \Sigma(r)$ with rate $r_n(\cdot)$. In other words, for $D = D(B, l)$,

$$\sup_{f \in \Sigma(t)} E_f \|\hat{f}_n - f\|_\infty \leq Dr_n(t) \qquad t = r, s. \tag{6}$$

We are now able to construct the confidence band C_n. The idea is to adapt the width of C_n with the likelihood that f be in $\Sigma(r)$ or in $\Sigma(s)$. Define

$$d_n = d(f_n, \Sigma(s)).$$

If d_n is greater than some constant times $r_n(r)$, Equation 5 shows that it is unlikely that f is in $\Sigma(s)$ (intuitively, $f_n \approx E_f f_n$), and thus we choose $\hat{f}_n \pm Lr_n(r)$. Elsewhere, we choose $\hat{f}_n \pm Lr_n(s)$. Accordingly, define C_n :

$$\hat{f}_n \pm Lr_n(r) \quad \text{if } d_n > \tau \quad \text{and} \quad \hat{f} \pm Lr_n(s) \quad \text{if } d_n \leq \tau, \tag{7}$$

where $\tau = \kappa r_n(r)$ and where κ and L are constants chosen below. Let us now prove that C_n is honest and adaptive.

2.3.2 Proof of honesty and adaptivity

We first prove that C_n is an honest confidence band for $\mathcal{P}_n$. Let $0 < \alpha < 1$. If $f \in \Sigma(s)$ we have, using adaptivity of $\hat{f}_n$ and Markov's inequality,

$$
\begin{aligned}
\inf_{f \in \Sigma(s)} P_f(f \in C_n) & \geq 1 - \sup_{f \in \Sigma(s)} P_f\left(\|\hat{f}_n - f\|_\infty > Lr_n(s)\right) \\
& \geq 1 - \frac{1}{Lr_n(s)} \sup_{f \in \Sigma(s)} E_f\|\hat{f}_n - f\|_\infty \\
& \geq 1 - \frac{D}{L}
\end{aligned}
$$

which is greater than $1 - \alpha$ for L large enough.

Now, if $f \in \tilde{\Sigma}(r, \rho_n)$, we have, using again Markov's inequality,

$$
\inf_{f \in \tilde{\Sigma}(r, \rho_n)} P_f(f \in C_n) \geq 1 - \frac{\sup_{f \in \tilde{\Sigma}(r, \rho_n)} E_f\|\hat{f}_n - f\|_\infty}{Lr_n(r)} - \sup_{f \in \tilde{\Sigma}(r, \rho_n)} P_f(d_n \leq \tau)
$$

and the first substracted term is smaller than $\alpha/2$ for L large enough. Now, $P_f(d_n \leq \tau)$ equals, for every $f \in \tilde{\Sigma}(r, \rho_n)$,

$$
\begin{aligned}
P_f\left(\inf_{g \in \Sigma(s)} \|f_n - g\|_\infty \leq \kappa r_n(r)\right) & \leq P_f\left(\inf_g \|f - g\|_\infty - \|f_n - E_f f_n\|_\infty - \|E_f f_n - f\|_\infty \leq \kappa r_n(r)\right) \\
& \leq P_f\left(\rho_n - \|E_f f_n - f\|_\infty - \kappa r_n(r) \leq \|f_n - E_f f_n\|_\infty\right) \\
& \leq P_f\left(\|f_n - E_f f_n\|_\infty \geq (\lambda - \kappa - b)r_n(r)\right) \\
& \leq c_1 n^{-\gamma_1} = o(1)
\end{aligned}
$$

thanks to Corollary 1, by choosing λ large enough. This completes the proof of honesty of the band. Let us now deal with adaptivity.

By definition of C_n we have
$$
|C_n| \leq Lr_n(r)
$$
so the case $f \in \tilde{\Sigma}(r, \rho_n)$ is immediate. If $f \in \Sigma(s)$ then,

$$
\begin{aligned}
E_f|C_n| & \leq Lr_n(r)P_f(d_n > \tau) + Lr_n(s) \\
& \leq Lr_n(r)P_f\left(\inf_{g \in \Sigma(s)} \|f_n - g\|_\infty > \kappa r_n(r)\right) + Lr_n(s) \\
& \leq Lr_n(r)P_f\left(\|f_n - f\|_\infty > \kappa r_n(r)\right) + Lr_n(s) \\
& \leq Lr_n(r)P_f\left(\|f_n - E_f f_n\|_\infty > (\kappa - b)r_n(r)\right) + Lr_n(s) \\
& \leq Lr_n(r)c_2 n^{-\gamma_2} + Lr_n(s) = O(r_n(s))
\end{aligned}
$$

thanks to Corollary 1, since γ_2 is sufficiently large if κ is chosen large enough.

3 Auxiliary results

The first section proves Proposition 1. It is the adaptation of the proof in Lepski and Tsybakov [12]. Lemma 2 is a standard procedure to lower bound the error ; Lemma 3 and Lemma 4 deal with the functions in $\mathcal{M}_n(j_n)$, and Lemma 5 uses them all to give the required bound, thanks to Bahr-Esseen's inequality [16]. The second section presents the estimators used to contruct the confidence bands and to prove the existence of adaptive estimators. The third section deals with the concentration inequality used in proving the bands are indeed honest and adaptive, and that the estimator we construct is indeed adaptive.

3.1 Proof of Proposition 1

Throughout this section, T_n is any test, $(j_n)_{n\geq 0}$ is a non negative real sequence such that $2^{-j_n r} = o(2^{-j_n^* r})$ as $n \to \infty$, and $M = c_0(2^{j_n} - 1)$. Recall the definition of $\mathcal{M}_n(j_n)$ and that $f_0 = 0$.

Lemma 2. *For all $0 < \eta < 1$,*

$$r(T_n, j_n) \geq (1 - \eta)P\left(\frac{1}{M}\sum_{m=1}^{M}\xi_m \geq 1 - \eta\right),$$

where $\xi_m = \frac{dP_m}{dP_0}(\varepsilon_1, ..., \varepsilon_n)$ and $P_m = P_{f_m}$.

Proof.

$$
\begin{aligned}
r(T_n, j_n) &= E_{f_0}(T_n) + \sup_{f \in \mathcal{M}_n(j_n)} E_f(1 - T_n) \\
&\geq E_{f_0}(T_n) + \frac{1}{M}\sum_{m=1}^{M} E_{f_m}(1 - T_n) \\
&\geq E_{f_0}(T_n) + E_{f_0}((1 - T_n)Z_n)
\end{aligned}
$$

where

$$Z_n = \frac{1}{M}\sum_{m=1}^{M}\frac{dP_m}{dP_0}(Y_1, ..., Y_n).$$

Let $0 < \eta < 1$.

$$
\begin{aligned}
\text{If } T_n = 1 \quad \text{then} \quad &T_n + (1 - T_n)Z_n = 1 \geq 1 - \eta \\
\text{else} \quad &T_n + (1 - T_n)Z_n = Z_n.
\end{aligned}
$$

So,

$$
\begin{aligned}
T_n + (1 - T_n)Z_n &\geq (T_n + (1 - T_n)Z_n)1_{Z_n \geq 1 - \eta} \\
&\geq (1 - \eta)1_{Z_n \geq 1 - \eta}.
\end{aligned}
$$

As a result,

$$
\begin{aligned}
r_n(T_n, j_n) &\geq (1 - \eta)P_{f_0}(Z_n \geq 1 - \eta) \\
&= (1 - \eta)P\left(\frac{1}{M}\sum_{m=1}^{M}\frac{dP_m}{dP_0}(\varepsilon_1, ..., \varepsilon_n) \geq 1 - \eta\right).
\end{aligned}
$$

$\square$

8

We want to show that $P\left(\frac{1}{M}\sum_{m=1}^{M}\xi_m \geq 1-\eta\right) \to 1$ as $n \to \infty$. It will be sufficient to conclude because this quantity is independent of T_n. To do this, note that

$$P\left(\frac{1}{M}\sum_{m=1}^{M}\xi_m \geq 1-\eta\right) = 1 - P\left(\frac{1}{M}\sum_{m=1}^{M}\xi_m < 1-\eta\right).$$

Now,

$$P\left(\frac{1}{M}\sum_{m=1}^{M}\xi_m < 1-\eta\right) = P\left(\frac{1}{M}\sum_{m=1}^{M}\xi_m - 1 < -\eta\right)$$

$$\leq P\left(\left|\frac{1}{M}\sum_{m=1}^{M}\xi_m - 1\right| > \eta\right)$$

and we therefore want to prove that the last quantity tends towards 0.

Lemma 3. *For $1 \leq m \leq M$,*

$$\xi_m = \exp\left(\frac{\alpha_m}{\sigma}\zeta_m - \frac{2\sigma^2}{\alpha_m^2}\right),$$

with $\zeta_m \sim \mathcal{N}(0,1)$ i.i.d. and

$$\alpha_m^2 = \sum_{i=1}^{n}(f_m(x_i))^2.$$

(As for the f_m's, we do not make explicit the dependence on n of α_m.)

Proof.

$$\xi_m = \prod_{i=1}^{n}\exp\left(-\frac{(\varepsilon_i - f_m(x_i))^2 - \varepsilon_i^2}{2\sigma^2}\right)$$

$$= \prod_{i=1}^{n}\exp\left(\frac{f_m(x_i)(2\varepsilon_i - f_m(x_i))}{2\sigma^2}\right)$$

$$= \exp\left(\frac{1}{2\sigma^2}\sum_{i=1}^{n}f_m(x_i)(2\varepsilon_i - f_m(x_i))\right)$$

$$= \exp\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}f_m(x_i)\varepsilon_i - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(f_m(x_i))^2\right).$$

Now, define

$$\zeta_m = \frac{1}{\alpha_m\sigma}\sum_{i=1}^{n}f_m(x_i)\varepsilon_i.$$

Therefore,

$$\xi_m = \exp(\frac{\alpha_m}{\sigma}\zeta_m - \frac{2\sigma^2}{\alpha_m^2}).$$

ζ_m is a gaussian random variable as a linear combination of independent gaussian random variables, and straightforward computations show that $\zeta_m \sim \mathcal{N}(0,1)$.

To prove that the ζ_m's are independent, we first note that the vector $\zeta_1, ..., \zeta_M$ is a gaussian vector. Indeed, any linear combination of its coordinates is a linear combination of the ε_i's and consequently a gaussian random variable as above. Thus, it is sufficient to prove that the covariances all equal 0.

But, for $m \neq m'$,

$$Cov(\zeta_m, \zeta_{m'}) = (\sigma^2\alpha_m\alpha_{m'})^{-1}\sum_{i=1}^{n}f_m(x_i)f_{m'}(x_i) = 0$$

since the f_m's supports are not overlapping. $\qquad\square$

Lemma 4. *For all* $1 \leq m \leq M$,

$$\alpha_m^2 = n2^{-j_n(2r+1)} + O(2^{j_n(1-2r)})$$

with the last term independent of m.

Proof. We drop the n subscript in j_n so as to simplify notations. Recall $supp(\psi_{jm}) \subset [0,1]$. First,

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} 2^j \psi^2(2^j x_i - m) - \int_{\mathbb{R}} \psi^2(t)\,dt &= \frac{1}{n}\sum_{i=1}^{n} 2^j \psi^2(2^j x_i - m) - 2^j \int_{[0,1]} \psi^2(2^j t - m)\,dt \\
&= 2^j \sum_{i=1}^{n} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \left(\psi^2(2^j x_i - m) - \psi^2(2^j t - m) \right)\,dt \\
&= 2^j \sum_{i=1}^{n} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \Big(\psi(2^j x_i - m) - \psi(2^j t - m) \Big) \\
&\qquad\qquad\qquad \Big(\psi(2^j x_i - m) + \psi(2^j t - m) \Big)\,dt \\
&\leq 2^{2j} \frac{||\psi'||_\infty}{n} 2||\psi||_\infty = K\frac{2^{2j}}{n}.
\end{aligned}
$$

Here we use the mean value theorem and the fact that for all $i \in [1,n]$ and $t \in [\frac{i-1}{n}, \frac{i}{n}]$, $|t - x_i| \leq \frac{1}{n}$.

Therefore,

$$\sum_{i=1}^{n} 2^j \psi^2(2^j x_i - m) = n \int_{\mathbb{R}} \psi^2(t)\,dt + O(2^{2j})$$

and recall $\int_{\mathbb{R}} \psi^2(t)\,dt = 1$.

Now,

$$
\begin{aligned}
\alpha_m^2 &= 2^{-j(2r+1)} \sum_{i=1}^{n} \psi_{jm}^2(x_i) \\
&= 2^{-j(2r+1)} \sum_{i=1}^{n} 2^j \psi^2(2^j x_i - m) \\
&= 2^{-j(2r+1)} \left(n + O\left(2^{2j}\right) \right) \\
&= n2^{-j(2r+1)} + O\left(2^{j(1-2r)}\right).
\end{aligned}
$$

$\square$

Lemma 5. *Let* $0 < v < 1$.

$$P\left(\left| \frac{1}{M}\sum_{m=1}^{M} \xi_m - 1 \right| > \eta \right) \leq c_1 \frac{\exp(c_2 n 2^{-j_n(2r+1)})}{(2^{j_n} - 1)^v}$$

where c_1 *and* c_2 *are positive constants.*

Proof. First note that for all $1 \leq m \leq M$, $E(\xi_m) = 1$, so we have, by Markov's inequality

$$P\left(\left|\frac{1}{M}\sum_{m=1}^{M}\xi_m - 1\right| > \eta\right) = P\left(\left|\frac{1}{M}\sum_{m=1}^{M}(\xi_m - E(\xi_m))\right| > \eta\right)$$

$$= P\left(\left|\sum_{m=1}^{M}(\xi_m - E(\xi_m))\right| > M\eta\right)$$

$$= P\left(\left|\sum_{m=1}^{M}(\xi_m - E(\xi_m))\right|^{1+v} > (M\eta)^{1+v}\right)$$

$$\leq \frac{E\left(\left|\sum_{m=1}^{M}(\xi_m - E(\xi_m))\right|^{1+v}\right)}{(M\eta)^{1+v}}.$$

Now, Bahr and Esseen recall that, for $1 \leq r \leq 2$ and x, y complex numbers, $|x+y|^r + |x-y|^r \leq 2(|x|^r + |y|^r)$. This, and the inequality which bears their names give

$$E\left(\left|\sum_{m=1}^{M}(\xi_m - E(\xi_m))\right|^{1+v}\right) \leq \frac{K}{2}\sum_{m=1}^{M}E\left(|\xi_m - E(\xi_m)|^{1+v}\right)$$

$$\leq K\sum_{m=1}^{M}E\left(|\xi_m|^{1+v} + 1\right)$$

$$\leq K\left(\sum_{m=1}^{M}E\left(|\xi_m|^{1+v}\right) + M\right)$$

which leads us to

$$P\left(\left|\frac{1}{M}\sum_{m=1}^{M}\xi_m - 1\right| > \eta\right) \leq K\left(\frac{\sum_{m=1}^{M}E\left(|\xi_m|^{1+v}\right)}{(M\eta)^{1+v}} + \frac{1}{M^v\eta^{1+v}}\right). \tag{8}$$

But, thanks to Lemma 4, we have

$$E\left(|\xi_m|^{1+v}\right) = \exp\left(\frac{\alpha_m^2}{2\sigma^2}(v+v^2)\right)$$

$$= \exp\left(\frac{1}{2\sigma^2}(v+v^2)\left(n2^{-j_n(2r+1)} + O\left(2^{j_n(1-2r)}\right)\right)\right)$$

$$= \exp\left(\frac{1}{2\sigma^2}(v+v^2)\left(n2^{-j_n(2r+1)}\right)\right)\exp\left(O\left(2^{j_n(1-2r)}\right)\right),$$

with $\exp(O(2^{j_n(1-2r)}))$ bounded independently of m thanks to the same lemma by a constant K', because $r > 1/2$.

Now, Equation 8 leads to

$$P\left(\left|\frac{1}{M}\sum_{m=1}^{M}\xi_m - 1\right| > \eta\right) \leq K\left(\frac{MK'\exp\left(\frac{1}{2\sigma^2}(v+v^2)\left(n2^{-j_n(2r+1)}\right)\right)}{(M\eta)^{1+v}} + \frac{1}{M^v\eta^{1+v}}\right)$$

$$\leq K\left(\frac{K'\exp\left(\frac{1}{2\sigma^2}(v+v^2)\left(n2^{-j_n(2r+1)}\right)\right)}{M^v\eta^{1+v}} + \frac{1}{M^v\eta^{1+v}}\right)$$

$$\leq K\max(K', 1)\frac{\exp\left(\frac{1}{2\sigma^2}(v+v^2)\left(n2^{-j_n(2r+1)}\right)\right)}{c_0^v(2^{j_n} - 1)^v\eta^{1+v}}$$

since $1 \leq \exp\left(\frac{1}{2\sigma^2}(v+v^2)\left(n2^{-j_n(2r+1)}\right)\right)$. $\qquad\qquad\qquad\qquad\qquad\qquad$ □

As a consequence, we just have to plug in the "value" of j_n in

$$P\left(\left|\frac{1}{M}\sum_{m=1}^{M}\xi_m - 1\right| > \eta\right) \leq c_1 \frac{\exp\left(c_2 n 2^{-j_n(2r+1)}\right)}{(2^{j_n}-1)^v}.$$

Now, $2^{-j_n r} = o(2^{-j_n^* r})$ as $n \to \infty$. But $2^{-j_n^* r} \simeq \left(\frac{\log n}{n}\right)^{\frac{r}{2r+1}}$ so $2^{-j_n(2r+1)} = o(\frac{\log n}{n})$ as $n \to \infty$ and therefore the argument of the exponential is negligible in front of $\log n$. Since $2^{j_n} \geq \left(\frac{n}{\log n}\right)^{\frac{1}{2r+1}}$, the ratio tends towards 0 as $n \to \infty$. The discussion before Lemma 3 shows this completes the proof of Proposition 1.

3.2 Estimators

3.2.1 Local polynomial estimator

We fix an integer $l \geq s$. t is any number such that $0 < t \leq l$. In this section, E stands for E_f. Reading through the proofs of Proposition 1.13 and Theorem 1.8 in Tsybakov [15], we get the following.

Theorem 3. *There exist $n_0 \in \mathbb{N}$, $c_1 > 0$ such that, for all $n \geq n_0$, for all $(\log n/n)^{1/(2l+1)} \geq h \geq \log n/n$, there exist functions $x \in [0,1] \mapsto W_{ni}(h,x)$, $1 \leq i \leq n$ verifying :*

- $\sup_{i,x} |W_{ni}(x)| \leq \frac{c_1}{nh}$

- $\sum_{i=1}^{n} |W_{ni}(x)| \leq c_1$

- $W_{ni}(x) = 0 \quad if \quad |x_i - x| > h$

- $x \mapsto W_{ni}(x)$ *is continuous.*

There exist further $c_2 > 0$ such that, for all $0 < t \leq l$, for all $f \in \Sigma(t)$, the local polynomial estimator of order l,

$$f_n(h)(x) = \sum_{i=1}^{n} W_{ni}(h,x)Y_i = \sum_{i=1}^{n} W_{ni}(x)Y_i \qquad\qquad (9)$$

satisfies :

$$\|Ef_n(h) - f\|_\infty \leq \frac{c_1\|f\|_{t,\infty}}{l!}h^t = B(h,f) \qquad\qquad (10)$$

$$E\|f_n(h) - Ef_n(h)\|_\infty^2 \leq c_2^2\frac{\log n}{nh} = c_2^2\sigma^2(h,n). \qquad\qquad (11)$$

For $h = h_n(t) = \left(\frac{\log n}{n}\right)^{\frac{1}{2t+1}}$, this implies that $f_n(h)$ satisfies

$$\sup_{f \in \Sigma(t)} E_f\|f_n(h_n(t)) - f\|_\infty \leq D\left(\frac{\log n}{n}\right)^{\frac{t}{2t+1}},$$

with D depending on B and l. This means that $f_n(h_n(t))$ is rate optimal over $\Sigma(t)$. Thanks to it, we shall now construct adaptive estimators, rate optimal over a range of regularity indexes, by proving Theorem 1.

12

3.2.2 Proof of Theorem 1 : existence of adaptive estimators

The proof is an adaptation of that of Giné and Nickl [8] in the density case, which uses Lepski's method. For each $0 < t \leq l$, we have an estimator $f_n(h_n(t))$ rate optimal over $\Sigma(t)$. We want to devise a procedure which allows us to choose h_n according only to the data, such that if $f \in \Sigma(t)$, h will be roughly $h(t)$. Lepski's method consists in discretising the set of all possible bandwiths, and choosing h_n so that, for all $h \simeq h_n(t) < h_n$, the distance between $f_n(h)$ and $f_n(h_n)$ is of the order the optimal rate for $\Sigma(t)$, $r_n(t)$.

Fix $\rho > 1$. Define

$$\mathcal{H} = \left\{ \rho^{-k} \,\middle|\, k \geq 0,\, \rho^{-k} > \frac{(\log n)^2}{n} \right\}$$

and

$$\hat{h}_n = \max \left\{ h \in \mathcal{H} \,\middle|\, \forall g < h,\, \|f_n(\hat{h}_n) - f_n(g)\|_\infty \leq \left(M \frac{\log n}{ng} \right)^{1/2} \right\} \tag{12}$$

with $M = 16(\sqrt{2\sigma c_1 K} + c_2)^2$; K is choosen later. $|\mathcal{H}| \simeq \log n$.
Indeed, for all $t > 0$, $\frac{\log n}{n} \leq h_n(t) \leq 1$ and we choose $(\log n)^2$ for practical reasons (this choice is not restrictive since indeed, what matters in this ratio is the power of n, not that of the logarithm).

The adaptive estimator is $\hat{f}_n = f_n(\hat{h}_n)$ but we will write $f_n(\hat{h}_n)$ in the proof. Fix now $t > 0$ and $f \in \Sigma(t)$. Define

$$h_f = \max \left\{ h \in \mathcal{H} \,\middle|\, B(h, f) \leq \frac{\sqrt{M}}{4} \sigma(h, n) \right\},$$

which verifies, thanks to Theorem 3,

$$h_f \simeq \left(\frac{\log n}{n} \right)^{\frac{1}{2t+1}}.$$

We now bound $E\|f_n(\hat{h}_n) - f\|_\infty$. We distinguish the cases $\{\hat{h}_n \geq h_f\}$ and $\{\hat{h}_n < h_f\}$. First,

$$E\| f_n(\hat{h}_n) - f\|_\infty I_{\hat{h}_n \geq h_f}$$

$$\leq E\left(\|f_n(\hat{h}_n) - f_n(h_f)\|_\infty + \|f_n(h_f) - Ef_n(h_f)\|_\infty + \|Ef_n(h_f) - f\|_\infty \right) I_{\hat{h}_n \geq h_f}$$

$$\leq E\left(\sum_{h \geq h_f} \|f_n(h) - f_n(h_f)\|_\infty I_{\hat{h}_n = h} \right) + E\|f_n(h_f) - Ef_n(h_f)\|_\infty + \|Ef_n(h_f) - f\|_\infty$$

$$\leq \sqrt{M}\sigma(h_f, n)P(\hat{h}_n \geq h_f) + E\|f_n(h_f) - Ef_n(h_f)\|_\infty + \|Ef_n(h_f) - f\|_\infty$$

$$\leq \sqrt{M}\sigma(h_f, n) + c_2\sigma(h_f, n) + \frac{\sqrt{M}}{4}\sigma(h_f, n) = O(\sigma(h_f, n))$$

13

Then,

$$E\|f_n(\hat{h}_n) - f\|_\infty I_{\hat{h}_n < h_f}$$

$$= \sum_{h \in \mathcal{H}, h < h_f} E\left(\|f_n(\hat{h}_n) - f\|_\infty I_{\hat{h}_n = h}\right)$$

$$\leq \sum_{h \in \mathcal{H}, h < h_f} E\left(\|f_n(\hat{h}_n) - Ef_n(\hat{h}_n)\|_\infty + \|Ef_n(\hat{h}_n) - f\|_\infty\right) I_{\hat{h}_n = h}$$

$$\leq \sum_{h \in \mathcal{H}, h < h_f} \left(E\left(\|f_n(\hat{h}_n) - Ef_n(\hat{h}_n)\|_\infty I_{\hat{h}_n = h}\right) + \|Ef_n(\hat{h}_n) - f\|_\infty E\left(I_{\hat{h}_n = h}\right)\right)$$

$$\leq \sum_{h \in \mathcal{H}, h < h_f} \left(c_2 \sigma\left(\frac{(\log n)^2}{n}, f\right) P\left(\hat{h}_n = h\right)^{1/2}\right) + B(h_f, f)P(\hat{h}_n < h_f)$$

$$\leq \frac{c_2}{\sqrt{\log n}} \sum_{h \in \mathcal{H}, h < h_f} P\left(\hat{h}_n = h\right)^{1/2} + B(h_f, f).$$

Now, for $h < h_f$, and writing $h^+ = \rho h$,

$$P\left(\hat{h}_n = h\right) \leq P\left(\text{there is } g \leq h, \text{ such that } h^+ \text{ and } g \text{ do not satisfy the inequality } 12\right)$$

$$\leq \sum_{g \leq h, g \in \mathcal{H}} P\left(\|f_n(h^+) - f_n(g)\|_\infty \geq \sqrt{M}\left(\frac{\log n}{ng}\right)^{1/2}\right)$$

$$\leq \sum_{g \leq h, g \in \mathcal{H}} P\left(\|f_n(h^+) - f_n(g)\|_\infty \geq \sqrt{M}\sigma(g, n)\right)$$

For $g \leq h$, we have

$$\|f_n(h^+) - f_n(g)\|_\infty$$

$$\leq \|f_n(h^+) - Ef_n(h^+)\|_\infty + \|Ef_n(h^+) - f\|_\infty$$

$$+ \|f - Ef_n(g)\|_\infty + \|Ef_n(g) - f_n(g)\|_\infty$$

$$\leq 2B(h_f, f) + \|f_n(h^+) - Ef_n(h^+)\|_\infty + \|Ef_n(g) - f_n(g)\|_\infty$$

since $B(h, f)$ increases with h,

$$\leq 2\frac{\sqrt{M}}{4}\sigma(h_f, n) + \|f_n(h^+) - Ef_n(h^+)\|_\infty + \|Ef_n(g) - f_n(g)\|_\infty$$

$$\leq \frac{\sqrt{M}}{2}\sigma(g, n) + \|f_n(h^+) - Ef_n(h^+)\|_\infty + \|Ef_n(g) - f_n(g)\|_\infty$$

since $\sigma(h, f)$ decreases with h, so that,

$$P\left(\|f_n(h^+) - f_n(g)\|_\infty \geq \sqrt{M}\sigma(g, n)\right)$$

$$\leq P\left(\|f_n(h^+) - Ef_n(h^+)\|_\infty + \|Ef_n(g) - f_n(g)\|_\infty \geq \sqrt{M}\sigma(g, n) - \frac{\sqrt{M}}{2}\sigma(g, n)\right)$$

$$\leq P\left(\|f_n(h^+) - Ef_n(h^+)\|_\infty \geq \frac{\sqrt{M}}{4}\sigma(h^+, n)\right) + P\left(\|f_n(g) - Ef_n(g)\|_\infty \geq \frac{\sqrt{M}}{4}\sigma(g, n)\right)$$

$$\leq 4\exp(-K\log n)$$

14

thanks to Proposition 3. Therefore,

$$P(\hat{h}_n = h) \leq \sum_{g \leq h, g \in \mathcal{H}} P\left(\|f_n(h^+) - f_n(g)\|_\infty \geq \sqrt{M}\sigma(g,n) \right)$$
$$\leq 4\exp(-K\log n)\log n,$$

and

$$E\|f_n(h_n) - f\|_\infty I_{h_n < h_f}$$
$$\leq \frac{c_2}{\sqrt{\log n}} \sum_{h \in \mathcal{H}, h < h_f} P\left(\hat{h}_n = h \right)^{1/2} + B(h_f, n)$$
$$\leq \frac{c_2}{\sqrt{\log n}} \left(4\exp(-K\log n)\log n \right)^{1/2} \log n + B(h_f, n)$$
$$= O(\sigma(h_f, n))$$

for K large enough, choosen independently of f or t, which completes the proof of Theorem 1.

3.3   A concentration inequality

The following proposition is a key result which allows us to prove both the honesty and adaptivity of C_n in Section 2.3 as well as the existence of the adaptive estimators. Indeed, it gives an upper bound for the concentration of $f_n(h)$ around its expectation. The proof relies on an inequality due to Borell [1], which may be found in Theorem 1.7 in [11] in the form we use it.

Theorem 4. *Let $G(x)$, $x \in T$, be a centered Gaussian process indexed by the countable set T, and such that $\sup_{x \in T} |G(x)| < \infty$ almost surely. Then $E\sup_{x \in T} |G(x)| < \infty$, and for every $r \geq 0$ we have*

$$P\left\{ \left| \sup_{x \in T} |G(x)| - E\sup_{x \in T} |G(x)| \right| \geq r \right\} \leq 2e^{-r^2/2\sigma_0^2}$$

where $\sigma_0^2 = \sup_{x \in T} E(G^2(x)) < \infty$.

Proposition 3. *Let $(\log n/n)^{1/(2l+1)} \geq h \geq \log n/n$ and c_1 and c_2 the constants in Theorem 3. Let $f_n(h)$ be the local polynomial estimator of order l for f. Let $G_n = f_n(h) - E(f_n(h))$. Then,*

$$\sigma_0^2 = E(G_n^2) \leq \frac{\sigma^2 c_1^2}{nh} \tag{13}$$

$$\forall u \geq c_2 \left(\frac{\log n}{nh} \right)^{1/2}, P\left(\|G_n\|_\infty \geq u \right) \leq 2\exp\left(-\frac{\log n}{2\sigma c_1} \left(\frac{u}{\left(\frac{\log n}{nh} \right)^{1/2}} - c_2 \right)^2 \right). \tag{14}$$

Proof. Obviously, Equation 1 and Equation 9 imply that $(G_n(x),\ 0 \leq x \leq 1)$, which satisfies for all $0 \leq x \leq 1$,

$$G_n(x) = f_n(x) - E(f_n(x))$$
$$= \sum_{i=1}^{n} W_{ni}(x)Y_i - E\left(\sum_{i=1}^{n} W_{ni}(x)Y_i \right)$$
$$= \sum_{i=1}^{n} W_{ni}(x)\varepsilon_i$$

is a gaussian process.

15

- Recall the definition of f_n and that the ε_i's are i.i.d. $\mathcal{N}(0, \sigma^2)$. Then,

$$
\begin{aligned}
\sigma_0^2 &= \sup_{x \in [0,1]} E(G_n^2(x)) \\
&= \sup_{x \in [0,1]} E \left(\sum_{i=1}^n W_{ni}(x) \varepsilon_i \right)^2 \\
&= \sup_{x \in [0,1]} \sum_{i=1}^n W_{ni}^2(x) \sigma^2 \\
&\leq \sigma^2 \|W_{ni}\|_\infty \sup_{x \in [0,1]} \sum_{i=1}^n |W_{ni}(x)| \\
&\leq \frac{\sigma^2 c_1^2}{nh}.
\end{aligned}
$$

- Then, we want to bound the probability

$$
\begin{aligned}
P\left(\|G_n\|_\infty \geq u\right) &= P\left(\|G_n\|_\infty - E\|G_n\|_\infty \geq u - \|G_n\|_\infty\right) \\
&\leq P\left(\left|\|G_n\|_\infty - E\|G_n\|_\infty\right| \geq u - \|G_n\|_\infty\right) \\
&= P\left(\left|\sup_{x \in [0,1] \cap \mathbb{Q}} |G_n(x)| - E \sup_{x \in [0,1] \cap \mathbb{Q}} |G_n(x)|\right| \geq u - \|G_n\|_\infty\right)
\end{aligned}
$$

since G_n is continuous. Now, if $u - \|G_n\|_\infty \geq 0$, we can apply Theorem 4 to $(G_n(x),\ x \in [0,1] \cap \mathbb{Q})$ and write

$$
P\left(\|G_n\|_\infty \geq u\right) \leq 2 \exp\left(-\frac{(u - E\|G_n\|_\infty)^2}{2\sigma_0^2}\right).
$$

Finally, thanks to 11 and 13, for $u \geq c_2 \left(\frac{\log n}{nh}\right)^{1/2}$,

$$
P\left(\|G_n\|_\infty \geq u\right) \leq 2 \exp\left(-\frac{\log n}{2\sigma c_1}\left(\frac{u}{\left(\frac{\log n}{nh}\right)^{1/2}} - c_2\right)^2\right).
$$

$\square$

In the proof of section 2.3 we only need the following result, obtained with $h = \left(\frac{\log n}{n}\right)^{1/(2r+1)}$.

Corollary 1. *Let $C \geq 0$. Take $u = (c_2 + C)r_n(r)$, then*

$$
P\left(\|G_n\|_\infty \geq u\right) \leq 2n^{-C^2/2\sigma c_1}.
$$

References

[1] BORELL, C. (1975). The Brunn-Minkowski inequality in Gauss space. *Invent. Math.* **30** 207-216.

[2] BROWN, L. D., LOW, M. G. (1996). Asympotic equivalence of nonparametric regression and white noise. *Ann. Stat.* **24** 2384-2398.

[3] BULL, A. D. (2011). Honest adaptive confidence bands and self-similar functions. To appear.

[4] CLAESKENS, G., VAN KEILEGOM, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Stat.* **31** 1852-1884.

[5] COHEN, A., DAUBECHIES, I. AND VIAL, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Compute. Harmon. Anal.* **1** 54-81.

[6] GENOVESE, C., WASSERMAN, L. (2008). Adaptive confidence bands. *Ann. Stat.* **36** 875–905.

[7] GINÉ, E., NICKL, R. (2010). Confidence bands in density estimation. *Ann. Stat.* **38** 1122–1170.

[8] GINÉ, E., NICKL, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probab. Theory Related Fields* **143** 569-596.

[9] GOLDENSHLUGER, A., LEPSKI, O. (2009). Structural adaptation via $\mathbb{L}_p$-norm oracle inequalitites. *Probab. Theory Related Fields* **143** 41-71.

[10] HOFFMANN, M. AND NICKL, N. (2011). On adaptive inference and confidence bands. *Ann. Stat.* **39** 2383-2409.

[11] LEDOUX, M. (2001). The Concentration of measure phenomenon. American Mathematical Society.

[12] LEPSKI, O. AND TSYBAKOV, A. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probability Theory and Related Fields.* **117-1** 17-48.

[13] LOW, M. G. (1997). On nonparametric confidence intervals. *Ann. Stat.* **25** 2547–2554.

[14] MEYER, Y. (1992). Wavelets and Operators. *Cambridge Studies in Advanced Mathematics* **37**. Cambridge Univ. Press, Cambridge.

[15] TSYBAKOV, A. (2008). Introduction to nonparametric estimation. Springer.

[16] VON BAHR, B., ESSEEN, C. (1965). Inequalities for the rth absolute moment of a sum of random variables $1 \leqq r \leqq 2$. *Ann. Stat.* **36** 299-303.