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

Karine Bertin* Nicolas Klutchnikoff†

July 21, 2014

Abstract

This paper is devoted to the estimation of the common marginal density function of weakly dependent stationary 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: classical econometrics models such as GARCH as well as dynamical systems and 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, Hölder spaces, Kernel estimation, Oracle inequality, Weakly dependent processes

1 Introduction

Let $\mathbb{X} = (X_i)_{i\in\mathbb{Z}}$ be a real-valued discrete time process admitting a common marginal density $f : \mathbb{R} \to \mathbb{R}$. The aim of this paper is to estimate f on the basis of an observation $\mathcal{X}_n = (X_1, \dots, X_n)$ of size $n \in \mathbb{N}^*$.

In their paper, Gannaz and Wintenberger (2010), proposed a nonlinear wavelet estimator and investigated its theoretical properties in the context of weak dependent data: under mild assumptions (in particular on the dependence), they proved that the proposed estimator achieves near minimax rates of convergence over a scale of Besov balls for integrated risks. In other

*CIMFAV, Universidad de Valparaíso, Av. Pedro Montt 2421, Valparaíso, Chile, tel/fax: 0056322995532

†Crest-Ensai and Université de Strasbourg

words, the constructed estimator is adaptive with respect to both the unknown smoothness parameter and the type of dependence.

Our main purpose is to prove similar results for pointwise risks. The main interest in considering such risks is to obtain estimators that adapt to the local behavior of the density function to be estimated. To do so, we follow a different strategy since our estimator is a kernel density estimator with a data-driven selection of the bandwidth. 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. Among other we mention the following papers: Comte and Genon-Catalot (2012), Doumic, Hoffmann, Reynaud-Bouret, and Rivoirard (2012), Bertin, Lacour, and Rivoirard (2014). We point out that the main assumption in these papers is that the observations are independent and identically distributed (i.i.d.).

Nevertheless there are at least two practical motivations to consider dependent data. First, obtaining estimators that are robust with respect to moderate deviations from the i.i.d. ideal model is of particular interest. Second, many econometric models 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.

Recall that the process $\mathbb{X}$, considered as a time dependent process (or time series), is called weak dependent if, roughly, the covariance between functionals of the past and the future of the process decreases as the gap from the past to the future increases. Several notions of dependence provide bounds of covariance for some types of functionals (for example bounded or Lipschitz functionals). In this paper, we focus our attention on the α–mixing condition introduced by Rosenblatt (1956), the λ–dependence defined by Doukhan and Wintenberger (2007) and the $\tilde{\phi}$–dependence (see Dedecker and Prieur, 2005).

Several papers deal with the estimation of the common marginal density of a weak dependent process. Tribouley and Viennet (1998) proposed $\mathbb{L}_p$–adaptive estimators under β–mixing conditions. Comte and Merlevède (2002) obtained similar results (up to a logarithmic factor) under the weaker assumption that the process is α–mixing. Gannaz and Wintenberger (2010) extend these previous results to a wide variety of weak dependent processes including λ–dependent processes and $\tilde{\phi}$–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.

Notice also that Ragache and Wintenberger (2006) studied kernel density estimators from a minimax point of view for risks in the sup–norm. In this paper no data-driven choice of the bandwidth is proposed.

Our contribution is as follows. We obtain the adaptive rate of convergence for pointwise risks over a large scale of Hölder spaces in several situations of dependence. To our best knowledge, this is the first adaptive result for pointwise density estimation in a dependence context. 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 a tuning constant that can be chosen arbitrarily small. This implies that this procedure can be directly implemented in the 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. Several notions of weak dependence are introduced as well as basic assumptions on the distribution of the observations. 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 6. A simulation study is performed in Section 5 using data similar to ones used in Gannaz and Wintenberger (2010) in order to compare both results. The proofs of the technical results are gathered in the appendix.

2 Model

In what follows $\mathbb{X} = (X_i)_{i\in\mathbb{Z}}$ is a real-valued discrete time process and the observation consists in the vector $\mathcal{X}_n = (X_1, \dots, X_n)$ of size $n \in \mathbb{N}^*$. Our goal is to estimate the marginal density f of $\mathbb{X}$ at a fixed point x_0. In this section, basic assumptions on the distribution of $\mathbb{X}$ are stated. In particular, assumptions on the marginal density f as well as on the type of dependence are exposed. To do so, we present classical definitions of weakly dependent processes. Examples of processes that satisfy the required assumptions are given in order to illustrate the variety of models that can be considered in this work.

In what follows, $\rho = (\rho_r)_{r\in\mathbb{N}^*}$ denotes sequences that tend to 0 as r goes to infinity.

2.1 Strongly mixing processes.

The α–mixing coefficients of the process $\mathbb{X}$ are defined, for $r \in \mathbb{N}^*$ by

$$\begin{aligned}\alpha_r(\mathbb{X}) &= \sup_{n\in\mathbb{Z}} \sup \{|\mathbf{P}(A \cap B) - \mathbf{P}(A)\mathbf{P}(B)| : (A, B) \in \mathcal{P}_n \times \mathcal{F}_{n+r}\} \\ &= \sup_{n\in\mathbb{Z}} \sup \{\mathrm{Cov}(Y, Z)/(4\|Y\|_\infty\|Z\|_\infty) : (Y, Z) \in \mathbb{L}_\infty(\mathcal{P}_n \otimes \mathcal{F}_{n+r})\} \quad (1)\end{aligned}$$

where $\mathcal{P}_k = \sigma(X_i : i \leq k)$ and $\mathcal{F}_\ell = \sigma(X_i : i \geq \ell)$ are respectively the σ–algebras of the past (before time $k \in \mathbb{Z}$) and the future (after time $\ell \in \mathbb{Z}$) of the process $\mathbb{X}$. Here $\mathbb{L}_\infty(\mathcal{A})$ denotes the essentially bounded functions that are measurable with respect to a σ–algebra $\mathcal{A}$. Note that the second line comes from a characterization of strongly mixing processes (see Bradley, 2007).

Definition 1 (Rosenblatt, 1956)**.** The process $\mathbb{X}$ is called strongly mixing at rate ρ if, for any $r \in \mathbb{N}^*$, we have $\alpha_r(\mathbb{X}) \leq \rho_r$.

Remark. As a direct consequence of (1), the following bounds on covariance terms can be derived: for any u, v and r in $\mathbb{N}^*$, any sequence of real numbers $s_1 \leq \ldots \leq s_u \leq s_u + r \leq t_1 \leq \ldots \leq t_v$ and any functions $g_1 \in \mathbb{L}_\infty(\mathbb{R}^u)$ and $g_2 \in \mathbb{L}_\infty(\mathbb{R}^v)$ we have

$$|\mathrm{Cov}\,(g_1(X_{s_1}, \ldots, X_{s_u}), g_2(X_{t_1}, \ldots, X_{t_v}))| \leq 4\|g_1\|_\infty\|g_2\|_\infty\rho_r. \quad (2)$$

2.2 Doukhan-Wintenberger weak dependence.

Doukhan and Wintenberger (2007) introduced the notion of λ–dependence in order to generalize the notions of η (and its causal counterpart θ) and κ–dependences. Let us introduce some notations used throughout the paper. For $u \in \mathbb{N}^*$, the set Λ_u consists in all functions $g : \mathbb{R}^u \to \mathbb{R}$ such that $\mathrm{Lip}(g) < +\infty$, where

$$\mathrm{Lip}(g) = \sup_{x\neq y} \frac{|g(x) - g(y)|}{\sum_{i=1}^u |x_i - y_i|}$$

is the Lipschitz constant of g. Moreover we say that g belongs to $\Lambda_u^{(1)}$ if g belongs to $\Lambda_u \cap \mathbb{L}_\infty(\mathbb{R}^u)$ and $\|g\|_\infty \leq 1$. Finally, we consider the function $\Psi : \mathbb{R}_+^2 \times \mathbb{N}^2 \to \mathbb{R}_+$ defined by

$$\Psi(L_1, L_2, u, v) = L_1 u + L_2 v + L_1 L_2 uv.$$

With these notations in hand, we are in position to define the notion of λ–dependence.

Definition 2 (Doukhan and Wintenberger, 2007)**.** The process $\mathbb{X}$ is called λ–dependent at rate ρ if, for any u, v and r in $\mathbb{N}^*$, any sequence of real numbers $s_1 \leq \ldots \leq s_u \leq s_u + r \leq t_1 \leq \ldots \leq t_v$ and any functions $g_1 \in \Lambda_u^{(1)}$ and $g_2 \in \Lambda_u^{(1)}$ we have

$$|\mathrm{Cov}\,(g_1(X_{s_1}, \ldots, X_{s_u}), g_2(X_{t_1}, \ldots, X_{t_v}))| \leq \Psi(\mathrm{Lip}(g_1), \mathrm{Lip}(g_2), u, v)\rho_r. \quad (3)$$

2.3 Dedecker-Prieur weak dependence.

This notion is well adapted to processes that can be written in terms of dynamical systems and associated Markov chains. Let us introduce some notations. Let $\mathcal{A}$ be a σ-algebra and Y a random variable which takes values into $\mathbb{R}^v$. Define:

$$\tilde{\phi}(\mathcal{A}, Y) = \sup\{\|\mathbf{E}(g(Y)|\mathcal{A}) - \mathbf{E}(g(Y))\|_\infty : g \in \Lambda_v, \ \mathrm{Lip}(g) \leq 1\},$$

and, for any $(k, r) \in \mathbb{N}^2$,

$$\widetilde{\phi}_{\mathbb{X}}(r) = \sup_{k \in \mathbb{N}} \max_{v \leq k} v^{-1} \sup_n \sup_{n+r \leq t_1 \leq \ldots \leq t_v} \widetilde{\phi}(\mathcal{P}_n, (X_{t_1}, \ldots, X_{t_v})).$$

Definition 3 (Dedecker and Prieur, 2005)**.** The process $\mathbb{X}$ is called $\widetilde{\phi}$–dependent at rate ρ if for any $r \in \mathbb{N}^*$, $\widetilde{\phi}_{\mathbb{X}}(r) \leq \rho_r$.

Remark. This type of dependence also provides bounds on covariance terms. For any u, v and r in $\mathbb{N}^*$, any sequence of real numbers $s_1 \leq \ldots \leq s_u \leq s_u + r \leq t_1 \leq \ldots \leq t_v$ and any functions $g_1 \in \mathbb{L}_\infty(\mathbb{R}^u)$ and $g_2 \in \Lambda_v$ we have

$$\begin{aligned} \mathrm{Cov}(g_1(X_{s_1}, \ldots, X_{s_u}), g_2(X_{t_1}, \ldots, X_{t_v})) &\leq v\mathbf{E}(g_1(X_{s_1}, \ldots, X_{s_u}))\,\mathrm{Lip}(g_2)\rho_r \\ &\leq v\|g_1\|_\infty \,\mathrm{Lip}(g_2)\rho_r. \quad (4) \end{aligned}$$

2.4 Basic assumptions

In this section, we make some basic assumptions on the distribution of the process $\mathbb{X}$.

Assumption 1. We assume that there exists a density function $f \in \mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_\infty(\mathbb{R})$ such that, for any $i \in \mathbb{Z}$, the distribution of X_i admits the density f with respect to the Lebesgue measure on $\mathbb{R}$ and that there exists a positive constant f_∞ such that $\sup_{x \in V_n(x_0)} f(x) \leq \mathrm{f}_\infty$ where $V_n(x_0) = [x_0 - \frac{1}{\log n}, x_0 + \frac{1}{\log n}]$.

Note that we assume that f belongs to $\mathbb{L}_\infty(\mathbb{R})$. Such an assumption is classical in density estimation (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).

Assumption 2. The process $\mathbb{X}$ is either strongly mixing, λ–dependent or $\widetilde{\phi}$–dependent at rate ρ satisfying

$$\rho_r \leq \mathfrak{c} \exp(-\mathfrak{a} r^{\mathfrak{b}}), \qquad r \in \mathbb{N}^*$$

where $\mathfrak{a}$, $\mathfrak{b}$ and $\mathfrak{c}$ are positive constants.

In other words, we are assuming that ρ decreases at a geometric rate. Under such an assumption, Ragache and Wintenberger (2006) obtained near-minimax upper bounds for kernel density estimators over smooth classes of densities. Moreover, Doukhan and Neumann (2007), Merlevède, Peligrad, and Rio (2009) proved Bernstein-type inequalities under similar assumptions.

Assumption 3. For any i and j in $\mathbb{Z}$ such that $i \neq j$, the vector (X_i, X_j) admits a density function $f_{i,j}$. Moreover, there exists a positive constant $f_{\infty,\infty}$ such that $\sup_{x,y \in V_n(x_0)} f_{i,j}(x,y) \leq f_{\infty,\infty}$, uniformly in i, j.

This condition is more technical and is used in Lemma 1. Unfortunately it is quite strong and it is not satisfied even for simple Markov processes. The following is a well-known counter-example. Consider a random variable X_0 uniformly distributed on $[0,1]$ and, for $k \in \mathbb{N}^*$

$$X_k = \frac{1}{2}\left(X_{k-1} + \xi_k\right)$$

where $(\xi_k)_{k \in \mathbb{N}^*}$ is an i.i.d. sequence of Bernoulli variables with parameter $1/2$. The process $\mathbb{X}$ is then strictly stationary but, on the one hand it is not α–mixing and, on the other hand, for all k, the distribution of (X_0, X_k) is degenerated. However, through a reversion of time we have $X_{k-1} = T(X_k)$ where

$$T(y) = 2y \pmod 1.$$

Result of Lemma 1 can be extended to other dynamical systems of this type if the transformation T obeys some properties. Let I be a compact subset of $\mathbb{R}$ and assume that T acts from I to itself. Moreover, let us denote $T^k = T \circ T^{k-1}$ with $T^1 = T$. For any $x \in I$, set also

$$T^k(x+) = \lim_{\substack{\eta > 0 \\ x+\eta \in I}} T^k(x+\eta) \quad \text{and} \quad T^k(x-) = \lim_{\substack{\eta > 0 \\ x-\eta \in I}} T^k(x-\eta), \qquad k \in \mathbb{N}^*.$$

We now define

$$D_{\pm}^{k}(T)=\left\{x\in \mathrm{Int}(I):T^{k}(x\pm)=x\right\}, \qquad k\in\mathbb{N}^{*}$$

and

$$D^{\infty}(T)=\bigcup_{k\in\mathbb{N}^{*}}\left(D_{+}^{k}(T)\cup D_{-}^{k}(T)\right).$$

Assumption 4. There exists a function T for which $D^{\infty}(T)$ is negligible with respect to the Lebesgue measure and such that for all $k\in\mathbb{N}^{*}$, $X_k=T^k(X_0)$. In this situation, we conventionally set $f_{\infty,\infty}=0$.

This assumption is also used in Prieur (2001b) and Ragache and Wintenberger (2006) for instance.

2.5 Bernoulli shifts

In this section, we focus our attention on the example of Bernoulli shifts. Let us consider the process $\mathbb{X}$ defined by:

$$X_i=H((\xi_{i-j})_{j\in\mathbb{Z}}), \quad i\in\mathbb{Z} \tag{5}$$

where $H:\mathbb{R}^{\mathbb{Z}}\to[0,1]$ is a measurable function and the variables ξ_i are i.i.d. and real-valued. In addition, assume that there exists a sequence $(\delta_r)_{r\in\mathbb{N}^*}$ such that

$$\mathbf{E}|H((\xi_i)_{i\in\mathbb{Z}})-H((\xi_i')_{i\in\mathbb{Z}})|\le\delta_r$$

where, for any $r\in\mathbb{N}^*$, $(\xi_i')_{i\in\mathbb{Z}}$ is an i.i.d. sequence such that $\xi_i'=\xi_i$ if $|i|\le r$ and ξ_i is independent of ξ_i' otherwise. It can be proved (see Doukhan and Louhichi, 1999) that such processes are strongly stationary and η–dependent with rate $\rho_r\le 2\delta_{[r/2]}$. Consequently they are also λ–dependent with the same rate. Remark also that δ_r can be evaluated under both regularity conditions on the function H and integrability conditions on the ξ_i, $i\in\mathbb{Z}$. Indeed, if we assume that there exist $b\in(0,1]$ and positive constants $(a_i)_{i\in\mathbb{Z}}$ such that $|H((x_i)_i)-H((y_i)_i)|\le\sum_{i\in\mathbb{Z}}a_i|x_i-y_i|^b$ with $\xi_i\in\mathbb{L}_b(\mathbb{R})$ for all $i\in\mathbb{Z}$ then

$$\delta_r=\sum_{|i|\ge r}a_i\mathbf{E}|\xi_i|^b.$$

Moreover, under the weaker condition that $(\xi_i)_{i\in\mathbb{Z}}$ is λ–dependent and stronger assumptions on H (see Doukhan and Wintenberger, 2007), the process $\mathbb{X}$ inherits the same properties. Finally, we point out that classical econometrics models such as ARCH or GARCH can be viewed as causal Bernoulli shifts (that is, they obey (5) with $j\in\mathbb{N}$).

3 Estimation procedure

In this section we describe the construction of our procedure which is based on the so-called Goldenshluger-Lepski method (GLM for short). It consists in 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 rule of thumb.

3.1 Kernel density estimators

First, we consider a convolution kernel $K : \mathbb{R} \to \mathbb{R}$ and we state some classical assumptions such a kernel may satisfy.

Assumption 5. The kernel K is compactly supported on $[-1, 1]$. Its Lipschitz constant $\mathrm{Lip}(K)$ is finite and $\int_{\mathbb{R}} K(x)\,\mathrm{d}x = 1$.

Assumption 6. There exists $m \in \mathbb{N}$ such that the kernel K is of order m. That is, for any $0 \leq \ell \leq m$, we have

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

where $\delta_{0,\ell}$ is the Kronecker symbol.

Second we consider, for any h in $(0, +\infty)$, the estimator $\hat{f}_h$ defined by,

$$\hat{f}_h(x_0) = \frac{1}{n}\sum_{i=1}^{n} K_h(x_0 - X_i) \quad \text{where} \quad K_h(\cdot) = h^{-1}K(h^{-1}\cdot).$$

3.2 Bandwidth selection

We consider the finite set of bandwidth $\mathcal{H}_n$ defined by

$$\mathcal{H}_n = \{2^{-k} : k \in \mathbb{N}\} \cap [h_*, h^*] \quad \text{where} \quad h_* = n^{-1}\exp\left(\sqrt{\log n}\right) \quad \text{and} \quad h^* = (\log n)^{-1}.$$

We also define, for any $h \in \mathcal{H}_n$ the observable (and measurable) quantity

$$\widehat{M}_n(h) = \sqrt{2q|\log h|}\left(\sqrt{\widehat{J}_n(h) + \frac{\delta}{nh}} + \sqrt{\max_{\substack{\mathfrak{h} \geq h \\ \mathfrak{h} \in \mathcal{H}_n}} \widehat{J}_n(\mathfrak{h}) + \frac{\delta}{nh}}\right) \tag{6}$$

where

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

and $\delta > 0$ is a parameter of the procedure that can be chosen as close to 0 as we want. Following Goldenshluger and Lepski (2011), we consider for any $h \in \mathcal{H}_n$

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

where $\{y\}_+ = \max(0, y)$ for any $y \in \mathbb{R}$ and $h \vee \mathfrak{h} = \max(h, \mathfrak{h})$. We select the bandwidth $\hat{h}(x_0)$ using the rule

$$\hat{h}(x_0) = \arg\min_{h\in\mathcal{H}_n} \left(A(h, x_0) + \widehat{M}_n(h) \right). \tag{7}$$

The final estimator is defined by

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

3.3 Comments

The main principle used here is the GLM. Let us briefly comment the main idea behind this construction. First, remark that, using Lemma 2 and the fact that $h \vee \mathfrak{h} \geq \mathfrak{h}$, for h and $\mathfrak{h}$ in $\mathcal{H}_n$ we have

$$\begin{aligned}\sqrt{\mathrm{Var}(\hat{f}_{\mathfrak{h}}(x_0))} + \sqrt{\mathrm{Var}(\hat{f}_{h\vee\mathfrak{h}}(x_0))} &\leq \sqrt{J_n(\mathfrak{h}) + o\left(\frac{1}{n\mathfrak{h}}\right)} + \sqrt{J_n(h\vee\mathfrak{h}) + o\left(\frac{1}{n(h\vee\mathfrak{h})}\right)} \\ &\leq \sqrt{J_n(\mathfrak{h}) + \frac{\delta}{n\mathfrak{h}}} + \sqrt{\max_{\substack{h\geq\mathfrak{h}\\ \mathfrak{h}\in\mathcal{H}_n}} J_n(h) + \frac{\delta}{n\mathfrak{h}}}.\end{aligned}$$

where

$$J_n(\mathfrak{h}) = \frac{1}{n}\int K_{\mathfrak{h}}^2(x_0 - x) f(x) dx.$$

Since, with high probability $\widehat{J}_n(\mathfrak{h})$ is close to $J_n(\mathfrak{h})$, the term $\widehat{M}_{\mathfrak{h}}$ can be viewed as a penalized observable upper bound of the left hand side of this inequality. This implies that, for any h and $\mathfrak{h}$, the term

$$\left\{ |\hat{f}_{\mathfrak{h}}(x_0) - \mathbf{E}\hat{f}_{\mathfrak{h}}(x_0)| + |\hat{f}_{h\vee\mathfrak{h}}(x_0) - \mathbf{E}\hat{f}_{h,\mathfrak{h}}(x_0)| - \widehat{M}_n(\mathfrak{h}) \right\}_+$$

is, in some sense, small (see the proof of Theorem 1 for a precise result). Thus, up to a small additive term, $A(h, x_0)$ is dominated by $|\mathbf{E}\hat{f}_{h\vee\mathfrak{h}}(x_0) - \mathbf{E}\hat{f}_{\mathfrak{h}}(x_0)|$.

Moreover, this term can be easily bounded. Indeed if $h \leq \mathfrak{h}$ then it equals 0. Otherwise, $\mathfrak{h} \leq h$ and:

$$\begin{aligned} |\mathbf{E}\hat{f}_{h\vee\mathfrak{h}}(x_0) - \mathbf{E}\hat{f}_{\mathfrak{h}}(x_0)| &\leq |\mathbf{E}\hat{f}_h(x_0) - f(x_0)| + |\mathbf{E}\hat{f}_{\mathfrak{h}}(x_0) - f(x_0)| \\ &\leq 2\max_{\mathfrak{h}\leq h}|K_{\mathfrak{h}} \star f(x_0) - f(x_0)|, \end{aligned} \tag{8}$$

where the notation $\star$ stands for the convolution. This implies that $A(h,x)$ can be roughly bounded by the bias term of $\hat{f}_h$ under regularity assumptions on f. As a direct consequence, the selection rule (7) can be interpreted as a trade-off between estimations of the bias term (that is $A(h,x_0)$) and of a penalized standard deviation term (that is $\widehat{M}_n(h)$).

4 Results

Our main goal in this paper is to prove that the proposed procedure of estimation adapts to the unknown smoothness of the marginal density f and is robust with respect to several types of weak dependence. The accuracy of an arbitrary estimator $\tilde{f}_n$ (that is, a function which is measurable with respect to the observation $\mathcal{X}_n$), is evaluated using its pointwise risk (with respect to the true marginal density) defined by

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

where $x_0 \in \mathbb{R}$ and $q > 0$ are fixed real numbers. Here, $\mathbf{E}$ denotes the expectation with respect to the distribution of the process $\mathbb{X}$.

We prove two results. The first one consists in an oracle-type inequality: under appropriate assumptions, our estimator performs almost as well as the best linear kernel estimator in the considered family. The second one states that this procedure achieves classical minimax rates of convergence (up to a multiplicative logarithmic factor) over a wide scale of Hölder spaces. The considered family of kernel estimators is rich enough and the oracle inequality precise enough to obtain such results.

4.1 Oracle type inequality

Theorem 1. *Under Assumptions 1, 2, 5 and either 3 or 4, we have:*

$$R_q^q(\hat{f}, f) \leq C_1^* \min_{h\in\mathcal{H}_n}\left\{\max_{\substack{\mathfrak{h}\leq h\\ \mathfrak{h}\in\mathcal{H}_n}}|K_{\mathfrak{h}} \star f(x_0) - f(x_0)|^q + \left(\frac{|\log h|}{nh}\right)^{q/2}\right\} \tag{9}$$

where C_1^* is a positive constant that depends only on $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$, f_∞, $\mathrm{f}_{\infty,\infty}$ and K.

Proof of Theorem 1 is postponed to Section 6

Remark. Let us briefly comment this result.

1) The right hand side term of (9) can be viewed as a tight upper bound for $\min_{h\in\mathcal{H}_n} \mathbf{E}|\hat{f}_h(x_0) - f(x_0)|^q$ since it is the sum of an approximation of the bias term and the standard deviation term (up to a multiplicative logarithmic term) of $\hat{f}_h$. That means that our procedure performs almost as well as the best kernel density estimator in the considered family.

2) The extra $|\log h|$ term in the standard deviation is (for pointwise estimation) a well-known unavoidable phenomenon in the i.i.d. case (see Lepski, 1990, Klutchnikoff, 2014, among others). Roughly speaking, this extra term is used to control rare events. This is also the case in our situation of dependence.

4.2 Adaptive result

Set s and $\mathfrak{L}$ be two positive real numbers. The Hölder class $\mathcal{C}^s(\mathfrak{L})$ consists in function $f : \mathbb{R} \to \mathbb{R}$ such that f is $m_s = \sup\{k \in \mathbb{N} : k < s\}$ times differentiable and that satisfies

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

Theorem 2. *Assume that Assumptions 1, 2, 5, 6 and either 3 or 4 hold. Assume moreover that $f \in \mathcal{C}^s(\mathfrak{L})$ for $0 < s \leq m+1$ and $\mathfrak{L} > 0$. Then there exists a constant C_2^* that depends only on $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$, f_∞, $\mathrm{f}_{\infty,\infty}$, m, $\mathfrak{L}$ and K such that:*

$$R_q(\hat{f}, f) \leq C_2^* \left(\frac{\log n}{n}\right)^{\frac{s}{2s+1}}.$$

This result is a direct consequence of Theorem 1, since it can be easily proved that

$$\max_{\substack{\mathfrak{h}\leq h\\ \mathfrak{h}\in\mathcal{H}_n}} |K_\mathfrak{h} \star f(x_0) - f(x_0)|^q \leq C_3^* h^{sq},$$

for any bandwidth $h > 0$, where C_3^* depends only on s, $\mathfrak{L}$, K and q. This implies that, for n large enough, there exists $h_n(s, \mathfrak{L}, K, q) \in \mathcal{H}_n$ such that the right hand side of (9) is bounded, up to a multiplicative constant, by the expected rate.

As far as we know, this result is the first theoretical pointwise adaptive result for the estimation of the marginal density in a context of weak dependence. Moreover, integrating the pointwise risk on a bounded domain, we obtain that our procedure converges adaptively at rate $(n^{-1}\log n)^{s/(2s+1)}$ in $\mathbb{L}_p$–norm ($p \neq \infty$) over Hölder balls. This extends the results of Gannaz and Wintenberger (2010).

5 Simulation study

In this section, we study the performance of our procedure from an empirical point of view, using simulated data. Several types of dependence are considered. In order to compare our results with those obtained in Gannaz and Wintenberger (2010), we focus our study to the examples considered in this paper.

For a given distribution function F, we will generate observations $(X_1, \ldots, X_n)$ with distribution F for three cases of dependence.

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 $[0, 1]$.

Case 2. The X_i are $\tilde{\phi}$–dependent variables given by $F^{-1}(G(Y_i))$ where $G(y) = \frac{2}{\pi}\arcsin(\sqrt{y})$ and the Y_i are given by

$$Y_1 = G^{-1}(U_1)$$

and, for $i \geq 2$

$$Y_i = T(Y_{i-1}) \qquad \text{where} \qquad T(y) = 4y(1-y).$$

Note that G is the invariant distribution of T (see Prieur, 2001a).

Case 3. The X_i are λ–dependent given by $F^{-1}(G(Y_i))$ where the $Y_i, i \in \mathbb{Z}$ are solution of the non-causal equation:

$$Y_i = 2(Y_{i-1} + Y_{i+1})/5 + 5\xi_i/21,$$

where $(\xi_i)_{i\in\mathbb{Z}}$ is an i.i.d. sequence of Bernoulli variables with parameter $1/2$. Here G is the marginal distribution of the Y_i (see Gannaz and Wintenberger, 2010, for more details). The simulation of the variables $(Y_i)_{i=1,\ldots,n}$ is obtained using the method developped in Doukhan and Truquet (2007).

We consider two different density functions to estimate. The first one (also considered in Gannaz and Wintenberger (2010)) is:

$$f_1(x) = \sin(x)I_{[0,2.5]}(x) + I_{[2.5,4]}.$$

	Case 1	Case 2	Case 3
f_1	0.01272	0.01384	0.04193
f_2	0.02706	0.05127	0.09536

Table 1: MISE for the two densities f_1 and f_2 and the three cases of dependence.

The second one is the density of the mixture of two normal distributions

$$f_2(x) = \frac{1}{2}\phi_{0,1}(x) + \frac{1}{2}\phi_{-5,0.25}(x),$$

where $\phi_{\mu,\sigma}$ stands for the density of a normal distribution with mean μ and standard deviation σ.

Our goal is to evaluate the quality of estimation of our procedure for observations of size $n = 1000$. To do so, we simulate 500 replications of each case (for the three types of dependence and the two different density functions). We approximate, using Monte-Carlo method, the mean integrated squared error (MISE) in order to compare our results with the ones obtained in Gannaz and Wintenberger (2010).

In our procedure, in (6), we take $\delta = 0.01$, we use the Epanechnikov kernel and, since the MISE is considered, we take $q = 2$. Our results are summarized in Table 1.

For the estimation of the function f_1, we outperform the results of Gannaz and Wintenberger (2010). In the first case, our MISE is about 8 times smaller, 5 times smaller in the second case and 2 times smaller in the last case.

In Figure 1, the function f_2 (in solid lines) and our estimators (in dashed lines) are represented for the three cases of dependence. Our procedure is able to catch the two different peaks of the mixture (whereas kernel estimators with global choice of bandwidth fail in such a situation). Performances of the MISE are quite good.

To conclude, in the considered examples, our procedure has similar or better performances than yet existing methods used for dependent data. An important point is that the choice of the bandwidth depends on explicit constants that can be used directly in the practice and do not need previous calibration.

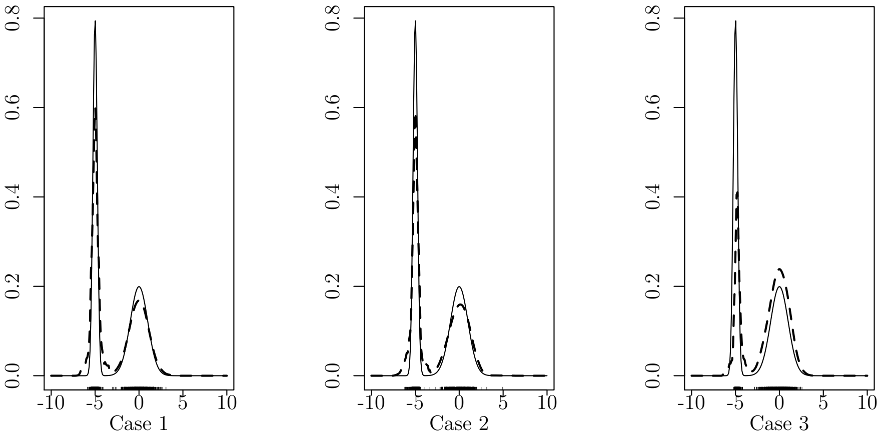

Figure 1: Behavior of our procedure to estimate the function f_2 in the three cases of dependence. In solid, the true function. In dashed lines, the estimator.

6 Proofs

6.1 Basic notations

For the sake of readability, we introduce in this section some conventions and notations that are used throughout the proofs. Moreover, here and later, we assume that Assumptions 1, 2, 5, and either 3 or 4 hold.

First, let us consider for any $h \in \mathcal{H}_n$ the functions g_h and $\bar{g}_h$ defined, for any $y \in \mathbb{R}$, by $g_h(y) = K_h(x_0 - y)$ and

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

Note that, using these notations we have:

$$\hat{f}_h(x_0) - \mathbf{E}\hat{f}_h(x_0) = \sum_{i=1}^{n} \bar{g}_h(X_i), \quad h \in \mathcal{H}_n.$$

Next, we introduce some constants. Let us consider:

$$C_1 = \|K\|_1^2(\mathrm{f}_\infty^2 + f_{\infty,\infty}), \qquad C_2 = \mathrm{f}_\infty\|K\|_2^2, \qquad \text{and} \qquad C_3 = 2\|K\|_\infty.$$

Moreover we define $L = \mathrm{Lip}(K)$ and

$$C_4 = 2C_3L + L^2, \quad C_5 = \max(C_3L, C_4, 4C_3^2), \qquad \text{and} \qquad C_6 = (2C_1)^{3/4}C_5^{1/4}.$$

6.2 Auxiliary results

In the three following lemmas, assume that Assumptions 1, 2, 5 and either 3 or 4 are satisfied. The first lemma provides bounds on covariance terms for functionals of the past and the future of the observations. The considered functionals depend on the kernel K.

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

$$D_1(h) = D_1(n,h) = \frac{C_3}{nh} \qquad \text{and} \qquad D_2(h) = D_2(n,h) = \frac{C_6}{n^2 h}.$$

Then for any u, v and r in $\mathbb{N}$, if $(s_1, \ldots, s_u, t_1, \ldots t_v) \in \mathbb{R}^{u+v}$ is such that $s_1 \leq \ldots, s_u \leq s_u + r \leq t_1 \leq \ldots \leq t_v$, we have

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

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

The following lemma provides a moment inequality for the classical kernel estimator.

Lemma 2. *We have*

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

Moreover for $\mathfrak{q} > 0$, we have

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

where $\mathfrak{C}_{\mathfrak{q}}$ is a positive constant. Here the $o(\cdot)$–terms depend only on $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$, f_∞, $\mathrm{f}_{\infty,\infty}$ and K.

The following result is an adaptation of the Bernstein-type inequality obtained by Doukhan and Neumann (2007).

Lemma 3 (Bernstein's inequality)**.** *Let ε_1 be a positive real number arbitrary small. Denote, for any $t \geq 0$,*

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

where

$$\sigma_n(h) = J_n(h) + \frac{\varepsilon_1}{nh} \tag{11}$$

and

$$B_n(h) = \frac{C_3 C_7}{nh} \tag{12}$$

where C_7 is a positive constant that depends only in $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$, f_∞, $\mathrm{f}_{\infty,\infty}$ and K. Then we have, for n large enough:

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

6.3 Proof of Theorem 1

Let ε_2 a be positive real number and let us denote $\gamma = q(1+\varepsilon_2)$. For convenience, we split the proof into several steps.

Step 1. Let us consider the random event

$$\mathcal{A} = \bigcap_{h \in \mathcal{H}_n} \left\{ \left| \frac{h}{n} \sum_{i=1}^{n} K_h^2(x_0 - X_i) - h \int K_h^2(x_0 - x) f(x) dx \right| \leq \frac{\delta}{2} \right\}$$

and the quantities Γ_1 and $\Gamma 2$ defined by :

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

and

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

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

$$\begin{aligned} R_q^q(\hat{f}, f) &\leq (*) + \left(R_{2q}^{2q}(\hat{f}, f)\mathbf{P}(\mathcal{A}^c)\right)^{1/2} \\ &\leq \Gamma_1 + \Gamma_2. \end{aligned}$$

We define

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

$$\widehat{M}_n^{(2)}(h) = \sqrt{2q|\log h| \left(\max_{\substack{\mathfrak{h} \geq h \\ \mathfrak{h} \in \mathcal{H}_n}} \widehat{J}_n(\mathfrak{h}) + \frac{\delta}{nh}\right)},$$

$$\mathfrak{M}_n(h, a) = \mathfrak{M}_n^{(1)}(h, a) + \mathfrak{M}_n^{(2)}(h, a),$$

$$\mathfrak{M}_n^{(1)}(h,a) = \sqrt{2q|\log h| \left(J_n(h) + \frac{a\delta}{nh} \right)},$$

and

$$\mathfrak{M}_n^{(2)}(h,a) = \sqrt{2q|\log h| \left(\max_{\substack{\mathfrak{h} \geq h \\ \mathfrak{h} \in \mathcal{H}_n}} J_n(\mathfrak{h}) + \frac{a\delta}{nh} \right)}.$$

Now note that if the event $\mathcal{A}$ holds, for i=1,2, we have:

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

Steps 2–5 are devoted to control the term Γ_1 whereas Γ_2 is upper bounded in Step 6.

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

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

We consider the first term of the right hand side of this inequality. The following inequalities are straightforward

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

Using similar arguments we can prove that

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

that leads, using (7), 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, for some positive constant c_q,

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

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

$$\begin{aligned} A(h, x_0) \leq \max_{\mathfrak{h} \in \mathcal{H}_n} &\left\{ \left| \mathbf{E}\hat{f}_{h\vee\mathfrak{h}}(x_0) - \mathbf{E}\hat{f}_{\mathfrak{h}}(x_0) \right| \right\}_+ \\ &+ \max_{\mathfrak{h} \in \mathcal{H}_n} \left\{ \left| \hat{f}_{h\vee\mathfrak{h}}(x_0) - \mathbf{E}\hat{f}_{h\vee\mathfrak{h}}(x_0) \right| - \widehat{M}_n^{(2)}(\mathfrak{h}) \right\}_+ \\ &+ \max_{\mathfrak{h} \in \mathcal{H}_n} \left\{ \left| \hat{f}_{\mathfrak{h}}(x_0) - \mathbf{E}\hat{f}_{\mathfrak{h}}(x_0) \right| - \widehat{M}_n^{(1)}(\mathfrak{h}) \right\}_+. \end{aligned}$$

Now define:

$$T_1 = \max_{\mathfrak{h}\in\mathcal{H}_n} \left\{ \left| \sum_{i=1}^n \bar{g}_{\mathfrak{h}}(X_i) \right| - \widehat{M}_n^{(1)}(\mathfrak{h}) \right\}_+$$

and, in a similar way,

$$T_2 = \max_{\mathfrak{h}\in\mathcal{H}_n} \left\{ \left| \sum_{i=1}^n \bar{g}_{h\vee\mathfrak{h}}(X_i) \right| - \widehat{M}_n^{(2)}(\mathfrak{h}) \right\}_+ .$$

Using these notations and (8), we obtain

$$A(h,x_0) \le 2 \max_{\substack{\mathfrak{h}\le h\\ \mathfrak{h}\in\mathcal{H}_n}} \left\{ |K_{\mathfrak{h}} \star f(x_0) - f(x_0)| \right\}_+ + T_1 + T_2. \tag{14}$$

Combining (13) and (14), and denoting by $E_h(x_0)$ the first term of the right hand side of (14), we obtain

$$\mathbf{E}\left(A^q(h,x_0)\mathbf{I}_{\mathcal{A}}\right) \le c_q \left(E_h^q(x_0) + \mathbf{E}\Big(T_1^q \mathbf{I}_{\mathcal{A}}\Big) + \mathbf{E}\Big(T_2^q \mathbf{I}_{\mathcal{A}}\Big) \right), \tag{15}$$

for some positive constant c_q.

Step 4. It remains to upper bound $\mathbf{E}(T_j^q \mathbf{I}_{\mathcal{A}})$ for $j=1,2$. To this aim, notice that,

$$\mathbf{E}(T_j^q \mathbf{I}_{\mathcal{A}}) \le \mathbf{E}\tilde{T}_j^q, \tag{16}$$

where

$$\tilde{T}_1 = \max_{\mathfrak{h}\in\mathcal{H}_n} \left\{ \left| \sum_{i=1}^n \bar{g}_{\mathfrak{h}}(X_i) \right| - \mathfrak{M}_n^{(1)}\left(\mathfrak{h}, \frac{1}{2}\right) \right\}_+ ,$$

and $\tilde{T}_2$ is defined similarly using the function $\bar{g}_{h\vee\mathfrak{h}}$ instead of $\bar{g}_{\mathfrak{h}}$ and $\mathfrak{M}_n^{(2)}$ instead of $\mathfrak{M}_n^{(1)}$. Now, we define $r(\cdot)$ by

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

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

$$2^{\mathfrak{d}-1} B_n(h)(2\gamma|\log h|)^{\mathfrak{d}} \le \varepsilon_3 \sqrt{2q|\log h| \frac{1}{nh}},$$

for ε_3 a positive real number. Moreover, we have

$$\begin{aligned}\sqrt{2\sigma_n^2(h)\gamma|\log h|} &\le \sqrt{2q|\log h| \left(J_n(h) + \frac{\varepsilon_1}{nh}\right)(1+\varepsilon_2)} \\ &\le \sqrt{2q|\log h| \left(J_n(h) + \frac{\varepsilon_1 + \varepsilon_2(C_2+\varepsilon_1)}{nh}\right)}.\end{aligned}$$

Last inequality comes from the fact that $J_n(h)$ is upper-bounded by $C_2/(nh)$. Now, we can choose ε_1, ε_2 and ε_3 such that

$$r(\gamma|\log h|) \leq \mathfrak{M}_n^{(1)}(h, \frac{1}{2}) \tag{17}$$

and, using similar arguments,

$$\sqrt{2\sigma_n^2(h \vee \mathfrak{h})\gamma|\log h|} + 2^{\mathfrak{d}-1} B_n(h \vee \mathfrak{h})\,(2\gamma|\log h|)^{\mathfrak{d}} \leq \mathfrak{M}_n^{(2)}(h, \frac{1}{2}). \tag{18}$$

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

$$\begin{aligned}
\mathbf{E}\tilde{T}_1^q &\leq \sum_{\mathfrak{h}\in\mathcal{H}_n} \int_0^\infty \mathbf{P}\left(\left|\sum_{i=1}^n \bar{g}_{\mathfrak{h}}(X_i)\right| \geq \mathfrak{M}_n^{(1)}(\mathfrak{h}, \frac{1}{2}) + t^{1/q}\right) dt \\
&\leq C \sum_{\mathfrak{h}\in\mathcal{H}_n} \int_0^\infty r'(u) r(u)^{q-1} \mathbf{P}\left(\left|\sum_{i=1}^n \bar{g}_{\mathfrak{h}}(X_i)\right| \geq r(\gamma|\log \mathfrak{h}|) + r(u)\right) du, \\
&\leq C \sum_{\mathfrak{h}\in\mathcal{H}_n} \int_0^\infty u^{-1}\lambda(u)^q \mathbf{P}\left(\left|\sum_{i=1}^n \bar{g}_{\mathfrak{h}}(X_i)\right| \geq \lambda(\gamma|\log \mathfrak{h}| + u)\right) du,
\end{aligned}$$

where $\lambda(\cdot)$ is defined by (10). Using Lemma 3, we obtain

$$\begin{aligned}
\mathbf{E}\tilde{T}_1^q &\leq C \sum_{\mathfrak{h}\in\mathcal{H}_n} \int_0^\infty u^{-1}\left(\sqrt{\sigma_n^2(\mathfrak{h})u} + B_n(\mathfrak{h})u^3\right)^q \exp\left\{-\frac{u}{2} - \frac{\gamma|\log \mathfrak{h}|}{2}\right\} du \\
&\leq C \sum_{\mathfrak{h}\in\mathcal{H}_n} \sigma_n^q(\mathfrak{h})\mathfrak{h}^{\gamma/2}.
\end{aligned}$$

Using the definitions of the quantities that appear in this equation and using (16), we readily obtain:

$$\mathbf{E}T_1^q \mathbf{I}_{\mathcal{A}} \leq C n^{-q/2} \sum_{k\in\mathbb{N}} (2^{(\gamma-q)/2})^{-k} \leq C n^{-q/2}. \tag{19}$$

Following the same lines and using (18), we obtain

$$\mathbf{E}T_2^q \mathbf{I}_{\mathcal{A}} \leq C n^{-q/2}. \tag{20}$$

Step 5. Lemma 2 implies that:

$$\mathbf{E}|\hat{f}_h(x_0) - f(x_0)|^q \leq c_q\left(|E_h(x_0)|^q + (nh)^{-q/2}\right) \tag{21}$$

for some positive constant c_q.

Combinning (13), (15), (19), (20),(21), we have:

$$\Gamma_1 \leq C^* \min_{h\in\mathcal{H}_n} \left\{ \max_{\substack{\mathfrak{h}\leq h\\ \mathfrak{h}\in\mathcal{H}_n}} |K_{\mathfrak{h}} \star f(x_0) - f(x_0)|^q + \left(\frac{|\log h|}{nh}\right)^{q/2} \right\} \tag{22}$$

where C^* is a positive constant that depends only on $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$, f_∞, $\mathrm{f}_{\infty,\infty}$ and K.
Step 6. Using Lemma 3 where in $\bar{g}_h$, K is replaced by K^2, we obtain that

$$\mathbf{P}\left(\left| \frac{h}{n}\sum_{i=1}^n K_h^2(x_0 - X_i) - h\int K_h^2(x_0 - x)f(x)dx \right| > \frac{\delta}{2} \right) \leq \exp(-C_1 n^2 h^2),$$

where C_1 is a constant that depends only on $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$, f_∞, $\mathrm{f}_{\infty,\infty}$ and δ. Then this implies that

$$\mathbf{P}\left(\mathcal{A}^c\right) \leq \frac{\log n}{\log 2} \exp\Big(- C_1 \exp(2\sqrt{\log n})\Big)$$

and then

$$\Gamma_2 = o\left(\frac{1}{n}\right). \tag{23}$$

Now, using (22) and (23), Theorem 1 follows.

A Proof of Lemma 1

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

$$\Upsilon_h(u,v) = \left| \mathrm{Cov}\left(\prod_{i=1}^u \bar{g}_h(X_{s_i}), \prod_{j=1}^v \bar{g}_h(X_{t_j}) \right) \right|.$$

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 $\ell = 2$ and let us denote $s = s_1$ and $t = t_1$. If Assumption 3 is fulfilled, we have, using the notation $m_h = \mathbf{E}g_h(X_1)$:

$$\begin{aligned}
n^2\Upsilon_h(u,v) &= \left|\mathrm{Cov}\left(g_h(X_s), g_h(X_t)\right)\right| \\
&= \left| \int_{\mathbb{R}^2} (g_h(x) - m_h)(g_h(y) - m_h) f_{s,t}(x,y)dxdy \right| \\
&\leq \left| \int_{\mathbb{R}^2} g_h(x)g_h(y) f_{s,t}(x,y)dxdy - m_h^2 \right| \\
&\leq (f_{\infty,\infty} + \mathrm{f}_\infty^2)\|g_h\|_1^2 \leq C_1,
\end{aligned}$$

Assume now that Assumption 4 holds true. Thanks to Lemma 2.3 in Prieur (2001b), we obtain

$$\begin{aligned}\Upsilon_h(u,v) &\leq (\mathbf{E}\bar{g}_h(X_1))^2\\ &\leq \left(\int_{\mathbb{R}} |\bar{g}_h(x)f(x)|\,dx\right)^2\\ &\leq \mathrm{f}_\infty^2\|K\|_1^2/n^2\\ &\leq C_1 n^{-2},\end{aligned}$$

since $f_{\infty,\infty}=0$ under Assumption 4. Thus, in both situations, the following bound holds true:

$$|\mathrm{Cov}\left(\bar{g}_h(X_s),\bar{g}_h(X_t)\right)| \leq C_1 n^{-2}. \tag{24}$$

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

$$\begin{cases} A = \mathbf{E}\left(\prod_{i=1}^u \bar{g}_h(X_{s_i})\prod_{i=1}^v \bar{g}_h(X_{t_j})\right)\\ B = \mathbf{E}\left(\prod_{i=1}^u \bar{g}_h(X_{s_i})\right)\mathbf{E}\left(\prod_{i=1}^v \bar{g}_h(X_{t_j})\right). \end{cases}$$

Remark that both A and B can be bounded, using (24), by

$$\|\bar{g}_\delta\|_\infty^{(u-2)+v}\,\mathrm{Cov}(\bar{g}_h(X_{s_1}),\bar{g}_h(X_{s_2})) \leq \left(\frac{C_3}{nh}\right)^{\ell-2}\frac{C_1}{n^2}.$$

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

$$\Upsilon_h(u,v) \leq \frac{2C_1}{n^2}\left(\frac{C_3}{nh}\right)^{\ell-2}. \tag{25}$$

Structural bound. In this part, we prove bounds for each type of weak dependence. Firstly, assume that the process is strongly mixing at rate ρ. Then, thanks to (2), we obtain that:

$$\begin{aligned}\Upsilon_h(u,v) &\leq 4\left(\frac{C_3}{nh}\right)^{\ell}\rho_r\\ &\leq \frac{1}{n^2}\left(\frac{C_3}{nh}\right)^{\ell-2}\frac{4C_3^2}{h^2}\Phi(u,v)\rho_r \end{aligned}\tag{26}$$

where $\Phi(u,v)=u+v+uv\geq 1$.

Secondly, let us assume that the process is λ–dependent at rate ρ. Using the fact that if $f \in \Lambda_1$ then $\mathrm{Lip}(f^{\otimes u}) \leq \mathrm{Lip}(f)$, we obtain, thanks to (3), that

$$\Upsilon_h(u,v) \leq \left(\frac{C_3}{nh}\right)^{\ell} \Psi\left(\frac{nh}{C_3}\mathrm{Lip}(\bar{g}_h), \frac{nh}{C_3}\mathrm{Lip}(\bar{g}_h), u, v\right)\rho_r$$

Let us remark that

$$\frac{nh}{C_3}\mathrm{Lip}(\bar{g}_h) = \frac{L}{C_3 h}.$$

Moreover, since $h \leq h^*$ we have

$$\Psi\left(\frac{nh}{C_3}\mathrm{Lip}(\bar{g}_h), \frac{nh}{C_3}\mathrm{Lip}(\bar{g}_h), u, v\right) \leq \frac{C_4}{C_3^2}\Phi(u,v)h^{-2}.$$

This implies that

$$\Upsilon_h(u,v) \leq \frac{1}{n^2}\left(\frac{C_3}{nh}\right)^{\ell-2}\frac{C_4}{h^4}\Phi(u,v)\rho_r. \tag{27}$$

Thirdly, we assume that the process is $\tilde{\phi}$–dependent at rate ρ. Thanks to (4), using the same arguments as above, we obtain

$$\begin{aligned}\Upsilon_h(u,v) &\leq v\frac{C_3^2}{n^2h^2}\left(\frac{C_3}{nh}\right)^{\ell-2}\frac{L}{C_3 h}\rho_r \\ &\leq \frac{1}{n^2}\left(\frac{C_3}{nh}\right)^{\ell-2}\frac{C_3 L}{h^4}\Phi(u,v)\rho_r. \end{aligned}\tag{28}$$

Last inequality follows from the fact that $h \leq h^*$.

Finally, using (26), (27) and (28) we obtain that, in each situation of dependence,

$$\Upsilon_h(u,v) \leq \frac{1}{n^2}\left(\frac{C_3}{nh}\right)^{\ell-2}\frac{C_5}{h^4}\Phi(u,v)\rho_r. \tag{29}$$

Conclusion. Now combining (25) and (29) we obtain:

$$\begin{aligned}\Upsilon_h(u,v) &\leq \frac{1}{n^2}\left(\frac{C_3}{nh}\right)^{\ell-2}(2C_1)^{3/4}\left(\frac{C_5}{h^4}\Phi(u,v)\rho_r\right)^{1/4} \\ &\leq \frac{C_6}{n^2h}\left(\frac{C_3}{nh}\right)^{\ell-2}\Phi(u,v)\rho_r^{1/4}.\end{aligned}$$

This proves Lemma 1.

B Proof of Lemma 2

Proof of this result can be readily adapted from the proof of Theorem 1 in Doukhan and Louhichi (2001) (using similar arguments that ones used in the proof of Lemma 1). The only thing to do is to bound explicitely the term

$$A_2(\bar{g}_h) = \mathbf{E}\left(\sum_{i=1}^{n} \bar{g}_h(X_i)\right)^2 .$$

Set $R = h^{\nu-1}$ where $0 < \nu < 1$. Remark that

$$\begin{aligned} A_2(\bar{g}_h) &= n\mathbf{E}\bar{g}_h(X_1)^2 + \sum_{i\neq j} \mathbf{E}\bar{g}_h(X_i)\bar{g}_h(X_j) \\ &= J_n(h) + 2\sum_{i=1}^{n-1}\sum_{r=1}^{n-i} \mathbf{E}\bar{g}_h(X_i)\bar{g}_h(X_{i+r}). \end{aligned}$$

Using Lemma 1 and (24), we obtain:

$$\begin{aligned} A_2(\bar{g}_h) &\leq J_n(h) + 2n\sum_{r=1}^{R} \frac{C_1}{n^2} + 2D_2^2(h) \sum_{r=R+1}^{n-1} (n-r)\Phi(1,1)\rho_r^{1/4} \\ &\leq J_n(h) + \frac{C}{nh}\left(h^\nu + \sum_{r=R+1}^{\infty} \rho_r^{1/4}\right). \end{aligned}$$

Since $h \leq h^*$ and using Assumption 2 we obtain:

$$A_2(\bar{g}_h) \leq J_n(h) + o\left(\frac{1}{nh}\right).$$

This equation, combined with the fact that $J_n(h) \leq C_2(nh)^{-1}$, completes the proof.

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, for n large enough, we have

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

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

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

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

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

This implies the results.

D Acknowledgments

The authors have been supported by Fondecyt project 1141258. Karine Bertin has been supported by the grant Anillo ACT-1112 CONICYT-PIA.

References

K. Bertin, C. Lacour, and V. Rivoirard. Adaptive estimation of conditional density function. Technical report, 2014. URL http://arxiv.org/abs/1312.7402.

Richard C. Bradley. *Introduction to strong mixing conditions. Vol. 1, 2, 3.* Kendrick Press, Heber City, UT, 2007. ISBN 0-9740427-6-5.

F. Comte and V. Genon-Catalot. Convolution power kernels for density estimation. *J. Statist. Plann. Inference*, 142(7):1698–1715, 2012. ISSN 0378-3758. doi: 10.1016/j.jspi.2012.02.038. URL http://dx.doi.org/10.1016/j.jspi.2012.02.038.

Fabienne Comte and Florence Merlevède. Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM Probab. Statist.*, 6:211–238, 2002. ISSN 1292-8100. doi: 10.1051/ps:2002012. URL http://dx.doi.org/10.1051/ps:2002012. New directions in time series analysis (Luminy, 2001).

Jérôme Dedecker and Clémentine Prieur. New dependence coefficients. Examples and applications to statistics. *Probab. Theory Related Fields*, 132(2):203–236, 2005. ISSN 0178-8051. doi: 10.1007/s00440-004-0394-3. URL http://dx.doi.org/10.1007/s00440-004-0394-3.

Paul Doukhan and Sana Louhichi. A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.*, 84(2):313–342, 1999. ISSN 0304-4149. doi: 10.1016/S0304-4149(99)00055-1. URL http://dx.doi.org/10.1016/S0304-4149(99)00055-1.

Paul Doukhan and Sana Louhichi. Functional estimation of a density under a new weak dependence condition. *Scand. J. Statist.*, 28(2):325–341, 2001. ISSN 0303-6898. doi: 10.1111/1467-9469.00240. URL http://dx.doi.org/10.1111/1467-9469.00240.

Paul Doukhan and Michael H. Neumann. Probability and moment inequalities for sums of weakly dependent random variables, with applications. *Stochastic Process. Appl.*, 117(7):878–903, 2007. ISSN 0304-4149. doi: 10.1016/j.spa.2006.10.011. URL http://dx.doi.org/10.1016/j.spa.2006.10.011.

Paul Doukhan and Lionel Truquet. A fixed point approach to model random fields. *ALEA Lat. Am. J. Probab. Math. Stat.*, 3:111–132, 2007. ISSN 1980-0436.

Paul Doukhan and Olivier Wintenberger. An invariance principle for weakly dependent stationary general models. *Probab. Math. Statist.*, 27(1):45–73, 2007. ISSN 0208-4147.

M. Doumic, M. Hoffmann, P. Reynaud-Bouret, and V. Rivoirard. Nonparametric Estimation of the Division Rate of a Size-Structured Population. *SIAM J. Numer. Anal.*, 50(2):925–950, 2012. ISSN 0036-1429. doi: 10.1137/110828344. URL http://dx.doi.org/10.1137/110828344.

Irène Gannaz and Olivier Wintenberger. Adaptive density estimation under weak dependence. *ESAIM Probab. Stat.*, 14:151–172, 2010. ISSN 1292-8100. doi: 10.1051/ps:2008025. URL http://dx.doi.org/10.1051/ps:2008025.

A. Goldenshluger and O. Lepski. On adaptive minimax density estimation on R^d. *Probab. Theory Related Fields*, 159(3-4):479–543, 2014. ISSN 0178-8051. doi: 10.1007/s00440-013-0512-1. URL http://dx.doi.org/10.1007/s00440-013-0512-1.

Alexander Goldenshluger and Oleg Lepski. Universal pointwise selection rule in multivariate function estimation. *Bernoulli*, 14(4):1150–1190, 2008. ISSN 1350-7265. doi: 10.3150/08-BEJ144. URL http://dx.doi.org/10.3150/08-BEJ144.

Alexander Goldenshluger and Oleg Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39(3):1608–1632, 2011. ISSN 0090-5364. doi: 10.1214/11-AOS883. URL http://dx.doi.org/10.1214/11-AOS883.

N. Klutchnikoff. Pointwise adaptive estimation of a multivariate function. *Mathematical Methods of Statistics*, 23(2):132–150, 2014. doi: 10.3103/S1066530714020045.

O. V. Lepski. A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(3):459–470, 1990. ISSN 0040-361X. doi: 10.1137/1135065. URL http://dx.doi.org/10.1137/1135065.

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

C. Prieur. *Applications statistiques de suites faiblement dépendantes et de systèmes dynamiques*. PhD thesis, CREST, 2001a.

Clémentine Prieur. Density estimation for one-dimensional dynamical systems. *ESAIM Probab. Statist.*, 5:51–76, 2001b. ISSN 1292-8100. doi: 10.1051/ps:2001102. URL http://dx.doi.org/10.1051/ps:2001102.

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

M. Rosenblatt. A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U. S. A.*, 42:43–47, 1956. ISSN 0027-8424.

Karine Tribouley and Gabrielle Viennet. $\mathbf{L}_p$ adaptive density estimation in a β mixing framework. *Ann. Inst. H. Poincaré Probab. Statist.*, 34(2):179–208, 1998. ISSN 0246-0203. doi: 10.1016/S0246-0203(98)80029-0. URL http://dx.doi.org/10.1016/S0246-0203(98)80029-0.