ISSN 1066-5307, Mathematical Methods of Statistics, 2008, Vol. 17, No. 2, pp. 87–112. © Allerton Press, Inc., 2008.

# Adaptive Density Deconvolution with Dependent Inputs

**F. Comte[1*], J. Dedecker[2**], and M. L. Taupin[3***]**

[1]*Université Paris Descartes, MAP5, UMR CNRS 8145, France*
[2]*Université Paris 6, Laboratoire de Statistique Théorique et Appliqué, France*
[3]*IUT de Paris et Université Paris Descartes, Laboratoire MAP5, UMR 8145, France*
Received September 28, 2007; in final form, March 21, 2008

**Abstract**—In the convolution model $Z_i = X_i + \varepsilon_i$, we give a model selection procedure to estimate the density of the unobserved variables $(X_i)_{1\le i\le n}$, when the sequence $(X_i)_{i\ge 1}$ is strictly stationary but not necessarily independent. This procedure depends on whether the density of the $\varepsilon_i$ is supersmooth or ordinary smooth. The rates of convergence of the penalized contrast estimators are the same as in the independent framework, and are minimax over most regularity classes on $\mathbb{R}$. Our results apply to mixing sequences, but also to many other dependent sequences. When the errors are supersmooth, the condition on the dependence coefficients is the minimal condition of that type ensuring that the sequence $(X_i)_{i\ge 1}$ is not a long-memory process.

**Key words:** adaptive estimation, deconvolution, dependence, mixing, penalized contrast, hidden Markov models.
**2000 Mathematics Subject Classification:** 62G07–62G20.

**DOI:** 10.3103/S1066530708020014

## 1. INTRODUCTION

The problem of estimating the density of identically distributed but not independent random variables $X_1, \dots, X_n$ when they are observed with an additive and independent noise is encountered in numerous contexts. This problem is described by the model

$$Z_i = X_i + \varepsilon_i \qquad \text{for} \quad i = 1, \dots, n, \tag{1.1}$$

where one observes $Z_1, \dots, Z_n$ and where $(\varepsilon_i)_{1\le i\le n}$ are independent and identically distributed (i.i.d.), and independent of $(X_i)_{1\le i\le n}$. When $(X_i)_{i\le 1\le n}$ is a Markov chain, the model (1.1) is a particular case of hidden Markov models, with an additive structure.

Our aim is the adaptive estimation of $g$, the common distribution of the unobserved variables $(X_i)_{1\le i\le n}$, when the density $f_\varepsilon$ of $\varepsilon_i$ is known. More precisely we shall construct an estimator of $g$ without any prior knowledge about its smoothness using the observations $(Z_i)_{i\le 1\le n}$ and the knowledge of the convolution kernel $f_\varepsilon$.

We shall assume that the known density $f_\varepsilon$ belongs to various collections of densities and that the dependence properties of the sequence $(X_i)_{i\ge 1}$ are described by appropriate dependence coefficients. More precisely, we consider two types of dependent sequences. We assume either that the sequence $(X_i)_{i\ge 1}$ is absolutely regular in the sense of Rozanov and Volkonskii [29], or that it is $\tau$-dependent in the sense of Dedecker and Prieur [11]. These dependence conditions are presented in Section 2 and motivated through various examples.

In density deconvolution, two factors determine the estimation accuracy. First, the smoothness of the density $g$ to be estimated, and secondly the smoothness of the error density, the worst rates of convergence being obtained for the smoothest error densities.

*E-mail: fabienne.comte@univ-paris5.fr
**E-mail: jerome.dedecker@upmc.fr
***E-mail: Marie-Luce.Taupin@univ-paris5.fr

We shall consider two classes of densities for $f_\varepsilon$: first the so-called supersmooth densities with exponential decay of their Fourier transform, and next the class of ordinary smooth densities with Fourier transform having a polynomial decay.

Let us briefly recall the previous results in the independent framework. To our knowledge, the first adaptive estimator has been proposed by Pensky and Vidakovic [22]. It is a wavelet estimator constructed *via* a thresholding procedure. This estimator achieves the minimax rates when $g$ belongs to a Sobolev class, but it fails to reach the minimax rates when both the error density and $g$ are supersmooth.

More recently, Comte *et al.* [9] have proposed an adaptive estimator of $g$ constructed by minimizing an appropriate penalized contrast function only depending on the observations and on $f_\varepsilon$. This estimator is minimax (sometimes within a negligible logarithmic factor) in all cases, where lower bounds have been previously known (i.e., in most cases). More precisely, the authors obtain non-asymptotic upper bounds for the Mean Integrated Square Error (MISE), which ensure an automatic trade-off between the bias term and the penalty term. Hence, the estimator automatically achieves the best rate obtained by the collection of non-penalized estimators when the (unknown) optimal space is selected (sometimes up to a negligible logarithmic factor). When both the density and the errors are supersmooth, this adaptive estimator significantly improves on the rates given by the adaptive estimator constructed in Pensky and Vidakovic [22], whereas both adaptive estimators have the same rate in the other cases. This improvement partly comes from the choice of the Shannon basis (see Section 3.2) instead of the wavelet basis considered in Pensky and Vidakovic [22].

In the dependent context, we follow the approach proposed in Comte *et al.* [9] to construct adaptive estimators of $g$. They are obtained by minimizing an appropriate penalized contrast function, with a penalty function depending on the known density $f_\varepsilon$. The adaptive estimators have the same rates as in the independent case under mild conditions on the dependence coefficients of $(X_i)_{i\geq 1}$. The important point here is that the penalty function is the same (or almost the same) as in the independent framework. More precisely, it has the same order as in the independent framework and even more important, it does not depend on the dependence coefficients of the sequence $(X_i)_{i\geq 1}$. Indeed, as compared with the independent case, the dependence induces additional terms in the risk bounds, but these terms are negligible with respect to the terms of the independent case. This explains why one can choose a penalty function which does not depend on the dependence coefficient. This fact is noteworthy when compared with density estimation in dependent context. Indeed, when the $(X_i)_{1\leq i\leq n}$ are observed (i.e., $\varepsilon_i = 0$), the threshold level proposed in Tribouley and Viennet [25] as well as the penalty function given in Comte and Merlevède [8] (see also our Proposition 5.1) both depend on the mixing coefficients of the sequence $(X_i)_{i\geq 1}$.

In Section 4 we deal with non-adaptive estimators. As usual, we show that the MISE of the minimum contrast estimator is bounded by a squared bias plus a variance term. The variance term can be split into two terms. The first and dominating term of the variance is exactly the variance of a density deconvolution estimator in the independent context. It is as usual related to $\int_{|x|\leq C_n} |f_\varepsilon^*(x)|^{-2}\,dx$, $C_n \to \infty$. The second and negligible term in the variance is the term involving the dependence structure of the sequence $(X_i)_{i\geq 1}$. The main consequence of this first result is that this non-adaptive estimator reaches the (minimax) rates of the i.i.d. case (as given in Fan [15], Butucea [4], and Butucea and Tsybakov [5]), as long as the dependence coefficients are summable. Moreover, even if the coefficients are not summable, there is no loss in the rate provided that the partial sums of the coefficients do not grow too fast with respect to $\int_{|x|\leq C_n} |f_\varepsilon^*(x)|^{-2}\,dx$. These results have to be compared with previously known results for non-adaptive density deconvolution in dependent contexts. For strongly mixing sequences in the sense of Rosenblatt [24], Masry [18] proposed a kernel-type estimator for the joint density $g_p$ of $(X_1, \ldots, X_p)$ when it exists. For the (pointwise) Mean Square Error, he obtains the same rates as in the i.i.d. case provided that the sequence of strong mixing coefficients decreases faster than $n^{-2}$ for ordinary smooth $f_\varepsilon$ and provided that it decreases faster than $n^{-1}$ for supersmooth $f_\varepsilon$. When $p = 1$, our assumption on the mixing coefficients is weaker, since we only need $\sum_{n>0} \alpha(n) < \infty$ in both cases (see our Remark 4.1).

In the main part (Section 5), we study the adaptive estimators. We show that the squared bias term and the variance term obtained in the upper bound of the MISE of the adaptive estimator are the same as in the independent case. The model selection procedure depends on whether the density $f_\varepsilon$ is supersmooth or ordinary smooth.

When $f_\varepsilon$ is supersmooth, an adaptive estimator is constructed with the same penalty as in the independent case. Its rate of convergence is exactly the same as in the independent case, provided that the dependence coefficients of $(X_i)_{i\geq 1}$ are summable. The main tools in this case are covariance inequalities for dependent variables and concentration inequalities. The case of supersmooth errors is particularly important, since it contains the case of Gaussian errors. It also contains the stochastic volatility model in which $\varepsilon_i \sim \log(\mathcal{N}(0,1)^2)$ (see van Es *et al.* [26, 27], Comte [6], Comte and Genon-Catalot [7]).

When $f_\varepsilon$ is ordinary smooth, an adaptive estimator is constructed with a penalty of the same order as in the independent context. Its rate of convergence is exactly the same as in the independent case. For ordinary smooth errors, the main tools are the coupling properties of the dependence coefficients (see Section 2.1). To use these properties, we need to consider a more restrictive type of dependence than for supersmooth errors, and we need to assume a polynomial decrease of the coefficients.

In both cases, super and ordinary smooth $f_\varepsilon$, the results hold for $\beta$-mixing and $\tau$-dependent random variables $(X_i)_{i\geq 1}$. To our knowledge, this is the first time that adaptive density deconvolution in a dependent context is considered. The robustness of this estimation procedure to dependence strongly uses the independence between $(X_i)_{1\leq i\leq n}$ and $(\varepsilon_i)_{i\leq 1\leq n}$ and the fact that the errors are i.i.d. random variables.

We refer to Comte *et al.* [9, 10] for practical implementation of the estimators and for the calibration of the constants in the penalty functions. In Comte *et al.* [10], the robustness of the procedure to various forms of dependence has been experimented in practice (see Tables 4 and 5 therein).

## 2. SOME MEASURES OF DEPENDENCE

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $Y$ be a random variable with values in a Banach space $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$. Denote by $\Lambda_\kappa(\mathbb{B})$ the set of $\kappa$-Lipschitz functions, i.e., the functions $f$ from $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ to $\mathbb{R}$ such that $|f(x)-f(y)| \leq \kappa\|x-y\|_{\mathbb{B}}$. Let $\mathcal{M}$ be a $\sigma$-algebra of $\mathcal{A}$. Let $\mathbb{P}_{Y|\mathcal{M}}$ be a conditional distribution of $Y$ given $\mathcal{M}$, let $\mathbb{P}_Y$ be the distribution of $Y$, and let $\mathcal{B}(\mathbb{B})$ be the Borel $\sigma$-algebra on $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$.

Define now

$$\beta(\mathcal{M}, \sigma(Y)) = \mathbb{E}\Big(\sup_{A\in\mathcal{B}(\mathbb{B})} |\mathbb{P}_{Y|\mathcal{M}}(A) - \mathbb{P}_Y(A)|\Big),$$

$$\text{and if } \mathbb{E}(\|Y\|) < \infty, \quad \tau(\mathcal{M}, Y) = \mathbb{E}\Big(\sup_{f\in\Lambda_1(\mathbb{B})} |\mathbb{P}_{Y|\mathcal{M}}(f) - \mathbb{P}_Y(f)|\Big).$$

The coefficient $\beta(\mathcal{M}, \sigma(Y))$ is the usual mixing coefficient introduced by Rozanov and Volkonskii [29]. The coefficient $\tau(\mathcal{M}, Y)$ has been introduced by Dedecker and Prieur [11].

Let $\mathbf{X} = (X_i)_{i\geq 1}$ be a strictly stationary sequence of real-valued random variables. For any $k \geq 0$, the coefficients $\beta_{\mathbf{X},1}(k)$ and $\tau_{\mathbf{X},1}(k)$ are defined by

$$\beta_{\mathbf{X},1}(k) = \beta(\sigma(X_1), \sigma(X_{1+k})), \tag{2.1}$$

$$\text{and if } \mathbb{E}(|X_1|) < \infty, \quad \tau_{\mathbf{X},1}(k) = \tau(\sigma(X_1), X_{1+k}). \tag{2.2}$$

On $\mathbb{R}^l$, we introduce the norm $\|x-y\|_{\mathbb{R}^l} = l^{-1}(|x_1-y_1| + \cdots + |x_l - y_l|)$. Let $\mathcal{M}_i = \sigma(X_k, 1 \leq k \leq i)$. The coefficients $\beta_{\mathbf{X},\infty}(k)$ and $\tau_{\mathbf{X},\infty}(k)$ are defined by

$$\beta_{\mathbf{X},\infty}(k) = \sup_{i\geq 1, l\geq 1} \sup\Big\{\beta(\mathcal{M}_i, \sigma(X_{i_1}, \dots, X_{i_l})),\ i+k \leq i_1 < \cdots < i_l\Big\},$$

$$\text{and if } \mathbb{E}(|X_1|) < \infty, \quad \tau_{\mathbf{X},\infty}(k) = \sup_{i\geq 1, l\geq 1} \sup\Big\{\tau(\mathcal{M}_i, (X_{i_1}, \dots, X_{i_l})),\ i+k \leq i_1 < \cdots < i_l\Big\}.$$

Let $Q_X$ be the generalized inverse of the tail function $x \to \mathbb{P}(|X_1| > x)$. We have the inequalities

$$\tau_{\mathbf{X},1}(k) \leq 2 \int_0^{\beta_{\mathbf{X},1}(k)} Q_X(u)\,du \qquad \text{and} \qquad \tau_{\mathbf{X},\infty}(k) \leq 2 \int_0^{\beta_{\mathbf{X},\infty}(k)} Q_X(u)\,du. \tag{2.3}$$

### 2.1. Coupling

We recall the coupling properties of these coefficients. Assume that $\Omega$ is rich enough, which means that there exists $U$ uniformly distributed over $[0,1]$ and independent of $\mathcal{M} \vee \sigma(X)$. There exist two $\mathcal{M} \vee \sigma(U) \vee \sigma(X)$-measurable random variables $X_1^*$ and $X_2^*$ distributed as $X$ and independent of $\mathcal{M}$ such that

$$\beta(\mathcal{M}, \sigma(X)) = \mathbb{P}(X \neq X_1^*) \qquad \text{and} \qquad \tau(\mathcal{M}, X) = \mathbb{E}(\|X - X_2^*\|_{\mathbb{B}}). \tag{2.4}$$

The first equality in (2.4) is due to Berbee [1], and the second one has been established in Dedecker and Prieur [11], Section 7.1.

### 2.2. Covariance Inequalities

Denote by $\|\cdot\|_{\infty,\mathbb{P}}$ the $\mathbb{L}^\infty(\Omega,\mathbb{P})$-norm. Let $X$, $Y$ be two real-valued random variables, and let $f$, $h$ be two measurable functions from $\mathbb{R}$ to $\mathbb{C}$. Then

$$|\operatorname{Cov}(f(Y), h(X))| \le 2\|f(Y)\|_{\infty,\mathbb{P}} \, \|h(X)\|_{\infty,\mathbb{P}} \, \beta(\sigma(X), \sigma(Y)), \tag{2.5}$$

and if $\mathrm{Lip}(h)$ is the Lipschitz coefficient of $h$,

$$|\operatorname{Cov}(f(Y), h(X))| \le \|f(Y)\|_{\infty,\mathbb{P}} \, \mathrm{Lip}(h) \, \tau(\sigma(Y), X). \tag{2.6}$$

Inequalities (2.5) and (2.6) follow from the coupling properties (2.4) by noting that if $X^*$ is distributed as $X$ and independent of $Y$,

$$\operatorname{Cov}(f(Y), h(X)) = \mathbb{E}\big(\overline{f(Y)}(h(X) - h(X^*))\big).$$

### 2.3. Examples

Examples of $\beta$-mixing sequences are well known (we refer to the books by Doukhan [13] and Bradley [3]). One of the most important examples is the following: a stationary, irreducible, aperiodic and positively recurrent Markov chain $(X_i)_{i\ge 1}$ is $\beta$-mixing, which means that $\beta_{\mathbf{X},\infty}(k)$ tends to zero as $k$ tends to infinity.

Unfortunately, many simple Markov chains are not $\beta$-mixing (and not even strongly mixing in the sense of Rosenblatt [24]). For instance, if $(\epsilon_i)_{i\ge 1}$ are i.i.d. with marginal such that

$$\mathbb{P}(\epsilon_1 = 1) = \mathbb{P}(\epsilon_1 = 0) = 1/2,$$

then the stationary solution $(X_i)_{i\ge 0}$ of the equation

$$X_n = \frac{1}{2}(X_{n-1} + \epsilon_n), \qquad X_0 \text{ independent of } (\epsilon_i)_{i\ge 1}, \tag{2.7}$$

is not $\beta$-mixing (and not even strongly mixing), since $\beta_{\mathbf{X},1}(k) = 1$ for any $k \ge 0$. By contrast, for this particular example, one has $\tau_{\mathbf{X},\infty}(k) \le 2^{-k}$. More generally, the coefficient $\tau_{\mathbf{X},\infty}(k)$ is easy to compute in many situations (see Dedecker and Prieur [11]). Let us recall some important examples:

**Linear processes.** Assume that $X_n = \sum_{j\ge 0} a_j \xi_{n-j}$, where $(\xi_i)_{i\in\mathbb{Z}}$ are i.i.d. One has the bounds

$$\tau_{\mathbf{X},\infty}(k) \le 2\mathbb{E}(|\xi_0|) \sum_{j\ge k} |a_j| \qquad \text{and} \qquad \tau_{\mathbf{X},\infty}(k) \le \sqrt{2\operatorname{Var}(\xi_0) \sum_{j\ge k} a_j^2}.$$

**Markov chains.** Let $(X_n)_{n\ge 0}$ be a stationary Markov chain such that $X_n = F(X_{n-1}, \xi_n)$ for some measurable function $F$ and some i.i.d. sequence $(\xi_i)_{i\ge 1}$ independent of $X_0$. Assume that there exists $\kappa < 1$ such that

$$\mathbb{E}\big(|F(x, \xi_0) - F(y, \xi_0)|\big) \le \kappa |x - y|.$$

Then one has the inequality

$$\tau_{\mathbf{X},\infty}(k) \le 2\mathbb{E}(|X_0|)\kappa^k.$$

An important example is $X_n = f(X_{n-1}) + \xi_n$ for some $\kappa$-Lipschitz function $f$.

**Expanding maps.** Let $T$ be a Borel-measurable map from $[0,1]$ to $[0,1]$. If $\nu$ is a $T$-invariant probability, the sequence $(Y_i = T^i)_{i\geq 0}$ of random variables from $([0,1],\nu)$ to $[0,1]$ is strictly stationary. Define the operator $K$ from $\mathbb{L}^1([0,1],\nu)$ to $\mathbb{L}^1([0,1],\nu)$ *via* the equality

$$\int_0^1 (Kh)(x)k(x)\,\nu(dx) = \int_0^1 h(x)(k\circ T)(x)\,\nu(dx),$$

where $h \in \mathbb{L}^1([0,1],\nu)$ and $k \in \mathbb{L}^\infty([0,1],\nu)$. It is easy to check that $(Y_1, Y_2, \ldots, Y_n)$ has the same distribution as $(X_n, X_{n-1}, \ldots, X_1)$, where $(X_i)_{i\in\mathbb{Z}}$ is a stationary Markov chain with invariant distribution $\nu$ and transition kernel $K$. If $T$ is uniformly expanding (see, for instance, the assumptions on p. 218 in Dedecker and Prieur [11]), then there exist $C > 0$ and $\rho$ in $]0,1[$ such that

$$\tau_{\mathbf{X},\infty}(k) \leq C\rho^k$$

(see Dedecker and Prieur, p. 230). Note that the Markov chain $(X_i)_{i\geq 1}$ is not $\beta$-mixing (and not even strongly mixing). Indeed, $\beta(\sigma(X_1),\sigma(X_n)) = \beta(\sigma(T^n),\sigma(T))$. Since $\sigma(T^n)\subset\sigma(T)$, it follows that

$$\beta(\sigma(X_1),\sigma(X_n)) \geq \beta(\sigma(T^n),\sigma(T^n)) = \beta(\sigma(T),\sigma(T))$$

and the latter is positive as soon as $\nu$ is nontrivial.

## 3. ASSUMPTIONS AND ESTIMATORS

For two complex-valued functions $u$ and $v$ in $\mathbb{L}_2(\mathbb{R})\cap\mathbb{L}_1(\mathbb{R})$, let

$$u^*(x) = \int e^{itx}u(t)\,dt, \quad u * v(x) = \int u(y)v(x-y)\,dy, \quad \text{and} \quad \langle u, v\rangle = \int u(x)\overline{v}(x)\,dx$$

with $\overline{z}$ the conjugate of a complex number $z$. We also use the notation

$$\|u\|_1 = \int |u(x)|\,dx, \quad \|u\|^2 = \int |u(x)|^2\,dx, \quad \text{and} \quad \|u\|_\infty = \sup_{x\in\mathbb{R}}|u(x)|.$$

### *3.1. Assumptions for Density Deconvolution*

The smoothness of $f_\varepsilon$ is described by assumption $(\mathbf{A}_1^\varepsilon)$ below.

$(\mathbf{A}_1^\varepsilon)$ There exist $\kappa_0' \geq \kappa_0 > 0$ and $\gamma \geq 0$, $\mu \geq 0$, $\delta \geq 0$ (with $\gamma > 0$ if $\delta = 0$) such that $f_\varepsilon^*$ satisfies

$$\kappa_0(x^2+1)^{-\gamma/2}\exp\{-\mu|x|^\delta\} \leq |f_\varepsilon^*(x)| \leq \kappa_0'(x^2+1)^{-\gamma/2}\exp\{-\mu|x|^\delta\}.$$

$(\mathbf{A}_2^\varepsilon)$ For all $x\in\mathbb{R}$, $f_\varepsilon^*(x)\neq 0$.

Note that $(\mathbf{A}_1^\varepsilon)$ implies $(\mathbf{A}_2^\varepsilon)$. Since $f_\varepsilon$ is known, the constants $\mu$, $\delta$, $\kappa_0$, and $\gamma$ defined in $(\mathbf{A}_1^\varepsilon)$ are also known. By convention, we set $\delta = 0$ if $\mu = 0$.

When $\delta = 0$ in $(\mathbf{A}_1^\varepsilon)$, $f_\varepsilon$ is usually called "ordinary smooth". When $\mu > 0$ and $\delta > 0$, $f_\varepsilon$ is called "supersmooth". Densities satisfying $(\mathbf{A}_1^\varepsilon)$ with $\delta > 0$ and $\mu > 0$ are infinitely differentiable. The standard examples for supersmooth densities are the following: Gaussian or Cauchy distributions are supersmooth of order $\gamma = 0$, $\delta = 2$ and $\gamma = 0$, $\delta = 1$ respectively. When $\varepsilon = \log(\eta^2)$ with $\eta \sim \mathcal{N}(0,1)$ as in van Es *et al.* [26, 27], then $\varepsilon$ is supersmooth with $\delta = 1$, $\gamma = 0$, and $\mu = \pi/2$. For ordinary smooth densities, one can cite, for instance, the double exponential (also called Laplace) distribution with $\delta = 0 = \mu$ and $\gamma = 2$. Although densities with $\delta > 2$ exist, they are difficult to express in a closed form. Nevertheless, our results hold for such densities.

Classically, the slowest rates of convergence for estimating $g$ are obtained for supersmooth error densities. In particular, when $\varepsilon$ is Gaussian and $g$ belongs to Sobolev classes, the minimax rates are negative powers of $\log n$ (see Fan [15]). Nevertheless, the rates are improved if $g$ has stronger smoothness properties described by the set

$$\mathcal{S}_{s,r,b}(C_1) = \left\{\psi \ \text{ such that } \int_{-\infty}^{+\infty} |\psi^*(x)|^2(x^2+1)^s\exp\{2b|x|^r\}\,dx \leq C_1\right\} \tag{3.1}$$

for $s$, $r$, $b$ non-negative numbers. By convention, we set $r = 0$ if $b = 0$.

Such smoothness classes are classically considered both in deconvolution and in density estimation without errors. When $r = 0$, (3.1) corresponds to a Sobolev ball. The functions in (3.1) with $r > 0$ and $b > 0$ are infinitely many times differentiable. They admit analytic continuation on a finite width strip when $r = 1$ and on the whole complex plane if $r = 2$.

Subsequently, the density $g$ is supposed to satisfy the following assumption.

$(\mathbf{A}_3^X)$ The density $g \in \mathbb{L}_2(\mathbb{R})$ and there exists $M_2 > 0$ such that $\int x^2 g^2(x)\,dx < M_2 < \infty$.

Assumption $(\mathbf{A}_3^X)$, which is due to the construction of the estimator, is quite unusual in density estimation. It already appears in density deconvolution in the independent framework in Comte *et al*. [9, 10]. It also appears in a slightly different way in Pensky and Vidakovic [22] who assume, instead of $(\mathbf{A}_3^X)$, that $\sup_{x\in\mathbb{R}} |x|g(x) < \infty$. It is important to note that Assumption $(\mathbf{A}_3^X)$ is very unrestrictive.

All densities having tails of order $|x|^{-(s+1)}$ as $x$ tends to infinity satisfy $(\mathbf{A}_3^X)$ only if $s > 1/2$. One can cite, for instance, the Cauchy distribution or all stable distributions with exponent $r > 1/2$ (see Devroye [12]). The Lévy distribution, with exponent $r = 1/2$ does not satisfy $(\mathbf{A}_3^X)$.

### *3.2. The Projection Spaces*

Let $\varphi(x) = \sin(\pi x)/(\pi x)$. For $m \in \mathbb{N}$ and $j \in \mathbb{Z}$, set $\varphi_{m,j}(x) = \sqrt{m}\varphi(mx - j)$. The collection of functions $\{\varphi_{m,j}\}_{j\in\mathbb{Z},m\in\mathbb{N}^*}$ is a basis of $\mathbb{L}^2(\mathbb{R})$ (see, e.g., Meyer [20], p. 22). For $m = 2^k$, $k \in \mathbb{N}$, it is known as the Shannon basis. Though we choose here integer values for $m$, a thinner grid would also be possible. Let us define

$$S_m = \overline{\operatorname{span}}\{\varphi_{m,j},\ j \in \mathbb{Z}\}, \qquad m \in \mathbb{N}.$$

The space $S_m$ is exactly the subspace of $\mathbb{L}_2(\mathbb{R})$ of functions having a Fourier transform with compact support contained in $[-\pi m, \pi m]$.

The orthogonal projection of $g$ on $S_m$ is $g_m = \sum_{j\in\mathbb{Z}} a_{m,j}(g)\varphi_{m,j}$, where $a_{m,j}(g) = \langle \varphi_{m,j}, g\rangle$. To obtain representations having a finite number of "coordinates", we introduce

$$S_m^{(n)} = \overline{\operatorname{span}}\left\{\varphi_{m,j}, |j| \le k_n\right\}$$

with integers $k_n$ to be specified later. The family $\{\varphi_{m,j}\}_{|j|\le k_n}$ is an orthonormal basis of $S_m^{(n)}$ and the orthogonal projection of $g$ on $S_m^{(n)}$ is given by $g_m^{(n)} = \sum_{|j|\le k_n} a_{m,j}(g)\varphi_{m,j}$.

### *3.3. Construction of the Minimum Contrast Estimators*

For an arbitrary fixed integer $m$, an estimator of $g$ belonging to $S_m^{(n)}$ is defined by

$$\hat{g}_m^{(n)} = \arg\min_{t\in S_m^{(n)}} \gamma_n(t), \tag{3.2}$$

where, for $t$ in $S_m^{(n)}$,

$$\gamma_n(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\,\|t\|^2 - 2u_t^*(Z_i)\right] \qquad \text{with} \quad u_t(x) = \frac{1}{2\pi}\left(\frac{t^*(-x)}{f_\varepsilon^*(x)}\right).$$

By using Parseval's and inverse Fourier formulae we obtain that $\mathbb{E}\left[u_t^*(Z_i)\right] = \langle t, g\rangle$, so that $\mathbb{E}(\gamma_n(t)) = \|t - g\|^2 - \|g\|^2$ is minimal when $t = g$. This shows that $\gamma_n(t)$ suits well for the estimation of $g$. Classical calculations show that

$$\hat{g}_m^{(n)} = \sum_{|j|\le k_n} \hat{a}_{m,j}\varphi_{m,j} \quad \text{with} \quad \hat{a}_{m,j} = \frac{1}{n}\sum_{i=1}^{n} u^*_{\varphi_{m,j}}(Z_i) \qquad \text{and} \qquad \mathbb{E}(\hat{a}_{m,j}) = \langle g, \varphi_{m,j}\rangle = a_{m,j}.$$

*3.4. Minimum Penalized Contrast Estimator*

As in the independent framework, the minimum penalized estimator of $g$ is defined as $\tilde{g} = \hat{g}_{\hat{m}_g}$, where $\hat{m}_g$ is chosen in a purely data-driven way. The main point of the estimation procedure lies in the choice of $m = \hat{m}_g$ for the estimators $\hat{g}_m$ from Section 3.3 in order to mimic the oracle parameter

$$\breve{m}_g = \arg\min_m \mathbb{E}\|\hat{g}_m - g\|_2^2. \tag{3.3}$$

For $m = 1, \ldots, m_n$ with $m_n \le n$, the model selection is performed in an automatic way, using the following penalized criteria

$$\tilde{g} = \hat{g}_{\hat{m}}^{(n)} \qquad \text{with} \quad \hat{m} = \arg\min_{m\in\{1,\ldots,m_n\}} \left[\gamma_n(\hat{g}_m^{(n)}) + \mathrm{pen}(m)\right], \tag{3.4}$$

where $\mathrm{pen}(m)$ is a penalty function given in the theorems that depends on $f_\varepsilon^*$ through $\Delta(m)$ defined by

$$\Delta(m) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{1}{|f_\varepsilon^*(x)|^2}\,dx. \tag{3.5}$$

The key point in the dependent context is to find a penalty function not depending on the mixing coefficients such that

$$\mathbb{E}\|\tilde{g} - g\|^2 \le C \inf_{m\in\{1,\ldots,m_n\}} \mathbb{E}\|\hat{g}_m - g\|^2.$$

## 4. RISK BOUNDS FOR THE MINIMUM CONTRAST ESTIMATORS $\hat{g}_m^{(n)}$

We focus here on non-adaptive estimation, starting with the presentation of general upper bounds for MISEs of the minimum contrast estimators $\hat{g}_m^{(n)}$.

**Proposition 4.1.** *If* $(\mathbf{A}_2^\varepsilon)$ *and* $(\mathbf{A}_3^X)$ *hold, then*

$$\mathbb{E}\|g - \hat{g}_m^{(n)}\|^2 \le \|g - g_m\|^2 + \frac{m^2(M_2+1)}{k_n} + \frac{2\Delta(m)}{n} + \frac{2R_{n,m}}{n},$$

*where*

$$R_{n,m} = \frac{1}{\pi}\sum_{k=2}^{n}\int_{-\pi m}^{\pi m} \left|\,\mathrm{Cov}(e^{ixX_1}, e^{ixX_k})\right| dx. \tag{4.1}$$

*Moreover,* $R_{n,m} \le \min(R_{n,m,\beta}, R_{n,m,\tau})$, *where*

$$R_{n,m,\beta} = 4m\sum_{k=1}^{n-1}\beta_{\mathbf{X},1}(k) \qquad \textit{and} \qquad R_{n,m,\tau} = \pi m^2\sum_{k=1}^{n-1}\tau_{\mathbf{X},1}(k).$$

**Remark 4.1.** The term $R_{n,m}$ can be easily bounded for many other dependent sequences. For instance, if $\alpha_{\mathbf{X},1} = \alpha(\sigma(X_1), \sigma(X_{1+k}))$ is the usual strong mixing coefficient of Rosenblatt [24], one has the upper bound $R_{n,m} \le 16m\sum_{k=1}^{n-1}\alpha_{\mathbf{X},1}(k)$. If $\mathbf{X}$ is a stationary sequence of associated random variables (see Esary *et al.* [14] for the definition), then $|\,\mathrm{Cov}(e^{ixX_1}, e^{ixX_k})| \le 4x^2\,\mathrm{Cov}(X_1, X_k)$, so that $R_{n,m} \le (8\pi^2/3)m^3\sum_{k=2}^{n}\mathrm{Cov}(X_1, X_k)$. For more about density deconvolution with associated inputs, we refer to the paper by Masry [19].

We now comment on the rates resulting from Proposition 4.1. As usual, the variance term $\Delta(m)/n$ depends on the rate of decay of the Fourier transform of $f_\varepsilon$. According to Butucea and Tsybakov [5], under $(\mathbf{A}_1^\varepsilon)$, we have

$$\lambda_1(f_\varepsilon, \kappa_0')\Gamma(m)(1+o(1)) \le \Delta(m) \le \lambda_1(f_\varepsilon, \kappa_0)\Gamma(m)(1+o(1)) \quad \text{as } m \to \infty, \tag{4.2}$$

where $\Gamma(m) = (1 + (\pi m)^2)^\gamma (\pi m)^{1-\delta} \exp\{2\mu(\pi m)^\delta\}$,

$$\lambda_1(f_\varepsilon, \kappa_0) = \frac{1}{\kappa_0^2 \pi R(\mu, \delta)}, \qquad \text{and} \quad R(\mu,\delta) = \mathbf{1}_{\{\delta=0\}} + 2\mu\delta \mathbf{1}_{\{\delta>0\}}. \tag{4.3}$$

If $(\mathbf{A}_1^\varepsilon)$ and $(\mathbf{A}_3^X)$ hold and if $k_n \geq n^2$, we have the upper bound

$$\mathbb{E}\|g - \hat{g}_m^{(n)}\|^2 \leq \|g - g_m\|^2 + \frac{m(M_2+1)}{n} + \frac{2\lambda_1(f_\varepsilon,\kappa_0)\Gamma(m)}{n} + \frac{2R_{n,m}}{n}. \tag{4.4}$$

Finally, since $g_m$ is the orthogonal projection of $g$ on $S_m$, we get that $g_m^* = g^* \mathbf{1}_{[-m\pi, m\pi]}$ and therefore

$$\|g - g_m\|^2 = \frac{1}{2\pi}\|g^* - g_m^*\|^2 = \frac{1}{2\pi}\int\limits_{|x|\geq \pi m} |g^*|^2(x)\,dx.$$

If $g$ belongs to the class $\mathcal{S}_{s,r,b}(C_1)$ defined in (3.1), then

$$\|g - g_m\|^2 \leq \frac{C_1}{2\pi}(m^2\pi^2 + 1)^{-s}\exp\{-2b\pi^r m^r\}.$$

Hence, according to (4.4), if $(\mathbf{A}_1^\varepsilon)$ and $(\mathbf{A}_3^X)$ hold and $k_n \geq n^2$, the risk of $\hat{g}_m^{(n)}$ is bounded by

$$\frac{C_1}{2\pi}(m^2\pi^2+1)^{-s}\exp\{-2b\pi^r m^r\} + \frac{2\lambda_1(f_\varepsilon,\kappa_0)(1+(\pi m)^2))^\gamma(\pi m)^{1-\delta}\exp\{2\mu\pi^\delta m^\delta\}}{n} + \frac{m(M_2+1)}{n} + \frac{2R_{n,m}}{n}.$$

If $\sum_{k>0}\beta_{\mathbf{X},1}(k) < \infty$, the remainder term $n^{-1}R_{n,m} + n^{-1}m(M_2+1)$ is of order $m/n$, which is negligible with respect to the main term $\Delta(m)/n$.

If $\sum_{k>0}\tau_{\mathbf{X},1}(k) < \infty$, the remainder terms $n^{-1}R_{n,m} + n^{-1}m(M_2+1)$ are of order $n^{-1}m^2$. Hence, provided that $\gamma > 1/2$ when $\delta = 0$ in $(\mathbf{A}_1^\varepsilon)$, the remainder terms $n^{-1}R_{n,m} + n^{-1}m^2(M_2+1)$ of order $m^2/n$ are negligible with respect to the main term $\Delta(m)/n$.

As in the independent case, we choose $\breve{m}$ as the minimizer of

$$(m^2\pi^2+1)^{-s}\exp\{-2b\pi^r m^r\} + \frac{(\pi m)^{2\gamma+1-\delta}\exp\{2\mu\pi^\delta m^\delta\}}{n}.$$

The behavior of $\breve{m}$ is recalled in Table 1. Hence the rate of convergence of $\hat{g}_{\breve{m}}^{(n)}$ is the same as in the i.i.d. case (see Table 1 below).

**Table 1.** Choice of $\breve{m}$ and corresponding rates for $g$ in $S_{s,r,b}(C_1)$ under $(\mathbf{A}_1^\varepsilon)$ and $(\mathbf{A}_3^X)$.

| | | $f_\varepsilon$: $\delta = 0$ ordinary smooth | $f_\varepsilon$: $\delta > 0$ supersmooth |
|---|---|---|---|
| $g$ | $r = 0$ Sobolev($s$) | $\pi\breve{m} = O(n^{1/(2s+2\gamma+1)})$<br>rate $= O(n^{-2s/(2s+2\gamma+1)})$<br>*minimax rate* | $\pi\breve{m} = [\log n/(2\mu+1)]^{1/\delta}$<br>rate $= O((\log n)^{-2s/\delta})$<br>*minimax rate* |
| | $r > 0$ $\mathcal{C}^\infty$ | $\pi\breve{m} = [\log n/2b]^{1/r}$<br>rate $= O\big(\frac{\log(n)^{(2\gamma+1)/r}}{n}\big)$<br>*minimax rate* | $\breve{m}$ solution of<br>$\breve{m}^{2s+2\gamma+1-r}\exp\{2\mu(\pi\breve{m})^\delta + 2b\pi^r\breve{m}^r\} = O(n)$<br>*minimax rate if $r < \delta$ and $s = 0$* |

When $r > 0, \delta > 0$, the value of $\breve{m}$ is not explicitly given. It is obtained as the solution of the equation

$$\breve{m}^{2s+2\gamma+1-r}\exp\{2\mu(\pi\breve{m})^{\delta}+2b\pi^r\breve{m}^r\}=O(n).$$

Consequently, the rate of $\hat{g}^{(n)}_{\breve{m}}$ is not explicit and depends on the ratio $r/\delta$. If $r/\delta$ or $\delta/r$ belongs to $]k/(k+1);(k+1)/(k+2)]$ with $k$ integer, the rate of convergence can be expressed as a function of $k$. We refer to Comte *et al.* [9] for further discussions about those rates. We refer to Lacour [17] for explicit formulae for the rates in the special case $r > 0, \delta > 0$.

## 5. RISK BOUNDS FOR ADAPTIVE ESTIMATORS

In the previous section, the construction of the estimators required the knowledge of the smoothness of $g$. We now come to adaptive estimation, without such prior knowledge.

For $\varpi > 1$ and $a > 1$, let $\mathrm{pen}(m)$ be defined by

$$\mathrm{pen}(m)=\begin{cases}4a\varpi\dfrac{\Delta(m)}{n} & \text{if}\quad 0\le\delta<1/3,\\ 4a\left(1+\dfrac{98\mu\lambda_2(f_\varepsilon,\kappa_0)}{\lambda_1(f_\varepsilon,\kappa_0')}(\pi m)^{\min((3\delta/2-1/2)_+,\delta)}\right)\dfrac{\Delta(m)}{n} & \text{if}\quad \delta\ge 1/3.\end{cases}\tag{5.1}$$

The constant $\lambda_1(f_\varepsilon,\kappa_0)$ is defined in (4.3) and $\lambda_2(f_\varepsilon,\kappa_0)$ is given by

$$\lambda_2(f_\varepsilon,\kappa_0)=\|f_\varepsilon\|\kappa_0^{-1}\sqrt{2\lambda_1(f_\varepsilon,\kappa_0)}\mathbf{1}_{0<\delta\le1}+2\lambda_1(f_\varepsilon,\kappa_0)\mathbf{1}_{\delta>1}.\tag{5.2}$$

In order to bound up $\mathrm{pen}(m)$, we suppose that

$$\pi m_n\le\begin{cases}n^{1/(2\gamma+1)} & \text{if}\quad\delta=0,\\ \left[\dfrac{\log n}{2\mu}+\dfrac{2\gamma+1-\delta}{2\delta\mu}\log\left(\dfrac{\log n}{2\mu}\right)\right]^{1/\delta} & \text{if}\quad\delta>0.\end{cases}\tag{5.3}$$

Subsequently we set

$$\kappa_a=(a+1)/(a-1)\qquad\text{and}\qquad C_a=\max(\kappa_a^2,2\kappa_a).\tag{5.4}$$

### 5.1. A First Bound in Adaptive Density Deconvolution

Theorem 5.1 gives a general bound, which holds under mild dependence conditions, for $f_\varepsilon$ being either ordinary smooth or supersmooth.

**Theorem 5.1.** *Assume that $f_\varepsilon$ satisfies $(\mathbf{A}_1^\varepsilon)$, that $g$ satisfies $(\mathbf{A}_3^X)$, and that $m_n$ satisfies* (5.3). *Consider the collection of estimators $\hat{g}^{(n)}_m$ defined by* (3.2) *with $k_n\ge n^2$ and $1\le m\le m_n$. Let* $\mathrm{pen}(m)$ *be defined by* (5.1). *The estimator $\tilde{g}=\hat{g}^{(n)}_{\hat{m}}$ defined by* (3.4) *satisfies*

$$\mathbb{E}(\|g-\tilde{g}\|^2)\le C_a\inf_{m\in\{1,\dots,m_n\}}\left[\|g-g_m\|^2+\mathrm{pen}(m)+\frac{m(M_2+1)}{n}\right]+\frac{\overline{C}(R_{n,m_n}+m_n)}{n},$$

*where $R_{n,m}$ is defined in* (4.1)*, $C_a$ is defined in* (5.4)*, and $\overline{C}$ is a constant depending on $f_\varepsilon$, $\varpi$, and $a$.*

Let us compare the rate of $\tilde{g}$ with the rate obtained in the independent framework. The term $\inf_{m\in\{1,\dots,m_n\}}[\|g-g_m\|^2+\mathrm{pen}(m)+m(M_2+1)/n]$ corresponds to the rate of $\tilde{g}$ when all variables are i.i.d.

The dependent context induces the additional term $n^{-1}(R_{n,m_n}+m_n)$. It is noteworthy that $R_{n,m_n}$ only involves the mild dependence coefficients $\beta_{\mathbf{X},1}(k)$ and $\tau_{\mathbf{X},1}(k)$ given in (2.1)–(2.2).

If these dependence coefficients are summable and the errors are supersmooth, $n^{-1}(R_{n,m_n}+m_n)$ is negligible and $\tilde{g}$ achieves the rate of the independent framework.

If $\varepsilon$ is ordinary smooth, the term $m_n/n$ may not be negligible (see the upper bound for $m_n$ given in (5.3) for $\delta=0$) and Theorem 5.1 does not allow to recover the rate of the independent case. To recover it, we will consider stronger dependence conditions.

### 5.2. Adaptive Density Deconvolution for Supersmooth $f_\varepsilon$

If $(\mathbf{A}_1^\varepsilon)$ holds for some $\delta > 0$, the following corollary is a straightforward consequence of Theorem 5.1.

**Corollary 5.1.** *Assume that $f_\varepsilon$ satisfies $(\mathbf{A}_1^\varepsilon)$ with $\delta > 0$, that $g$ satisfies $(\mathbf{A}_3^X)$, and that $m_n$ satisfies* (5.3)*. Let* $\mathrm{pen}(m)$ *be defined by* (5.1)*. Consider the collection of estimators $\hat{g}_m^{(n)}$ defined by* (3.2) *with $k_n \geq n^2$ and $1 \leq m \leq m_n$.*

(1) *If $\sum_{k>0} \beta_{\mathbf{X},1}(k) < \infty$, the estimator $\tilde{g} = \hat{g}_{\hat{m}}^{(n)}$ defined by* (3.4) *satisfies*

$$\mathbb{E}(\|g - \tilde{g}\|^2) \leq C_a \inf_{m \in \{1,\ldots,m_n\}} \left[ \|g - g_m\|^2 + \mathrm{pen}(m) + \frac{m(M_2+1)}{n} \right] + \frac{\overline{C}(\log n)^{1/\delta}}{n},$$

*where $C_a$ is defined in* (5.4) *and $\overline{C}$ is a constant depending on $f_\varepsilon$, $\varpi$, $a$, and $\sum_{k>0} \beta_{\mathbf{X},1}(k)$.*

(2) *If $\sum_{k>0} \tau_{\mathbf{X},1}(k) < \infty$, the estimator $\tilde{g} = \hat{g}_{\hat{m}}^{(n)}$ defined by* (3.4) *satisfies*

$$\mathbb{E}(\|g - \tilde{g}\|^2) \leq C_a \inf_{m \in \{1,\ldots,m_n\}} \left[ \|g - g_m\|^2 + \mathrm{pen}(m) + \frac{m(M_2+1)}{n} \right] + \frac{\overline{C}(\log n)^{2/\delta}}{n},$$

*where $C_a$ is defined in* (5.4) *and $\overline{C}$ is a constant depending on $f_\varepsilon$, $a$, and $\sum_{k>0} \tau_{\mathbf{X},1}(k)$.*

Corollary 5.1 requires important comments. The terms involving a power of $\log n$ are negligible with respect to $\inf_{m \in \{1,\ldots,m_n\}}[\|g - g_m\|^2 + \mathrm{pen}(m) + m(M_2+1)/n]$. The risk of $\tilde{g}$ is of order $\inf_{m \in \{1,\ldots,m_n\}}[\|g - g_m\|^2 + \mathrm{pen}(m)]$, that is of the best order as in the independent framework. The penalty does not depend on the dependence coefficients and is the same as in the independent framework.

As a conclusion, we see that the adaptive estimator $\tilde{g}$ constructed with the same penalty as in the independent framework still achieves the best rate under mild conditions on the dependence coefficients.

### 5.3. Adaptive Density Deconvolution for Ordinary Smooth $f_\varepsilon$

For $a > 1$ and $\varpi > 1$, define $\mathrm{pen}(m)$ by

$$\mathrm{pen}(m) = 4a\varpi \frac{\Delta(m)}{n}. \tag{5.5}$$

When $\delta = 0$, we have the following result, which cannot be deduced from Theorem 5.1.

**Theorem 5.2.** *Assume that $f_\varepsilon$ satisfies $(\mathbf{A}_1^\varepsilon)$ with $\delta = 0$, that $g$ satisfies $(\mathbf{A}_3^X)$, and that $m_n$ satisfies* (5.3)*. Let* $\mathrm{pen}(m)$ *be defined by* (5.5)*. Consider the collection of estimators $\hat{g}_m^{(n)}$ defined by* (3.2) *with $k_n \geq n^2$ and $1 \leq m \leq m_n$.*

(1) *If $\beta_{\mathbf{X},\infty}(k) = O(k^{-(1+\theta)})$ for some $\theta > (2\gamma+3)/(2\gamma+1)$, then the estimator $\tilde{g} = \hat{g}_{\hat{m}}^{(n)}$ defined by* (3.4) *satisfies*

$$\mathbb{E}(\|g - \tilde{g}\|^2) \leq C_a \inf_{m \in \{1,\ldots,m_n\}} \left[ \|g - g_m\|^2 + \mathrm{pen}(m) + \frac{m(M_2+1)}{n} \right] + \frac{\overline{C}}{n}, \tag{5.6}$$

*where $C_a$ is defined in* (5.4) *and $\overline{C}$ is a constant depending on $f_\varepsilon$, $a$, $\varpi$, and $\sum_{k>0} \beta_{\mathbf{X},\infty}(k)$.*

(2) *If $\gamma > 1/2$ in $(\mathbf{A}_1^\varepsilon)$ and $\tau_{\mathbf{X},\infty}(k) = O(k^{-(1+\theta)})$ for some $\theta > (2\gamma+5)/(2\gamma+1)$, then the estimator $\tilde{g} = \hat{g}_{\hat{m}}^{(n)}$ defined by* (3.4) *satisfies* (5.6)*, where $\overline{C}$ is a constant depending on $f_\varepsilon$, $a$, $\varpi$, and $\sum_{k>0} \tau_{\mathbf{X},\infty}(k)$.*

**Remark 5.1.** Note that the condition for $\beta_{\mathbf{X},\infty}(k)$ is realized for any $\gamma > 0$ provided $\theta > 3$. In the same way, the condition for $\tau_{\mathbf{X},\infty}(k)$ is realized for any $\gamma > 1/2$ provided $\theta > 3$. In both cases, the condition on $\theta$ is weaker as $\gamma$ increases. In other words, the smoother is $f_\varepsilon$, the weaker is the condition on the dependence coefficients.

**Remark 5.2.** This result shows that the adaptive estimator $\tilde{g}$ constructed with the same penalty as in the independent framework still achieves the rate of the independent case but under stronger dependence conditions than those considered in Theorem 5.1. Indeed, it involves the dependence coefficients $\beta_{\mathbf{X},\infty}(k)$ and $\tau_{\mathbf{X},\infty}(k)$ and it requires stronger constraints on their rate of decay.

**Remark 5.3.** For $m$ large enough, the penalty function given in (5.5) is an upper bound of more precise penalty functions which depend on the dependence coefficients. More precisely, let $a > 1$ and $\varpi > 1$ be the two arbitrary constants in (5.5). Under the assumptions of (1) in Theorem 5.2, for $\varpi_1 \in ]1, \varpi[$ and $\varpi_2 > 1$, let pen$(m)$ be defined by

$$\text{pen}(m) = \frac{4a\varpi_1\Delta(m)}{n} + \frac{32a\varpi_2\Big(1 + 4\sum_{k=1}^{n}\beta_{\mathbf{X},1}(k)\Big)m}{n}. \tag{5.7}$$

Under the assumptions of (2) in Theorem 5.2, for $\varpi_1 \in ]1, \varpi[$, let pen$(m)$ be defined by

$$\text{pen}(m) = \frac{4a\varpi_1\Delta(m)}{n} + \frac{32a[1 + 38\log m]\big(m + \pi\sum_{k=1}^{n}\tau_{\mathbf{X},1}(k)m^2\big)}{n}. \tag{5.8}$$

In both cases, the estimator $\tilde{g} = \hat{g}^{(n)}_{\hat{m}}$ defined by (3.4) satisfies (5.6). Remark 5.3 follows from the proof of Theorem 5.2.

### *5.4. Case without Noise*

The case without noise corresponds to the usual density estimation problem, where the $X_i$'s are observed. One can deduce from the proofs of Proposition 4.1, Theorem 5.2, and Remark 5.3 a result for density estimation without errors on the whole real line, that is when $X_i$ is observed. Indeed the proofs can be adapted to this case by considering that $\varepsilon = 0$, $Z = X$, and by taking $f_\varepsilon^* \equiv 1$ in all steps. It follows that $u_t^*(Z_i) = t(X_i)$ and the contrast $\gamma_n$ simply becomes

$$\gamma_{n,X}(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^{n} t(X_i). \tag{5.9}$$

Let $k_n \geq n^2$, $a > 1$, and $\varpi > 1$, and consider as before

$$\hat{g}^{(n)}_m = \arg\min_{t\in S_m^{(n)}} \gamma_{n,X}(t), \qquad \text{pen}(m) = 32a\varpi\left(1 + 4\sum_{k=1}^{n}\beta_{\mathbf{X},1}(k)\right)\frac{m}{n}, \tag{5.10}$$

and

$$\hat{m} = \arg\min_{m\in\{1,\ldots,n\}}[\gamma_{n,X}(g^{(n)}_m) + \text{pen}(m)]. \tag{5.11}$$

**Proposition 5.1.** *Assume that $\varepsilon = 0$. Let $k_n \geq n^2$. Then*

(1)

$$\mathbb{E}\|g - \hat{g}^{(n)}_m\|^2 \leq \|g - g_m\|^2 + \frac{m(M_2+3)}{n} + \frac{2R_{n,m}}{n}.$$

(2) *If $\beta_{\mathbf{X},\infty} = O(k^{-(1+\theta)})$ for some $\theta > 3$, then the estimator $\tilde{g} = \hat{g}_{\hat{m}}$ defined by* (5.10) *and* (5.11) *satisfies*

$$\mathbb{E}(\|g - \tilde{g}\|^2) \leq C_a \inf_{m\in\{1,\ldots,n\}}\left[\|g - g_m\|^2 + \text{pen}(m) + \frac{m(M_2+1)}{n}\right] + \frac{\overline{C}}{n},$$

*where $C_a$ is defined in* (5.4) *and $\overline{C}$ is a constant depending on $a$, $\varpi$, and $\sum_{k>0}\beta_{\mathbf{X},\infty}(k)$.*

The result (1) shows that if $\sum_{k>0}\beta_{\mathbf{X},1}(k)<\infty$, one obtains the same bounds (and the same rates) as in the i.i.d. case. However, if $\sum_{k>0}\tau_{\mathbf{X},1}(k)<\infty$, the result (1) still holds but the term $n^{-1}R_{n,m}$ is of order $m^2/n$ and the rate for $\hat{g}_m^{(n)}$ is worse than in the i.i.d. case.

The result (2) shows that this estimation procedure also works in density estimation without errors. It allows us to estimate a density on the whole real line and to reach the usual rates of convergence by using a penalty of the classical order $m/n$. This remark is valid in the $\beta$-mixing framework and in the case of independent $X_i$'s. We refer to Pensky [21] and Rigollet [23] for recent results in adaptive density estimation on the whole real line in the i.i.d. case.

## 6. PROOFS

### *6.1. Proof of Proposition* 4.1

This proof follows the same lines as in the independent framework (see Comte *et al.* [9]). The main difference lies in the control of the variance term. We keep the same notation as in Section 3.3. According to (3.2), for any given $m$ belonging to $\{1,\dots,m_n\}$, $\hat{g}_m^{(n)}$ satisfies $\gamma_n(\hat{g}_m^{(n)})-\gamma_n(g_m^{(n)})\le 0$. For a random variable $Y$ with density $f_Y$ and any function $\psi$ such that $\psi(Y)$ is integrable, let

$$\nu_{n,Y}(\psi)=\frac{1}{n}\sum_{i=1}^n[\psi(Y_i)-\langle\psi,f_Y\rangle],\quad \text{so that}\quad \nu_{n,Z}(u_t^*)=\frac{1}{n}\sum_{i=1}^n\big[u_t^*(Z_i)-\langle t,g\rangle\big]. \tag{6.1}$$

Since

$$\gamma_n(t)-\gamma_n(s)=\|t-g\|^2-\|s-g\|^2-2\nu_{n,Z}(u_{t-s}^*), \tag{6.2}$$

we infer that

$$\|g-\hat{g}_m^{(n)}\|^2\le\|g-g_m^{(n)}\|^2+2\nu_{n,Z}\big(u^*_{\hat{g}_m^{(n)}-g_m^{(n)}}\big). \tag{6.3}$$

Writing that $\hat{a}_{m,j}-a_{m,j}=\nu_{n,Z}(u^*_{\varphi_{m,j}})$, we obtain

$$\nu_{n,Z}\big(u^*_{\hat{g}_m^{(n)}-g_m^{(n)}}\big)=\sum_{|j|\le k_n}(\hat{a}_{m,j}-a_{m,j})\nu_{n,Z}(u^*_{\varphi_{m,j}})=\sum_{|j|\le k_n}\big[\nu_{n,Z}(u^*_{\varphi_{m,j}})\big]^2.$$

Consequently, $\mathbb{E}\|g-\hat{g}_m^{(n)}\|^2\le\|g-g_m^{(n)}\|^2+2\sum_{j\in\mathbb{Z}}\mathbb{E}[(\nu_{n,Z}(u^*_{\varphi_{m,j}}))^2]$. According to Comte *et al.* [9],

$$\|g-g_m^{(n)}\|^2=\|g-g_m\|^2+\|g_m-g_m^{(n)}\|^2\le\|g-g_m\|^2+\frac{(\pi m)^2(M_2+1)}{k_n}. \tag{6.4}$$

The variance term is studied by using that for $f\in\mathbb{L}_1(\mathbb{R})$,

$$\nu_{n,Z}(f^*)=\int\nu_{n,Z}(e^{ix\cdot})f(x)\,dx. \tag{6.5}$$

Now, we use (6.5) and apply Parseval's formula to obtain

$$\mathbb{E}\bigg(\sum_{j\in\mathbb{Z}}(\nu_{n,Z}(u^*_{\varphi_{m,j}}))^2\bigg)=\frac{1}{4\pi^2}\sum_{j\in\mathbb{Z}}\mathbb{E}\bigg(\int\frac{\varphi^*_{m,j}(-x)}{f^*_\varepsilon(x)}\nu_{n,Z}(e^{ix\cdot})\,dx\bigg)^2=\frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{\mathbb{E}|\nu_{n,Z}(e^{ix\cdot})|^2}{|f^*_\varepsilon(x)|^2}\,dx. \tag{6.6}$$

Since $\nu_{n,Z}$ involves centered and stationary variables,

$$\begin{aligned}\mathbb{E}|\nu_{n,Z}(e^{ix\cdot})|^2=\operatorname{Var}|\nu_{n,Z}(e^{ix\cdot})|&=\frac{1}{n^2}\bigg(\sum_{k=1}^n\operatorname{Var}(e^{ixZ_k})+\sum_{1\le k\ne l\le n}\operatorname{Cov}(e^{ixZ_k},e^{ixZ_l})\bigg)\\&=\frac{1}{n}\operatorname{Var}(e^{ixZ_1})+\frac{1}{n^2}\sum_{1\le k\ne l\le n}\operatorname{Cov}(e^{ixZ_k},e^{ixZ_l}).\end{aligned} \tag{6.7}$$

Since $(X_i)_{i\geq 1}$ and $(\varepsilon_i)_{i\geq 1}$ are independent, we have $\mathbb{E}(e^{ixZ_k}) = f_\varepsilon^*(x)g^*(x)$, so that

$$\mathrm{Cov}(e^{ixZ_k}, e^{ixZ_l}) = \mathbb{E}(e^{ix(Z_l-Z_k)}) - |\mathbb{E}(e^{ixZ_k})|^2 = \mathbb{E}(e^{ix(Z_l-Z_k)}) - |f_\varepsilon^*(x)g^*(x)|^2.$$

Next, by independence of $X$ and $\varepsilon$, we write, for $k \neq l$,

$$\mathbb{E}(e^{ix(Z_l-Z_k)}) = \mathbb{E}(e^{ix(X_l-X_k)})\mathbb{E}(e^{ix(\varepsilon_l-\varepsilon_k)}) = \mathbb{E}(e^{ix(X_l-X_k)})|f_\varepsilon^*(x)|^2,$$

and consequently

$$\mathrm{Cov}(e^{ixZ_k}, e^{ixZ_l}) = \mathrm{Cov}(e^{ixX_k}, e^{ixX_l})|f_\varepsilon^*(x)|^2. \tag{6.8}$$

From (6.7), (6.8) and the stationarity of $(X_i)_{i\geq 1}$, we obtain that

$$\mathbb{E}|\nu_{n,Z}(e^{ix\cdot})|^2 \leq \frac{1}{n} + \frac{2}{n}\sum_{k=2}^{n} \big|\mathrm{Cov}(e^{ixX_1}, e^{ixX_k})\big||f_\varepsilon^*(x)|^2. \tag{6.9}$$

The first part of Proposition 4.1 follows from the stationarity of the $X_i$'s, and from (6.3), (6.4), (6.6), and (6.9).

Let us prove that $R_{n,m} \leq \min(R_{n,m,\beta}, R_{n,m,\tau})$, where $R_{n,m,\beta}$ and $R_{n,m,\tau}$ are defined in Proposition 4.1. Using the inequalities (2.5) and (2.6), we obtain the bounds

$$\big|\mathrm{Cov}(e^{ixX_1}, e^{ixX_k})\big| \leq 2\beta_{\mathbf{X},1}(k-1) \qquad \text{and} \qquad \big|\mathrm{Cov}(e^{ixX_1}, e^{ixX_k})\big| \leq |x|\tau_{\mathbf{X},1}(k-1)$$

(for the last inequality, note that $t \to e^{ixt}$ is $|x|$-Lipschitz). The result easily follows. □

### *6.2. Proof of Theorem* 5.1

By definition, $\tilde g$ satisfies

$$\gamma_n(\tilde g) + \mathrm{pen}(\hat m) \leq \gamma_n(g_m) + \mathrm{pen}(m) \qquad \text{for all} \quad m \in \{1,\ldots,m_n\}.$$

Therefore, by using (6.2) we obtain that

$$\|\tilde g - g\|^2 \leq \|g_m^{(n)} - g\|^2 + 2\nu_{n,Z}(u^*_{\tilde g - g_m^{(n)}}) + \mathrm{pen}(m) - \mathrm{pen}(\hat m).$$

If $t = t_1 + t_2$ with $t_1$ in $S_m^{(n)}$ and $t_2$ in $S_{m'}^{(n)}$, $t^*$ has its support in $[-\pi\max(m,m'), \pi\max(m,m')]$ and $t$ belongs to $S^{(n)}_{\max(m,m')}$. Set $B_{m,m'}(0,1) = \{t \in S^{(n)}_{\max(m,m')} \mid \|t\| = 1\}$. For $\nu_{n,Z}$ defined in (6.1) we get

$$|\nu_{n,Z}(u^*_{\tilde g - g_m^{(n)}})| \leq \|\tilde g - g_m^{(n)}\| \sup_{t\in B_{m,\hat m}(0,1)} |\nu_{n,Z}(u_t^*)|.$$

Using that $2uv \leq a^{-1}u^2 + av^2$ for any $a > 1$, leads to

$$\|\tilde g - g\|^2 \leq \|g_m^{(n)} - g\|^2 + a^{-1}\|\tilde g - g_m^{(n)}\|^2 + a\sup_{t\in B_{m,\hat m}(0,1)} (\nu_{n,Z}(u_t^*))^2 + \mathrm{pen}(m) - \mathrm{pen}(\hat m).$$

Now, according to Lemma 7.1, write that $\nu_{n,Z}(u_t^*) = \nu_n^{(1)}(t) + \nu_{n,X}(t)$, where

$$\nu_n^{(1)}(t) = n^{-1}\sum_{i=1}^{n}\big[u_t^*(Z_i) - \mathbb{E}(u_t^*(Z_i) \mid \sigma(X_i,\, i\geq 1))\big] = n^{-1}\sum_{i=1}^{n}[u_t^*(Z_i) - t(X_i)]. \tag{6.10}$$

Consequently,

$$\|\tilde g - g\|^2 \leq \|g_m^{(n)} - g\|^2 + a^{-1}\|\tilde g - g_m^{(n)}\|^2 + 2a\sup_{t\in B_{m,\hat m}(0,1)}\big(\nu_n^{(1)}(t)\big)^2 + 2a\sup_{t\in B_{m,\hat m}(0,1)}\big(\nu_{n,X}(t)\big)^2$$
$$+ \mathrm{pen}(m) - \mathrm{pen}(\hat m).$$

Hence by writing that $\|\tilde g - g_m^{(n)}\|^2 \leq (1+\kappa_a^{-1})\|\tilde g - g\|^2 + (1+\kappa_a)\|g - g_m^{(n)}\|^2$ with $\kappa_a$ defined in (5.4), we have

$$\|\tilde g - g\|^2 \leq \kappa_a^2\|g_m^{(n)} - g\|^2 + 2a\kappa_a\sup_{t\in B_{m,\hat m}(0,1)}(\nu_n^{(1)}(t))^2 + 2a\kappa_a\sup_{t\in B_{m,\hat m}(0,1)}(\nu_{n,X}(t))^2$$
$$+ \kappa_a(\mathrm{pen}(m) - \mathrm{pen}(\hat m)).$$

Choose some positive function $p(m, m')$ such that

$$2ap(m, m') \le \mathrm{pen}(m) + \mathrm{pen}(m'). \tag{6.11}$$

For this function $p(m, m')$ we have

$$\begin{aligned}\|\tilde{g} - g\|^2 &\le \kappa_a^2 \|g - g_m^{(n)}\|^2 + 2\kappa_a \,\mathrm{pen}(m) + 2a\kappa_a \sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_{n,X}(t))^2 + 2a\kappa_a W_n(m, \hat{m}) \\ &\le \kappa_a^2 \|g - g_m^{(n)}\|^2 + 2\kappa_a \,\mathrm{pen}(m) + 2a\kappa_a \sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_{n,X}(t))^2 \\ &\quad + 2a\kappa_a \sum_{m'=1}^{m_n} W_n(m, m'), \end{aligned} \tag{6.12}$$

where

$$W_n(m, m') := \Big[ \sup_{t \in B_{m,m'}(0,1)} |\nu_n^{(1)}(t)|^2 - p(m, m') \Big]_+. \tag{6.13}$$

The main parts of the proof are the two following points:

**1)** Study of $W_n(m, m')$. We look for $p(m, m')$ such that for a constant $A_1$,

$$\sum_{m'=1}^{m_n} \mathbb{E}(W_n(m, m')) \le \frac{A_1}{n}. \tag{6.14}$$

**2)** Study of $\sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_{n,X}(t))^2$. We prove that

$$\mathbb{E}\Big[ \sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_{n,X}(t))^2 \Big] \le \frac{m_n + R_{n,m_n}}{n}, \tag{6.15}$$

where $R_{n,m}$ is defined in (4.1). Combining (6.12), (6.14), and (6.15), we infer that, for all $1 \le m \le m_n$,

$$\mathbb{E}\|g - \tilde{g}\|^2 \le \kappa_a^2 \|g - g_m^{(n)}\|^2 + 2\kappa_a \,\mathrm{pen}(m) + \frac{2a\kappa_a (m_n + R_{n,m_n})}{n} + \frac{2a\kappa_a A_1}{n}.$$

If we denote $C_a = \max(\kappa_a^2, 2\kappa_a)$, this can also be written

$$\begin{aligned}\mathbb{E}\|g - \tilde{g}\|^2 &\le C_a \inf_{m \in \{1,\dots,m_n\}} \left[ \|g - g_m^{(n)}\|^2 + \|g_m^{(n)} - g_m\| + \mathrm{pen}(m) \right] + \frac{2a\kappa_a (m_n + R_{n,m_n})}{n} + \frac{2a\kappa_a A_1}{n} \\ &\le C_a \inf_{m \in \{1,\dots,m_n\}} \left[ \|g - g_m\|^2 + \frac{(M_2 + 1)m^2}{k_n} + \mathrm{pen}(m) \right] + \frac{2a\kappa_a (m_n + R_{n,m_n})}{n} + \frac{2a\kappa_a A_1}{n}. \end{aligned}$$

*Proof of* (6.14). We start by writing $\mathbb{E}(W_n(m, m')) = \mathbb{E}[\sup_{t \in B_{m,m'}(0,1)} |\nu_n^{(1)}(t)|^2 - p(m, m')]_+$ as

$$\mathbb{E}\Big\{ \mathbb{E}_{\mathbf{X}} \Big[ \sup_{t \in B_{m,m'}(0,1)} |\nu_n^{(1)}(t)|^2 - p(m, m') \Big]_+ \Big\},$$

where $\mathbb{E}_{\mathbf{X}}(Y)$ denotes the conditional expectation $\mathbb{E}(Y \mid \sigma(X_i,\ i \ge 0))$. The point is that, conditionally on $\sigma(X_i,\ i \ge 0)$, the random variables $u_t^*(Z_i) - \mathbb{E}(u_t^*(Z_i) \mid \sigma(X_i,\ i \ge 0))$ are centered, independent but non-identically distributed. We proceed as in the independent case (see Comte *et al.* [9]), by applying the following lemma to the expectation $\mathbb{E}_{\mathbf{X}}[\sup_{t \in B_{m,m'}(0,1)} |\nu_n^{(1)}(t)|^2 - p(m, m')]_+$.

**Lemma 6.1.** *Let $Y_1, \dots, Y_n$ be independent random variables and let $\mathcal{F}$ be a countable class of uniformly bounded measurable functions. Then for $\xi^2 > 0$*

$$\mathbb{E}\Big[ \sup_{f \in \mathcal{F}} |\nu_{n,Y}(f)|^2 - 2(1 + 2\xi^2)H^2 \Big]_+ \le \frac{2}{K_1} \left( \frac{v}{n} e^{-K_1 \xi^2 \frac{nH^2}{v}} + \frac{49 M_1^2}{4K_1 n^2 C^2(\xi^2)} e^{-\frac{2\sqrt{2}K_1 C(\xi^2)\xi}{7\sqrt{2}} \frac{nH}{M_1}} \right)$$

*with* $C(\xi^2) = (\sqrt{1+\xi^2} - 1) \wedge 1$, $K_1 = 1/6$, *and*

$$\sup_{f \in \mathcal{F}} \|f\|_\infty \leq M_1, \qquad \mathbb{E}\Big[\sup_{f \in \mathcal{F}} |\nu_{n,Y}(f)|\Big] \leq H, \qquad \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{k=1}^{n} \mathrm{Var}(f(Y_k)) \leq v.$$

The proof of this inequality can be found in Section 7. It follows from a concentration Inequality in Klein and Rio [16] and arguments that can be found in Birgé and Massart [2]. Usual density arguments show that this result can be applied to the class of functions $\mathcal{F} = B_{m,m'}(0,1)$. Applying Lemma 6.1, one has the bound

$$\mathbb{E}_{\mathbf{X}}\Big[\sup_{t \in B_{m,m'}(0,1)} |\nu_n^{(1)}(t)|^2 - 2(1+2\xi^2)H^2\Big]_+ \leq \frac{2}{K_1}\left(\frac{v}{n} e^{-K_1 \xi^2 \frac{nH^2}{v}} + \frac{49 M_1^2}{4K_1 n^2 C^2(\xi^2)} e^{-\frac{2\sqrt{2}K_1 C(\xi^2)\xi}{7}\frac{nH}{M_1}}\right),$$

where

$$\sup_{t \in B_{m,m'}(0,1)} \|u_t^*(Z_1)\|_\infty \leq M_1, \quad \mathbb{E}_{\mathbf{X}}\Big[\sup_{t \in B_{m,m'}(0,1)} |\nu_n^{(1)}(t)|\Big] \leq H, \quad \sup_{t \in B_{m,m'}} \frac{1}{n}\sum_{k=1}^{n} \mathrm{Var}_{\mathbf{X}}(u_t^*(Z_k)) \leq v.$$

Let $m^* = \max(m, m')$. Applying Lemma 7.3 of Section 7, we propose to take

$$H^2 = H^2(m^*) = \frac{\Delta(m^*)}{n}, \quad M_1 = M_1(m^*) = \sqrt{nH^2}, \quad \text{and} \quad v = v(m^*) = \frac{\sqrt{\Delta_2(m^*, f_Z)}}{2\pi}$$

with, for $f_Z$ denoting the density of $Z_1$,

$$\Delta_2(m, f_Z) = \int_{-\pi m}^{\pi m}\int_{-\pi m}^{\pi m} \frac{|f_Z^*(x-y)|^2}{|f_\varepsilon^*(x) f_\varepsilon^*(y)|^2}\, dx\, dy. \tag{6.16}$$

From the definition (6.13) of $W_n(m, m')$, by taking $p(m, m') = 2(1+2\xi^2)H^2(m^*)$, we get that

$$\mathbb{E}(W_n(m,m')) \leq \mathbb{E}\Big\{\mathbb{E}_{\mathbf{X}}\Big[\sup_{t \in B_{m,m'}(0,1)} |\nu_n^{(1)}(t)|^2 - 2(1+2\xi^2)H^2(m^*)\Big]_+\Big\}. \tag{6.17}$$

According to the condition (6.11), we thus take $\mathrm{pen}(m) = 2ap(m,m) = 4n^{-1}a(1+2\xi^2)\Delta(m)$, where $\xi^2$ is suitably chosen in the control of the sum of the right-hand side of (6.17). Set $m_0$ such that for $m^* \geq m_0$

$$(1/2)\lambda_1(f_\varepsilon, \kappa_0')\Gamma(m^*) \leq \Delta(m^*) \leq 2\lambda_1(f_\varepsilon, \kappa_0)\Gamma(m^*), \tag{6.18}$$

where $\Gamma(m)$ is defined in (4.2) and $\lambda_1(f_\varepsilon, \kappa_0)$ and $\lambda_1(f_\varepsilon, \kappa_0')$ are defined in (4.3). We split the sum over $m'$ in two parts and write

$$\sum_{m'=1}^{m_n} \mathbb{E}(W_n(m,m')) = \sum_{m' | m^* < m_0} \mathbb{E}(W_n(m,m')) + \sum_{m' | m^* \geq m_0} \mathbb{E}(W_n(m,m')). \tag{6.19}$$

By applying Lemma 6.1 and (6.18), we get the global bound $\mathbb{E}_{\mathbf{X}}(W_n(m,m')) \leq K[I(m^*) + II(m^*)]$, where $I(m^*)$ and $II(m^*)$ are defined by

$$I(m^*) = \frac{v(m^*)}{n}\exp\Big\{-K_1\xi^2 \frac{\Delta(m^*)}{v(m^*)}\Big\} \quad \text{and} \quad II(m^*) = \frac{\Delta(m^*)}{n^2}\exp\Big\{-\frac{2\sqrt{2}K_1 \xi C(\xi^2)}{7}\sqrt{n}\Big\}.$$

Since $I$ and $II$ do not depend on the $X_i$'s, we infer that $\mathbb{E}(W_n(m,m')) \leq K[I(m^*) + II(m^*)]$.

When $m^* \leq m_0$, with $m_0$ finite, we see that for all $m \in \{1, \ldots, m_n\}$,

$$\sum_{m' | m^* \leq m_0} \mathbb{E}(W_n(m,m')) \leq \frac{C(m_0)}{n}.$$

We now come to the sum over $m'$ such that $m^* > m_0$.

When $\delta > 1$ we use a rough bound for $\Delta_2(m, f_Z)$ given by $\sqrt{\Delta_2(m, f_Z)} \leq 2\pi n H^2(m)$.

When $0 \leq \delta \leq 1$, write that

$$\Delta_2(m, f_Z) \leq \left\| |f_\varepsilon^*|^{-2} \mathbf{1}_{[-\pi m, \pi m]} \right\|_\infty \Delta(m) \|f_Z^*\|^2 (2\pi).$$

Under $(\mathbf{A}_1^\varepsilon)$ with $0 < \delta < 1$, we use that $\|f_Z^*\|^2 \leq \|f_\varepsilon^*\|^2 < \infty$, that $\sqrt{2\pi}\|f_\varepsilon^*\| = \|f_\varepsilon\|$ and apply (6.18) to infer that for $m^* \geq m_0$,

$$v(m^*) = \frac{\sqrt{\Delta_2(m^*, f_Z)}}{2\pi} \leq \lambda_2(f_\varepsilon, \kappa_0)\Gamma_2(m^*), \tag{6.20}$$

where $\lambda_2(f_\varepsilon, \kappa_0)$ is defined in (5.2) and

$$\Gamma_2(m) = (1 + (\pi m)^2)^\gamma (\pi m)^{\min((1/2-\delta/2),(1-\delta))} \exp(2\mu(\pi m)^\delta) = (\pi m)^{-(1/2-\delta/2)_+}\Gamma(m). \tag{6.21}$$

Under $(\mathbf{A}_1^\varepsilon)$ with $\delta = 0$, we use that $\|f_Z^*\|^2 \leq \|g^*\|^2 < \infty$, that $\sqrt{2\pi}\|g^*\| = \|g\|$ and apply (6.18) to infer that for $m^* \geq m_0$,

$$v(m^*) = \frac{\sqrt{\Delta_2(m^*, f_Z)}}{2\pi} \leq \|g\|\kappa_0^{-1}\sqrt{2\lambda_1(f_\varepsilon, \kappa_0)}\Gamma_2(m^*). \tag{6.22}$$

Combining (6.18), (6.20), and (6.22), we obtain that for $m^* \geq m_0$,

$$I(m^*) \leq \frac{\lambda_2(\kappa_0)\Gamma_2(m^*)}{n} \exp\left\{ -\frac{K_1\xi^2\lambda_1(f_\varepsilon, \kappa_0')}{2\lambda_2(\kappa_0)}(\pi m^*)^{(1/2-\delta/2)_+} \right\}$$

$$\text{and} \qquad II(m^*) \leq \frac{\Delta(m^*)}{n^2} \exp\left\{ -\frac{2\sqrt{2}K_1\xi C(\xi^2)\sqrt{n}}{7} \right\}$$

with

$$\lambda_2(\kappa_0) = \|g\|\kappa_0^{-1}\sqrt{2\lambda_1(f_\varepsilon, \kappa_0)}\mathbf{1}_{\delta=0} + \lambda_2(f_\varepsilon, \kappa_0).$$

• Study of $\sum_{m'|m^* \geq m_0} II(m^*)$. According to the choice of $v(m^*)$, $H^2(m^*)$, and $M_1(m^*)$, we have

$$\begin{aligned}\sum_{m'|m^* \geq m_0} II(m^*) &\leq \sum_{m'=1}^{m_n} \frac{\Delta(m^*)}{n^2} \exp\left\{ \frac{-2\sqrt{2}K_1\xi C(\xi^2)\sqrt{n}}{7} \right\} \\ &\leq \frac{\Delta(m_n)m_n}{n^2} \exp\left\{ \frac{-2\sqrt{2}K_1\xi C(\xi^2)\sqrt{n}}{7} \right\}.\end{aligned}$$

Since under (5.3), $\Delta(m_n)/n$ is bounded, we deduce that $\sum_{m'|m^* \geq m_0} II(m^*) \leq C/n$.

• Study of $\sum_{m'|m^* \geq m_0} I(m^*)$. Denote $\psi = 2\gamma + \min(1/2 - \delta/2, 1 - \delta)$, $\omega = (1/2 - \delta/2)_+$, and $K' = K_1\lambda_1(f_\varepsilon, \kappa_0')/(2\lambda_2(\kappa_0))$. For $a, b \geq 1$, we have that

$$\begin{aligned}\max(a,b)^\psi e^{2\mu\pi^\delta \max(a,b)^\delta} e^{-K'\xi^2 \max(a,b)^\omega} &\leq \left(a^\psi e^{2\mu\pi^\delta a^\delta} + b^\psi e^{2\mu\pi^\delta b^\delta}\right) e^{-(K'\xi^2/2)(a^\omega + b^\omega)} \\ &\leq a^\psi e^{2\mu\pi^\delta a^\delta} e^{-(K'\xi^2/2)a^\omega} e^{-(K'\xi^2/2)b^\omega} + b^\psi e^{2\mu\pi^\delta b^\delta} e^{-(K'\xi^2/2)b^\omega}.\end{aligned} \tag{6.23}$$

Consequently,

$$\begin{aligned}\sum_{m'|m^* \geq m_0} I(m^*) &\leq \sum_{m'=1}^{m_n} \frac{\lambda_2(\kappa_0)\Gamma_2(m^*)}{n} \exp\left\{ -\frac{K_1\xi^2(\lambda_1(f_\varepsilon, \kappa_0')}{2\lambda_2(\kappa_0)}(\pi m^*)^{(1/2-\delta/2)_+} \right\} \\ &\leq \frac{2\lambda_2(\kappa_0)\Gamma_2(m)}{n} \exp\left\{ -\frac{K'\xi^2}{2}(\pi m)^{(1/2-\delta/2)_+} \right\} \sum_{m'=1}^{m_n} \exp\left\{ -\frac{K'\xi^2}{2}(\pi m')^{(1/2-\delta/2)_+} \right\} \\ &\quad + \sum_{m'=1}^{m_n} \frac{2\lambda_2(\kappa_0)\Gamma_2(m')}{n} \exp\left\{ -\frac{K'\xi^2}{2}(\pi m')^{(1/2-\delta/2)_+} \right\}.\end{aligned} \tag{6.24}$$

**Case** $0 \le \delta < 1/3$. In that case, since $\delta < (1/2 - \delta/2)_+$, any choice of $\xi^2 > 0$ ensures that the quantity $\Gamma_2(m)\exp\{-(K'\xi^2/2)m^{1/2-\delta/2}\}$ is bounded, and thus the first term in (6.24) is bounded by $C/n$. Clearly $n^{-1}\sum_{m'=1}^{m_n}\Gamma_2(m')\exp\{-(K'/2)(m')^{1/2-\delta/2}\}$ is bounded by $\tilde{C}/n$, and consequently $\sum_{m'|m^*\ge m_0} I(m^*) \le D/n$. According to (6.11), the result follows by choosing $\mathrm{pen}(m) = 2ap(m,m) = 4a(1+2\xi^2)n^{-1}\Delta(m) = 4a\varpi\Delta(m)/n$.

**Case** $\delta = 1/3$. In that case $\delta = (1/2-\delta/2)_+$ and $\lambda_2(\kappa_0) = \lambda_2(f_\varepsilon, \kappa_0)$. According to (6.24), we choose $\xi^2$ such that $(K'\xi^2/2)(\pi m)^\delta = (2+\epsilon)\mu(\pi m)^\delta$ for some $\epsilon > 0$, for instance,

$$\xi^2 = 49\mu\lambda_2(f_\varepsilon,\kappa_0)/\lambda_1(f_\varepsilon,\kappa_0').$$

Arguing as for the case $0 \le \delta < 1/3$, this choice ensures that $\sum_{m'|m^*\ge m_0} I(m^*) \le D/n$, and consequently (6.14) holds. The result follows by taking $p(m,m') = 2(1+2\xi^2)\Delta(m^*)/n$, and $\mathrm{pen}(m) = 2ap(m,m) = 4a(1+2\xi^2)\Delta(m)/n$.

**Case** $\delta > 1/3$. In that case $\delta > (1/2-\delta/2)_+$ and $\lambda_2(\kappa_0) = \lambda_2(f_\varepsilon,\kappa_0)$. Choose $\xi^2(m)$ such that $(K'\xi^2/2)(\pi m)^{(1/2-\delta/2)_+} = (2+\epsilon)\mu\pi^\delta m^\delta$ for some $\epsilon > 0$. For instance, the choice

$$\xi^2(m) = 49\mu\lambda_2(f_\varepsilon,\kappa_0)(\pi m)^{\delta-(1/2-\delta/2)_+}/\lambda_1(f_\varepsilon,\kappa_0')$$

ensures that $\sum_{m'|m^*\ge m_0} I(m^*) \le D/n$, so that (6.14) holds. The result follows by choosing $p(m,m') = 2(1+2\xi^2(m^*))\Delta(m^*)/n$, associated to $\mathrm{pen}(m) = 2ap(m,m) = 4a(1+2\xi^2(m))\Delta(m)/n$.

*Proof of* (6.15). Since $\max(m,\hat{m}) \le m_n$, according to (6.5),

$$\begin{aligned}\sup_{t\in B_{m,\hat{m}}(0,1)} \mathbb{E}(\nu_{n,X}(t))^2 &\le \sup_{t\in S_{m_n},\|t\|=1} \mathbb{E}\left(\frac{1}{2\pi}\int \nu_{n,X}(e^{ix\cdot})t^*(-x)\,dx\right)^2 \\ &\le \frac{1}{2\pi}\int_{-\pi m_n}^{\pi m_n} \mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^n e^{ixX_k}\right)dx \\ &\le \frac{m_n}{n} + \frac{1}{\pi n}\int_{-\pi m_n}^{\pi m_n}\sum_{k=2}^n \left|\mathrm{Cov}(e^{ixX_1}, e^{ixX_k})\right| dx\end{aligned}$$

and Theorem 5.1 is proved. □

## 6.3. *Proof of Theorem* 5.2 (1)

We use the coupling argument recalled in Section 2.1 to construct approximating variables for the $X_i$'s. For $n = 2p_nq_n + r_n$, $0 \le r_n < q_n$, and $\ell = 0,\ldots,p_n-1$, denote by

$$E_\ell = (X_{2\ell q_n+1},\ldots,X_{(2\ell+1)q_n}), \qquad F_\ell = (X_{(2\ell+1)q_n+1},\ldots,X_{(2\ell+2)q_n}),$$
$$E_\ell^\star = (X^\star_{2\ell q_n+1},\ldots,X^\star_{(2\ell+1)q_n}), \qquad F_\ell^\star = (X^\star_{(2\ell+1)q_n+1},\ldots,X^\star_{(2\ell+2)q_n}).$$

The variables $E_\ell^\star$ and $F_\ell^\star$ are such that

— $E_\ell^\star$, $E_\ell$, $F_\ell^\star$, and $F_\ell$ are identically distributed,

— $\mathbb{P}(E_\ell \ne E_\ell^\star) \le \beta_{\mathbf{X},\infty}(q_n)$ and $\mathbb{P}(F_\ell \ne F_\ell^\star) \le \beta_{\mathbf{X},\infty}(q_n)$,

— The variables $(E_\ell^\star)_{0\le\ell\le p_n-1}$ are i.i.d., and so are the variables $(F_\ell^\star)_{0\le\ell\le p_n-1}$.

Without loss of generality and for the sake of simplicity we assume that $r_n = 0$. For $\kappa_a$ defined in (5.4), we start with

$$\begin{aligned}\|\tilde{g} - g\|^2 &\le \kappa_a^2\|g_m^{(n)} - g\|^2 + 2a\kappa_a \sup_{t\in B_{m,\hat{m}}(0,1)} (\nu_n^{(1)}(t))^2 + 2a\kappa_a \sup_{t\in B_{m,\hat{m}}(0,1)} (\nu_{n,X}(t))^2 \\ &\quad + \kappa_a(\mathrm{pen}(m) - \mathrm{pen}(\hat{m})) \\ &\le \kappa_a^2\|g_m^{(n)} - g\|^2 + 2a\kappa_a \sup_{t\in B_{m,\hat{m}}(0,1)} (\nu_n^{(1)}(t))^2 + 4a\kappa_a \sup_{t\in B_{m,\hat{m}}(0,1)} (\nu_{n,X}^\star(t))^2 \\ &\quad + 4a\kappa_a \sup_{t\in B_{m,\hat{m}}(0,1)} (\nu_{n,X}(t) - \nu_{n,X}^\star(t))^2 + \kappa_a(\mathrm{pen}(m) - \mathrm{pen}(\hat{m})),\end{aligned}$$

where $\nu_{n,X}^\star(t)$ is defined as $\nu_{n,X}(t)$ with $X_i^\star$ instead of $X_i$. Choose $p_1(m,m')$ and $p_2(m,m')$ such that

$$2ap_1(m,m') \le [\mathrm{pen}_1(m) + \mathrm{pen}_1(m')] \qquad \text{and} \qquad 4ap_2(m,m') \le [\mathrm{pen}_2(m) + \mathrm{pen}_2(m')]$$

for $\mathrm{pen}(m) = \mathrm{pen}_1(m) + \mathrm{pen}_2(m)$. It follows that

$$\begin{aligned}\|\tilde{g} - g\|^2 &\le \kappa_a^2\|g - g_m^{(n)}\|^2 + 2\kappa_a\,\mathrm{pen}(m) + 4a\kappa_a W_{n,X}^\star(m,\hat{m}) \\ &\quad + 4a\kappa_a \sup_{t\in B_{m,\hat{m}}(0,1)} (\nu_{n,X}(t) - \nu_{n,X}^\star(t))^2 + 2a\kappa_a W_n(m,\hat{m}) \\ &\le \kappa_a^2\|g - g_m^{(n)}\|^2 + 2\kappa_a\,\mathrm{pen}(m) + 4a\kappa_a \sum_{m'=1}^{m_n} W_{n,X}^\star(m,m') \\ &\quad + 2a\kappa_a \sum_{m'=1}^{m_n} W_n(m,m') + 4a\kappa_a \sup_{t\in B_{m,\hat{m}}(0,1)} (\nu_{n,X}(t) - \nu_{n,X}^\star(t))^2, \end{aligned} \tag{6.25}$$

where

$$W_n(m,m') := \Big[\sup_{t\in B_{m,m'}(0,1)} |\nu_n^{(1)}(t)|^2 - p_1(m,m')\Big]_+, \tag{6.26}$$

$$W_{n,X}^\star(m,m') := \Big[\sup_{t\in B_{m,m'}(0,1)} |\nu_{n,X}^\star(t)|^2 - p_2(m,m')\Big]_+. \tag{6.27}$$

The main parts of the proof consist in the three following points :

**1)** Study of $W_n(m,m')$. We look for $p_1(m,m')$ such that for a constant $A_2$,

$$\sum_{m'=1}^{m_n} \mathbb{E}(W_n(m,m')) \le \frac{A_2}{n}. \tag{6.28}$$

**2)** Study of $W_{n,X}^\star(m,m')$. We look for $p_2(m,m')$ such that for a constant $A_3$,

$$\sum_{m'=1}^{m_n} \mathbb{E}(W_{n,X}^\star(m,m')) \le \frac{A_3}{n}. \tag{6.29}$$

**3)** Study of $\sup_{t\in B_{m,\hat{m}}(0,1)}(\nu_{n,X}(t) - \nu_{n,X}^\star(t))^2$. We prove that

$$\mathbb{E}\Big[\sup_{t\in B_{m,\hat{m}}(0,1)} (\nu_{n,X}^\star(t) - \nu_{n,X}(t))^2\Big] \le 4\beta_{\mathbf{X},\infty}(q_n)m_n \le \frac{A_4}{n}. \tag{6.30}$$

*Proof of* (6.28). The proof of (6.28) for ordinary smooth errors ($\delta = 0$ in $(\mathbf{A}_1^\varepsilon)$) is the same as the proof of (6.14) by taking $p_1(m,m') = p(m,m')$ with $p(m,m')$ as in the proof of (6.14) and $\xi_1^2 > 0$. Hence we choose $\mathrm{pen}_1(m) = 2ap_1(m,m) = 4a(1 + 2\xi_1^2)\Delta(m)/n$.

*Proof of* (6.29). We proceed as in the independent case by applying Lemma 6.1. Let $m^* = \max(m,m')$. The process $W_{n,X}^\star(m,m')$ must be split into two terms $(W_{n,1,X}^\star(m,m') + W_{n,2,X}^\star(m,m'))/2$ involving respectively the odd and even blocks, which are of the same type. More precisely, $W^\star_{n,k,X}(m,m')$ is defined, for $k=1,2$, by

$$W^\star_{n,k,X}(m,m') = \Big[\sup_{t\in B_{m,m'}(0,1)}\Big|\frac{1}{p_nq_n}\sum_{\ell=1}^{p_n}\sum_{i=1}^{q_n}\big(t(X^\star_{(2\ell+k-1)q_n+i}) - \langle t,g\rangle\big)\Big|^2 - p_{2,k}(m,m')\Big]_+,$$

and $p_2(m,m') = 2p_{2,1}(m,m') + 2p_{2,2}(m,m')$.

We only study $W^\star_{n,1,X}(m,m')$ and conclude for $W^\star_{n,2,X}(m,m')$ by using analogous arguments and by choosing $p_{2,1}(m,m') = p_{2,2}(m,m')$. The study of $W^\star_{n,1,X}(m,m')$ consists in applying Lemma 6.1 to $\nu^\star_{n,1,X}(t)$ defined by

$$\nu^\star_{n,1,X}(t) = \frac{1}{p_n}\sum_{\ell=1}^{p_n}\nu^\star_{q_n,\ell,X}(t) \qquad \text{with} \quad \nu^\star_{q_n,\ell,X}(t) = \frac{1}{q_n}\sum_{j=1}^{q_n} t(X^\star_{2\ell q_n+j}) - \langle t,g\rangle,$$

considered as the sum of $p_n$ independent random variables $\nu^\star_{q_n,\ell,X}(t)$. Denote by $M_1^\star(m^*)$, $H^\star(m^*)$, and $v^\star(m^*)$ quantities such that

$$\sup_{t\in B_{m,m'}(0,1)}\|\nu^\star_{q_n,\ell,X}(t)\|_\infty \le M_1^\star(m^*),$$
$$\mathbb{E}\big(\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|\big) \le H^\star(m^*),$$
$$\text{and} \qquad \sup_{t\in B_{m,m'}(0,1)}\mathrm{Var}(\nu^\star_{q_n,\ell,X}(t)) \le v^\star(m^*).$$

Lemma 7.5 leads to the choices $M_1^\star(m^*) = \sqrt{m^*}$,

$$(H^\star(m^*))^2 = \frac{\big(1+4\sum_{k=1}^n \beta_{\mathbf{X},1}(k)\big)m^*}{n}, \qquad \text{and} \qquad v^\star(m^*) = \frac{8\big(\sum_{k=0}^{q_n}(k+1)\beta_{\mathbf{X},1}(k)\|g\|_\infty m^*\big)^{1/2}}{q_n}.$$

Take $\xi_2^2 > 0$. There exists $m_0 = m_0(\xi_1^2,\xi_2^2)$ such that for $m^* \ge m_0$,

$$2(1+2\xi_2^2)(H^\star(m^*))^2 \le \xi_1^2\Delta(m^*)/(4n).$$

Then we take $p_{2,1}(m,m') = \xi_1^2\Delta(m^*)/(4n)$, and get

$$\begin{aligned}
\sum_{m'=1}^{m_n}\mathbb{E}(W^\star_{n,1,X}(m,m')) &= \sum_{m'|m^*\le m_0}\mathbb{E}(W^\star_{n,1,X}(m,m')) + \sum_{m'|m^*>m_0}\mathbb{E}(W^\star_{n,1,X}(m,m'))\\
&\le \sum_{m'|m^*\le m_0}\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2 - 2(1+2\xi_2^2)(H^\star(m^*))^2\Big]_+\\
&\quad + \sum_{m'|m^*\le m_0}|p_{21}(m,m') - 2(1+2\xi_2^2)(H^\star(m^*))^2|\\
&\quad + \sum_{m'|m^*>m_0}\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2 - 2(1+2\xi_2^2)(H^\star(m^*))^2\Big]_+.
\end{aligned}$$

It follows that

$$\begin{aligned}
\sum_{m'=1}^{m_n}\mathbb{E}(W^\star_{n,1,X}(m,m')) &\le 2\sum_{m'=1}^{m_n}\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2 - 2(1+2\xi_2^2)(H^\star(m^*))^2\Big]_+\\
&\quad + \sum_{m'|m^*\le m_0}|p_{2,1}(m,m') - 2(1+2\xi_2^2)(H^\star(m^*))^2|\\
&\le 2\sum_{m'=1}^{m_n}\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2 - 2(1+2\xi_2^2)(H^\star(m^*))^2\Big]_+ + \frac{C(m_0)}{n}.
\end{aligned}$$

We apply Lemma 6.1 to $\mathbb{E}\big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2-2(1+2\xi_2^2)(H^\star(m^*))^2\big]_+$ and obtain

$$\sum_{m'=1}^{m_n}\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2-2(1+2\xi_2^2)(H^\star(m^*))^2\Big]_+\le K\sum_{m'\ge1}[I^\star(m^*)+II^\star(m^*)]$$

with $I^\star(m^*)$ and $II^\star(m^*)$ defined by

$$I^\star(m^*)=\frac{m^*}{n}\exp\big\{-K_2\sqrt{m^*}\big\}\qquad\text{and}\qquad II^\star(m^*)=\frac{q_n^2m^*}{n^2}\exp\Big\{-\frac{\sqrt{2}K_1\xi C(\xi)}{7}\frac{\sqrt{n}}{q_n}\Big\},$$

where $K_2=(K_1/32)(1+4\sum_{k=1}^n\beta_{\mathbf{X},1}(k))/\sqrt{\|g\|_\infty\sum_{k=0}^{q_n}(k+1)\beta_{\mathbf{X},1}(k)}$. If we take $q_n=[n^c]$, for $c$ in $]0,1/2[$, then

$$\sum_{m'\ge1}I\star(m^*)\le\frac{C}{n}\qquad\text{and}\qquad\sum_{m'\ge1}II^\star(m^*)\le\frac{C}{n}.$$

Finally,

$$\sum_{m'=1}^{m_n}\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2-2(1+2\xi_2^2)(H^\star(m^*))^2\Big]_+\le\frac{C_1}{n}$$

and

$$\sum_{m'=1}^{m_n}\mathbb{E}[W^\star_{n,X}(m,m')]\le2\sum_{m'=1}^{m_n}\mathbb{E}[W^\star_{n,1,X}(m,m')+W^\star_{n,2,X}(m,m')]\le\frac{C_2}{n}.$$

The result follows by choosing

$$p_2(m,m')=2p_{2,1}(m,m')+2p_{2,2}(m,m')=\xi_1^2\Delta(m^*)/n,\qquad \mathrm{pen}_2(m)=4ap_2(m,m),$$

and

$$\mathrm{pen}(m)=\mathrm{pen}_1(m)+\mathrm{pen}_2(m)=4a(1+2\xi_1^2)\Delta(m)/n+4a\xi_1^2\Delta(m)/n=4a\varpi\Delta(m)/n,$$

where $\varpi=1+3\xi_1^2$. The penalty function appearing in Remark 5.3, formula (5.7), follows from the choice

$$\begin{aligned}p_{21}(m,m')=p_{22}(m,m')&=\frac{2(1+2\xi_2^2)\Big(1+4\sum_{k=1}^\infty\beta_{\mathbf{X},1}(k)\Big)m^*}{n}\\&=\frac{2(1+2\xi_2^2)\Big(1+4\sum_{k=1}^\infty\beta_{\mathbf{X},1}(k)\Big)m^*}{n},\\ \mathrm{pen}_2(m)&=4ap_2(m,m)=4a(2p_{2,1}(m,m)+2p_{2,2}(m,m))=16ap_{2,1}(m,m)\\&=\frac{32a(1+2\xi_2^2)\big(1+4\sum_{k=1}^\infty\beta_{\mathbf{X},1}(k)\big)m}{n},\end{aligned}$$

and

$$\mathrm{pen}(m)=\frac{4a(1+2\xi_1^2)\Delta(m)}{n}+\frac{32a(1+2\xi_2^2)\big(1+4\sum_{k=1}^\infty\beta_{\mathbf{X},1}(k)\big)m}{n}.$$

*Proof of* (6.30). A rough bound is obtained by writing that

$$\begin{aligned}\sup_{t\in B_{m,\hat m}(0,1)}|\nu^\star_{n,X}(t)-\nu_{n,X}(t)|^2&=\sup_{t\in S^{(n)}_{\max(m,\hat m)},\|t\|\le1}|\nu^\star_{n,X}(t)-\nu_{n,X}(t)|^2\\&\le\sup_{t\in S_{m_n},\|t\|\le1}|\nu^\star_{n,X}(t)-\nu_{n,X}(t)|^2.\end{aligned}$$

According to (6.5),

$$\nu^\star_{n,X}(t) - \nu_{n,X}(t) = \frac{1}{2\pi}\int [\nu^\star_{n,X}(e^{ix\cdot}) - \nu_{n,X}(e^{ix\cdot})] t^*(-x)\, dx.$$

Since $|\nu_{n,X}(e^{ix\cdot}) - \nu^\star_{n,X}(e^{ix\cdot})| \le 2$, we have

$$\sup_{t\in B_{m,\hat{m}}(0,1)} |\nu^\star_{n,X}(t) - \nu_{n,X}(t)|^2 \le \sup_{t\in S_{m_n}, \|t\|\le 1} \frac{1}{4\pi^2}\left| \int [\nu^\star_{n,X}(e^{ix\cdot}) - \nu_{n,X}(e^{ix\cdot})] t^*(-x)\, dx\right|^2$$

$$\le \frac{1}{2\pi}\int_{-\pi m_n}^{\pi m_n} |\nu^\star_{n,X}(e^{ix\cdot}) - \nu_{n,X}(e^{ix\cdot})|^2\, dx \le \frac{1}{\pi}\int_{-\pi m_n}^{\pi m_n} |\nu^\star_{n,X}(e^{ix\cdot}) - \nu_{n,X}(e^{ix\cdot})|\, dx.$$

According to the properties of the coupling,

$$\mathbb{E}\Big[\sup_{t\in B_{m,\hat{m}}(0,1)} |\nu^\star_{n,X}(t) - \nu_{n,X}(t)|^2\Big] \le \frac{1}{\pi}\int_{-\pi m_n}^{\pi m_n} \mathbb{E}|\nu^\star_{n,X}(e^{ix\cdot}) - \nu_{n,X}(e^{ix\cdot})|\, dx \le 4\beta_{\mathbf{X},\infty}(q_n) m_n.$$

For ordinary smooth errors, according to (5.3), $m_n \le n^{1/(2\gamma+1)}$. It follows that if we choose $q_n$ such that $\beta_{\mathbf{X},\infty}(q_n) = O(n^{-(2\gamma+2)/(2\gamma+1)})$, then $\beta_{\mathbf{X},\infty}(q_n) m_n = O(n^{-1})$. For $q_n = [n^c]$ and $\beta_{\mathbf{X},\infty}(n) = O(n^{-1-\theta})$ we obtain the condition $n^{-c(1+\theta)} = O(n^{-(2\gamma+2)/(2\gamma+1)})$. If $\theta > (2\gamma+3)/(2\gamma+1)$, one can find $c < 1/2$ such that this condition is satisfied.

### *6.4. Proof of Theorem* 5.2 (2)

We proceed as in the $\beta$-mixing case, by using the coupling argument given in Section 2.1. The variables $E_\ell$, $E^\star_\ell$, $F_\ell$, $F^\star_\ell$ are constructed as in Section 6.3 and are such that

— $E^\star_\ell$, $E_\ell$, $F^\star_\ell$ and $F_\ell$ are identically distributed,

— $\sum_{i=1}^{q_n} \mathbb{E}(|X_{2\ell q_n+i} - X^\star_{2\ell q_n+i}|) \le q_n \tau_{\mathbf{X},\infty}(q_n)$ and $\sum_{i=1}^{q_n} \mathbb{E}(|X_{(2\ell+1)q_n+i} - X^\star_{(2\ell+1)q_n+i}|) \le q_n \tau_{\mathbf{X},\infty}(q_n)$,

— The variables $(E^\star_\ell)_{0\le\ell\le p_n-1}$ are i.i.d., and so are the variables $(F^\star_\ell)_{0\le\ell\le p_n-1}$.

Without loss of generality and for the sake of simplicity we assume that $r_n = 0$. As for the proof of Theorem 5.2 under **2)**, we start with (6.25). Hence the proof consists of the following steps:

**1)** Study of $W_n(m,m')$. We look for $p_1(m,m')$ such that for a constant $K_2$,

$$\sum_{m'=1}^{m_n} \mathbb{E}(W_n(m,m')) \le \frac{K_2}{n}. \tag{6.31}$$

**2)** Study of $W^\star_{n,X}(m,m')$. We look for $p_2(m,m')$ such that for a constant $K_3$,

$$\sum_{m'=1}^{m_n} \mathbb{E}(W^\star_{n,X}(m,m')) \le \frac{K_3}{n}. \tag{6.32}$$

**3)** Study of $\sup_{t\in B_{m,\hat{m}}(0,1)}(\nu_{n,X}(t) - \nu^\star_{n,X}(t))^2$. We prove that

$$\mathbb{E}\Big[\sup_{t\in B_{m,\hat{m}}(0,1)} \big(\nu^\star_{n,X}(t) - \nu_{n,X}(t)\big)^2\Big] \le \pi\tau_{\mathbf{X},\infty}(q_n) {m_n}^2 \le \frac{K_4}{n}. \tag{6.33}$$

*Proof of* (6.31). The proof of (6.31) for ordinary smooth errors is the same as the proof of (6.14) and leads to the choice $\mathrm{pen}_1(m) = 4a(1+2\xi_1^2)\Delta(m)/n$.

*Proof of* (6.32). As in the proof of (6.29), we apply Lemma 6.1 with

$$(H^\star(m^*))^2 = \frac{\left(m^* + \pi\sum_{k=1}^{n-1}\tau_{\mathbf{X},1}(k)(m^*)^2\right)}{n}, \qquad M_1^\star(m^*) = m^*,$$
$$\text{and} \qquad v^\star(m^*) = \frac{\left(m^* + \pi\sum_{k=1}^{n-1}\tau_{\mathbf{X},1}(k)(m^*)^2\right)}{q_n}.$$

We take $\xi_2^2 = \xi_2^2(m) = (3/K_1+1)\log m$. In the same way as for the proof of Theorem 5.2 (1), we use that there exists $m_0 = m_0(\xi_1^2, \xi_2^2)$ such that for $m^* \geq m_0$,

$$2(1+2\xi_2^2(m^*))(H^\star(m^*))^2 \leq \xi_1^2\Delta(m^*)/(4n),$$

where $\xi_1^2$ is in $\mathrm{pen}_1(m)$. Then we take $p_{21}(m,m') = \xi_1^2\Delta(m^*)/(4n)$ and get

$$\sum_{m'=1}^{m_n}\mathbb{E}(W^\star_{n,1,X}(m,m')) \leq 2\sum_{m'=1}^{m_n}\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2 - 2(1+2\xi_2^2(m^*))(H^\star(m^*))^2\Big]_+ + \frac{C(m_0)}{n}.$$

We now apply Lemma 6.1 to $\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2 - 2(1+2\xi_2^2(m^*))(H^\star(m^*))^2\Big]_+$ and obtain

$$\sum_{m'=1}^{m_n}\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2 - 2(1+2\xi_2^2(m^*))(H^\star(m^*))^2\Big]_+ \leq K\sum_{m'\geq 1}^{m_n}[I^\star(m^*) + II^\star(m^*)]$$

with $I^\star(m^*)$ and $II^\star(m^*)$ now defined by

$$I^\star(m^*) = \frac{m^{*2}}{n}\exp\{-K_1\xi^2(m^*)\}$$
$$\text{and} \qquad II^\star(m^*) = \frac{q_n^2 m^{*2}}{n^2}\exp\left\{-\frac{\sqrt{2}K_1\xi C(\xi)\left(1+\pi\sum_{k=1}^{n}\tau_{\mathbf{X},1}(k)\right)}{7}\frac{\sqrt{n}}{q_n}\right\}.$$

If we take $q_n = [n^c]$, with $c$ in $]0, 1/2[$, then

$$\sum_{m'} I^\star(m^*) \leq \frac{C}{n} \qquad \text{and} \qquad \sum_{m'\geq 1} II^\star(m^*) \leq \frac{C}{n}.$$

Finally, $\sum_{m'=1}^{m_n}\mathbb{E}[W_n^\star(m,m')] \leq 2\sum_{m'=1}^{m_n}\mathbb{E}[W^\star_{n,1,X}(m,m') + W^\star_{n,2,X}(m,m')] \leq C/n$. The result follows by choosing

$$p_2(m,m') = 2p_{2,1}(m,m') + 2p_{2,2}(m,m') = \xi_1^2\Delta(m^*)/n,$$
$$\mathrm{pen}_2(m) = 4ap_2(m,m) = 4a\xi_1^2\Delta(m^*)/n,$$
$$\text{and} \qquad \mathrm{pen}(m) = \mathrm{pen}_1(m) + \mathrm{pen}_2(m) = 4a(1+2\xi_1^2)\Delta(m)/n + 4a\xi_1^2\Delta(m)/n = 4a\varpi\Delta(m)/n,$$

where $\varpi = 1 + 3\xi_1^2$. The penalty in Remark 5.3, formula (5.8), follows from the choices

$$p_{21}(m,m') = p_{22}(m,m') = \frac{2\left[1+2(\frac{3}{K_1}+1)\log(m^*)\right]\left(m^* + 2\pi\sum_{k=1}^{n-1}\tau_{\mathbf{X},1}(k)(m^*)^2\right)}{n},$$
$$\mathrm{pen}_2(m) = 4a(2p_{2,1}(m,m) + 2p_{2,2}(m,m)) = 16ap_{2,1}(m,m)$$
$$= \frac{32a\left[1+2(\frac{3}{K_1}+1)\log m\right]\left(m + 2\pi\sum_{k=1}^{n-1}\tau_{\mathbf{X},1}(k)(m)^2\right)}{n},$$

and

$$\mathrm{pen}(m) = \frac{4a(1+2\xi_1^2)\Delta(m)}{n} + \frac{32a\left[1+2(\frac{3}{K_1}+1)\log m\right]\left(m + 2\pi\sum_{k=1}^{n-1}\tau_{\mathbf{X},1}(k)m^2\right)}{n}.$$

*Proof of* (6.33). The proof of (6.33) is similar to the proof of (6.15). Since $|e^{-ixt} - e^{-ixs}| \le |x||t-s|$, one has

$$\sum_{i=1}^{q_n} \mathbb{E}(|e^{-iX_{2\ell q_n+i}} - e^{-iX^\star_{2\ell q_n+i}}|) \le q_n|x|\tau_{\mathbf{X},\infty}(q_n).$$

It follows that

$$\mathbb{E}\Big[\sup_{t\in B_{m,\hat{m}}(0,1)} |\nu^\star_{n,X}(t) - \nu_{n,X}(t)|^2\Big] \le \frac{1}{\pi}\int_{-\pi m_n}^{\pi m_n} \mathbb{E}|\nu^\star_{n,X}(e^{ix\cdot}) - \nu_{n,X}(e^{ix\cdot})|\,dx \le \pi\tau_{\mathbf{X},\infty}(q_n){m_n}^2.$$

For ordinary smooth errors, according to (5.3), $m_n^2 \le n^{2/(2\gamma+1)}$. It follows that if we choose $q_n$ such that $\tau_{\mathbf{X},\infty}(q_n) = O(n^{-(2\gamma+3)/(2\gamma+1)})$, then $\tau_{\mathbf{X},\infty}(q_n)m_n^2 = O(n^{-1})$.

For $q_n = [n^c]$ and $\tau_{\mathbf{X},\infty}(n) = O(n^{-1-\theta})$ we obtain the condition $n^{-c(1+\theta)} = O(n^{-(2\gamma+3)/(2\gamma+1)})$. If $\theta > (2\gamma+5)/(2\gamma+1)$, one can find $c < 1/2$ such that this condition is satisfied. $\square$

## 7. TECHNICAL LEMMAS

**Lemma 7.1.** *If we denote by $\nu_{n,X}(t)$ the quantity defined by* (6.1)*, then*

$$n^{-1}\sum_{k=1}^{n} \mathbb{E}\big(u_t^*(Z_k) \mid \sigma(X_i,\, i\ge 0)\big) - \langle t, g\rangle = \nu_{n,X}(t).$$

The proof of Lemma 7.1, rather straightforward, is omitted.

**Lemma 7.2.** *Let $\Delta(m)$ be defined by* (3.5)*. We have the equalities*

$$\sum_{j\in\mathbb{Z}} \big|u^*_{\varphi_{m,j}}(z)\big|^2 = (2\pi)^{-1} m \int \left|\frac{\varphi^*(x)}{f_\varepsilon^*(xm)}\right|^2 dx = \Delta(m).$$

**Lemma 7.3.** *Let $\nu_{n,Z}(u_t^*)$, $\Delta(m)$, and $\Delta_2(m, f_Z)$ be defined in* (6.1), (3.5)*, and* (6.16)*. Then*

$$\sup_{t\in B_{m,m'}(0,1)} \|u_t^*\|_\infty \le \sqrt{\Delta(m^*)}, \qquad \mathbb{E}[\sup_{t\in B_{m,m'}(0,1)} |\nu_{n,Z}(u_t^*)|] \le \sqrt{\Delta(m^*)/n},$$

$$\textit{and} \qquad \sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(u_t^*(Z_1)) \le \sqrt{\Delta_2(m^*, f_Z)}/(2\pi).$$

We refer to Comte *et al*. [9] for the proofs of Lemmas 7.2 and 7.3.

**Lemma 7.4.** $\|\sum_{j\in\mathbb{Z}} |\varphi_{m,j}|^2\|_\infty \le m.$

*Proof*. Write

$$\sum_{j\in\mathbb{Z}} |\varphi_{m,j}(x)|^2 = \frac{1}{(2\pi)^2}\sum_{j\in\mathbb{Z}}\left|\int e^{-iux}\varphi^*_{m,j}(u)\,du\right|^2 = \frac{m}{(2\pi)^2}\sum_{j\in\mathbb{Z}}\left|\int e^{-ixum}e^{iju}\varphi^*(u)\,du\right|^2.$$

We conclude by applying Parseval's Formula, which implies that

$$\sum_{j\in\mathbb{Z}} |\varphi_{m,j}(x)|^2 = (2\pi)^{-1} m\int |\varphi^*(u)|^2\,du = m.$$

**Lemma 7.5.** *For $B_{m,m'}(0,1) = \{t \in S_{m\vee m'} \mid \|t\|_2 = 1\}$, we have, for $m^* = m \vee m'$,*

$$\sup_{t\in B_{m,m'}(0,1)} \|t\|_\infty \le \sqrt{m^*}, \qquad \mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)} |\nu^\star_{n,1,X}(t)|\Big] \le \sqrt{\frac{(1+4\sum_{k=1}^n \beta_{\mathbf{X},1}(k))m^*}{n}},$$

$$\textit{and} \qquad \sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(\nu^\star_{q_n,\ell,X}(t)) \le \frac{\big[2\|g\|_\infty(1+32\sum_{k=1}^n (1+k)\beta_{\mathbf{X},1}(k))\big]^{1/2}\sqrt{m^*}}{q_n}.$$

*Proof*. For $t$ in $B_{m,m'}(0,1)$, with $m^* = m\vee m'$, one has $t = \sum_{j\in\mathbb{Z}} b_{m^*,j}\varphi_{m^*,j}$. Applying the Cauchy–Schwarz Inequality and Lemma 7.4 we obtain

$$\sup_{t\in B_{m,m'}(0,1)} \|t\|_\infty \le \Big\|\sum_{j\in\mathbb{Z}} |\varphi_{m^*,j}|^2\Big\|_\infty^{1/2} \le \sqrt{m^*}.$$

Now, using again the Cauchy–Schwarz Inequality,

$$\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)} |\nu^\star_{n,1,X}(t)|\Big] \le \mathbb{E}\Big[\sqrt{\sum_{j\in\mathbb{Z}} (\nu^\star_{n,1X}(\varphi_{m^*,j}))^2}\ \Big] \le \sqrt{\sum_{j\in\mathbb{Z}} \mathrm{Var}(\nu^\star_{n,1,X}(\varphi_{m^*,j}))}.$$

By analogy with (6.6), we write

$$\mathbb{E}\Big(\sum_{j\in\mathbb{Z}} \big(\nu^\star_{n,1,X}(\varphi_{m,j})\big)^2\Big) = \frac{1}{4\pi^2}\sum_{j\in\mathbb{Z}} \mathbb{E}\Big(\int \varphi^*_{m,j}(-x)\nu^\star_{n,1,X}(e^{ix\cdot})\,dx\Big)^2 = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} \mathbb{E}|\nu^\star_{n,1,X}(e^{ix\cdot})|^2\,dx.$$

This yields

$$\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)} |\nu^\star_{n,1,X}(t)|\Big] \le \frac{\big(1+4\sum_{k=1}^n \beta_{\mathbf{X},1}(k)\big)m^*}{n}.$$

Finally, we apply Viennet's [28] variance inequality (see Theorem 2.1, p. 472, and Lemma 4.2, p. 481). Hence there exist some measurable functions $b_k$ such that $0 \le b_k \le 1$ and $\mathbb{E}\big[(\sum_{k=1}^n b_k(X_1))^2\big] \le \sum_{k\ge1}(1+k)\beta_{\mathbf{X},1}(k)$, for which

$$\sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(\nu_{q_n,\ell,X}(t)) \le \sup_{t\in B_{m,m'}(0,1)} \frac{1}{q_n}\int \Big(1+4\sum_{k=1}^{q_n} b_k\Big) t^2(x)g(x)\,dx.$$

Consequently

$$\begin{aligned}\sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(\nu_{q_n,\ell,X}(t)) &\le \sup_{t\in B_{m,m'}(0,1)} \frac{1}{q_n}\|t\|_\infty\|g\|_\infty^{1/2}\Big[\int\Big(1+4\sum_{k=1}^{q_n} b_k\Big)^2 g(x)\,dx\Big]^{1/2}\\ &\le \sqrt{2\|g\|_\infty\Big(1+32\sum_{k=1}^{q_n}(1+k)\beta_{\mathbf{X},1}(k)\Big)}\,\frac{\sqrt{m^*}}{q_n}.\end{aligned}$$

*Proof of Lemma* 6.1. Starting with the concentration inequality given in Klein and Rio [16] and arguing as in Birgé and Massart [2] (see the proof of their Corollary 2, p. 354) we obtain the upper bound

$$\mathbb{P}\Big(\sup_{f\in\mathcal{F}} |\nu_{n,Y}(f)| \ge (1+\eta)H + \lambda\Big) \le 2\exp\Big[-K_1 n\Big(\frac{\lambda^2}{v} \wedge \frac{4\lambda(\eta\wedge1)}{7M_1}\Big)\Big], \tag{7.1}$$

where $K_1 = 1/6$. It remains to integrate this inequality as follows: define the nonnegative random variable $X = [\sup_{f\in\mathcal{F}} |\nu_{n,Y}(f)|^2 - 2(1+2\epsilon)H^2]_+$. We have

$$\mathbb{E}(X) = \int_0^{+\infty} \mathbb{P}\Big(\sup_{f\in\mathcal{F}} |\nu_{n,Y}(f)|^2 \ge 2(1+2\epsilon)H^2 + \tau\Big)\,d\tau$$

$$= \int_0^{+\infty} \mathbb{P}\Big(\sup_{f\in\mathcal{F}} |\nu_{n,Y}(f)| \geq \sqrt{2(1+\epsilon)H^2 + 2(\epsilon H^2 + \tau/2)}\Big)\, d\tau$$

$$\leq \int_0^{+\infty} \mathbb{P}\Big(\sup_{f\in\mathcal{F}} |\nu_{n,Y}(f)| \geq \sqrt{(1+\epsilon)}H + \sqrt{\epsilon H^2 + \tau/2}\Big)\, d\tau.$$

Taking $\eta = (\sqrt{1+\epsilon} - 1)$ and $C(\epsilon) = (1 \wedge \eta)$ we obtain

$$\mathbb{E}(X) \leq \int_0^{+\infty} e^{-\frac{K_1 n}{v}(\epsilon H^2 + \tau/2)}\, d\tau + \int_0^{+\infty} e^{-\frac{4K_1 n C(\epsilon)}{7M_1\sqrt{2}}(\sqrt{\epsilon}H + \sqrt{\tau/2})}\, d\tau$$

$$\leq e^{-K_1\epsilon\frac{nH^2}{v}} \int_0^{+\infty} e^{-\frac{K_1 n}{2v}\tau}\, d\tau + e^{-\frac{2\sqrt{2}K_1 C(\epsilon)\sqrt{\epsilon}}{7}\frac{nH}{M_1}} \int_0^{+\infty} e^{-\frac{2K_1 C(\epsilon) n\sqrt{\tau}}{7M_1}}\, d\tau.$$

Using that for any positive constant $C$, $\int_0^{+\infty} e^{-Cx}\, dx = 1/C$, and $\int_0^{+\infty} e^{-C\sqrt{x}}\, dx = 2/C^2$, we obtain

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}} |\nu_{n,Y}(f)|^2 - 2(1+2\epsilon)H^2\Big]_+ \leq \frac{2}{K_1}\Big(\frac{v}{n} e^{-K_1\epsilon\frac{nH^2}{v}} + \frac{49M_1^2}{4K_1 n^2 C^2(\epsilon)} e^{-\frac{2\sqrt{2}K_1 C(\epsilon)\sqrt{\epsilon}}{7}\frac{nH}{M_1}}\Big).$$

□

## REFERENCES


1. H. Berbee, *Random Walks with Stationary Increments and Renewal Theory*, in *Mathematical Centre Tracts*, Vol. 112 (Mathematisch Centrum, Amsterdam, 1979).
2. L. Birgé and P. Massart, "Minimum Contrast Estimators on Sieves: Exponential Bounds and Rates of Convergence", Bernoulli **4**, 329–375 (1998).
3. R. C. Bradley, *Introduction to Strong Mixing Conditions* (Kendrick Press, Heber City, UT, 2007), Vol. 1.
4. C. Butucea, "Deconvolution of Supersmooth Densities with Smooth Noise", Canadian J. Statist. **32**, 181–192 (2004).
5. C. Butucea and A. B. Tsybakov, "Sharp Optimality for Density Deconvolution with Dominating Bias", Teor. Veroyatn. Primen. **52**, 111–128 (2007).
6. F. Comte, "Kernel Deconvolution of Stochastic Volatility Models", J. Time Ser. Anal. **25**, 563–582 (2004).
7. F. Comte and V. Genon-Catalot, "Penalized Projection Estimator for Volatility Density", Scand. J. Statist. **33**, 875–893 (2006).
8. F. Comte and F. Merlevède, "Adaptive Estimation of the Stationary Density of Discrete and Continuous Time Mixing Processes", ESAIM Probab. Statist. (Electronic) **6**, 211–238 (2002).
9. F. Comte, Y. Rozenholc, and M. L. Taupin, "Penalized Contrast Estimator for Adaptive Density Deconvolution", Canadian J. Statist. **34**, 431–452 (2006).
10. F. Comte, Y. Rozenholc, M. L. and Taupin, "Finite Sample Penalization in Adaptive Density Deconvolution", J. Statist. Comput. Simul. **77**, 977–1000 (2007).
11. J. Dedecker and C. Prieur, "New Dependence Coefficients. Examples and Applications to Statistics", Probab. Theory Rel. Fields **132**, 203–236 (2005).
12. L. Devroye, *Non-Uniform Random Variate Generation* (Springer, New York etc., 1986).
13. P. Doukhan, *Mixing: Properties and Examples*, in *Lecture Notes in Statistics*, Vol. 85 (New York, Springer, 1994).
14. J. D. Esary, F. Proschan, and D. W. Walkup, "Association of Random Variables, with Applications", Ann. Math. Statist. **38**, 1466–1474 (1967).
15. J. Fan, "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems", Ann. Statist. **19**, 1257–1272 (1991).
16. T. Klein and E. Rio, "Concentration around the Mean for Maxima of Empirical Processes", Ann. Probab. **33**, 1060–1077 (2005).
17. C. Lacour, "Rates of Convergence for Nonparametric Deconvolution", C. R. Math. Acad. Sci. Paris Sér. I Math. **342**, 877–882 (2006).
18. E. Masry, "Strong Consistency and Rates for Deconvolution of Multivariate Densities of Stationary Processes", Stochastic Processes Appl. **47**, 53–74 (1993).

19. E. Masry, "Deconvolving Multivariate Kernel Density Estimates from Contaminated Associated Observations", IEEE Trans. Inform. Theory **49**, 2941–2952 (2003).
20. Y. Meyer, *Ondelettes et opérateurs I: Ondelettes*, in *Actualités Mathématiques* (Hermann, Paris, 1990).
21. M. Pensky, "Estimation of a Smooth Density Function Using Meyer-Type Wavelets", Statist. Decis. **17**, 111–123 (1999).
22. M. Pensky and B. Vidakovic, "Adaptive Wavelet Estimator for Nonparametric Density Deconvolution", Ann. Statist. **27**, 2033–2053 (1999).
23. P. Rigollet, "Adaptive Density Estimation Using the Blockwise Stein Method", Bernoulli **12**, 351–370 (2006).
24. M. Rosenblatt, "A Central Limit Theorem and a Strong Mixing Condition", Proc. Nat. Acad. Sci. USA **42**, 43–47 (1956).
25. K. Tribouley and G. Viennet, "$\mathbb{L}_p$ Adaptive Density Estimation in a $\beta$ Mixing Framework", Ann. Inst. H. Poincaré, Probab. Statist. **34**, 179–208 (1998).
26. B. van Es, P. Spreij, and H. van Zanten, "Nonparametric Volatility Density Estimation", Bernoulli **9**, 451–465 (2003).
27. B. van Es, P. Spreij, and H. van Zanten, "Nonparametric Volatility Density Estimation for Discrete Time Models", J. Nonparam. Statist. **17**, 237–251 (2005).
28. G. Viennet, "Inequalities for Absolutely Regular Sequences: Application to Density Estimation", Probab. Theory Rel. Fields **107**, 467–492 (1997).
29. V. A. Volkonskii and Yu. A. Rozanov, "Some Limit Theorems for Random Functions. I", Theory Probab. Appl. **4**, 178–197 (1960).