Pointwise Adaptive Estimation of the Marginal Density of a Weakly Dependent Process

Karine Bertin[a,*], Nicolas Klutchnikoff[b]

[a]CIMFAV, Universidad de Valparaíso, General Cruz 222, Valparaíso, Chile
[b]IRMAR, Université de Rennes 2, CNRS, UEB, Campus Villejean, Place du recteur Henri le Moal CS 24307, 35043 Rennes cedex, France

Abstract

This paper is devoted to the estimation of the common marginal density function of weakly dependent processes. The accuracy of estimation is measured using pointwise risks. We propose a data-driven procedure using kernel rules. The bandwidth is selected using the approach of Goldenshluger and Lepski and we prove that the resulting estimator satisfies an oracle type inequality. The procedure is also proved to be adaptive (in a minimax framework) over a scale of Hölder balls for several types of dependence: strong mixing processes, λ-dependent processes or i.i.d. sequences can be considered using a single procedure of estimation. Some simulations illustrate the performance of the proposed method.

Keywords:
Adaptive minimax rates, Density estimation, Kernel estimation, Oracle inequality, Weakly dependent processes

1. Introduction

Let $\mathbb{X} = (X_i)_{i\in\mathbb{Z}}$ be a real-valued weakly dependent process admitting a common marginal density $f : \mathbb{R} \to \mathbb{R}$. We consider the problem of estimating f at a fixed point x_0 based on observation of $X_1, \ldots, X_n$ with $n \in \mathbb{N}^*$. The accuracy of an estimator $\tilde{f}_n$ is evaluated using the pointwise risk defined, for

*Corresponding author.
Email addresses: `karine.bertin@uv.cl` (Karine Bertin), `nicolas.klutchnikoff@univ-rennes2.fr` (Nicolas Klutchnikoff)

fixed $x_0 \in \mathbb{R}$ and $q > 0$, by

$$R_q(\tilde{f}_n, f) = \left(\mathbf{E}|\tilde{f}_n(x_0) - f(x_0)|^q\right)^{1/q},$$

where $\mathbf{E}$ denotes the expectation with respect to the distribution of the process $\mathbb{X}$. The main interest in considering such risks is to obtain estimators that adapt to the local behavior of the density function to be estimated.

The aim of this paper is to obtain adaptive estimators of f on Hölder classes of regularity $s > 0$ for this risk and different types of weakly dependent processes.

In the independent and identically distributed (i.i.d.) case, the minimax rate of convergence is $n^{-s/(2s+1)}$ (see Tsybakov, 2009, and references therein). Adaptive procedures based on the classical Lepki procedure (see Lepski, 1990) have been obtained (see Butucea, 2000, 2001) with rates of the form $(\log n/n)^{s/(2s+1)}$.

In the context of dependent data, Ragache and Wintenberger (2006) as well as Rio (2000) studied kernel density estimators from a minimax point of view for pointwise risks. They obtained the same rate of convergence as in the independent case when the coefficients of dependence decay at a geometric rate. Several papers deal with the adaptive estimation of the common marginal density of a weakly dependent process. Tribouley and Viennet (1998) and Comte and Merlevède (2002) proposed $\mathbb{L}_p$–adaptive estimators under α–mixing or β–mixing conditions that converge at the previously mentioned rates. Gannaz and Wintenberger (2010) extend these results to a wide variety of weakly dependent processes including λ–dependent processes. Note that, in these papers, the proposed procedures are based on nonlinear wavelet estimators and only integrated risks are considered. Moreover, the thresholds are not explicitly defined since they depend on an unknown multiplicative constant. As a consequence, such methods can not be used directly for practical purposes.

Our main purpose is to prove similar results for pointwise risks. We propose here a kernel density estimator with a data-driven selection of the bandwidth where the selection rule is performed using the so-called Goldenshluger-Lepski method (see Goldenshluger and Lepski, 2008, 2011, 2014). This method was successfully used in different contexts such as in Comte and Genon-Catalot (2012); Doumic et al. (2012); Bertin et al. (2014); Rebelles (2015), but only with i.i.d. observations. However there are at least two practical motivations to consider dependent data. Firstly, obtaining estimators that are robust with respect to slight perturbations from the i.i.d. ideal model can be useful. Secondly, many econometric models (such as ARCH or GARCH) deal with dependent data that admit a common marginal density. These two motivations suggest to consider a class of dependent data as large as possible and to find a single procedure of estimation that adapts to each situation of dependence.

Our contribution is the following. We obtain the adaptive rate of convergence for pointwise risks over a large scale of Hölder spaces in several situations of dependence, such as α–mixing introduced by Rosenblatt (1956) and the λ–dependence defined by Doukhan and Wintenberger (2007). This partially generalizes previous results obtained in i.i.d. case by Butucea (2000) and Rebelles (2015). To the best of our knowledge, this is the first adaptive result for pointwise density estimation in the context of dependent data. To establish it, we prove an oracle type inequality: the selected estimator performs almost as well as the best estimator in a given large finite family of kernel estimators. Our data-driven procedure depends only on explicit quantities. This implies that this procedure can be directly implemented in practice. As a direct consequence, we get a new method to choose an accurate local bandwidth for kernel estimators.

The rest of this paper is organized as follows. Section 2 is devoted to the presentation of our model and of the assumptions on the process $\mathbb{X}$. The construction of our procedure of estimation is developed in Section 3. The main results of the paper are stated in Section 4 whereas their proofs are postponed to Section 7. Examples of processes that can be considered in our framework are given in Section 5. A simulation study is performed in Section 6 to illustrate the performance of our method in comparison to other classical estimation procedures. The proofs of the technical results are presented in the appendix.

2. Model

In this section, we present the assumptions made on the distribution of the random process $\mathbb{X} = (X_i)_{i\in\mathbb{Z}}$. These assumptions depend on five positive real parameters $(\mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \mathfrak{B}, \mathfrak{C})$. In Section 5, we offer some classical examples that satisfy these assumptions.

Assumption 1. *The X_i's are identically distributed. Moreover, the distribution of X_0 admits a density function $f : \mathbb{R} \to \mathbb{R}_+$ (with respect to the Lebesgue* *measure) that satisfies:*

$$\sup_{x \in V_n(x_0)} f(x) \leq \mathfrak{B},$$

where

$$V_n(x_0) = [x_0 - h^*, x_0 + h^*] \qquad \textit{with} \qquad h^* = \exp\Big(-\sqrt{\log n}\Big).$$

This assumption deals with the distribution of the marginals of the process $\mathbb{X}$ and is classical in the i.i.d. case (see Goldenshluger and Lepski, 2011, and references therein). Note also that stationarity of $\mathbb{X}$ is not assumed. Thus, re-sampled processes of stationary processes can be considered in this study as in Ragache and Wintenberger (2006).

To state the second assumption, we introduce some notations used throughout the paper. Let u and v be two positive integers and set $i \in \mathbb{Z}^u$ and $j \in \mathbb{Z}^v$. We define the random vector $\mathbb{X}_i = (X_{i_1}, \ldots, X_{i_u})$ valued in $\mathbb{R}^u$ as well as the gap between j and i by $\gamma(j,i) = \min(j) - \max(i) \in \mathbb{Z}$. The functional class $\mathbb{G}_u$ consists of real-valued functions g defined on $\mathbb{R}^u$ such that

$$\operatorname{Supp}(g) \subseteq (V_n(x_0))^u,$$

$$\|g\|_{\infty,u} = \sup_{x \in \mathbb{R}^u} |g(x)| < \infty,$$

and

$$\operatorname{Lip}_u(g) = \sup_{x \neq y} \frac{|g(x) - g(y)|}{\sum_{i=1}^u |x_i - y_i|} < +\infty.$$

We now define the sequence $\rho(\mathbb{X}) = (\rho_r(\mathbb{X}))_{r \in \mathbb{N}^*}$ by

$$\rho_r(\mathbb{X}) = \sup_{u,v \in \mathbb{N}^*} \sup_{\substack{(i,j) \in \mathbb{Z}^u \times \mathbb{Z}^v \\ \gamma(j,i) \geq r}} \sup_{g \in \mathbb{G}_u} \sup_{\tilde{g} \in \mathbb{G}_v} \frac{\left|\operatorname{Cov}\Big(g(\mathbb{X}_i), \tilde{g}(\mathbb{X}_j)\Big)\right|}{\Psi(u,v,g,\tilde{g})},$$

where $\Psi(u,v,g,\tilde{g}) = \max(\Psi_1(u,v,g,\tilde{g}), \Psi_2(u,v,g,\tilde{g}))$ with

$$\Psi_1(u,v,g,\tilde{g}) = 4\|g\|_{\infty,u}\|\tilde{g}\|_{\infty,v}$$

and

$$\Psi_2(u,v,g,\tilde{g}) = u\|\tilde{g}\|_{\infty,v} \operatorname{Lip}_u(g) + v\|g\|_{\infty,u} \operatorname{Lip}_v(\tilde{g}) + uv \operatorname{Lip}_u(g) \operatorname{Lip}_v(\tilde{g}).$$

Assumption 2. *The sequence $\rho(\mathbb{X})$ is such that, for any $r \in \mathbb{N}$:*

$$\rho_r(\mathbb{X}) \leq \mathfrak{c} \exp(-\mathfrak{a} r^{\mathfrak{b}}).$$

It is worth mentioning that $\rho(\mathbb{X})$ is a measure of dependence of the process $\mathbb{X}$. Our assumption ensures that $\mathbb{X}$ is a weakly dependent process in the sense introduced by Doukhan and Louhichi (1999) and includes strong mixing processes as well as λ-weak-dependent processes. For a broader picture of weakly-dependent processes, as well as examples and applications, we refer the reader to Dedecker et al. (2007) and references therein. Our assumption guarantees that the decay occurs at a geometric rate. In this context, Doukhan and Neumann (2007) and Merlevède et al. (2009) proved Bernstein-type inequalities that are used in the proof of Theorem 1.

Our last assumption deals with the joint distribution of the random pairs (X_i, X_j) for $i, j \in \mathbb{Z}$.

Assumption 3. *We assume that:*

$$\sup_{i \neq j} \sup_{g, \tilde{g} \in \mathbb{G}_1} \mathbf{E}\Big(g(X_i)\tilde{g}(X_j)\Big) \leq \mathfrak{C}\|g\|_1\|\tilde{g}\|_1,$$

where

$$\|g\|_1 = \int_{\mathbb{R}} |g(t)|dt.$$

In their paper, Gannaz and Wintenberger (2010) also made an assumption on the joint distributions. They assume that, for any i, j in $\mathbb{Z}$, the distribution of the random vector (X_i, X_j) admits a density function $f_{i,j} : \mathbb{R}^2 \to \mathbb{R}$ with respect to the Lebesgue measure that satisfies

$$\sup_{i \neq j} \sup_{(x,y) \in (V_n(x_0))^2} |f_{i,j}(x, y)| \leq \mathfrak{C}.$$

It is readily seen that this assumption implies Assumption 3.

3. Estimation procedure

In this section, we describe the construction of our procedure which is based on the so-called Goldenshluger-Lepski method. It consists of selecting, in a data driven way, an estimator in a given family of linear kernel density estimators. Consequently, our method offers a new approach to select an optimal bandwidth for kernel estimators in order to estimate the marginal density of a process in several situations of weak dependence. This leads to a procedure of estimation which is well adapted to inhomogeneous smoothness of the underlying marginal density. Notice also that our procedure is completely data-driven: it depends only on explicit constants that do not need to be calibrated by simulations or using the so-called rule of thumb.

Kernel density estimators.. A function $K : \mathbb{R} \to \mathbb{R}$ is called a kernel if K has a compact support in $[-1, 1]$, its Lipschitz constant $\mathrm{Lip}(K)$ is finite and $\int_{\mathbb{R}} K(x)\,\mathrm{d}x = 1$. As usual, to obtain the minimax rates of convergence for smooth densities (see Theorem 2), we must consider high order kernels. More precisely, we say that K is of order m if for any $1 \le \ell \le m$, we have

$$\int_{\mathbb{R}} K(x)x^{\ell}\,\mathrm{d}x = 0.$$

Following Rosenblatt et al. (1956) and Parzen (1962), we consider kernel density estimators $\hat{f}_h$ defined, for $h > 0$, by:

$$\hat{f}_h(x_0) = \frac{1}{n}\sum_{i=1}^{n} K_h(x_0 - X_i),$$

where $K_h(\cdot) = h^{-1}K(h^{-1}\cdot)$. In the following paragraph, we propose a data-driven procedure to select the bandwidth h in the finite subset $\mathscr{H}_n$ of $(0, 1)$ defined by $\mathscr{H}_n = \{2^{-k} : k \in \mathbb{N}\} \cap [h_*, h^*]$. Here $h_* = n^{-1}\exp\left(\sqrt{\log n} - 1\right)$ and h^* is defined in Assumption 1.

Bandwidth selection.. Following Goldenshluger and Lepski (2011), we first define for $h \in \mathscr{H}_n$ the two following quantities

$$A(h, x_0) = \max_{\mathfrak{h} \in \mathscr{H}_n} \left\{ |\hat{f}_{h \vee \mathfrak{h}}(x_0) - \hat{f}_{\mathfrak{h}}(x_0)| - \widehat{M}_n(h, \mathfrak{h}) \right\}_+$$

and

$$\widehat{M}_n(h) = \sqrt{2q|\log h| \left(\widehat{J}_n(h) + \frac{\delta_n}{nh} \right)}, \tag{1}$$

where $\delta_n = (\log n)^{-1/2}$, $\{y\}_+ = \max(0, y)$ denotes the nonnegative part of $y \in \mathbb{R}$ and $h \vee \mathfrak{h} = \max(h, \mathfrak{h})$ for any $h, \mathfrak{h} \in \mathscr{H}_n$. We also consider

$$\widehat{M}_n(h, \mathfrak{h}) = \widehat{M}_n(\mathfrak{h}) + \widehat{M}_n(h \vee \mathfrak{h}),$$

and

$$\widehat{J}_n(h) = \frac{1}{n^2}\sum_{i=1}^{n} K_h^2(x_0 - X_i).$$

Our procedure consists of selecting the bandwidth $\hat{h}(x_0)$ such that

$$\hat{h}(x_0) = \arg\min_{h \in \mathscr{H}_n} \left(A(h, x_0) + \widehat{M}_n(h) \right). \tag{2}$$

The final estimator of $f(x_0)$ is then defined by

$$\hat{f}(x_0) = \hat{f}_{\hat{h}(x_0)}(x_0).$$

Comments.. Our procedure selects in (2) a data-driven bandwidth $\hat{h}(x_0)$ that makes a trade-off between the quantities $A(h, x_0)$ and $\widehat{M}_n(h)$. This can be viewed as an empirical version of the classical trade-off between a bias term and a penalized standard deviation term.

The quantity $\widehat{M}_n(h)$ can be viewed as a penalized upper bound of the standard deviation of the estimator $\hat{f}_h$. Indeed, Lemma 2 combined with arguments given in the proof of Theorem 1 in Section 7, implies that, for n large enough

$$\mathrm{Var}(\hat{f}_h(x_0)) \leq J_n(h) + \frac{\delta_n}{nh} \leq \hat{J}_n(h) + \frac{C\delta_n}{nh}$$

where C is some positive constant and

$$J_n(h) = \frac{1}{n} \int K_h^2(x_0 - x) f(x) dx \tag{3}$$

would be the variance of $\hat{f}_h(x_0)$ if the data were i.i.d.

Moreover, the quantity $A(h, x_0)$ is a rough estimator of the bias term of $\hat{f}_h$. Indeed (see proof of Theorem 1), we have

$$A(h, x_0) \leq 2E_h(x_0) + o_{\mathbf{P}}\left(\frac{1}{\sqrt{nh}}\right),$$

where $E_h(x_0) = \max_{\mathfrak{h} \leq h} |K_{\mathfrak{h}} \star f(x_0) - f(x_0)|$ is a nondecreasing version of the bias term which is of the same order of the bias of $\hat{f}_h$ over Hölder balls.

4. Results

In this section, we state two results, the proofs of which are postponed to Section 7. To do so, we introduce some notations. Here and after, $\mathscr{L}(\mathbb{X})$ denotes the distribution of the process $\mathbb{X}$. For $\vartheta = (\mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \mathfrak{B}, \mathfrak{C}) \in \Theta = (0, \infty)^5$, we say that $\mathscr{L}(\mathbb{X})$ belongs to the *model* $\mathscr{P}_\vartheta$ if Assumptions 1–3 are fulfilled with these parameters. For $s > 0$ and $\mathfrak{L} > 0$, the Hölder ball $\mathscr{C}^s(\mathfrak{L})$ contains the functions $f : \mathbb{R} \to \mathbb{R}$ such that f is $m_s = \sup\{k \in \mathbb{N} : k < s\}$ times differentiable and

$$|f^{(m_s)}(x) - f^{(m_s)}(y)| \leq \mathfrak{L}|x - y|^{s - m_s}, \qquad x, y \in \mathbb{R}.$$

The distribution $\mathscr{L}(\mathbb{X}) \in \mathscr{P}_\vartheta$ belongs to $\mathscr{P}_\vartheta(s, \mathfrak{L})$ if the marginal density f of the process $\mathbb{X}$ belongs to $\mathscr{C}^s(\mathfrak{L})$.

Theorem 1. *Set $\vartheta \in \Theta$ and assume that $\mathscr{L}(\mathbb{X})$ belongs to $\mathscr{P}_\vartheta$. Let K be an arbitrary kernel. Then, for any $n \geq 1$:*

$$R_q^q(\hat{f}, f) \leq \min_{h \in \mathscr{H}_n} \left\{ C_1^* E_h^q(x_0) + C_2^* \left(\frac{|\log h|}{nh} \right)^{q/2} \right\} + \frac{C_3^*}{\delta_n n^{q/2}} \tag{4}$$

where C_1^* is a positive constant that depends only on q while C_2^* and C_3^* are positive constant that depends only on ϑ, q and K. The expressions of C_1^* and C_2^* are given in (21).

Remark. This theorem gives a nonasymptotic oracle-type inequality. First, note that, for any $h \in \mathscr{H}_n$, the reminder term $C_3^*(\delta_n n^{q/2})^{-1}$ is negligible with respect to the penalized standard deviation term $C_2^*(|\log h|/(nh))^{q/2}$. Moreover, since $E_h(x_0)$ is a nondecreasing version of the bias of $\hat{f}_h$, the right hand side of (4) can be viewed as a tight upper bound for $\min_{h\in\mathscr{H}_n} \mathbf{E}|\hat{f}_h(x_0) - f(x_0)|^q$. That means that our procedure performs almost as well as the best kernel density estimator in the family $\{\hat{f}_h : h \in \mathscr{H}_n\}$. As usual for oracle-type inequalities (see Goldenshluger and Lepski, 2011), the constants C_2^* and C_3^* depend on the parameter ϑ while our estimation procedure does not.

Theorem 2. *Set $\vartheta \in \Theta$, $s > 0$ and $\mathfrak{L} > 0$ and let K be a kernel of order $m \geq m_s$. There exists a constant C_4^* that depends only on ϑ, s, $\mathfrak{L}$, K and q such that:*

$$\sup_{\mathscr{L}(\mathbb{X}) \in \mathscr{P}_\vartheta(s,\mathfrak{L})} \left(\mathbf{E}|\hat{f}(x_0) - f(x_0)|^q \right)^{1/q} \leq C_4^* \left(\frac{\log n}{n} \right)^{\frac{s}{2s+1}}.$$

This result is a direct consequence of Theorem 1, since it can be easily proved that, for any $h > 0$,

$$\sup_{f \in \mathscr{C}^s(\mathfrak{L},\mathfrak{B})} E_h^q(x_0) \leq C h^{sq}, \tag{5}$$

where C depends only on s, $\mathfrak{L}$, K and q. This implies that, for n large enough, there exists $h_n(s, \mathfrak{L}, K, q) \in \mathscr{H}_n$ such that the right hand side of (4) is bounded, up to a multiplicative constant, by $(\log n/n)^{s/(2s+1)}$.

Remark. As far as we know this result is the first theoretical pointwise adaptive result for the estimation of the marginal density in the context of weak dependence. The minimax rate of convergence in the i.i.d. case is known to be $n^{-s/(2s+1)}$ (see Hasminskii and Ibragimov, 1990, among others). Using Lemma 2 (see Section 7) and (5), the minimax rate of convergence remains unchanged in our dependence framework. More precisely we have the following proposition.

Proposition 1. *Set $\vartheta \in \Theta$, $s > 0$ and $\mathfrak{L} > 0$ and let K be a kernel of order $m \geq m_s$. Define $h_n(s) = n^{-\frac{1}{2s+1}}$. There exists a constant C_5^* that depends only on ϑ, s, $\mathfrak{L}$, K and q such that the kernel estimator with kernel K and bandwidth $h_n(s)$ satisfies*

$$\sup_{\mathscr{L}(\mathbb{X}) \in \mathscr{P}_\vartheta(s,\mathfrak{L})} \left(\mathbf{E} |\hat{f}_{h_n(s)}(x_0) - f(x_0)|^q \right)^{1/q} \leq C_5^* n^{-\frac{s}{2s+1}}.$$

Doukhan and Louhichi (1999) proved a similar result. Moreover it is well-known (see Lepski, 1990; Tsybakov, 1998; Butucea, 2001; Klutchnikoff, 2014; Rebelles, 2015) that for pointwise risks the extra logarithmic factor $\log n$ is an unavoidable price to pay for adaptation even in the i.i.d. framework. Consequently, our procedure is adaptive to both the regularity of f and the structure of dependence.

5. Examples

To illustrate the variety of situations that can be considered in our framework, we propose simple examples of processes $\mathbb{Y} = (Y_i)_{i \in \mathbb{Z}}$ that satisfy our assumptions. Note however, that more general processes can be considered (see Dedecker et al., 2007, for details).

Infinite moving average (continuous innovations).. Let $(\xi_i)_{i \in \mathbb{Z}}$ be a sequence of i.i.d. random variables such that the distribution of ξ_0 admits a bounded density function $p_\xi(\cdot)$ and satisfy $\mathbf{E}|\xi_0|^b < +\infty$ for some $b > 0$. For any $i \in \mathbb{Z}$, we define:

$$Y_i = \sum_{j \in \mathbb{Z}} a_j \xi_{i-j}$$

where $(a_j)_{j \in \mathbb{Z}}$ is a sequence of deterministic positive real numbers that satisfies:

$$a_j = a_0 \tau^{|j|}$$

with $a_0 > 0$ and $0 < \tau < 1$. It is well-known (see Doukhan and Louhichi, 1999) that such processes are strongly stationary and weak-dependent.

First, note that, for any bounded continuous function ϕ we have:

$$\mathbf{E}\phi(Y_i) = \int_{\mathbb{R}} \phi(y) \frac{1}{a_0} \mathbf{E}\left(p_\xi \left(\frac{y - (Y_i - a_0\xi_i)}{a_0} \right) \right) dy$$

Thus the marginal density of Y_i is bounded by $\|p_\xi\|_\infty / a_0$. Hence Assumption 1 is fulfilled.

Next, in a similar way, it can be proved that the distribution of the random pair (Y_i, Y_j) admits the following density :

$$f_{i,j}(u,v) = \frac{1}{\Delta}\mathbf{E}\left[p_\xi\left(\frac{a_0(u-A) - a_{i-j}(v-B)}{\Delta} \right) p_\xi \left(\frac{a_0(v-B) - a_{i-j}(u-A)}{\Delta} \right) \right]$$

where $A = Y_i - a_0\xi_i - a_{i-j}\xi_j$, $B = Y_j - a_0\xi_j - a_{i-j}\xi_i$ and $\Delta = a_0^2 - a_{i-j}^2$. This implies that Assumption 3 is fulfilled with $\mathfrak{C} = \|p_\xi\|_\infty^2 / (a_0^2 - a_1^2) < +\infty$.

Last, Doukhan and Louhichi (1999) proved that:

$$\rho_r(\mathbb{Y}) \leq 2 \sum_{i \geq [r/2]} a_i \mathbf{E}|\xi_i|^b,$$

where $[r/2]$ is the integer part of $r/2$. This implies that Assumption 2 is fulfilled with $\mathfrak{a} = |\log \tau|/4$, $\mathfrak{b} = 1$ and $\mathfrak{c} = 2a_0\mathbf{E}|\xi_0|^b$.

Infinite moving average (Bernoulli innovations).. We consider the following auto-regressive and non-causal model defined by:

$$Y_i = 2(Y_{i-1} + Y_{i+1})/5 + 5\xi_i/21, \quad i \in \mathbb{Z}, \tag{6}$$

where $(\xi_i)_{i\in\mathbb{Z}}$ is an i.i.d. sequence of Bernoulli variables with parameter $1/2$. A stationary solution of (6) can be written as:

$$Y_i = \sum_{j\in\mathbb{Z}} a_j \xi_{i-j},$$

with $a_j = (3 \cdot 2^{|j|})^{-1}$. This implies that the marginal distribution of Y_i is that of $(U + U' + \xi_0)/3$ where U and U' are independent uniform variables on $[0, 1]$. Thus, Assumption 1 is fulfilled. To prove that Assumption 3 is also fulfilled we write:

$$Y_i = a_0 V + \Sigma_1 + a_{|j-i|} V'$$
$$Y_j = a_{|j-i|} V + \Sigma_2 + a_0 V'$$

where V and V' are independent uniform variables on $[0,1]$, $\Sigma_1 = \sum_{k=0}^{|j-i|} a_k \xi_k$ and $\Sigma_2 = \sum_{k=0}^{|j-i|} a_k \xi_{k-|j-i|}$. This implies that the random pair (Y_i, Y_j) admits the following density function:

$$f_{i,j}(x,y) = \frac{1}{a_0^2 - a_{|j-i|}^2} \mathbf{E}\left[\mathbf{I}_{[0,1]}\left(\Phi(x-\Sigma_1, y-\Sigma_2)\right) \mathbf{I}_{[0,1]}\left(\Phi(y-\Sigma_2, x-\Sigma_1)\right)\right]$$

where $\Phi(z,z') = (a_0 z - a_{j-i} z')(a_0^2 - a_{j-i}^2)^{-1}$. Thus, Assumption 3 holds with $\mathfrak{C} = (a_0^2 - a_1^2)^{-1}$. Finally Assumption 2 is fulfilled using the same arguments as above.

GARCH processes.. Let α, β and γ be positive real numbers. Let also $(\xi_i)_{i\in\mathbb{Z}}$ be i.i.d. innovations with bounded marginal density $p_\xi(\cdot)$. The process $\mathbb{Y} = (Y_i)_{i\in\mathbb{Z}}$ defined by

$$Y_i = \sigma_i \xi_i \qquad \text{with} \qquad \sigma_i^2 = \gamma + \alpha Y_{i-1}^2 + \beta \sigma_{i-1}^2. \tag{7}$$

is called a GARCH$(1,1)$ process. It is well known (see Dedecker et al., 2007) that, as soon as $\alpha < 1$, there exists a stationary solution that admits a marginal density function and which is weakly-dependent with a geometric rate. That is, Assumption 2 is satisfied.

Now, denote by $(\mathscr{F}_i)_{i\in\mathbb{Z}}$ the natural filtration associated with the process $\mathbb{Y}$. Let ϕ be a bounded continuous function. We have:

$$\begin{aligned} \mathbf{E}\phi(Y_i) &= \mathbf{E}\Big(\mathbf{E}(\phi(Y_i) \mid \mathscr{F}_{i-1})\Big) \\ &= \mathbf{E}\Big(\mathbf{E}(\phi(\sigma_i \xi_i) \mid \mathscr{F}_{i-1})\Big). \end{aligned}$$

Now remark that, since $\sigma_i \in \mathscr{F}_{i-1}$ and ξ_i is independent of $\mathscr{F}_{i-1}$, we have:

$$\mathbf{E}\phi(Y_i) = \int_{\mathbb{R}} \phi(x) \mathbf{E}\left(\frac{1}{\sigma_i} p_\xi\left(\frac{x}{\sigma_i}\right)\right) dx.$$

This implies that Assumption 1 is fulfilled with $\mathfrak{B} = \|p_\xi\|_\infty / \sqrt{\gamma}$.

In a similar way, for $i, j \in \mathbb{Z}$ such that $i < j$ and g, $\tilde{g}$ in $\mathbb{G}_1^*$ we have:

$$\begin{aligned} \mathbf{E}\Big(g(Y_i)\tilde{g}(Y_j)\Big) &= \mathbf{E}\Big(g(Y_i)\mathbf{E}\Big(\tilde{g}(\sigma_j \xi_j) \mid \mathscr{F}_i\Big)\Big) \\ &= \mathbf{E}\Big(g(Y_i)\mathbf{E}\Big(\tilde{g}(\sigma_j \xi_j) \mid \mathscr{F}_i\Big)\Big) \\ &= \mathbf{E}\Big(g(Y_i)\mathbf{E}\Big(\mathbf{E}\Big(\tilde{g}(\sigma_j \xi_j) \mid \mathscr{F}_{j-1}\Big) \mid \mathscr{F}_i\Big)\Big). \end{aligned}$$

This easily implies that Assumption 3 is fulfilled with $\mathfrak{C} = \|p_\xi\|_\infty^2 / \gamma$.

Remark. Note that augmented GARCH processes (see Duan, 1997) could also be considered. These processes, that generalize several GARCH-type processes (see Aue et al., 2006), are defined by

$$Y_i = \sigma_i \xi_i \qquad \text{with} \qquad \Lambda(\sigma_i^2) = g(\xi_{i-1}) + c(\xi_{i-1})\Lambda(\sigma_{i-1}^2),$$

where Λ, c and g are real-valued functions and Λ^{-1} exists. Under conditions on the functions c and g given in Aue et al. (2006), there exists a stationary solution. Moreover, following the same lines as above, it can be deduced that Assumptions 1 and 3 are satisfied as soon as $\inf_{x,y} \Lambda^{-1}\left(c(x)\Lambda(y) + g(x)\right) \geq \gamma > 0$.

Regular transformations.. Let $T : \mathbb{R} \to \mathbb{R}$ be a differentiable function and let $y_0 \in \mathbb{R}$. Assume that, for any $y \in \mathbb{R}$, $0 \leq T'(y) \leq \tau$ and $\min_{y \in V_n(y_0)} T'(y) \geq \tau^{-1}$ for some constant $\tau > 1$. Let $\mathbb{Y} = (Y_i)_{i\in\mathbb{Z}}$ be a process such that $\mathscr{L}(\mathbb{Y})$ belongs to $\mathscr{P}_\vartheta$ for some $\vartheta \in \Theta$. We consider the new process $\mathbb{X} = (X_i)_{i\in\mathbb{Z}}$ defined by $X_i = T(Y_i)$ for any $i \in \mathbb{Z}$. Proposition 2.1 in Dedecker et al. (2007) states that the process $\mathbb{X}$ is again weakly-dependent at geometric rate. That is, Assumption 2 is fulfilled for some parameters. Now, if we define

$$V_n(x_0) = [x_0 - \tau^{-1}h^*, x_0 + \tau^{-1}h^*] \subset T(V_n(y_0)),$$

Assumption 1 is fulfilled with the parameter $\tau\mathfrak{B}$. Now remark that, if $g \in \mathbb{G}_1$ then $g \circ T$ also belongs to $\mathbb{G}_1$. This implies that Assumption 3 is fulfilled with the constant $\mathfrak{C}(\tau\mathfrak{B})^2$.

6. Simulation study

In this section, we study the performance of our procedure using simulated data. More precisely, we aim at estimating three density functions, for three types of dependent processes. In each situation, we study the accuracy of our procedure as well as classical competitive methods, calculating empirical risks using $p = 10000$ Monte-Carlo replications. In the following, we detail our simulation scheme and comment the obtained results.

Simulation scheme.. We consider three density functions to be estimated. The first one is:

$$f_1(x) = 1.28\left(\sin((3\pi/2 - 1)x)I_{[0,0.65]}(x) + I_{(0.65,0.95]}(x) + cI_{[0,1]}(x)\right),$$

where c is a positive constant such that f_1 is a density. The second one is the density of a mixture of three normal distributions restricted to the support $[0, 1]$, defined by:

$$f_2(x) = \left(\frac{1}{2}\phi_{0.5,0.1}(x) + \frac{1}{4}\phi_{0.6,0.01}(x) + \frac{1}{4}\phi_{0.65,0.015}(x) + c\right) I_{[0,1]}(x),$$

where $\phi_{\mu,\sigma}$ stands for the density of a normal distribution with mean μ and standard deviation σ and c is a positive constant such that f_2 is a density. Note that very similar densities were also considered in Gannaz and Wintenberger (2010). The third one is:

$$f_3(x) = \sum_{k=1}^{5} \left(2 - 40\left|x - \frac{k}{10} + \frac{1}{20}\right|\right) I_{(\frac{k-1}{10},\frac{k}{10}]}(x) + I_{(0.5,1]}(x).$$

The function f_1 is very smooth except in the discontinuity point $x = 0.65$. The function f_2 is a classical example where rule-of-thumb bandwidths do not work. The third function has several spikes in $[0, 0.5]$ and is constant on $[0.5, 1]$. As a consequence, a global choice of bandwidth can fail to catch the two different behaviors of the function. The three densities are bounded from above (Assumption 1 is then satisfied) and their inverse cumulative distribution functions are Lipschitz.

We simulate data $(X_1, \ldots, X_n)$ with density $f \in \{f_1, f_2, f_3\}$ in three cases of dependence. Denote by F the cumulative distribution function of f.

Case 1. The X_i are independent variables given by $F^{-1}(U_i)$ where the U_i are i.i.d. uniform variables on $[0, 1]$. Assumptions 2 and 3 are clearly satisfied.

Case 2. The X_i are λ–dependent given by $F^{-1}(G(Y_i))$ where Y_i is defined by (6) and G is the cumulative distribution function of Y_i.

Case 3. The X_i are given by $F^{-1}(G(Y_i))$ where $\mathbb{Y} = (Y_i)_{i\in\mathbb{Z}}$ is an ARCH(1) process given by (7) where the $(\xi_i)_{i\in\mathbb{Z}}$ are i.i.d. standard normal variables, $\alpha = 0.5$, $\beta = 0$ and $\gamma = 0.5$. In this case, the function G is estimated using the empirical distribution function of a simulated process $\tilde{\mathbb{Y}}$ independent of $\mathbb{Y}$ with the same distribution.

		Case 1	Case 2	Case 3
f_1	GL	0.036	0.033	0.044
	CV	0.027	0.034	0.049
	RT	0.036	0.040	0.054
f_2	GL	0.181	0.203	0.222
	CV	0.079	0.116	0.162
	RT	0.965	0.975	0.971
f_3	GL	0.090	0.098	0.118
	CV	0.172	0.180	0.190
	RT	0.263	0.266	0.286

Table 1: Mean of $\widehat{ISE}$ for the three densities, the three cases of dependence and the three procedures GL, CV and RT.

Quality criteria.. For each density function $f \in \{f_1, f_2, f_3\}$ and each case of dependence, we simulate $p = 10000$ sequences of observations $(X_1, \ldots, X_n)$ with $n = 1000$. Given an estimation procedure, we calculate p estimators $\hat{f}^{(1)}, \ldots, \hat{f}^{(p)}$. We consider the empirical integrated square error:

$$\widehat{ISE} = \frac{1}{p} \sum_{j=1}^{p} \int_{[0,1]} \left(f(x) - \hat{f}^{(j)}(x) \right)^2 dx.$$

Comparison with other estimation procedures.. We propose to compare in this simulation study the following procedures.

- Our procedure (GL) $\hat{f}$ performed with, in (1), $q = 2$ and $\delta_n = (\log n)^{-1/2}$.
- The leave-one-out cross validation (CV) performed on the family of kernel estimators given in Subsection 3.
- The kernel procedure with bandwidth given by the rule-of-thumb (RT).
- The procedure developed by Gannaz and Wintenberger (2010).

In the first three procedures, we use the uniform kernel. Our results are summarized in Table 1.

		$\phi(n) = n$	$\phi(n) = \lfloor n^{3/2} \rfloor$	$\phi(n) = n^2$
f_1	Case 2	0.033	0.034	0.037
	Case 3	0.044	0.042	0.038
f_2	Case 2	0.203	0.192	0.187
	Case 3	0.222	0.220	0.191
f_3	Case 2	0.098	0.091	0.092
	Case 3	0.118	0.109	0.099

Table 2: Mean of $\widehat{ISE}$ for GL procedure for the three densities, case 2 and case 3 of dependence and 3 types of subsampling.

For the estimation of the function f_1, our procedure gives better results than the CV or RT methods in cases of dependence (2 and 3). We also outperform the results of Gannaz and Wintenberger (2010) for case 1 and 2 where the ISE was around 0.09 (case 3 was not considered in their paper). For the estimation of f_2 which is quite smooth, the cross validation method is about two times better than GL method and as expected, the RT method does not work. For the estimation of f_3 that contains several peaks, the GL procedure is about two times better than the CV method. We also apply our procedure with Epanechnikov and Gaussian kernels and the obtained results are quite similar. Even if the Gaussian kernel does not satisfy the assumption of compact support, its performances are slightly better and it could be used for practical purposes.

Sensitivity to non-stationarity. We simulate non-stationary processes by subsampling the stationary processes considered in the previous paragraph for the 3 functions and case 2 and case 3. More precisely, the processes $X_{\phi(n)}$ are generated for three cases of functions $\phi(n)$ and our results for $p = 10000$ Monte-Carlo replications and sample size 1000 are obtained in Table 2. As expected, the behavior of our estimator is not oversensitive to subsampling.

Conclusion.. To conclude, in the considered examples, our procedure has similar or better performances than already existing methods used for dependent data. Moreover, it gives better results when the density function to be estimated presents irregularities. This illustrates the fact that our method adapts locally to the irregularities of the function thanks to the use of local bandwidths. An other important point is that the choice of the bandwidth depends on explicit constants that can be used directly in practice and do not need previous calibration. Additionally, our GL procedure is about 25 times faster than cross-validation.

7. Proofs

7.1. Basic notation

Here and later we fix $\vartheta \in \Theta$ and $\mathbb{X}$ is a process such that $\mathscr{L}(\mathbb{X}) \in \mathscr{P}_\vartheta$. For the sake of readability, we introduce some conventions and notations used throughout the proofs.

For any $h \in \mathscr{H}_n$ and $y \in \mathbb{R}$ we define $g_h(y) = K_h(x_0 - y)$ and

$$\bar{g}_h(y) = \frac{g_h(y) - \mathbf{E} g_h(X_1)}{n},$$

so that $\hat{f}_h(x_0) - \mathbf{E}\hat{f}_h(x_0) = \sum_{i=1}^n \bar{g}_h(X_i)$. We also introduce some constants:

$$C_1 = \|K\|_1^2(\mathfrak{B}^2 + \mathfrak{C}), \qquad C_2 = \mathfrak{B}\|K\|_2^2, \qquad C_3 = 2\|K\|_\infty,$$

$$L = \mathrm{Lip}(K), \quad C_4 = 2C_3L + L^2, \qquad C_5 = (2C_1)^{3/4} C_4^{1/4},$$

and:

$$C_6 = \sup_{n\geq 2}\left(2C_1\delta_n^{-2}(h^*)^{3/4} + (6C_5)\left(\sum_{r=1}^{\infty}\rho_r^{1/8}\right)\mathfrak{c}^{1/8}\delta_n^{-2}\exp\Big(-\frac{\mathfrak{a}}{8}\exp\Big(\frac{\mathfrak{b}}{4}\sqrt{\log n}\Big)\Big)\right)$$

which is a finite constant.

7.2. Auxiliary results

Lemma 1. *For any $h \in \mathscr{H}_n$, we define*

$$D_1(h) = D_1(n,h) = \frac{C_3}{nh} \qquad \text{and} \qquad D_2(h) = D_2(n,h) = \frac{C_5}{n^2 h}.$$

Then for any u, v and r in $\mathbb{N}$, if $(i_1, \ldots, i_u, j_1, \ldots j_v) \in \mathbb{Z}^{u+v}$ is such that $i_1 \leq \ldots, i_u \leq i_u + r \leq j_1 \leq \ldots \leq j_v$, we have

$$\left|\mathrm{Cov}\left(\prod_{k=1}^{u}\bar{g}_h(X_{i_k}), \prod_{m=1}^{v}\bar{g}_h(X_{j_m})\right)\right| \leq \Phi(u,v) D_1^{u+v-2}(h) D_2(h)\rho_r^{1/4},$$

where $\Phi(u,v) = u + v + uv$.

Lemma 2. *We have, for any $n \in \mathbb{N} \setminus \{0,1\}$:*

$$\mathbf{E}\left(\left|\sum_{i=1}^{n} \bar{g}_h(X_i)\right|^2\right) \leq J_n(h) + \frac{C_6 \delta_n^2}{nh} \leq \frac{C_2 + C_6 \delta_n^2}{nh}, \tag{8}$$

where J_n is defined by (3). Moreover for $\mathfrak{q} > 0$, we have

$$\mathbf{E}\left(\left|\sum_{i=1}^{n} \bar{g}_h(X_i)\right|^{\mathfrak{q}}\right) \leq \mathfrak{C}_{\mathfrak{q}} (nh)^{-\mathfrak{q}/2}, \tag{9}$$

where $\mathfrak{C}_{\mathfrak{q}}$ is a positive constant that depends on ϑ, $\mathfrak{q}$ and K.

Lemma 3 (Bernstein's inequality)**.** *We have:*

$$\mathbf{P}\left(\left|\sum_{i=1}^{n} \bar{g}_h(X_i)\right| \geq \lambda(t)\right) \leq \exp(-t/2)$$

where,

$$\lambda(t) = \sigma_n(h)\sqrt{2t} + B_n(h)\,(2t)^{2+1/\mathfrak{b}} \qquad t \geq 0 \tag{10}$$

$$\sigma_n(h) = J_n(h) + \frac{C_6 \delta_n^2}{nh}, \tag{11}$$

$$B_n(h) = \frac{C_7}{nh} \tag{12}$$

and C_7 is a positive constant that depends only on ϑ and K.

7.3. Proof of Theorem 1

Let us denote $\gamma_n = q(1 + (4C_2)^{-1}\delta_n)$. For convenience, we split the proof into several steps.

Step 1. Let us consider the random event

$$\mathscr{A} = \bigcap_{h \in \mathscr{H}_n} \left\{ \left| \widehat{J}_n(h) - J_n(h) \right| \leq \frac{\delta_n}{2nh} \right\}$$

and the quantities Γ_1 and Γ_2 defined by :

$$\Gamma_1 = \mathbf{E}\left|\hat{f}(x_0) - f(x_0)\right|^q \mathbf{I}_{\mathscr{A}}$$

and

$$\Gamma_2 = \left(\max_{h \in \mathscr{H}_n} R_{2q}^{2q}(\hat{f}_h, f)\mathbf{P}(\mathscr{A}^c)\right)^{1/2}$$

where $\mathbf{I}_{\mathscr{A}}$ is the characteristic function of the set $\mathscr{A}$. Using Cauchy-Schwarz inequality, it follows that:

$$R_q^q(\hat{f}, f) \leq \Gamma_1 + \Gamma_2.$$

We define

$$\mathfrak{M}_n(h, a) = \sqrt{2q|\log h| \left(J_n(h) + \frac{a\delta_n}{nh} \right)}.$$

Now note that if the event $\mathscr{A}$ holds, we have:

$$\mathfrak{M}_n\left(h, \frac{1}{2}\right) \leq \widehat{M}_n(h) \leq \mathfrak{M}_n\left(h, \frac{3}{2}\right).$$

Step 2. Let $h \in \mathscr{H}_n$ be an arbitrary bandwidth. Using triangular inequality we have:

$$|\hat{f}(x_0) - f(x_0)| \leq |\hat{f}_{\hat{h}}(x_0) - \hat{f}_{h\vee\hat{h}}(x_0)| + |\hat{f}_{h\vee\hat{h}}(x_0) - \hat{f}_h(x_0)| + |\hat{f}_h(x_0) - f(x_0)|.$$

If $h \vee \hat{h} = \hat{h}$, the first term of the right hand side of this inequality is equal to 0, and if $h \vee \hat{h} = h$, it satisfies

$$\begin{aligned}
|\hat{f}_{\hat{h}}(x_0) - \hat{f}_{h\vee\hat{h}}(x_0)| &\leq \left\{|\hat{f}_{\hat{h}}(x_0) - \hat{f}_{h\vee\hat{h}}(x_0)| - \widehat{M}_n(h, \hat{h})\right\}_+ + \widehat{M}_n(h, \hat{h}) \\
&\leq \max_{\mathfrak{h}\in\mathscr{H}_n} \left\{|\hat{f}_{\mathfrak{h}}(x_0) - \hat{f}_{h\vee\mathfrak{h}}(x_0)| - \widehat{M}_n(h, \mathfrak{h})\right\}_+ + \widehat{M}_n(h, \hat{h}) \\
&\leq A(h, x_0) + \widehat{M}_n(\hat{h}) + \widehat{M}_n(h).
\end{aligned}$$

Applying the same reasoning to the term $|\hat{f}_{h\vee\hat{h}}(x_0) - \hat{f}_h(x_0)|$ and using (2), this leads to

$$|\hat{f}(x_0) - f(x_0)| \leq 2\Big(A(h, x_0) + \widehat{M}_n(h)\Big) + |\hat{f}_h(x_0) - f(x_0)|.$$

Using this equation, we obtain that

$$\Gamma_1 \leq 3^{q-1}\left(2^q\mathbf{E}\Big(A^q(h, x_0)\mathbf{I}_{\mathscr{A}}\Big) + 2^q\mathfrak{M}_n^q\left(h, \frac{3}{2}\right) + R_q^q(\hat{f}_h, f))\right). \tag{13}$$

Step 3. Now, we upper bound $A(h, x_0)$. Using basic inequalities we have:

$$\begin{aligned}
A(h, x_0) \leq \max_{\mathfrak{h}\in\mathscr{H}_n} \left\{\left|\mathbf{E}\hat{f}_{h\vee\mathfrak{h}}(x_0) - \mathbf{E}\hat{f}_{\mathfrak{h}}(x_0)\right|\right\}_+ + \max_{\mathfrak{h}\in\mathscr{H}_n} \left\{\left|\hat{f}_{\mathfrak{h}}(x_0) - \mathbf{E}\hat{f}_{\mathfrak{h}}(x_0)\right| - \widehat{M}_n(\mathfrak{h})\right\}_+ \\
+ \max_{\mathfrak{h}\in\mathscr{H}_n} \left\{\left|\hat{f}_{h\vee\mathfrak{h}}(x_0) - \mathbf{E}\hat{f}_{h\vee\mathfrak{h}}(x_0)\right| - \widehat{M}_n(h\vee\mathfrak{h})\right\}_+ \\
\leq \max_{\mathfrak{h}\in\mathscr{H}_n} \left\{\left|\mathbf{E}\hat{f}_{h\vee\mathfrak{h}}(x_0) - \mathbf{E}\hat{f}_{\mathfrak{h}}(x_0)\right|\right\}_+ + 2T,
\end{aligned}$$

where

$$T = \max_{\mathfrak{h}\in\mathscr{H}_n}\left\{\left|\hat{f}_{\mathfrak{h}}(x_0) - \mathbf{E}\hat{f}_{\mathfrak{h}}(x_0)\right| - \widehat{M}_n(\mathfrak{h})\right\}_+ = \max_{\mathfrak{h}\in\mathscr{H}_n}\left\{\left|\sum_{i=1}^n \bar{g}_{\mathfrak{h}}(X_i)\right| - \widehat{M}_n(\mathfrak{h})\right\}_+.$$

Then, we obtain

$$A(h,x_0) \leq 2E_h(x_0) + 2T,$$

which leads to:

$$\mathbf{E}\left(A^q(h,x_0)\mathbf{I}_{\mathscr{A}}\right) \leq 2^{2q-1}\left(E_h^q(x_0) + \mathbf{E}\Big(T^q\mathbf{I}_{\mathscr{A}}\Big)\right). \tag{14}$$

Step 4. It remains to upper bound $\mathbf{E}(T^q\mathbf{I}_{\mathscr{A}})$. To this aim, notice that,

$$\mathbf{E}(T^q\mathbf{I}_{\mathscr{A}}) \leq \mathbf{E}\tilde{T}^q,$$

where

$$\tilde{T} = \max_{\mathfrak{h}\in\mathscr{H}_n}\left\{\left|\sum_{i=1}^n \bar{g}_{\mathfrak{h}}(X_i)\right| - \mathfrak{M}_n\left(\mathfrak{h},\frac{1}{2}\right)\right\}_+.$$

Now, we define $r(\cdot)$ by

$$r(u) = \sqrt{2\sigma_n^2(h)u} + 2^{\mathfrak{d}-1}B_n(h)\,(2u)^{\mathfrak{d}}, \qquad u \geq 0$$

where $B_n(h)$ and $\sigma_n(h)$ are given by (11) and (12) and $\mathfrak{d} = 2 + \mathfrak{b}^{-1}$. Since $h \geq h_* = n^{-1}\exp(\sqrt{\log n})$, we have, for n large enough:

$$2^{\mathfrak{d}-1}B_n(h)(2\gamma_n|\log h|)^{\mathfrak{d}} \leq \delta_n\sqrt{\frac{2q|\log h|}{nh}}. \tag{15}$$

Moreover, we have, for n large enough:

$$\begin{aligned}\sqrt{2\sigma_n^2(h)\gamma_n|\log h|} &\leq \sqrt{2q|\log h|\left(J_n(h) + \frac{C_6\delta_n^2}{nh}\right)\left(1 + \frac{\delta_n}{4C_2}\right)} \\ &\leq \sqrt{2q|\log h|\left(J_n(h) + \frac{\delta_n}{3nh}\right)}.\end{aligned} \tag{16}$$

Last inequality comes from the fact that $J_n(h)$ is upper bounded by $C_2/(nh)$. Now, using (15) and (16), we obtain, for n large enough:

$$r(\gamma_n|\log h|) \leq \mathfrak{M}_n\left(h,\frac{1}{2}\right). \tag{17}$$

Thus, doing the change of variables $t=(r(u))^q$ and thanks to (17), we obtain for n large enough:

$$\begin{aligned}\mathbf{E}\tilde{T}^q &\leq \sum_{\mathfrak{h}\in\mathscr{H}_n}\int_0^\infty \mathbf{P}\left(\left|\sum_{i=1}^n \bar{g}_{\mathfrak{h}}(X_i)\right| \geq \mathfrak{M}_n\left(\mathfrak{h},\frac{1}{2}\right)+t^{1/q}\right)dt\\ &\leq q\sum_{\mathfrak{h}\in\mathscr{H}_n}\int_0^\infty r'(u)r(u)^{q-1}\mathbf{P}\left(\left|\sum_{i=1}^n \bar{g}_{\mathfrak{h}}(X_i)\right|\geq r(\gamma_n|\log\mathfrak{h}|)+r(u)\right)du,\\ &\leq q\mathfrak{d}2^{(\mathfrak{d}-1)q}\sum_{\mathfrak{h}\in\mathscr{H}_n}\int_0^\infty u^{-1}\lambda(u)^q\mathbf{P}\left(\left|\sum_{i=1}^n\bar{g}_{\mathfrak{h}}(X_i)\right|\geq\lambda(\gamma_n|\log\mathfrak{h}|+u)\right)du,\end{aligned}$$

where $\lambda(\cdot)$ is defined by (10). Now, combining the above inequality with Lemma 3, we obtain for all n:

$$\begin{aligned}\mathbf{E}\tilde{T}^q &\leq \varkappa_1\sum_{\mathfrak{h}\in\mathscr{H}_n}\int_0^\infty u^{-1}\left(\sqrt{\sigma_n^2(\mathfrak{h})u}+B_n(\mathfrak{h})u^3\right)^q\exp\left\{-\frac{u}{2}-\frac{\gamma_n|\log\mathfrak{h}|}{2}\right\}du\\ &\leq \varkappa_2\sum_{\mathfrak{h}\in\mathscr{H}_n}\sigma_n^q(\mathfrak{h})\mathfrak{h}^{\gamma_n/2}\leq\varkappa_3 n^{-q/2}\delta_n^{-1}\end{aligned}\tag{18}$$

where $\varkappa_1$, $\varkappa_2$ and $\varkappa_3$ are absolute constants depending only on ϑ and K.
Step 5. Lemma 2 implies that:

$$\mathbf{E}|\hat{f}_h(x_0)-f(x_0)|^q\leq 2^{q-1}\left(|E_h(x_0)|^q+\mathfrak{C}_q(nh)^{-q/2}\right).\tag{19}$$

Combining (13), (14), (18) and (19), we have:

$$\Gamma_1\leq\min_{h\in\mathscr{H}_n}\left\{C_1^*\max_{\substack{\mathfrak{h}\leq h\\ \mathfrak{h}\in\mathscr{H}_n}}|K_{\mathfrak{h}}\star f(x_0)-f(x_0)|^q+C_2^*\left(\frac{|\log h|}{nh}\right)^{q/2}\right\}+\varkappa_3 n^{-q/2}\delta_n^{-1}\tag{20}$$

where

$$C_1^*=3^{q-1}2^{q-1}(1+2^{2q})\text{ and }\quad C_2^*=3^{q-1}\left(2^q\left(2q\left(C_2+\frac{3}{2}\right)\right)^{q/2}+2^{q-1}\mathfrak{C}_q\right).\tag{21}$$

Step 6. Using Lemma 3 where in $\bar{g}_h$, K is replaced by K^2, we obtain that

$$\mathbf{P}\left(\left|\frac{h}{n}\sum_{i=1}^n K_h^2(x_0-X_i)-h\int K_h^2(x_0-x)f(x)dx\right|>\frac{\delta_n}{2}\right)\leq\exp(-\varkappa_4(nh)^{1/\mathfrak{d}}\delta_n^2),$$

where $\varkappa_4$ is a constant that depends only on ϑ and K. Since $\#(\mathscr{H}_n) \leq \log n / \log 2$ and

$$(nh)^{1/\mathfrak{d}} \delta_n^2 \geq \frac{\exp\left(- \sqrt{\log n}\right)}{\log n},$$

we deduce that

$$\Gamma_2 \leq \varkappa_5 n^{-q/2} \tag{22}$$

where $\varkappa_5$ is a positive constant that depends only on ϑ and K. Now, using (20) and (22), Theorem 1 follows.

Appendix A. Proof of Lemma 1

In order to prove this lemma, we derive two different bounds for the term

$$\Upsilon_h(u,v) = \left| \mathrm{Cov}\left(\prod_{k=1}^{u} \bar{g}_h(X_{i_k}), \prod_{m=1}^{v} \bar{g}_h(X_{j_m}) \right) \right|.$$

The first bound is obtained by a direct calculation whereas the second one is obtained thanks to the dependence structure of the observations. For the sake of readability, we denote $\ell = u + v$ throughout this proof.

Direct bound. The proof of this bound is composed of two steps. First, we assume that $\ell = 2$, then the general case $\ell \geq 3$ is considered.

Assume that $\ell = 2$. If Assumptions 1 and 3 are fulfilled, we have

$$\begin{aligned} n^2 \Upsilon_h(u,v) &\leq |\mathbf{E}\left(g_h(X_i) g_h(X_j)\right)| + \left(\mathbf{E} g_h(X_1)\right)^2 \\ &\leq (\mathfrak{C} + \mathfrak{B}^2) \|g_h\|_1^2 \leq C_1. \end{aligned}$$

Then, we have

$$|\mathrm{Cov}\left(\bar{g}_h(X_i), \bar{g}_h(X_j)\right)| \leq C_1 n^{-2}. \tag{A.1}$$

Let us now assume that $\ell \geq 3$. Without loss of generality, we can assume that $u \geq 2$ and $v \geq 1$. We have:

$$\Upsilon_h(u,v) \leq A + B$$

where

$$\begin{cases} A = \mathbf{E}\left(\prod_{k=1}^{u} \bar{g}_h(X_{i_k}) \prod_{m=1}^{v} \bar{g}_h(X_{j_m}) \right) \\ B = \mathbf{E}\left(\prod_{k=1}^{u} \bar{g}_h(X_{i_k}) \right) \mathbf{E}\left(\prod_{m=1}^{v} \bar{g}_h(X_{j_m}) \right). \end{cases}$$

Remark that both A and B can be bounded, using (A.1), by

$$\|\bar{g}_h\|_\infty^{(u-2)+v}\,\mathrm{Cov}(\bar{g}_h(X_{i_1}),\bar{g}_h(X_{i_2})) \le \left(\frac{C_3}{nh}\right)^{\ell-2}\frac{C_1}{n^2}.$$

This implies our first bound, for all $\ell \ge 2$:

$$\Upsilon_h(u,v) \le \frac{2C_1}{n^2}\left(\frac{C_3}{nh}\right)^{\ell-2}. \tag{A.2}$$

Structural bound. Using Assumption 2, we obtain that

$$\Upsilon_h(u,v) \le \Psi\left(u,v,\bar{g}_h^{\otimes u},\bar{g}_h^{\otimes v}\right)\rho_r.$$

Now using that

$$\left\|\frac{nh\bar{g}_h}{C_3}\right\|_\infty \le 1$$

and

$$\mathrm{Lip}\left(\frac{nh\bar{g}_h}{C_3}\right)^{\otimes u} \le \mathrm{Lip}\left(\frac{nh\bar{g}_h}{C_3}\right) \le \frac{L}{C_3 h},$$

we obtain, since $h \le h^*$, that

$$\Upsilon_h(u,v) \le \left(\frac{C_3}{nh}\right)^{\ell}\Phi(u,v)\frac{C_4}{C_3^2h^2}\rho_r.$$

This implies that

$$\Upsilon_h(u,v) \le \frac{1}{n^2}\left(\frac{C_3}{nh}\right)^{\ell-2}\frac{C_4}{h^4}\Phi(u,v)\rho_r. \tag{A.3}$$

Conclusion. Now combining (A.2) and (A.3) we obtain:

$$\begin{aligned}\Upsilon_h(u,v) &\le \frac{1}{n^2}\left(\frac{C_3}{nh}\right)^{\ell-2}(2C_1)^{3/4}\left(\frac{C_4}{h^4}\Phi(u,v)\rho_r\right)^{1/4}\\ &\le \frac{C_5}{n^2h}\left(\frac{C_3}{nh}\right)^{\ell-2}\Phi(u,v)\rho_r^{1/4}.\end{aligned}$$

This proves Lemma 1.

Appendix B. Proof of Lemma 2

We first prove that (8) holds. Let $n \geq 3$ be an arbitrary integer. Then,

$$\begin{aligned}\mathbf{E}\left(\sum_{i=1}^{n}\bar{g}_h(X_i)\right)^2 &= n\mathbf{E}\bar{g}_h(X_1)^2 + \sum_{i\neq j}\mathbf{E}\bar{g}_h(X_i)\bar{g}_h(X_j)\\ &= J_n(h) + 2\sum_{i=1}^{n-1}\sum_{r=1}^{n-i}\mathbf{E}\bar{g}_h(X_i)\bar{g}_h(X_{i+r}).\end{aligned}$$

Let R be the integer part of $h^{-1/4}$. Using Lemma 1 and (A.1) we obtain:

$$\begin{aligned}\mathbf{E}\left(\sum_{i=1}^{n}\bar{g}_h(X_i)\right)^2 &\leq J_n(h) + 2n\sum_{r=1}^{R}\frac{C_1}{n^2} + 2D_2(h)\sum_{r=R+1}^{n-1}(n-r)\Phi(1,1)\rho_r^{1/4}\\ &\leq J_n(h) + \frac{1}{nh}\left((2C_1)h^{3/4} + (6C_5)\sum_{r=R+1}^{\infty}\rho_r^{1/4}\right)\\ &\leq J_n(h) + \frac{\delta_n^2}{nh}\left(2C_1\delta_n^{-2}(h^*)^{3/4} + (6C_5)\sum_{r=1}^{\infty}\rho_r^{1/8}\left(\rho_{R+1}^{1/8}\delta_n^{-2}\right)\right)\\ &\leq J_n(h) + \frac{C_6\delta_n^2}{nh}.\end{aligned}$$

This equation, combined with the fact that $J_n(h) \leq C_2(nh)^{-1}$, completes the proof of (8). Now, we prove (9). Set $\mathfrak{q} > 0$ and define $k_{\mathfrak{q}}$ as the smallest integer greater than or equal to $\mathfrak{q}$. Using Jensen 's inequality and basic calculations, we have

$$\mathbf{E}\left(\left|\sum_{i=1}^{n}\bar{g}_h(X_i)\right|^{\mathfrak{q}}\right) \leq \left((2k_{\mathfrak{q}})!A_{2k_{\mathfrak{q}}}\right)^{\mathfrak{q}/(2k_{\mathfrak{q}})}$$

where, for any integer ℓ greater than 1, we define

$$A_\ell = \sum_{t\in\mathscr{T}_\ell}\mathbf{E}|\bar{g}_h(X_{t_1})\cdots\bar{g}_h(X_{t_\ell})|.$$

Here $\mathscr{T}_\ell$ is the set of all $t = (t_1, \ldots, t_\ell) \in \mathbb{N}^\ell$ such that $1 \leq t_1 \leq \ldots \leq t_\ell \leq n$. Following the proof of Theorem 1 in Doukhan and Louhichi (2001) we obtain:

$$A_\ell \leq \frac{C_\ell^*}{(nh)^{\ell-1}} + \sum_{m=2}^{\ell-2}A_mA_{\ell-m}$$

where

$$C_\ell^* = \left(\ell + \frac{\ell^2}{2}\right) C_3^{\ell-2} C_5 \sum_{r\geq 1} (1+r)^{\ell-2} \rho_r^{1/4}$$

is a constant that depends only on ℓ, ϑ and K. Using (8) we readily obtain that:

$$A_2 \leq \mathbf{E}\left(\sum_{i=1}^n \bar{g}_h(X_i)\right)^2 \leq \frac{C_2 + C_6\delta_n^2}{nh}.$$

Moreover, by induction on ℓ, we prove that there exists an absolute constant C_ℓ^{**} that depends on the same parameters as C_ℓ^* such that, for odd ℓ:

$$A_\ell \leq C_\ell^{**}\left(\frac{1}{nh}\right)^{\frac{\ell+1}{2}}$$

and for even ℓ greater than 3:

$$A_\ell \leq a_\ell \left(\frac{C_2 + C_6\delta_n^2}{nh}\right)^{\ell/2} + C_\ell^{**}\left(\frac{1}{nh}\right)^{\ell/2+1}$$

where a_ℓ is such that $a_2 = 1$ and

$$a_\ell = \sum_{\substack{m=2 \\ m \text{ even}}}^{\ell-2} a_m a_{\ell-m}.$$

This implies that

$$\mathbf{E}\left(\left|\sum_{i=1}^n \bar{g}_h(X_i)\right|^{\mathfrak{q}}\right) \leq \mathfrak{C}_{\mathfrak{q}}\left(\frac{1}{nh}\right)^{q/2}$$

where

$$\mathfrak{C}_{\mathfrak{q}} = \left((2\mathfrak{q}+1)!\left(a_{2k_{\mathfrak{q}}}(C_2 + C_6)^{k_{\mathfrak{q}}} + C_{2k_{\mathfrak{q}}}^{**}\right)\right)^{1/2}.$$

Appendix C. Proof of Lemma 3

First, let us remark that Lemma 6.2 in Gannaz and Wintenberger (2010) and Assumption 2 imply that there exist positive constants L_1 and L_2 (that depend on $\mathfrak{a}$, $\mathfrak{b}$ and $\mathfrak{c}$) such that, for any $k \in \mathbb{N}$ we have,

$$\sum_{r\in\mathbb{N}} (1+r)^k \rho_r^{1/4} \leq L_1 L_2^k (k!)^{1/\mathfrak{b}}.$$

This implies that, using Lemma 2, one can apply the Bernstein-type inequality obtained by Doukhan and Neumann (2007, see Theorem 1). First, remark that, using Lemma 2, we have

$$\mathbf{E}\left(\sum_{i=1}^{n}\bar{g}_h(X_i)\right)^2 \le \sigma_n(h) \quad \text{and} \quad B_n(h) = \frac{2L_2C_3}{nh}.$$

where the theoretical expression of $B_n(h)$ is given in Doukhan and Neumann (2007). Let us now denote $\mathfrak{d} = 2 + \mathfrak{b}^{-1}$. We obtain:

$$\mathbf{P}\left(\left|\sum_{i=1}^{n}\bar{g}_h(X_i)\right| \ge u\right) \le \exp\left(-\frac{u^2/2}{\sigma_n^2(h) + B_n^{\frac{1}{\mathfrak{d}}}(h)u^{\frac{2\mathfrak{d}-1}{\mathfrak{d}}}}\right).$$

Now, let us remark that, on the one hand $\lambda(t) \ge \sigma_n(h)\sqrt{2t}$ and thus $\lambda^2(t) \ge 2\sigma_n^2(h)t$. On the other hand, $\lambda^{2+\frac{1-2\mathfrak{d}}{\mathfrak{d}}}(t) \ge 2B_n^{\frac{1}{\mathfrak{d}}}(h)t$ and thus $\lambda^2(t) \ge (2B_n^{\frac{1}{\mathfrak{d}}}(h)t)\lambda^{\frac{2\mathfrak{d}-1}{a\mathfrak{d}}}(t)$. This implies that $\lambda^2(t) \ge t(\sigma_n^2(h)+B_n^{\frac{1}{\mathfrak{d}}}(h)\lambda^{\frac{2\mathfrak{d}-1}{\mathfrak{d}}}(t))$ and thus, finally:

$$\exp\left(-\frac{\lambda^2(t)}{\sigma_n^2(h) + B_n^{\frac{1}{\mathfrak{d}}}(h)\lambda^{\frac{2\mathfrak{d}-1}{\mathfrak{d}}}(t)}\right) \le \exp(-t/2).$$

This implies the results.

Acknowledgments

The authors have been supported by Fondecyt project 1141258, Mathamsud project 16-MATH-03 SIDRE and ECOS project C15E05.

References

Aue, A., Berkes, I., Horváth, L., et al., 2006. Strong approximation for the sums of squares of augmented garch sequences. Bernoulli 12 (4), 583–608.

Bertin, K., Lacour, C., Rivoirard, V., 2014. Adaptive estimation of conditional density function. To appear in Annales de l'Institut Henri Poincaré. Probabilités et Statistiques.
URL http://arxiv.org/abs/1312.7402

Butucea, C., 2000. Two adaptive rates of convergence in pointwise density estimation. Math. Methods Statist. 9 (1), 39–64.

Butucea, C., 2001. Exact adaptive pointwise estimation on sobolev classes of densities. ESAIM: Probability and Statistics 5, 1–31.

Comte, F., Genon-Catalot, V., 2012. Convolution power kernels for density estimation. J. Statist. Plann. Inference 142 (7), 1698–1715.
URL http://dx.doi.org/10.1016/j.jspi.2012.02.038

Comte, F., Merlevède, F., 2002. Adaptive estimation of the stationary density of discrete and continuous time mixing processes. ESAIM Probab. Statist. 6, 211–238, new directions in time series analysis (Luminy, 2001).
URL http://dx.doi.org/10.1051/ps:2002012

Dedecker, J., Doukhan, P., Lang, G., León R., J. R., Louhichi, S., Prieur, C., 2007. Weak dependence: with examples and applications. Vol. 190 of Lecture Notes in Statistics. Springer, New York.

Doukhan, P., Louhichi, S., 1999. A new weak dependence condition and applications to moment inequalities. Stochastic Process. Appl. 84 (2), 313–342.
URL http://dx.doi.org/10.1016/S0304-4149(99)00055-1

Doukhan, P., Louhichi, S., 2001. Functional estimation of a density under a new weak dependence condition. Scand. J. Statist. 28 (2), 325–341.
URL http://dx.doi.org/10.1111/1467-9469.00240

Doukhan, P., Neumann, M. H., 2007. Probability and moment inequalities for sums of weakly dependent random variables, with applications. Stochastic Process. Appl. 117 (7), 878–903.
URL http://dx.doi.org/10.1016/j.spa.2006.10.011

Doukhan, P., Wintenberger, O., 2007. An invariance principle for weakly dependent stationary general models. Probab. Math. Statist. 27 (1), 45–73.

Doumic, M., Hoffmann, M., Reynaud-Bouret, P., Rivoirard, V., 2012. Nonparametric Estimation of the Division Rate of a Size-Structured Population. SIAM J. Numer. Anal. 50 (2), 925–950.
URL http://dx.doi.org/10.1137/110828344

Duan, J.-C., 1997. Augmented garch (p, q) process and its diffusion limit. Journal of Econometrics 79 (1), 97–127.

Gannaz, I., Wintenberger, O., 2010. Adaptive density estimation under weak dependence. ESAIM Probab. Stat. 14, 151–172.
URL http://dx.doi.org/10.1051/ps:2008025

Goldenshluger, A., Lepski, O., 2008. Universal pointwise selection rule in multivariate function estimation. Bernoulli 14 (4), 1150–1190.
URL http://dx.doi.org/10.3150/08-BEJ144

Goldenshluger, A., Lepski, O., 2011. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. Ann. Statist. 39 (3), 1608–1632.
URL http://dx.doi.org/10.1214/11-AOS883

Goldenshluger, A., Lepski, O., 2014. On adaptive minimax density estimation on R^d. Probab. Theory Related Fields 159 (3-4), 479–543.
URL http://dx.doi.org/10.1007/s00440-013-0512-1

Hasminskii, R., Ibragimov, I., 1990. On density estimation in the view of Kolmogorov's ideas in approximation theory. Ann. Statist. 18 (3), 999–1010.
URL http://dx.doi.org/10.1214/aos/1176347736

Klutchnikoff, N., 2014. Pointwise adaptive estimation of a multivariate function. Math. Methods Statist. 23 (2), 132–150.
URL http://dx.doi.org/10.3103/S1066530714020045

Lepski, O. V., 1990. A problem of adaptive estimation in Gaussian white noise. Teor. Veroyatnost. i Primenen. 35 (3), 459–470.
URL http://dx.doi.org/10.1137/1135065

Merlevède, F., Peligrad, M., Rio, E., 2009. Bernstein inequality and moderate deviations under strong mixing conditions. In: High dimensional probability V: the Luminy volume. Vol. 5 of Inst. Math. Stat. Collect. Inst. Math. Statist., Beachwood, OH, pp. 273–292.
URL http://dx.doi.org/10.1214/09-IMSCOLL518

Parzen, E., 1962. On estimation of a probability density function and mode. The annals of mathematical statistics 33 (3), 1065–1076.

Ragache, N., Wintenberger, O., 2006. Convergence rates for density estimators of weakly dependent time series. In: Dependence in probability and statistics. Vol. 187 of Lecture Notes in Statist. Springer, New York, pp. 349–372.
URL http://dx.doi.org/10.1007/0-387-36062-X_16

Rebelles, G., 2015. Pointwise adaptive estimation of a multivariate density under independence hypothesis. Bernoulli 21 (4), 1984–2023.
URL http://dx.doi.org/10.3150/14-BEJ633

Rio, E., 2000. Théorie asymptotique des processus aléatoires faiblement dépendants. Vol. 31 of Mathématiques & Applications (Berlin) [Mathematics & Applications]. Springer-Verlag, Berlin.

Rosenblatt, M., 1956. A central limit theorem and a strong mixing condition. Proc. Nat. Acad. Sci. U. S. A. 42, 43–47.

Rosenblatt, M., et al., 1956. Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics 27 (3), 832–837.

Tribouley, K., Viennet, G., 1998. $\mathbf{L}_p$ adaptive density estimation in a β mixing framework. Ann. Inst. H. Poincaré Probab. Statist. 34 (2), 179–208.
URL http://dx.doi.org/10.1016/S0246-0203(98)80029-0

Tsybakov, A. B., 1998. Pointwise and sup-norm sharp adaptive estimation of functions on the sobolev classes. The Annals of Statistics 26 (6), 2420–2469.

Tsybakov, A. B., 2009. Introduction to nonparametric estimation. Springer Series in Statistics. Springer, New York, revised and extended from the 2004 French original, Translated by Vladimir Zaiats.
URL http://dx.doi.org/10.1007/b13794