Available online at www.sciencedirect.com
ScienceDirect
Ann. I. H. Poincaré – PR 43 (2007) 571–597
ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.elsevier.com/locate/anihpb

Adaptive estimation of the transition density of a Markov chain

Claire Lacour

Laboratoire MAP5, Université Paris 5, 45, rue des Saints-Pères, 75270 Paris Cedex 06, France

Received 23 January 2006; received in revised form 17 July 2006; accepted 12 September 2006

Available online 15 December 2006

Abstract

In this paper a new estimator for the transition density π of an homogeneous Markov chain is considered. We introduce an original contrast derived from regression framework and we use a model selection method to estimate π under mild conditions. The resulting estimate is adaptive with an optimal rate of convergence over a large range of anisotropic Besov spaces $B_{2,\infty}^{(\alpha_1,\alpha_2)}$. Some simulations are also presented.
© 2006 Elsevier Masson SAS. All rights reserved.

Résumé

Dans cet article, on considère un nouvel estimateur de la densité de transition π d'une chaîne de Markov homogène. Pour cela, on introduit un contraste original issu de la théorie de la régression et on utilise une méthode de sélection de modèles pour estimer π sous des conditions peu restrictives. L'estimateur obtenu est adaptatif et la vitesse de convergence est optimale pour une importante classe d'espaces de Besov anisotropes $B_{2,\infty}^{(\alpha_1,\alpha_2)}$. On présente également des simulations.
© 2006 Elsevier Masson SAS. All rights reserved.

MSC: 62M05; 62G05; 62H12

Keywords: Adaptive estimation; Transition density; Markov chain; Model selection; Penalized contrast

1. Introduction

We consider (X_i) a homogeneous Markov chain. The purpose of this paper is to estimate the transition density of such a chain. This quantity allows to comprehend the form of dependence between variables and is defined by $\pi(x,y)\,\mathrm{d}y = P(X_{i+1} \in \mathrm{d}y \mid X_i = x)$. It enables also to compute other quantities, like $\mathbb{E}[F(X_{i+1}) \mid X_i = x]$ for example. As many authors, we choose for this a nonparametric approach. Roussas [25] first studies an estimator of the transition density of a Markov chain. He proves the consistency and the asymptotic normality of a kernel estimator for chains satisfying a strong condition known as Doeblin's hypothesis. In Bosq [9], an estimator by projection is studied in a mixing framework and the consistence is also proved. Basu and Sahoo [5] establish a Berry–Essen inequality for a kernel estimator under an assumption introduced by Rosenblatt, weaker than the Doeblin's hypothesis. Athreya and Atuncar [2] improve the result of Roussas since they only need the Harris recurrence of the Markov chain. Other authors are interested in the estimation of the transition density in the non-stationary case: Doukhan and Ghindès [16]

E-mail address: lacour@math-info.univ-paris5.fr (C. Lacour).

0246-0203/$ – see front matter © 2006 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.anihpb.2006.09.003

bound the integrated risk for any initial distribution. In [18], recursive estimators for a non-stationary Markov chain are described. More recently, Clemençon [11] computes the lower bound of the minimax L^p risk and describes a quotient estimator using wavelets. Lacour [20] finds an estimator by projection with model selection that reaches the optimal rate of convergence.

All these authors have estimated π by observing that $\pi = g/f$ where g is the density of (X_i, X_{i+1}) and f the stationary density. If $\hat{g}$ and $\hat{f}$ are estimators of g and f, then an estimator of π can be obtained by writing $\hat{\pi} = \hat{g}/\hat{f}$. But this method has the drawback that the resulting rate of convergence depends on the regularity of f. And the stationary density f can be less regular than the transition density.

The aim here is to find an estimator $\tilde{\pi}$ of π from the observations $X_1, \dots, X_{n+1}$ such that the order of the L^2 risk depends only on the regularity of π and is optimal.

Clémençon [11] introduces an estimation procedure based on an analogy with the regression framework using the thresholding of wavelets coefficients for regular Markov chains. We propose in this paper an other method based on regression, which improves the rate and has the advantage to be really computable. Indeed, this method allows to reach the optimal rate of convergence, without the logarithmic loss obtained by Clémençon [11] and can be applied to β-mixing Markov chains (the notion of "regular" Markov chains in [11] is equivalent to Φ-mixing and is then a stronger assumption). We use model selection via penalization as described in [4] with a new contrast inspired by the classical regression contrast. To deal with the dependence we use auxiliary variables X_i^* as in [27]. But contrary to most cases in such estimation procedure, our penalty does not contain any mixing term and is entirely computable.

In addition, we consider transition densities belonging to anisotropic Besov spaces, i.e. with different regularities with respect to the two directions. Our projection spaces (piecewise polynomials, trigonometric polynomials or wavelets) have different dimensions in the two directions and the procedure selects automatically both well fitted dimensions. A lower bound for the rate of convergence on anisotropic Besov balls is proved, which shows that our estimation procedure is optimal in a minimax sense.

The paper is organized as follows. First, we present the assumptions on the Markov chain and on the collections of models. We also give examples of chains and models. Section 3 is devoted to estimation procedure and the link with classical regression. The bound on the empirical risk is established in Section 4 and the L^2 control is studied in Section 5. We compute both upper bound and lower bound for the mean integrated squared error. In Section 6, some simulation results are given. The proofs are gathered in the last section.

2. Assumptions

2.1. Assumptions on the Markov chain

We consider an irreducible Markov chain (X_n) taking its values in the real line $\mathbb{R}$. We suppose that (X_n) is positive recurrent, i.e. it admits a stationary probability measure μ (for more details, we refer to [21]). We assume that the distribution μ has a density f with respect to the Lebesgue measure and that the transition kernel $P(x, A) = P(X_{i+1} \in A | X_i = x)$ has also a density, denoted by π. Since the number of observations is finite, π is estimated on a compact set $A = A_1 \times A_2$ only. More precisely, the Markov process is supposed to satisfy the following assumptions:

A1. (X_n) is irreducible and positive recurrent.
A2. The distribution of X_0 is equal to μ , thus the chain is (strictly) stationary.
A3. The transition density π is bounded on A, i.e. $\|\pi\|_\infty := \sup_{(x,y)\in A} |\pi(x, y)| < \infty$.
A4. The stationary density f verifies $\|f\|_\infty := \sup_{x\in A_1} |f(x)| < \infty$ and there exists a positive real f_0 such that, for all x in A_1, $f(x) \geqslant f_0$.
A5. The chain is geometrically β-mixing ($\beta_q \leqslant \mathrm{e}^{-\gamma q}$), or arithmetically β-mixing ($\beta_q \leqslant q^{-\gamma}$).

Since (X_i) is a stationary Markov chain, the β-mixing is very explicit, the mixing coefficients can be written:

$$\beta_q = \int \left\| P^q(x, \cdot) - \mu \right\|_{\mathrm{TV}} f(x)\, \mathrm{d}x \tag{1}$$

where $\|\cdot\|_{\mathrm{TV}}$ is the total variation norm (see [15]).

Notice that we distinguish the sets A_1 and A_2 in this work because the two directions x and y in $\pi(x, y)$ do not play the same role, but in practice A_1 and A_2 will be equal and identical or close to the value domain of the chain.

2.2. *Examples of chains*

A lot of processes verify the previous assumptions, as (classical or more general) autoregressive processes, or diffusions. Here we give a nonexhaustive list of such chains.

2.2.1. *Diffusion processes*

We consider the process $(X_{i\Delta})_{1\leqslant i\leqslant n}$ where $\Delta > 0$ is the observation step and $(X_t)_{t\geqslant 0}$ is defined by

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$$

where W is the standard Brownian motion, b is a locally bounded Borel function and σ an uniformly continuous function. We suppose that the drift function b and the diffusion coefficient σ satisfy the following conditions, given in [24] (Proposition 1):

(1) there exists λ_-, λ_+ such that $\forall x \neq 0$, $0 < \lambda_- < \sigma^2(x) < \lambda_+$,
(2) there exists $M_0 \geqslant 0, \alpha > -1$ and $r > 0$ such that

$$\forall |x| \geqslant M_0, \quad xb(x) \leqslant -r|x|^{\alpha+1}.$$

Then, if X_0 follows the stationary distribution, the discretized process $(X_{i\Delta})_{1\leqslant i\leqslant n}$ satisfies Assumptions A1–A5. Note that the mixing is geometrical as soon as $\alpha \geqslant 0$. The continuity of the transition density ensures that Assumption A3 holds. Moreover, we can write

$$f(x) = \frac{1}{M\sigma^2(x)} \exp\left[2\int_0^x \frac{b(u)}{\sigma^2(u)}\,\mathrm{d}u\right]$$

with M such that $\int f = 1$. Consequently Assumption A4 is verified with

$$\|f\|_\infty \leqslant \frac{1}{M\lambda_-} \exp\left[\frac{2}{\lambda_-} \sup_{x\in A_1} \int_0^x b(u)\,\mathrm{d}u\right] \quad \text{and} \quad f_0 \geqslant \frac{1}{M\lambda_+} \exp\left[\frac{2}{\lambda_+} \inf_{x\in A_1} \int_0^x b(u)\,\mathrm{d}u\right].$$

2.2.2. *Nonlinear AR(1) processes*

Let us consider the following process

$$X_n = \varphi(X_{n-1}) + \varepsilon_{X_{n-1},n}$$

where $\varepsilon_{x,n}$ has a positive density l_x with respect to the Lebesgue measure, which does not depend on n. We suppose that the following conditions are verified:

(1) There exist $M > 0$ and $\rho < 1$ such that, for all $|x| > M$, $|\varphi(x)| < \rho|x|$ and $\sup_{|x|\leqslant M} |\varphi(x)| < \infty$.
(2) There exist $l_0 > 0$, $l_1 > 0$ such that $\forall x, y$ $l_0 \leqslant l_x(y) \leqslant l_1$.

Then Mokkadem [22] proves that the chain is Harris recurrent and geometrically ergodic. It implies that Assumptions A1 and A5 are satisfied. Moreover $\pi(x, y) = l_x(y - \varphi(x))$ and $f(y) = \int f(x)\pi(x, y)\,\mathrm{d}x$ and then Assumptions A3, A4 hold with $f_0 \geqslant l_0$ and $\|f\|_\infty \leqslant \|\pi\|_\infty \leqslant l_1$.

2.2.3. *ARX(1,1) models*

The nonlinear process ARX(1,1) is defined by

$$X_n = F(X_{n-1}, Z_n) + \xi_n$$

where F is bounded and (ξ_n), (Z_n) are independent sequences of i.i.d. random variables with $\mathbb{E}|\xi_n| < \infty$. We suppose that the distribution of Z_n has a positive density l with respect to the Lebesgue measure. Assume that there exist $\rho < 1$, a locally bounded and measurable function $h : \mathbb{R} \mapsto \mathbb{R}^+$ such that $\mathbb{E}h(Z_n) < \infty$ and positive constants M, c such that

$$\forall \left|(u, v)\right| > M \quad \left|F(u, v)\right| < \rho|u| + h(v) - c \quad \text{and} \quad \sup_{|x|\leqslant M} \left|F(x)\right| < \infty.$$

Then Doukhan [15] proves (p. 102) that (X_n) is a geometrically β-mixing process. We can write

$$\pi(x, y) = \int l(z) f_\xi\big(y - F(x, z)\big)\,\mathrm{d}z$$

where f_ξ is the density of ξ_n. So, if we assume furthermore that there exist $a_0, a_1 > 0$ such that $a_0 \leqslant f_\xi \leqslant a_1$, then Assumptions A3, A4 are verified with $f_0 \geqslant a_0$ and $\|f\|_\infty \leqslant \|\pi\|_\infty \leqslant a_1$.

2.2.4. ARCH processes

The model is

$$X_{n+1} = F(X_n) + G(X_n)\varepsilon_{n+1}$$

where F and G are continuous functions and for all x, $G(x) \neq 0$. We suppose that the distribution of ε_n has a positive density l with respect to the Lebesgue measure and that there exists $s \geqslant 1$ such that $\mathbb{E}|\varepsilon_n|^s < \infty$. The chain (X_n) satisfies Assumptions A1 and A5 if (see [1]):

$$\limsup_{|x|\to\infty} \frac{|F(x)| + |G(x)|(\mathbb{E}|\varepsilon_n|^s)^{1/s}}{|x|} < 1. \tag{2}$$

In addition, we assume that $\forall x \quad l_0 \leqslant l(x) \leqslant l_1$. Then Assumption A3 is verified with $\|\pi\|_\infty \leqslant l_1 / \inf_{x\in A_1} G(x)$. And Assumption A4 holds with $f_0 \geqslant l_0 \int f G^{-1}$ and $\|f\|_\infty \leqslant l_1 \int f G^{-1}$.

2.3. Assumptions on the models

In order to estimate π, we need to introduce a collection $\{S_m, m \in \mathcal{M}_n\}$ of spaces, that we call models. For each $m = (m_1, m_2)$, S_m is a space of functions with support in A defined from two spaces: F_{m_1} and H_{m_2}. F_{m_1} is a subspace of $(L^2 \cap L^\infty)(\mathbb{R})$ spanned by an orthonormal basis $(\varphi_j^m)_{j\in J_m}$ with $|J_m| = D_{m_1}$ such that, for all j, the support of φ_j^m is included in A_1. In the same way H_{m_2} is a subspace of $(L^2 \cap L^\infty)(\mathbb{R})$ spanned by an orthonormal basis $(\psi_k^m)_{k\in K_m}$ with $|K_m| = D_{m_2}$ such that, for all k, the support of ψ_k^m is included in A_2. Here j and k are not necessarily integers, it can be couples of integers as in the case of a piecewise polynomial space. Then, we define

$$S_m = F_{m_1} \otimes H_{m_2} = \left\{t,\ t(x, y) = \sum_{j\in J_m}\sum_{k\in K_m} a_{j,k}^m \varphi_j^m(x)\psi_k^m(y)\right\}.$$

The assumptions on the models are the following:

M1. For all m_2, $D_{m_2} \leqslant n^{1/3}$ and $\mathcal{D}_n := \max_{m\in\mathcal{M}_n} D_{m_1} \leqslant n^{1/3}$.

M2. There exist positive reals ϕ_1, ϕ_2 such that, for all u in F_{m_1}, $\|u\|_\infty^2 \leqslant \phi_1 D_{m_1} \int u^2$, and for all v in H_{m_2}, $\sup_{x\in A_2} |v(x)|^2 \leqslant \phi_2 D_{m_2} \int v^2$. By letting $\phi_0 = \sqrt{\phi_1\phi_2}$, that leads to

$$\forall t \in S_m \quad \|t\|_\infty \leqslant \phi_0 \sqrt{D_{m_1} D_{m_2}}\, \|t\| \tag{3}$$

where $\|t\|^2 = \int_{\mathbb{R}^2} t^2(x, y)\,\mathrm{d}x\,\mathrm{d}y$.

M3. $D_{m_1} \leqslant D_{m_1'} \Rightarrow F_{m_1} \subset F_{m_1'}$ and $D_{m_2} \leqslant D_{m_2'} \Rightarrow H_{m_2} \subset H_{m_2'}$.

The first assumption guarantees that $\dim S_m = D_{m_1} D_{m_2} \leqslant n^{2/3} \leqslant n$ where n is the number of observations. The condition M2 implies a useful link between the L^2 norm and the infinite norm. The third assumption ensures that, for m and m' in $\mathcal{M}_n$, $S_m + S_{m'}$ is included in a model (since $S_m + S_{m'} \subset S_{m''}$ with $D_{m_1''} = \max(D_{m_1}, D_{m_1'})$ and $D_{m_2''} = \max(D_{m_2}, D_{m_2'})$). We denote by $\mathcal{S}$ the space with maximal dimension among the $(S_m)_{m\in\mathcal{M}_n}$. Thus for all m in $\mathcal{M}_n$, $S_m \subset \mathcal{S}$.

2.4. Examples of models

We show here that Assumptions M1–M3 are not too restrictive. Indeed, they are verified for the spaces F_{m_1} (and H_{m_2}) spanned by the following bases (see [4]):

- Trigonometric basis: for $A=[0,1]$, $\langle\varphi_0,\dots,\varphi_{m_1-1}\rangle$ with $\varphi_0=\mathbb{1}_{[0,1]}$, $\varphi_{2j}(x)=\sqrt{2}\cos(2\pi jx)\ \mathbb{1}_{[0,1]}(x)$, $\varphi_{2j-1}(x)=\sqrt{2}\sin(2\pi jx)\mathbb{1}_{[0,1]}(x)$ for $j\geqslant 1$. For this model $D_{m_1}=m_1$ and $\phi_1=2$ hold.
- Histogram basis: for $A=[0,1]$, $\langle\varphi_1,\dots,\varphi_{2^{m_1}}\rangle$ with $\varphi_j=2^{m_1/2}\mathbb{1}_{[(j-1)/2^{m_1},j/2^{m_1}[}$ for $j=1,\dots,2^{m_1}$. Here $D_{m_1}=2^{m_1}$, $\phi_1=1$.
- Regular piecewise polynomial basis: for $A=[0,1]$, polynomials of degree $0,\dots,r$ (where r is fixed) on each interval $[(l-1)/2^D,l/2^D[$, $l=1,\dots,2^D$. In this case, $m_1=(D,r)$, $J_m=\{j=(l,d),\ 1\leqslant l\leqslant 2^D, 0\leqslant d\leqslant r\}$, $D_{m_1}=(r+1)2^D$. We can put $\phi_1=\sqrt{r+1}$.
- Regular wavelet basis: $\langle\Psi_{lk}, l=-1,\dots,m_1, k\in\Lambda(l)\rangle$ where $\Psi_{-1,k}$ points out the translates of the father wavelet and $\Psi_{lk}(x)=2^{l/2}\Psi(2^lx-k)$ where Ψ is the mother wavelet. We assume that the support of the wavelets is included in A_1 and that Ψ_{-1} belongs to the Sobolev space W_2^r.

3. Estimation procedure

3.1. Definition of the contrast

To estimate the function π, we define the contrast

$$\gamma_n(t)=\frac{1}{n}\sum_{i=1}^{n}\left[\int_{\mathbb{R}} t^2(X_i,y)\,\mathrm{d}y-2t(X_i,X_{i+1})\right]. \tag{4}$$

We choose this contrast because

$$\mathbb{E}\gamma_n(t)=\|t-\pi\|_f^2-\|\pi\|_f^2$$

where

$$\|t\|_f^2=\int_{\mathbb{R}^2} t^2(x,y)f(x)\,\mathrm{d}x\,\mathrm{d}y.$$

Therefore $\gamma_n(t)$ is the empirical counterpart of the $\|\cdot\|_f$-distance between t and f and the minimization of this contrast comes down to minimize $\|t-\pi\|_f$. This contrast is new but is actually connected with the one used in regression problems, as we will see in the next subsection.

We want to estimate π by minimizing this contrast on S_m. Let $t(x,y)=\sum_{j\in J_m}\sum_{k\in K_m}a_{j,k}\varphi_j^m(x)\psi_k^m(y)$ a function in S_m. Then, if A_m denotes the matrix $(a_{j,k})_{j\in J_m,k\in K_m}$,

$$\forall j_0\ \forall k_0\quad \frac{\partial\gamma_n(t)}{\partial a_{j_0,k_0}}=0\Leftrightarrow G_mA_m=Z_m,$$

where

$$\begin{cases} G_m=\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\varphi_j^m(X_i)\varphi_l^m(X_i)\right)_{j,l\in J_m}, \\ Z_m=\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\varphi_j^m(X_i)\psi_k^m(X_{i+1})\right)_{j\in J_m,\ k\in K_m}. \end{cases}$$

Indeed,

$$\frac{\partial\gamma_n(t)}{\partial a_{j_0,k_0}}=0\Leftrightarrow\sum_{j\in J_m}a_{j,k_0}\frac{1}{n}\sum_{i=1}^{n}\varphi_j^m(X_i)\varphi_{j_0}^m(X_i)=\frac{1}{n}\sum_{i=1}^{n}\varphi_{j_0}^m(X_i)\psi_{k_0}^m(X_{i+1}). \tag{5}$$

We cannot define a unique minimizer of the contrast $\gamma_n(t)$, since G_m is not necessarily invertible. For example, G_m is not invertible if there exists j_0 in J_m such that there is no observation in the support of φ_{j_0} (G_m has a null column). This phenomenon happens when localized bases (as histogram bases or piecewise polynomial bases) are used. However, the following proposition will enable us to define an estimator:

Proposition 1.

$$\forall j_0\ \forall k_0 \quad \frac{\partial \gamma_n(t)}{\partial a_{j_0,k_0}} = 0 \Leftrightarrow \forall y \quad \big(t(X_i,y)\big)_{1\leqslant i\leqslant n} = P_W\left(\left(\sum_k \psi_k^m(X_{i+1})\psi_k^m(y)\right)_{1\leqslant i\leqslant n}\right)$$

where P_W denotes the orthogonal projection on $W = \{(t(X_i,y))_{1\leqslant i\leqslant n},\ t\in S_m\}$.

Thus the minimization of $\gamma_n(t)$ leads to a unique vector $(\hat{\pi}_m(X_i,y))_{1\leqslant i\leqslant n}$ defined as the projection of $(\sum_k \psi_k(X_{i+1})\psi_k(y))_{1\leqslant i\leqslant n}$ on W. The associated function $\hat{\pi}_m(\cdot,\cdot)$ is not defined uniquely but we can choose a function $\hat{\pi}_m$ in S_m whose values at (X_i,y) are fixed according to Proposition 1. For the sake of simplicity, we denote

$$\hat{\pi}_m = \arg\min_{t\in S_m} \gamma_n(t).$$

This underlying function is more a theoretical tool and the estimator is actually the vector $(\hat{\pi}_m(X_i,y))_{1\leqslant i\leqslant n}$. This remark leads to consider the risk defined with the empirical norm

$$\|t\|_n = \left(\frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}} t^2(X_i,y)\,\mathrm{d}y\right)^{1/2}. \tag{6}$$

This norm is the natural distance in this problem and we can notice that if t is deterministic with support included in $A_1\times\mathbb{R}$

$$f_0\|t\|^2 \leqslant \mathbb{E}\|t\|_n^2 = \|t\|_f^2 \leqslant \|f\|_\infty \|t\|^2$$

and then the mean of this empirical norm is equivalent to the L^2 norm $\|\cdot\|$.

3.2. Link with classical regression

Let us fix k in K_m and let

$$Y_{i,k} = \psi_k^m(X_{i+1}) \quad \text{for } i\in\{1,\dots,n\},$$
$$t_k(x) = \int t(x,y)\psi_k^m(y)\,\mathrm{d}y \quad \text{for all } t \text{ in } L^2(\mathbb{R}^2).$$

Actually, $Y_{i,k}$ and t_k depend on m but we do not mention this for the sake of simplicity. For the same reason, we denote in this subsection ψ_k^m by ψ_k and φ_j^m by φ_j. Then, if t belongs to S_m,

$$\begin{aligned} t(x,y) &= \sum_{j\in J_m}\sum_{k\in K_m}\left(\int t(x',y')\varphi_j(x')\psi_k(y')\,\mathrm{d}x'\mathrm{d}y'\right)\varphi_j(x)\psi_k(y) \\ &= \sum_{k\in K_m}\sum_{j\in J_m}\left(\int t_k(x')\varphi_j(x')\,\mathrm{d}x'\right)\varphi_j(x)\psi_k(y) = \sum_{k\in K_m} t_k(x)\psi_k(y) \end{aligned}$$

and then, by replacing this expression of t in $\gamma_n(t)$, we obtain

$$\begin{aligned} \gamma_n(t) &= \frac{1}{n}\sum_{i=1}^n\left[\int \sum_{k,k'} t_k(X_i)t_{k'}(X_i)\psi_k(y)\psi_{k'}(y)\,\mathrm{d}y - 2\sum_k t_k(X_i)\psi_k(X_{i+1})\right] \\ &= \frac{1}{n}\sum_{i=1}^n\sum_{k\in K_m}\left[t_k^2(X_i) - 2t_k(X_i)Y_{i,k}\right] = \frac{1}{n}\sum_{i=1}^n\sum_{k\in K_m}\left[t_k(X_i)-Y_{i,k}\right]^2 - Y_{i,k}^2. \end{aligned}$$

Consequently

$$\min_{t\in S_m}\gamma_n(t) = \sum_{k\in K_m}\min_{t_k\in F_{m_1}}\frac{1}{n}\sum_{i=1}^n\left[t_k(X_i)-Y_{i,k}\right]^2 - Y_{i,k}^2.$$

We recognize, for all k, the least squares contrast, which is used in regression problems. Here the regression function is $\pi_k = \int \pi(\cdot, y)\psi_k(y)\,\mathrm{d}y$ which verifies

$$Y_{i,k} = \pi_k(X_i) + \varepsilon_{i,k} \tag{7}$$

where

$$\varepsilon_{i,k} = \psi_k(X_{i+1}) - \mathbb{E}\big[\psi_k(X_{i+1})|X_i\big]. \tag{8}$$

The estimator $\hat{\pi}_m$ can be written as $\sum_{k\in K_m} \hat{\pi}_k(x)\psi_k(y)$ where $\hat{\pi}_k$ is the classical least squares estimator for the regression model (7) (as previously, only the vector $(\hat{\pi}_k(X_i))_{1\leqslant i\leqslant n}$ is uniquely defined).

This regression model is used in Clémençon [11] to estimate the transition density. In the same manner, we could use here the contrast $\gamma_n^{(k)}(t) = \frac{1}{n}\sum_{i=1}^n [\psi_k(X_{i+1}) - t(X_i)]^2$ to take advantage of analogy with regression. This method allows to have a good estimation of the projection of π on some S_m by estimating first each π_k, but does not provide an adaptive method. Model selection requires a more global contrast, as described in (4).

3.3. Definition of the estimator

We have then an estimator of π for all S_m. Let now

$$\hat{m} = \arg\min_{m\in\mathcal{M}_n} \big\{\gamma_n(\hat{\pi}_m) + \mathrm{pen}(m)\big\}$$

where pen is a penalty function to be specified later. Then we can define $\tilde{\pi} = \hat{\pi}_{\hat{m}}$ and compute the empirical mean integrated squared error $\mathbb{E}\|\pi - \tilde{\pi}\|_n^2$ where $\|\cdot\|_n$ is the empirical norm defined in (6).

4. Calculation of the risk

For a function h and a subspace S, let

$$d(h,S) = \inf_{g\in S}\|h-g\| = \inf_{g\in S}\left(\iint \big|h(x,y) - g(x,y)\big|^2\,\mathrm{d}x\,\mathrm{d}y\right)^{1/2}.$$

With an inequality of Talagrand [26], we can prove the following result.

Theorem 2. *We consider a Markov chain satisfying Assumptions* A1–A5 (*with* $\gamma > 14$ *in the case of an arithmetical mixing*). *We consider* $\tilde{\pi}$ *the estimator of the transition density* π *described in Section* 3 *with models verifying Assumptions* M1–M3 *and the following penalty*:

$$\mathrm{pen}(m) = K_0\|\pi\|_\infty \frac{D_{m_1}D_{m_2}}{n} \tag{9}$$

where K_0 *is a numerical constant. Then*

$$\mathbb{E}\|\pi\mathbb{1}_A - \tilde{\pi}\|_n^2 \leqslant C \inf_{m\in\mathcal{M}_n}\big\{d^2(\pi\mathbb{1}_A, S_m) + \mathrm{pen}(m)\big\} + \frac{C'}{n}$$

where $C = \max(5\|f\|_\infty, 6)$ *and* C' *is a constant depending on* $\phi_1, \phi_2, \|\pi\|_\infty, f_0, \|f\|_\infty, \gamma$.

The constant K_0 in the penalty is purely numerical (we can choose $K_0 = 45$). We observe that the term $\|\pi\|_\infty$ appears in the penalty although it is unknown. Nevertheless it can be replaced by any bound of $\|\pi\|_\infty$. Moreover, it is possible to use $\|\hat{\pi}\|_\infty$ where $\hat{\pi}$ is some estimator of π. This method of random penalty (specifically with infinite norm) is successfully used in [7] and [12] for example, and can be applied here even if it means considering π regular enough. This is proved in Appendix A.

It is relevant to notice that the penalty term does not contain any mixing term and is then entirely computable. It is in fact related to martingale properties of the underlying empirical processes. The constant K_0 is a fixed universal numerical constant; for practical purposes, it is adjusted by simulations.

We are now interested in the rate of convergence of the risk. We consider that π restricted to A belongs to the anisotropic Besov space on A with regularity $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$. Note that if π belongs to $B^{\boldsymbol{\alpha}}_{2,\infty}(\mathbb{R}^2)$, then π restricted to

A belongs to $B^{\boldsymbol{\alpha}}_{2,\infty}(A)$. Let us recall the definition of $B^{\boldsymbol{\alpha}}_{2,\infty}(A)$. Let e_1 and e_2 be the canonical basis vectors in $\mathbb{R}^2$ and for $i = 1, 2$, $A^r_{h,i} = \{x \in \mathbb{R}^2; x, x + he_i, \ldots, x + rhe_i \in A\}$. Next, for x in $A^r_{h,i}$, let

$$\Delta^r_{h,i} g(x) = \sum_{k=0}^{r} (-1)^{r-k} \binom{r}{k} g(x + khe_i)$$

the rth difference operator with step h. For $t > 0$, the directional moduli of smoothness are given by

$$\omega_{r_i,i}(g, t) = \sup_{|h| \leqslant t} \left(\int_{A^{r_i}_{h,i}} \left| \Delta^{r_i}_{h,i} g(x) \right|^2 \mathrm{d}x \right)^{1/2}.$$

We say that g is in the Besov space $B^{\boldsymbol{\alpha}}_{2,\infty}(A)$ if

$$\sup_{t>0} \sum_{i=1}^{2} t^{-\alpha_i} \omega_{r_i,i}(g, t) < \infty$$

for r_i integers larger than α_i. The transition density π can thus have different smoothness properties with respect to different directions. The procedure described here allows an adaptation of the approximation space to each directional regularity. More precisely, if $\alpha_2 > \alpha_1$ for example, the estimator chooses a space of dimension $D_{m_2} = D_{m_1}^{\alpha_1/\alpha_2} < D_{m_1}$ for the second direction, where π is more regular. We can thus write the following corollary.

Corollary 3. *We suppose that π restricted to A belongs to the anisotropic Besov space $B^{\boldsymbol{\alpha}}_{2,\infty}(A)$ with regularity $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ such that $\alpha_1 - 2\alpha_2 + 2\alpha_1\alpha_2 > 0$ and $\alpha_2 - 2\alpha_1 + 2\alpha_1\alpha_2 > 0$. We consider the spaces described in Section* 2.4 (*with the regularity r of the polynomials and the wavelets larger than $\alpha_i - 1$*). *Then, under the assumptions of Theorem* 2,

$$\mathbb{E}\|\pi \mathbb{1}_A - \tilde{\pi}\|_n^2 = \mathrm{O}\big(n^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+2}}\big).$$

where $\bar{\alpha}$ is the harmonic mean of α_1 and α_2.

The harmonic mean of α_1 and α_2 is the real $\bar{\alpha}$ such that $2/\bar{\alpha} = 1/\alpha_1 + 1/\alpha_2$. Note that the condition $\alpha_1 - 2\alpha_2 + 2\alpha_1\alpha_2 > 0$ is ensured as soon as $\alpha_1 \geqslant 1$ and the condition $\alpha_2 - 2\alpha_1 + 2\alpha_1\alpha_2 > 0$ as soon as $\alpha_2 \geqslant 1$.

Thus we obtain the rate of convergence $n^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+2}}$, which is optimal in the minimax sense (see Section 5.3 for the lower bound).

5. L^2 control

5.1. Estimation procedure

Although the empirical norm is the more natural in this problem, we are interested in a L^2 control of the risk. For this, the estimation procedure must be modified. We truncate the previous estimator in the following way:

$$\tilde{\pi}^* = \begin{cases} \tilde{\pi} & \text{if } \|\tilde{\pi}\| \leqslant k_n, \\ 0 & \text{else} \end{cases} \tag{10}$$

with $k_n = n^{2/3}$.

5.2. Calculation of the L^2 risk

We obtain in this framework a result similar to Theorem 2.

Theorem 4. *We consider a Markov chain satisfying Assumptions* A1–A5 (*with $\gamma > 20$ in the case of an arithmetical mixing*). *We consider $\tilde{\pi}^*$ the estimator of the transition density π described in Section* 5.1. *Then*

$$\mathbb{E}\|\tilde{\pi}^* - \pi \mathbb{1}_A\|^2 \leqslant C \inf_{m \in \mathcal{M}_n} \big\{ d^2(\pi \mathbb{1}_A, S_m) + \mathrm{pen}(m) \big\} + \frac{C'}{n}$$

where $C = \max(36 f_0^{-1}\|f\|_\infty + 2, 36 f_0^{-1})$ *and* C' *is a constant depending on* $\phi_1, \phi_2, \|\pi\|_\infty, \|\pi\|, f_0, \|f\|_\infty, \gamma$.

If π is regular, we can state the following corollary:

Corollary 5. *We suppose that the restriction of* π *to* A *belongs to the anisotropic Besov space* $B^{\boldsymbol{\alpha}}_{2,\infty}(A)$ *with regularity* $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ *such that* $\alpha_1 - 2\alpha_2 + 2\alpha_1\alpha_2 > 0$ *and* $\alpha_2 - 2\alpha_1 + 2\alpha_1\alpha_2 > 0$. *We consider the spaces described in Section* 2.4 (*with the regularity* r *of the polynomials and the wavelets larger than* $\alpha_i - 1$). *Then, under the assumptions of Theorem* 4,

$$\mathbb{E}\|\pi \mathbb{1}_A - \tilde{\pi}^*\|^2 = \mathrm{O}\big(n^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+2}}\big).$$

where $\bar{\alpha}$ *is the harmonic mean of* α_1 *and* α_2.

The same rate of convergence is then achieved with the L^2 norm instead of the empirical norm. And the procedure allows to adapt automatically the two dimensions of the projection spaces to the regularities α_1 and α_2 of the transition density π. If $\alpha_1 = 1$ we recognize the rate $n^{-\frac{\alpha_2}{3\alpha_2+1}}$ established by Birgé [6] with metrical arguments. The optimality is proved in the following subsection.

If $\alpha_1 = \alpha_2 = \alpha$ ("classical" Besov space), then $\bar{\alpha} = \alpha$ and our result is thus an improvement of the one of Clémençon [11], whose procedure achieves only the rate $(\log(n)/n)^{\frac{2\alpha}{2\alpha+2}}$ and allows to use only wavelets. We can observe that in this case, the condition $\alpha_1 - 2\alpha_2 + 2\alpha_1\alpha_2 > 0$ is equivalent to $\alpha > 1/2$ and so is verified if the function π is regular enough.

Actually, in the case $\alpha_1 = \alpha_2$, an estimation with isotropic spaces ($D_{m_1} = D_{m_2}$) is preferable. Indeed, in this framework, the models are nested and so we can consider spaces with larger dimension ($D_m^2 \leqslant n$ instead of $D_m^2 \leqslant n^{2/3}$). Then Corollary 3 is valid whatever $\alpha > 0$. Moreover, for the arithmetic mixing, assumption $\gamma > 6$ is sufficient.

5.3. Lower bound

We denote by $\|\cdot\|_A$ the norm in $L^2(A)$, i.e. $\|g\|_A = (\int_A |g|^2)^{1/2}$. We set

$$\mathcal{B} = \{\pi \text{ transition density on } \mathbb{R} \text{ of a positive recurrent Markov chain such that } \|\pi\|_{B^{\boldsymbol{\alpha}}_{2,\infty}(A)} \leqslant L\}$$

and $\mathbb{E}_\pi$ the expectation corresponding to the distribution of $X_1, \ldots, X_n$ if the true transition density of the Markov chain is π and the initial distribution is the stationary distribution.

Theorem 6. *There exists a positive constant* C *such that, if* n *is large enough,*

$$\inf_{\hat{\pi}_n} \sup_{\pi \in \mathcal{B}} \mathbb{E}_\pi \|\hat{\pi}_n - \pi\|_A^2 \geqslant C n^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+2}}$$

where the infimum is taken over all estimators $\hat{\pi}_n$ *of* π *based on the observations* $X_1, \ldots, X_{n+1}$.

So the lower bound in [11] is generalized for the case $\alpha_1 \neq \alpha_2$. It shows that our procedure reaches the optimal minimax rate, whatever the regularity of π, without needing to know $\boldsymbol{\alpha}$.

6. Simulations

To evaluate the performance of our method, we simulate a Markov chain with a known transition density and then we estimate this density and compare the two functions for different values of n. The estimation procedure is easy, we can decompose it in some steps:

- find the coefficients matrix A_m for each $m = (m_1, m_2)$,
- compute $\gamma_n(\hat{\pi}_m) = \mathrm{Tr}({}^t A_m G_m A_m - 2\,{}^t Z_m A_m)$,
- find $\hat{m}$ such that $\gamma_n(\hat{\pi}_m) + \mathrm{pen}(m)$ is minimum,
- compute $\hat{\pi}_{\hat{m}}$.

For the first step, we use two different kinds of bases: the histogram bases and the trigonometric bases, as described in Section 2.4. We renormalize these bases so that they are defined on the estimation domain A instead of $[0,1]^2$. For the third step, we choose $\text{pen}(m) = 0.5 D_{m_1} D_{m_2}/n$.

We consider three Markov chains:

• An autoregressive process defined by $X_{n+1} = aX_n + b + \varepsilon_{n+1}$, where the ε_n are i.i.d. centered Gaussian random variables with variance σ^2. The stationary distribution of this process is a Gaussian with mean $b/(1-a)$ and with variance $\sigma^2/(1-a^2)$. The transition density is $\pi(x,y) = \varphi(y - ax - b)$ where $\varphi(z) = 1/(\sigma\sqrt{2\pi})\exp(-z^2/2\sigma^2)$ is the density of a standard Gaussian. Here we choose $a = 0.5$, $b = 3$, $\sigma = 1$ and we note this process AR(1). It is estimated on $[4,8]^2$.

• A discrete radial Ornstein–Uhlenbeck process, i.e. the Euclidean norm of a vector (ξ^1, ξ^2, ξ^3) whose components are i.i.d. processes satisfying, for $j = 1,2,3$, $\xi^j_{n+1} = a\xi^j_n + \beta\varepsilon^j_n$ where ε^j_n are i.i.d. standard Gaussian. This process is studied in detail in [10]. Its transition density is

$$\pi(x,y) = \mathbb{1}_{y>0} \exp\left(-\frac{y^2 + a^2x^2}{2\beta^2}\right) I_{1/2}\left(\frac{axy}{\beta^2}\right)\frac{y}{\beta^2}\sqrt{\frac{y}{ax}}$$

where $I_{1/2}$ is the Bessel function with index $1/2$. The stationary density of this chain is

$$f(x) = \mathbb{1}_{x>0} \exp\{-x^2/2\rho^2\} 2x^2/\big(\rho^3\sqrt{2\pi}\big)$$

with $\rho^2 = \beta^2/(1-a^2)$. We choose $a = 0.5$, $\beta = 3$ and we denote this process by $\sqrt{\text{CIR}}$ since it is the square root of a Cox–Ingersoll–Ross process. The estimation domain is $[2,10]^2$.

• An ARCH process defined by $X_{n+1} = \sin(X_n) + (\cos(X_n) + 3)\varepsilon_{n+1}$ where the ε_{n+1} are i.i.d. standard Gaussian. We verify that the condition (2) is satisfied. Here the transition density is

$$\pi(x,y) = \varphi\left(\frac{y - \sin(x)}{\cos(x) + 3}\right)\frac{1}{\cos(x) + 3}$$

and we estimate this chain on $[-6,6]^2$.

We can illustrate the results by some figures. Fig. 1 shows the surface $z = \pi(x,y)$ and the estimated surface $z = \tilde{\pi}(x,y)$. We use a histogram basis and we see that the procedure chooses different dimensions on the abscissa and on the ordinate since the estimator is constant on rectangles instead of squares. Fig. 2 presents sections of this kind of surfaces for the AR(1) process estimated with trigonometric bases. We can see the curves $z = \pi(4.6, y)$ versus $z = \tilde{\pi}(4.6, y)$ and the curves $z = \pi(x, 5)$ versus $z = \tilde{\pi}(x, 5)$. The second section shows that it may exist some edge effects due to the mixed control of the two directions.

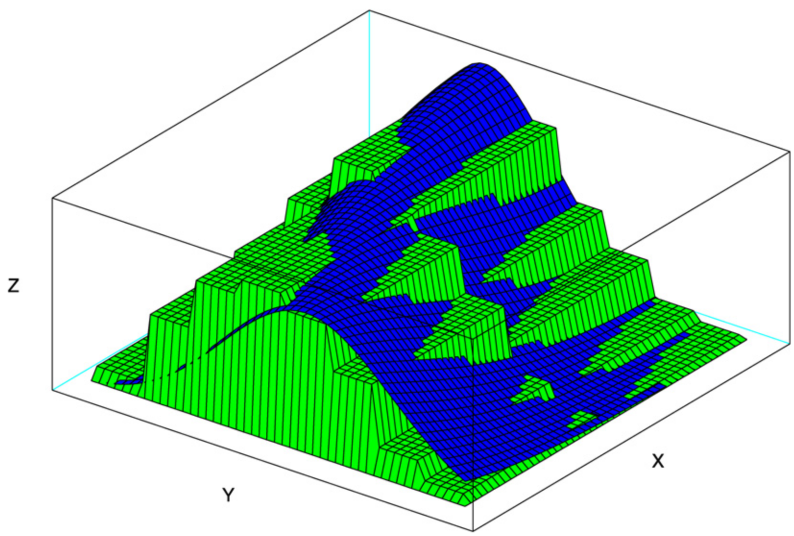
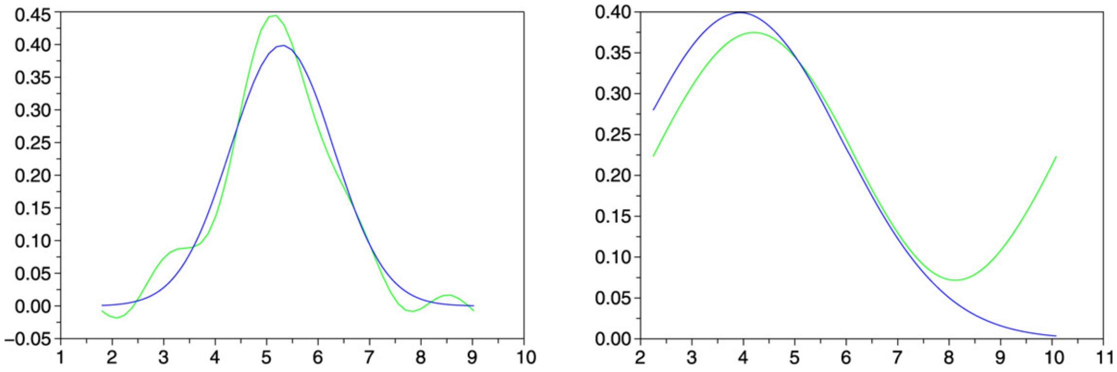


Fig. 1. Estimator (light surface) and true function (dark surface) for a $\sqrt{\text{CIR}}$ process estimated with a histogram basis, $n = 1000$.



$x = 4.6$ $y = 5$

Fig. 2. Sections for AR(1) process estimated with a trigonometric basis, $n = 1000$, dark line: true function, light line: estimator.

Table 1
Empirical risk $\mathbb{E}\|\pi - \tilde{\pi}\|_n^2$ for simulated data with $\mathrm{pen}(m) = 0.5 D_{m_1} D_{m_2}/n$, averaged over $N = 200$ samples

Law	n					
	50	100	250	500	1000	basis
AR(1)	0.067	0.055	0.043	0.038	0.033	H
	0.096	0.081	0.063	0.054	0.045	T
$\sqrt{\mathrm{CIR}}$	0.026	0.023	0.019	0.016	0.014	H
	0.019	0.015	0.009	0.007	0.006	T
ARCH	0.031	0.027	0.016	0.015	0.014	H
	0.020	0.012	0.008	0.007	0.007	T

H: histogram basis, T: trigonometric basis.

Table 2
L^2 risk $\mathbb{E}\|\pi - \tilde{\pi}^*\|^2$ for simulated data with $\mathrm{pen}(m) = 0.5 D_{m_1} D_{m_2}/n$, averaged over $N = 200$ samples

Law	n					
	50	100	250	500	1000	basis
AR(1)	0.242	0.189	0.132	0.109	0.085	H
	0.438	0.357	0.253	0.213	0.180	T
$\sqrt{\mathrm{CIR}}$	0.152	0.130	0.094	0.066	0.054	H
	0.152	0.123	0.072	0.052	0.046	T
ARCH	0.367	0.303	0.168	0.156	0.144	H
	0.249	0.137	0.096	0.092	0.090	T

H: histogram basis, T: trigonometric basis.

Table 3
$L^2(f(x)\,\mathrm{d}x\,\mathrm{d}y)$ risk $\mathbb{E}\|\pi - \tilde{\pi}^*\|_f^2$ for simulated data with $\mathrm{pen}(m) = 0.5 D_{m_1} D_{m_2}/n$, averaged over $N = 200$ samples

Law	n					
	50	100	250	500	1000	basis
AR(1)	0.052	0.038	0.026	0.020	0.015	H
	0.081	0.069	0.046	0.037	0.031	T
$\sqrt{\mathrm{CIR}}$	0.016	0.014	0.010	0.006	0.004	H
	0.018	0.012	0.008	0.006	0.004	T

H: histogram basis, T: trigonometric basis.

For more precise results, empirical risk and L^2 risk are given respectively in Tables 1 and 2.

We observe that the results are better when we consider the empirical norm. It was expectable, given that this norm is adapted to the studied problem. Actually the better norm to evaluate the distance between π and its estimator is the norm $\|\cdot\|_f$. Table 3 shows that the errors in this case are very satisfactory.

So the results are roughly good but we cannot pretend that a basis among the others gives better results. We can then imagine a mixed strategy, i.e. a procedure which uses several kinds of bases and which can choose the best basis. These techniques are successfully used in a regression framework by Comte and Rozenholc [13,14].

7. Proofs

7.1. Proof of Proposition 1

Equality (5) yields, by multiplying by $\psi_{k_0}^m(y)$,

$$\sum_{j\in J_m} a_{j,k_0}\sum_{i=1}^n \varphi_j^m(X_i)\psi_{k_0}^m(y)\varphi_{j_0}^m(X_i)=\sum_{i=1}^n \varphi_{j_0}^m(X_i)\psi_{k_0}^m(X_{i+1})\psi_{k_0}^m(y).$$

Then, we sum over k_0 in K_m:

$$\sum_{i=1}^n t(X_i,y)\varphi_{j_0}^m(X_i)=\sum_{i=1}^n\sum_{k_0\in K_m}\psi_{k_0}^m(X_{i+1})\psi_{k_0}^m(y)\varphi_{j_0}^m(X_i).$$

If we multiply this equality by $a'_{j_0,k}\psi_k^m(y)$ and if we sum over $k\in K_m$ and $j_0\in J_m$, we obtain

$$\sum_{i=1}^n\left[t(X_i,y)-\sum_{k_0\in K_m}\psi_{k_0}^m(X_{i+1})\psi_{k_0}^m(y)\right]\sum_{k\in K_m}\sum_{j_0\in J_m}a'_{j_0,k}\varphi_{j_0}^m(X_i)\psi_k^m(y)=0$$

i.e.

$$\sum_{i=1}^n\left[t(X_i,y)-\sum_{k_0\in K_m}\psi_{k_0}^m(X_{i+1})\psi_{k_0}^m(y)\right]u(X_i,y)=0$$

for all u in S_m. So the vector $(t(X_i,y)-\sum_{k\in K_m}\psi_k^m(X_{i+1})\psi_k^m(y))_{1\leqslant i\leqslant n}$ is orthogonal to each vector in W. Since $t(X_i,y)$ belongs to W, the proposition is proved.

7.2. Proof of Theorem 2

For ρ a real larger than 1, let

$$\Omega_\rho=\left\{\forall t\in\mathcal{S}\ \|t\|_f^2\leqslant\rho\|t\|_n^2\right\}.$$

In the case of an arithmetical mixing, since $\gamma>14$, there exists a real c such that

$$\begin{cases}0<c<\dfrac{1}{6},\\[2mm] \gamma c>\dfrac{7}{3}.\end{cases}$$

We set in this case $q_n=\frac{1}{2}\lfloor n^c\rfloor$. In the case of a geometrical mixing, we set $q_n=\frac{1}{2}\lfloor c\ln(n)\rfloor$ where c is a real larger than $7/3\gamma$.

For the sake of simplicity, we suppose that $n=4p_nq_n$, with p_n an integer. Let for $i=1,\dots,n/2$, $U_i=(X_{2i-1},X_{2i})$.

Let

$$\begin{cases}A_l=(U_{2lq_n+1},\dots,U_{(2l+1)q_n}), & l=0,\dots,p_n-1,\\ B_l=(U_{(2l+1)q_n+1},\dots,U_{(2l+2)q_n}), & l=0,\dots,p_n-1.\end{cases}$$

We use now the mixing assumption A5. As in Viennet [27] we can build a sequence (A_l^*) such that

$$\begin{cases} A_l \text{ and } A_l^* \text{ have the same distribution,} \\ A_l^* \text{ and } A_{l'}^* \text{ are independent if } l \neq l', \\ P(A_l \neq A_l^*) \leqslant \beta_{2q_n}. \end{cases}$$

In the same way, we build (B_l^*) and we define for any $l \in \{0, \dots, p_n - 1\}$,

$$A_l^* = \left(U_{2lq_n+1}^*, \dots, U_{(2l+1)q_n}^*\right), \qquad B_l^* = \left(U_{(2l+1)q_n+1}^*, \dots, U_{(2l+2)q_n}^*\right)$$

so that the sequence $(U_1^*, \dots, U_{n/2}^*)$ and then the sequence $(X_1^*, \dots, X_n^*)$ are well defined.

Let now $V_i = (X_{2i}, X_{2i+1})$ for $i = 1, \dots, n/2$ and

$$\begin{cases} C_l = (V_{2lq_n+1}, \dots, V_{(2l+1)q_n}), & l = 0, \dots, p_n - 1, \\ D_l = (V_{(2l+1)q_n+1}, \dots, V_{(2l+2)q_n}), & l = 0, \dots, p_n - 1. \end{cases}$$

We can build $(V_1^{**}, \dots, V_{n/2}^{**})$ and then $(X_2^{**}, \dots, X_{n+1}^{**})$ such that

$$\begin{cases} C_l \text{ and } C_l^{**} \text{ have the same distribution,} \\ C_l^{**} \text{ and } C_{l'}^{**} \text{ are independent if } l \neq l', \\ P(C_l \neq C_l^{**}) \leqslant \beta_{2q_n}. \end{cases}$$

We put $X_{n+1}^* = X_{n+1}$ and $X_1^{**} = X_1$. Now let

$$\Omega^* = \left\{\forall i \ X_i = X_i^* = X_i^{**}\right\} \quad \text{and} \quad \Omega_\rho^* = \Omega_\rho \cap \Omega^*.$$

We denote by π_m the orthogonal projection of π on S_m. Now,

$$\mathbb{E}\|\tilde{\pi} - \pi \mathbb{1}_A\|_n^2 = \mathbb{E}\left(\|\tilde{\pi} - \pi \mathbb{1}_A\|_n^2 \mathbb{1}_{\Omega_\rho^*}\right) + \mathbb{E}\left(\|\tilde{\pi} - \pi \mathbb{1}_A\|_n^2 \mathbb{1}_{\Omega_\rho^{*c}}\right) \tag{11}$$

To bound the first term, we observe that for all s, t

$$\gamma_n(t) - \gamma_n(s) = \|t - \pi\|_n^2 - \|s - \pi\|_n^2 - 2Z_n(t - s)$$

where

$$Z_n(t) = \frac{1}{n} \sum_{i=1}^n \left\{ t(X_i, X_{i+1}) - \int_{\mathbb{R}} t(X_i, y)\pi(X_i, y)\,\mathrm{d}y \right\}.$$

Since $\|t - \pi\|_n^2 = \|t - \pi \mathbb{1}_A\|_n^2 + \|\pi \mathbb{1}_{A^c}\|_n^2$, we can write

$$\gamma_n(t) - \gamma_n(s) = \|t - \pi \mathbb{1}_A\|_n^2 - \|s - \pi \mathbb{1}_A\|_n^2 - 2Z_n(t - s).$$

The definition of $\hat{m}$ gives, for some fixed $m \in \mathcal{M}_n$,

$$\gamma_n(\tilde{\pi}) + \mathrm{pen}(\hat{m}) \leqslant \gamma_n(\pi_m) + \mathrm{pen}(m).$$

And then

$$\begin{aligned} \|\tilde{\pi} - \pi \mathbb{1}_A\|_n^2 &\leqslant \|\pi_m - \pi \mathbb{1}_A\|_n^2 + 2Z_n(\tilde{\pi} - \pi_m) + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}) \\ &\leqslant \|\pi_m - \pi \mathbb{1}_A\|_n^2 + 2\|\tilde{\pi} - \pi_m\|_f \sup_{t \in B_f(\hat{m})} Z_n(t) + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}) \end{aligned}$$

where, for all m', $B_f(m') = \{t \in S_m + S_{m'}, \ \|t\|_f = 1\}$. Let θ a real larger than 2ρ and $p(\cdot, \cdot)$ a function such that $\theta p(m, m') \leqslant \mathrm{pen}(m) + \mathrm{pen}(m')$. Then

$$\begin{aligned} \|\tilde{\pi} - \pi \mathbb{1}_A\|_n^2 \mathbb{1}_{\Omega_\rho^*} \leqslant{}& \|\pi_m - \pi \mathbb{1}_A\|_n^2 + \frac{1}{\theta}\|\tilde{\pi} - \pi_m\|_f^2 \mathbb{1}_{\Omega_\rho^*} + 2\,\mathrm{pen}(m) \\ &+ \theta \sum_{m' \in \mathcal{M}_n} \left[\sup_{t \in B_f(m')} Z_n^2(t) - p(m, m') \right]_+ \mathbb{1}_{\Omega_\rho^*}. \end{aligned} \tag{12}$$

But $\|\tilde{\pi} - \pi_m\|_f^2 \mathbb{1}_{\Omega_\rho^*} \leqslant \rho\|\tilde{\pi} - \pi_m\|_n^2 \mathbb{1}_{\Omega_\rho^*} \leqslant 2\rho\|\tilde{\pi} - \pi \mathbb{1}_A\|_n^2 \mathbb{1}_{\Omega_\rho^*} + 2\rho\|\pi \mathbb{1}_A - \pi_m\|_n^2$.

Then, inequality (12) becomes

$$\|\tilde\pi-\pi\mathbb{1}_A\|_n^2\mathbb{1}_{\Omega_\rho^*}\left(1-\frac{2\rho}{\theta}\right)\leqslant\left(1+\frac{2\rho}{\theta}\right)\|\pi_m-\pi\mathbb{1}_A\|_n^2+2\,\mathrm{pen}(m)$$
$$+\theta\sum_{m'\in\mathcal{M}_n}\Big[\sup_{t\in B_f(m')}Z_n^2(t)-p(m,m')\Big]_+\mathbb{1}_{\Omega_\rho^*}$$

so

$$\mathbb{E}\big(\|\tilde\pi-\pi\mathbb{1}_A\|_n^2\mathbb{1}_{\Omega_\rho^*}\big)\leqslant\frac{\theta+2\rho}{\theta-2\rho}\mathbb{E}\|\pi\mathbb{1}_A-\pi_m\|_n^2+\frac{2\theta}{\theta-2\rho}\,\mathrm{pen}(m)$$
$$+\frac{\theta^2}{\theta-2\rho}\sum_{m'\in\mathcal{M}_n}\mathbb{E}\Big(\Big[\sup_{t\in B_f(m')}Z_n^2(t)-p(m,m')\Big]_+\mathbb{1}_{\Omega_\rho^*}\Big).\tag{13}$$

We now use the following proposition:

Proposition 7. *Let*

$$p(m,m')=10\|\pi\|_\infty\frac{D(m,m')}{n}$$

where $D(m,m')$ denotes the dimension of $S_m+S_{m'}$. Then, under the assumptions of Theorem 2*, there exists a constant C_1 such that*

$$\sum_{m'\in\mathcal{M}_n}\mathbb{E}\Big(\Big[\sup_{t\in B_f(m')}Z_n^2(t)-p(m,m')\Big]_+\mathbb{1}_{\Omega^*}\Big)\leqslant\frac{C_1}{n}.\tag{14}$$

Then, with $\theta=3\rho$, inequalities (13) and (14) yield

$$\mathbb{E}\big(\|\tilde\pi-\pi\mathbb{1}_A\|_n^2\mathbb{1}_{\Omega_\rho^*}\big)\leqslant 5\|f\|_\infty\|\pi_m-\pi\mathbb{1}_A\|^2+6\,\mathrm{pen}(m)+\frac{9\rho C_1}{n}.\tag{15}$$

The penalty term $\mathrm{pen}(m)$ has to verify $\mathrm{pen}(m)+\mathrm{pen}(m')\geqslant 30\rho\|\pi\|_\infty D(m,m')/n$ i.e. $30\rho\|\pi\|_\infty\dim(S_m+S_{m'})\leqslant\mathrm{pen}(m)+\mathrm{pen}(m')$. We choose $\rho=3/2$ and so $\mathrm{pen}(m)=45\|\pi\|_\infty D_{m_1}D_{m_2}/n$.

To bound the second term in (11), we recall (see Section 3) that $(\hat\pi_{\hat m}(X_i,y))_{1\leqslant i\leqslant n}$ is the orthogonal projection of $(\sum_k\psi_k(X_{i+1})\psi_k(y))_{1\leqslant i\leqslant n}$ on

$$W=\big\{\big(t(X_i,y)\big)_{1\leqslant i\leqslant n},\ t\in S_{\hat m}\big\}$$

where $\psi_k=\psi_k^{\hat m}$. Thus, since P_W denotes the orthogonal projection on W, using (7), (8)

$$\begin{aligned}\big(\hat\pi_{\hat m}(X_i,y)\big)_{1\leqslant i\leqslant n}&=P_W\bigg(\Big(\sum_k\psi_k(X_{i+1})\psi_k(y)\Big)_{1\leqslant i\leqslant n}\bigg)\\&=P_W\bigg(\Big(\sum_k\pi_k(X_i)\psi_k(y)\Big)_{1\leqslant i\leqslant n}\bigg)+P_W\bigg(\Big(\sum_k\varepsilon_{i,k}\psi_k(y)\Big)_{1\leqslant i\leqslant n}\bigg)\\&=P_W\big(\pi\mathbb{1}_A(X_i,y)\big)_{1\leqslant i\leqslant n}+P_W\bigg(\Big(\sum_k\varepsilon_{i,k}\psi_k(y)\Big)_{1\leqslant i\leqslant n}\bigg).\end{aligned}$$

We denote by $\|\cdot\|_{\mathbb{R}^n}$ the Euclidean norm in $\mathbb{R}^n$, by X the vector $(X_i)_{1\leqslant i\leqslant n}$ and by ε_k the vector $(\varepsilon_{i,k})_{1\leqslant i\leqslant n}$. Thus

$$\begin{aligned}\|\pi\mathbb{1}_A-\hat\pi_{\hat m}\|_n^2&=\frac1n\int\bigg\|\pi\mathbb{1}_A(X,y)-P_W\big(\pi\mathbb{1}_A(X,y)\big)-P_W\Big(\sum_k\varepsilon_k\psi_k(y)\Big)\bigg\|_{\mathbb{R}^n}^2\mathrm{d}y\\&=\frac1n\int\big\|\pi\mathbb{1}_A(X,y)-P_W\big(\pi\mathbb{1}_A(X,y)\big)\big\|_{\mathbb{R}^n}^2\mathrm{d}y+\frac1n\int\bigg\|P_W\Big(\sum_k\varepsilon_k\psi_k(y)\Big)\bigg\|_{\mathbb{R}^n}^2\mathrm{d}y\end{aligned}$$

$$\leqslant \frac{1}{n}\int \|\pi \mathbb{1}_A(X,y)\|_{\mathbb{R}^n}^2 \,\mathrm{d}y + \frac{1}{n}\int \Big\| \sum_k \varepsilon_k \psi_k(y) \Big\|_{\mathbb{R}^n}^2 \mathrm{d}y$$

$$\leqslant \frac{1}{n}\sum_{i=1}^n \|\pi\|_\infty \int \pi(X_i,y)\,\mathrm{d}y + \frac{1}{n}\sum_{i=1}^n \int \Big[\sum_k \varepsilon_{i,k}\psi_k(y)\Big]^2 \mathrm{d}y$$

$$\leqslant \|\pi\|_\infty + \frac{1}{n}\sum_{i=1}^n \sum_k \varepsilon_{i,k}^2.$$

But Assumption M2 implies $\| \sum_{k\in K_{\hat{m}}} \psi_k^2\|_\infty \leqslant \phi_2 D_{\hat{m}_2}$. So, using (8),

$$\varepsilon_{i,k}^2 \leqslant 2\psi_k^2(X_{i+1}) + 2\mathbb{E}\big[\psi_k(X_{i+1})|X_i\big]^2$$

and

$$\sum_k \varepsilon_{i,k}^2 \leqslant 2\sum_k \psi_k^2(X_{i+1}) + 2\mathbb{E}\Big[\sum_k \psi_k^2(X_{i+1})|X_i\Big] \leqslant 4\phi_2 D_{\hat{m}_2}.$$

Thus we obtain

$$\|\pi\mathbb{1}_A - \hat{\pi}_{\hat{m}}\|_n^2 \leqslant \|\pi\|_\infty + 4\phi_2 D_{\hat{m}_2} \leqslant \|\pi\|_\infty + 4\phi_2 n^{1/3} \tag{16}$$

and, by taking the expectation, $\mathbb{E}(\|\pi\mathbb{1}_A - \hat{\pi}_{\hat{m}}\|_n^2 \mathbb{1}_{\Omega_\rho^{*c}}) \leqslant (\|\pi\|_\infty + 4\phi_2 n^{1/3}) P(\Omega_\rho^{*c})$.

We now remark that $P(\Omega_\rho^{*c}) = P(\Omega^{*c}) + P(\Omega_\rho^c \cap \Omega^*)$. In the geometric case $\beta_{2q_n} \leqslant \mathrm{e}^{-\gamma c \ln(n)} \leqslant n^{-\gamma c}$ and in the other case $\beta_{2q_n} \leqslant (2q_n)^{-\gamma} \leqslant n^{-\gamma c}$. Then

$$P(\Omega^{*c}) \leqslant 4p_n\beta_{2q_n} \leqslant n^{1-c\gamma}.$$

But we have choose c such that $c\gamma > 7/3$ and so $P(\Omega^{*c}) \leqslant n^{-4/3}$. Now we will use the following proposition:

Proposition 8. *Let $\rho > 1$. Then, under the assumptions of Theorems 2 or 4, there exists $C_2 > 0$ such that*

$$P\big(\Omega_\rho^c \cap \Omega^*\big) \leqslant C_2/n^{7/3}.$$

This proposition implies that $\mathbb{E}(\|\pi\mathbb{1}_A - \hat{\pi}_{\hat{m}}\|_n^2 \mathbb{1}_{\Omega_\rho^{*c}}) \leqslant C_3/n$.

Now we use (15) and we observe that this inequality holds for all m in $\mathcal{M}_n$, so

$$\mathbb{E}\|\tilde{\pi} - \pi\mathbb{1}_A\|_n^2 \leqslant C \inf_{m\in\mathcal{M}_n} \big(\|\pi\mathbb{1}_A - \pi_m\|^2 + \mathrm{pen}(m)\big) + \frac{C_4}{n}$$

with $C = \max(5\|f\|_\infty, 6)$.

7.3. Proof of Corollary 3

To control the bias term, we use the following lemma

Lemma 9. *Let π_A belong to $B_{2,\infty}^{\boldsymbol{\alpha}}(A)$. We consider that S'_m is one of the following spaces on A:*

- *a space of piecewise polynomials of degrees bounded by $s_i > \alpha_i - 1$ ($i = 1, 2$) based on a partition with rectangles of vertices $1/D_{m_1}$ and $1/D_{m_2}$,*
- *a linear span of $\{\phi_\lambda\psi_\mu, \lambda \in \bigcup_0^{m_1} \Lambda(j), \mu \in \bigcup_0^{m_2} M(k)\}$ where $\{\phi_\lambda\}$ and $\{\psi_\mu\}$ are orthonormal wavelet bases of respective regularities $s_1 > \alpha_1 - 1$ and $s_2 > \alpha_2 - 1$ (here $D_{m_i} = 2^{m_i}$, $i = 1, 2$),*
- *the space of trigonometric polynomials with degree smaller than D_{m_1} in the first direction and smaller than D_{m_2} in the second direction.*

Let π'_m be the orthogonal projection of π_A on S'_m. Then, there exists a positive constant C_0 such that

$$\left(\int_A |\pi_A - \pi'_m|^2\right)^{1/2} \leqslant C_0\left[D_{m_1}^{-\alpha_1} + D_{m_2}^{-\alpha_2}\right].$$

Proof. It is proved in [19] for S'_m a space of wavelets or polynomials and in [23] (pp. 191 and 200) for a space of trigonometric polynomials that

$$\left(\int_A |\pi_A - \pi'_m|^2\right)^{1/2} \leqslant C\left[\omega_{s_1+1,1}\left(\pi, D_{m_1}^{-1}\right) + \omega_{s_2+1,2}\left(\pi, D_{m_2}^{-1}\right)\right].$$

The definition of $B^{\boldsymbol{\alpha}}_{2,\infty}(A)$ implies $(\int_A |\pi_A - \pi'_m|^2)^{1/2} \leqslant C_0[D_{m_1}^{-\alpha_1} + D_{m_2}^{-\alpha_2}]$. □

If we choose for S'_m the set of the restrictions to A of the functions of S_m and π_A the restriction of π to A, we can apply Lemma 9. But π'_m is also the restriction to A of π_m so that

$$\|\pi \mathbb{1}_A - \pi_m\| \leqslant C_0\left[D_{m_1}^{-\alpha_1} + D_{m_2}^{-\alpha_2}\right].$$

According to Theorem 2

$$\mathbb{E}\|\tilde{\pi} - \pi \mathbb{1}_A\|_n^2 \leqslant C'' \inf_{m \in \mathcal{M}_n} \left\{ D_{m_1}^{-2\alpha_1} + D_{m_2}^{-2\alpha_2} + \frac{D_{m_1} D_{m_2}}{n} \right\}.$$

In particular, if m^* is such that $D_{m_1^*} = \lfloor n^{\frac{\alpha_2}{\alpha_1+\alpha_2+2\alpha_1\alpha_2}} \rfloor$ and $D_{m_2^*} = \lfloor (D_{m_1^*})^{\alpha_1/\alpha_2} \rfloor$ then

$$\mathbb{E}\|\tilde{\pi} - \pi \mathbb{1}_A\|_n^2 \leqslant C''' \left\{ D_{m_1^*}^{-2\alpha_1} + \frac{D_{m_1^*}^{1+\alpha_1/\alpha_2}}{n} \right\} = \mathrm{O}\left(n^{-\frac{2\alpha_1\alpha_2}{\alpha_1+\alpha_2+2\alpha_1\alpha_2}}\right).$$

But the harmonic mean of α_1 and α_2 is $\bar{\alpha} = 2\alpha_1\alpha_2/(\alpha_1+\alpha_2)$. Then $\mathbb{E}\|\tilde{\pi} - \pi \mathbb{1}_A\|_n^2 = \mathrm{O}(n^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+2}})$.

The condition $D_{m_1} \leqslant n^{1/3}$ allows this choice of m only if $\frac{\alpha_2}{\alpha_1+\alpha_2+2\alpha_1\alpha_2} < \frac{1}{3}$ i.e. if $\alpha_1 - 2\alpha_2 + 2\alpha_1\alpha_2 > 0$. In the same manner, the condition $\alpha_2 - 2\alpha_1 + 2\alpha_1\alpha_2 > 0$ must be verified.

7.4. *Proof of Theorem 4*

We use the same notations as for the proof of Theorem 2. Let us write

$$\mathbb{E}\|\tilde{\pi}^* - \pi \mathbb{1}_A\|^2 = B_1 + B_2 + B_3$$

with

$$\begin{cases} B_1 = \mathbb{E}\left(\|\tilde{\pi}^* - \pi \mathbb{1}_A\|^2 \mathbb{1}_{\Omega_\rho^*} \mathbb{1}_{\|\tilde{\pi}\| \leqslant k_n}\right), \\ B_2 = \mathbb{E}\left(\|\tilde{\pi}^* - \pi \mathbb{1}_A\|^2 \mathbb{1}_{\Omega_\rho^*} \mathbb{1}_{\|\tilde{\pi}\| > k_n}\right), \\ B_3 = \mathbb{E}\left(\|\tilde{\pi}^* - \pi \mathbb{1}_A\|^2 \mathbb{1}_{\Omega_\rho^{*c}}\right). \end{cases}$$

To bound the first term, we observe that for all $m \in \mathcal{M}_n$, on Ω_ρ^*, $\|\tilde{\pi} - \pi_m\|^2 \leqslant f_0^{-1}\rho\|\tilde{\pi} - \pi_m\|_n^2$. Then

$$\begin{aligned} \|\tilde{\pi} - \pi \mathbb{1}_A\|^2 \mathbb{1}_{\Omega_\rho^*} &\leqslant 2\|\tilde{\pi} - \pi_m\|^2 \mathbb{1}_{\Omega_\rho^*} + 2\|\pi_m - \pi \mathbb{1}_A\|^2 \\ &\leqslant 2 f_0^{-1}\rho\|\tilde{\pi} - \pi_m\|_n^2 \mathbb{1}_{\Omega_\rho^*} + 2\|\pi_m - \pi \mathbb{1}_A\|^2 \\ &\leqslant 2 f_0^{-1}\rho\left\{2\|\tilde{\pi} - \pi \mathbb{1}_A\|_n^2 \mathbb{1}_{\Omega_\rho^*} + 2\|\pi_m - \pi \mathbb{1}_A\|_n^2\right\} + 2\|\pi_m - \pi \mathbb{1}_A\|^2. \end{aligned}$$

Thus

$$B_1 \leqslant \mathbb{E}\left(\|\tilde{\pi} - \pi \mathbb{1}_A\|^2 \mathbb{1}_{\Omega_\rho^*}\right) \leqslant 4 f_0^{-1}\rho\mathbb{E}\left(\|\tilde{\pi} - \pi \mathbb{1}_A\|_n^2 \mathbb{1}_{\Omega_\rho^*}\right) + \left(4 f_0^{-1}\rho\|f\|_\infty + 2\right)\|\pi_m - \pi \mathbb{1}_A\|^2.$$

But, using (15), we obtain

$$B_1 \leqslant \left(24 f_0^{-1}\rho\|f\|_\infty + 2\right)\|\pi_m - \pi\mathbb{1}_A\|^2 + 24 f_0^{-1}\rho\,\mathrm{pen}(m) + 36 f_0^{-1}\rho^2\frac{C_1}{n}.$$

Since $\rho = 3/2$, by setting $C = \max(36 f_0^{-1}\|f\|_\infty + 1, 36 f_0^{-1})$,

$$B_1 \leqslant C\left\{\|\pi_m - \pi\mathbb{1}_A\|^2 + \mathrm{pen}(m)\right\} + \frac{81 f_0^{-1}C_1}{n}$$

for all $m \in \mathcal{M}_n$.

Next, the definition of $\tilde{\pi}^*$ and the Markov inequality provide

$$B_2 \leqslant \mathbb{E}\left(\|\pi\mathbb{1}_A\|^2\mathbb{1}_{\Omega_\rho^*}\mathbb{1}_{\|\tilde{\pi}\|>k_n}\right) \leqslant \|\pi\|^2\frac{\mathbb{E}(\|\tilde{\pi}\|^2\mathbb{1}_{\Omega_\rho^*})}{k_n^2}. \tag{17}$$

But $\|\tilde{\pi}\|^2\mathbb{1}_{\Omega_\rho^*} \leqslant \rho f_0^{-1}\|\tilde{\pi}\|_n^2 \leqslant 2\rho f_0^{-1}(\|\tilde{\pi} - \pi\mathbb{1}_A\|_n^2 + \|\pi\mathbb{1}_A\|_n^2)$. Now we use (16) to state

$$\begin{aligned}
\|\tilde{\pi}\|^2\mathbb{1}_{\Omega_\rho^*} &\leqslant 2\rho f_0^{-1}\left(\|\pi\|_\infty + 4\phi_2 n^{1/3} + \|\pi\mathbb{1}_A\|_n^2\right)\\
&\leqslant 2\rho f_0^{-1}\left(\|\pi\|_\infty + 4\phi_2 n^{1/3} + \frac{1}{n}\sum_{i=1}^n \|\pi\|_\infty \int \pi(X_i, y)\,\mathrm{d}y\right)\\
&\leqslant 2\rho f_0^{-1}\left(2\|\pi\|_\infty + 4\phi_2 n^{1/3}\right).
\end{aligned}$$

Then, since $k_n = n^{2/3}$, (17) becomes

$$B_2 \leqslant \|\pi\|^2\frac{2\rho f_0^{-1}(2\|\pi\|_\infty + 4\phi_2 n^{1/3})}{k_n^2} \leqslant 4\rho f_0^{-1}\|\pi\|^2\left(\frac{\|\pi\|_\infty}{n^{4/3}} + \frac{2\phi_2}{n}\right).$$

Lastly

$$B_3 \leqslant \mathbb{E}\left(2\left(\|\tilde{\pi}^*\|^2 + \|\pi\mathbb{1}_A\|^2\right)\mathbb{1}_{\Omega_\rho^{*c}}\right) \leqslant 2\left(k_n^2 + \|\pi\|^2\right)P\left(\Omega_\rho^{*c}\right).$$

We now remark that $P(\Omega_\rho^{*c}) = P(\Omega^{*c}) + P(\Omega_\rho^c \cap \Omega^*)$. In the geometric case $\beta_{2q_n} \leqslant \mathrm{e}^{-\gamma c\ln(n)} \leqslant n^{-\gamma c}$ and in the other case $\beta_{2q_n} \leqslant (2q_n)^{-\gamma} \leqslant n^{-\gamma c}$. Then

$$P\left(\Omega^{*c}\right) \leqslant 4p_n\beta_{2q_n} \leqslant n^{1-c\gamma}.$$

But, if $\gamma > 20$ in the arithmetic case, we can choose c such that $c\gamma > 10/3$ and so $P(\Omega^{*c}) \leqslant n^{-7/3}$. Then, using Proposition 8,

$$B_3 \leqslant 2\left(n^{4/3} + \|\pi\|^2\right)\frac{1 + C_2}{n^{7/3}} \leqslant \frac{2(C_2+1)(1+\|\pi\|^2)}{n}.$$

7.5. Proof of Theorem 6

Let ψ be a very regular wavelet with compact support. For $J = (j_1, j_2) \in \mathbb{Z}^2$ to be chosen below and $K = (k_1, k_2) \in \mathbb{Z}^2$, we set

$$\psi_{JK}(x, y) = 2^{(j_1+j_2)/2}\psi\left(2^{j_1}x - k_1\right)\psi\left(2^{j_2}y - k_2\right).$$

Let $\pi_0(x, y) = c_0\mathbb{1}_B(y)$ with B a compact set such that $A \subset B \times B$ and $|B| \geqslant 2|A|^{1/2}/L$, and $c_0 = |B|^{-1}$. So π_0 is a transition density with $\|\pi_0\|_{B_{2,\infty}^{\boldsymbol{\alpha}}(A)} \leqslant L/2$. Now we set R_J the maximal subset of $\mathbb{Z}^2$ such that

$$\mathrm{Supp}(\psi_{JK}) \subset A \quad \forall K \in R_J, \qquad \mathrm{Supp}(\psi_{JK}) \cap \mathrm{Supp}(\psi_{JK'}) = \emptyset \quad \text{if } K \neq K'.$$

The cardinal of R_J is $|R_J| = c2^{j_1+j_2}$, with c a positive constant which depends only on A and the support of ψ. Let, for all $\varepsilon = (\varepsilon_K) \in \{-1, 1\}^{|R_J|}$,

$$\pi_\varepsilon = \pi_0 + \frac{1}{\sqrt{n}}\sum_{K \in R_J}\varepsilon_K\psi_{JK}.$$

Let us denote by $\mathcal{G}$ the set of all such π_ε. Since $\int \psi = 0$ and π_0 is a transition density, for all x in $\mathbb{R}$,

$$\int \pi_\varepsilon(x, y)\,\mathrm{d}y = 1.$$

Additionally $\pi_\varepsilon(x, y) = \pi_0(x, y) \geqslant 0$ if $(x, y) \notin A$, and if $(x, y) \in A$: $\pi_\varepsilon \geqslant c_0 - 2^{(j_1+j_2)/2}\|\psi\|_\infty^2/\sqrt{n}$ and then $\pi_\varepsilon(x, y) \geqslant c_0/2 > 0$ as soon as

$$\left(\frac{2^{j_1+j_2}}{n}\right)^{1/2} \leqslant \frac{c_0}{2\|\psi\|_\infty^2}. \tag{18}$$

Thus, if (18) holds, $\pi_\varepsilon(x, y) \geqslant (c_0/2)\mathbb{1}_B(y)$ for all x, y. It implies that the underlying Markov chain is Doeblin recurrent and then positive recurrent. We verify that $f = c_0\mathbb{1}_B$ is the stationary density. To prove that $\pi_\varepsilon \in \mathcal{B}$, we still have to compute $\|\pi_\varepsilon\|_{B^{\boldsymbol{\alpha}}_{2,\infty}(A)}$. Hochmuth [19] proves that for ψ smooth enough $\|\sum_{K\in R_J} \varepsilon_K \psi_{JK}\|_{B^{\boldsymbol{\alpha}}_{2,\infty}(A)} \leqslant (2^{j_1\alpha_1} + 2^{j_2\alpha_2})\|\sum_{K\in R_J} \varepsilon_K\psi_{JK}\|_A$. Since

$$\left\|\sum_{K\in R_J} \varepsilon_K\psi_{JK}\right\|_A^2 = \sum_{K\in R_J} |\varepsilon_K|^2 = c2^{j_1+j_2},$$

then

$$\|\pi_\varepsilon\|_{B^{\boldsymbol{\alpha}}_{2,q}(A)} \leqslant \frac{L}{2} + \frac{2^{j_1\alpha_1} + 2^{j_2\alpha_2}}{\sqrt{n}} c^{1/2}2^{(j_1+j_2)/2}.$$

From now on, we suppose that Condition C is verified where

$$\text{Condition C:}\quad \frac{(2^{j_1\alpha_1} + 2^{j_2\alpha_2})2^{(j_1+j_2)/2}}{\sqrt{n}} \leqslant \frac{L}{2c^{1/2}}.$$

It entails in particular that (18) holds if j_1 and j_2 are great enough. Then for all ε, $\pi_\varepsilon \in \mathcal{B}$. We now use Lemma 10.2 p. 160 in Härdle et al. [17]. The likelihood ratio can be written

$$\Lambda_n(\pi_{\varepsilon_{*K}}, \pi_\varepsilon) = \prod_{i=1}^n \frac{\pi_{\varepsilon_{*K}}(X_i, X_{i+1})}{\pi_\varepsilon(X_i, X_{i+1})}.$$

Note that $\pi_\varepsilon(X_i, X_{i+1}) > 0$ P_{π_ε}- and $P_{\pi_{\varepsilon_{*K}}}$-almost surely (actually the chain "lives" on B). Then

$$\log\bigl(\Lambda_n(\pi_{\varepsilon_{*K}}, \pi_\varepsilon)\bigr) = \sum_{i=1}^n \log\left(1 - \frac{2}{\sqrt{n}}\frac{\varepsilon_K\psi_{JK}(X_i, X_{i+1})}{\pi_\varepsilon(X_i, X_{i+1})}\right).$$

We set $U_{JK}(X_i, X_{i+1}) = -\varepsilon_K\psi_{JK}(X_i, X_{i+1})/\pi_\varepsilon(X_i, X_{i+1})$ so that

$$\begin{aligned}\log\bigl(\Lambda_n(\pi_{\varepsilon_{*K}}, \pi_\varepsilon)\bigr) &= \sum_{i=1}^n \log\left(1 + \frac{2}{\sqrt{n}}U_{JK}(X_i, X_{i+1})\right)\\ &= \sum_{i=1}^n \left\{\theta\left(\frac{2}{\sqrt{n}}U_{JK}(X_i, X_{i+1})\right) + \frac{2}{\sqrt{n}}U_{JK}(X_i, X_{i+1}) - \frac{2}{n}U_{JK}^2(X_i, X_{i+1})\right\}\\ &= u_n + v_n - w_n\end{aligned}$$

with θ the function defined by $\theta(u) = \log(1+u) - u + u^2/2$. Now we prove the three following assertions

1° $\mathbb{E}_{\pi_\varepsilon}(|u_n|) = \mathbb{E}_{\pi_\varepsilon}(|\sum_{i=1}^n \theta(\frac{2}{\sqrt{n}}U_{JK}(X_i, X_{i+1}))|) \underset{n\to\infty}{\to} 0$,

2° $\mathbb{E}_{\pi_\varepsilon}(w_n) = \mathbb{E}_{\pi_\varepsilon}(\frac{2}{n}\sum_{i=1}^n U_{JK}^2(X_i, X_{i+1})) \leqslant 4$,

3° $\mathbb{E}_{\pi_\varepsilon}(v_n^2) = \mathbb{E}_{\pi_\varepsilon}(\frac{4}{n}|\sum_{i=1}^n U_{JK}(X_i, X_{i+1})|^2) \leqslant 8$.

1°: First we observe that

$$\left\|\frac{2}{\sqrt{n}}U_{JK}\right\|_\infty \leqslant \frac{2}{\sqrt{n}}\frac{2^{(j_1+j_2)/2}\|\psi\|_\infty^2}{c_0/2} = \mathrm{O}\left(\frac{2^{(j_1+j_2)/2}}{\sqrt{n}}\right)$$

and $2^{(j_1+j_2)}/n \to 0$ since Condition C holds. So there exists some integer n_0 such that $\forall n \geqslant n_0$, $\forall x, y$,

$$\left|\theta\left(2U_{JK}(x,y)/\sqrt{n}\right)\right| \leqslant \left|2U_{JK}(x,y)/\sqrt{n}\right|^3.$$

But

$$\begin{aligned}\iint \left|\frac{2U_{JK}(x,y)}{\sqrt{n}}\right|^3 f(x)\pi_\varepsilon(x,y)\,\mathrm{d}x\,\mathrm{d}y &= \frac{8}{n\sqrt{n}}\iint \frac{|\psi_{JK}(x,y)|^3}{\pi_\varepsilon(x,y)^2} f(x)\,\mathrm{d}x\,\mathrm{d}y \\ &\leqslant \frac{8}{n\sqrt{n}}\frac{2^{(j_1+j_2)/2}\|\psi\|_\infty^2 c_0}{(c_0/2)^2}\iint \psi_{JK}(x,y)^2\,\mathrm{d}x\,\mathrm{d}y \\ &\leqslant \frac{32\|\psi\|_\infty^2}{c_0 n}\left(\frac{2^{(j_1+j_2)}}{n}\right)^{1/2}.\end{aligned}$$

Then

$$\mathbb{E}_{\pi_\varepsilon}|u_n| \leqslant \sum_{i=1}^n \frac{32\|\psi\|_\infty^2}{c_0 n}\left(\frac{2^{(j_1+j_2)}}{n}\right)^{1/2} \underset{n\to\infty}{\longrightarrow} 0.$$

2°: We bound the expectation of $U_{JK}(X_i, X_{i+1})^2$:

$$\mathbb{E}_{\pi_\varepsilon}\left(U_{JK}(X_i,X_{i+1})^2\right) = \iint \frac{\psi_{JK}^2(x,y)}{\pi_\varepsilon(x,y)} f(x)\,\mathrm{d}x\,\mathrm{d}y \leqslant c_0 \iint_A \frac{\psi_{JK}^2(x,y)}{c_0/2}\,\mathrm{d}x\,\mathrm{d}y \leqslant 2. \tag{19}$$

And then $\mathbb{E}_{\pi_\varepsilon}(w_n) = \mathbb{E}_{\pi_\varepsilon}((2/n)\sum_{i=1}^n U_{JK}(X_i,X_{i+1})^2) \leqslant 4$.

3°: We observe that $\mathbb{E}_{\pi_\varepsilon}(U_{JK}(X_i,X_{i+1})|X_1,\ldots,X_i) = 0$ and thus $\sum_{i=1}^n U_{JK}(X_i,X_{i+1})$ is a martingale. A classic property of square integrable martingales involves

$$E_{\pi_\varepsilon}\left(\left[\sum_{i=1}^n U_{JK}(X_i,X_{i+1})\right]^2\right) = \sum_{i=1}^n \mathbb{E}_{\pi_\varepsilon}\left[U_{JK}(X_i,X_{i+1})^2\right].$$

Thus, using (19), $\mathbb{E}_{\pi_\varepsilon}(v_n^2) = (4/n)\sum_{i=1}^n \mathbb{E}_{\pi_\varepsilon}[U_{JK}(X_i,X_{i+1})^2] \leqslant 8$.

We deduce easily from the three previous assertions 1°, 2° and 3° that there exists $\lambda > 0$ and p_0 such that $P_{\pi_\varepsilon}(\Lambda_n(\pi_{\varepsilon_{*K}},\pi_\varepsilon) > \mathrm{e}^{-\lambda}) \geqslant p_0$. Thus, according to Lemma 10.2 in [17],

$$\max_{\pi_\varepsilon\in\mathcal{G}} \mathbb{E}_{\pi_\varepsilon}\|\hat{\pi}_n - \pi_\varepsilon\|_A^2 \geqslant \frac{|R_J|}{2}\delta^2\,\mathrm{e}^{-\lambda}p_0$$

where $\delta = \inf_{\varepsilon\neq\varepsilon'}\|\pi_\varepsilon - \pi_{\varepsilon'}\|_A/2 = \|\varepsilon_K\psi_{JK}/\sqrt{n}\|_A = 1/\sqrt{n}$.

Now for all n we choose $J = J(n) = (j_1(n), j_2(n))$ such that

$$c_1/2 \leqslant 2^{j_1}n^{-\frac{\alpha_2}{\alpha_1+\alpha_2+2\alpha_1\alpha_2}} \leqslant c_1 \quad \text{and} \quad c_2/2 \leqslant 2^{j_2}n^{-\frac{\alpha_1}{\alpha_1+\alpha_2+2\alpha_1\alpha_2}} \leqslant c_2$$

with c_1 and c_2 such that $(c_1^{\alpha_1} + c_2^{\alpha_2})\sqrt{c_1c_2} \leqslant L/(2c^{1/2})$ so that Condition C is satisfied. Moreover, we have

$$|R_J|\delta^2 \geqslant \frac{cc_1c_2}{4}n^{\frac{\alpha_2+\alpha_1}{\alpha_1+\alpha_2+2\alpha_1\alpha_2}-1} \geqslant \frac{cc_1c_2}{4}n^{\frac{-2\alpha_1\alpha_2}{\alpha_1+\alpha_2+2\alpha_1\alpha_2}}.$$

Thus

$$\max_{\pi_\varepsilon\in\mathcal{G}} \mathbb{E}_{\pi_\varepsilon}\|\hat{\pi}_n - \pi_\varepsilon\|_A^2 \geqslant \frac{c\mathrm{e}^{-\lambda}p_0c_1c_2}{8}n^{\frac{-2\alpha_1\alpha_2}{\alpha_1+\alpha_2+2\alpha_1\alpha_2}}.$$

And then for all estimator

$$\sup_{\pi\in\mathcal{B}} \mathbb{E}_\pi\|\hat{\pi}_n - \pi\|_A^2 \geqslant Cn^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+2}}$$

with $C = c\mathrm{e}^{-\lambda}p_0c_1c_2/8$.

7.6. Proof of Proposition 7

Let

$$\begin{cases} \Gamma_i(t) = t(X_i, X_{i+1}) - \int t(X_i, y)\pi(X_i, y)\,\mathrm{d}y, \\ \Gamma_i^*(t) = t(X_i^*, X_{i+1}^*) - \int t(X_i^*, y)\pi(X_i^*, y)\,\mathrm{d}y, \\ \Gamma_i^{**}(t) = t(X_i^{**}, X_{i+1}^{**}) - \int t(X_i^{**}, y)\pi(X_i^{**}, y)\,\mathrm{d}y. \end{cases}$$

We now define $Z_n^*(t)$:

$$Z_n^*(t) = \frac{1}{n}\sum_{i \text{ odd}} \Gamma_i^*(t) + \frac{1}{n}\sum_{i \text{ even}} \Gamma_i^{**}(t).$$

Let us remark that $Z_n^*(t)\mathbb{1}_{\Omega^*} = Z_n(t)\mathbb{1}_{\Omega^*}$. Next we split each of these terms:

$$Z_{n,1}^*(t) = \frac{1}{n}\sum_{l=0}^{p_n-1} \sum_{i=4lq_n+1,\ i \text{ odd}}^{2(2l+1)q_n-1} \Gamma_i^*(t), \qquad Z_{n,2}^*(t) = \frac{1}{n}\sum_{l=0}^{p_n-1} \sum_{i=2(2l+1)q_n+1,\ i \text{ odd}}^{2(2l+2)q_n-1} \Gamma_i^*(t),$$

$$Z_{n,3}^*(t) = \frac{1}{n}\sum_{l=0}^{p_n-1} \sum_{i=4lq_n+2,\ i \text{ even}}^{2(2l+1)q_n} \Gamma_i^{**}(t), \qquad Z_{n,4}^*(t) = \frac{1}{n}\sum_{l=0}^{p_n-1} \sum_{i=2(2l+1)q_n+2,\ i \text{ even}}^{2(2l+2)q_n} \Gamma_i^{**}(t).$$

We use the following lemma:

Lemma 10. (Talagrand [26]) *Let $\mathcal{U}_0, \ldots, \mathcal{U}_{N-1}$ i.i.d. variables and $(\zeta_t)_{t\in B}$ a set of functions. Let $G(t) = \frac{1}{N}\sum_{l=0}^{N-1} \zeta_t(\mathcal{U}_l)$. We suppose that*

$$(1)\ \sup_{t\in B}\|\zeta_t\|_\infty \leqslant M_1, \quad (2)\ \mathbb{E}\Big(\sup_{t\in B}\big|G(t)\big|\Big) \leqslant H, \quad (3)\ \sup_{t\in B}\mathrm{Var}\big[\zeta_t(\mathcal{U}_0)\big] \leqslant v.$$

Then, there exists $K > 0$, $K_1 > 0$, $K_2 > 0$ such that

$$\mathbb{E}\Big[\sup_{t\in B} G^2(t) - 10H^2\Big]_+ \leqslant K\left[\frac{v}{N}\mathrm{e}^{-K_1\frac{NH^2}{v}} + \frac{M_1^2}{N^2}\mathrm{e}^{-K_2\frac{NH}{M_1}}\right].$$

Here $N = p_n$, $B = B_f(m')$ and for $l \in \{0, \ldots, p_n - 1\}$, $\mathcal{U}_l = (X_{4lq_n+1}^*, \ldots, X_{2(2l+1)q_n}^*)$,

$$\zeta_t(x_1, \ldots, x_{2q_n}) = \frac{1}{q_n}\sum_{i=1,\ i \text{ odd}}^{2q_n-1} t(x_i, x_{i+1}) - \int t(x_i, y)\pi(x_i, y)\,\mathrm{d}y.$$

Then

$$G(t) = \frac{1}{p_n}\sum_{l=0}^{p_n-1} \frac{1}{q_n}\sum_{i=4lq_n+1,\ i \text{ odd}}^{2(2l+1)q_n-1} \Gamma_i^*(t) = 4Z_{n,1}^*(t).$$

We now compute M_1, H and v.

(1) We recall that $S_m + S_{m'}$ is included in the model $S_{m''}$ with dimension $\max(D_{m_1}, D_{m_1'})\max(D_{m_2}, D_{m_2'})$.

$$\begin{aligned} \sup_{t\in B}\|\zeta_t\|_\infty &\leqslant \sup_{t\in B}\|t\|_\infty \frac{1}{q_n}\sum_{i=1,\ i \text{ odd}}^{2q_n-1}\left(1 + \int \pi(x_i, y)\,\mathrm{d}y\right) \\ &\leqslant 2\phi_0\sqrt{\max(D_{m_1}, D_{m_1'})\max(D_{m_2}, D_{m_2'})}\,\|t\| \leqslant \frac{2\phi_0}{f_0} n^{1/3}. \end{aligned}$$

Then we set $M_1 = (2\phi_0/f_0)n^{1/3}$.

(2) Since A_0 and A_0^* have the same distribution, $\zeta_t(\mathcal{U}_0) = \frac{1}{q_n}\sum_{i=1,\ i\ \text{odd}}^{2q_n-1} \Gamma_i^*(t)$ has the same distribution than $\frac{1}{q_n}\sum_{i=1,\ i\ \text{odd}}^{2q_n-1} \Gamma_i(t)$. We observe that $\mathbb{E}(\Gamma_i(t)|X_i) = 0$ and then for all set I

$$\mathbb{E}\left(\left[\sum_{i\in I}\Gamma_i(t)\right]^2\right) = \mathbb{E}\left(\sum_{i,j\in I}\Gamma_i(t)\Gamma_j(t)\right) = 2\mathbb{E}\left(\sum_{j<i}\mathbb{E}\left[\Gamma_i(t)\Gamma_j(t)|X_i\right]\right) + \sum_{i\in I}\mathbb{E}\left[\Gamma_i^2(t)\right]$$
$$= 2\mathbb{E}\left(\sum_{j<i}\Gamma_j(t)\mathbb{E}\left[\Gamma_i(t)|X_i\right]\right) + \sum_{i\in I}\mathbb{E}\left[\Gamma_i^2(t)\right] = \sum_{i\in I}\mathbb{E}\left[\Gamma_i^2(t)\right].$$

In particular

$$\operatorname{Var}\left[\zeta_t(\mathcal{U}_0)\right] = \mathbb{E}\left(\left[\frac{1}{q_n}\sum_{i=1,\ i\ \text{odd}}^{2q_n-1}\Gamma_i(t)\right]^2\right) = \frac{1}{q_n^2}\sum_{i=1,\ i\ \text{odd}}^{2q_n-1}\mathbb{E}\left[\Gamma_i^2(t)\right]$$
$$\leqslant \frac{1}{q_n^2}\sum_{i=1,\ i\ \text{odd}}^{2q_n-1}\mathbb{E}\left[t^2(X_i, X_{i+1})\right] \leqslant \frac{1}{q_n}\|\pi\|_\infty\|t\|_f^2.$$

Then $v = \|\pi\|_\infty/q_n$.

(3) Let $(\bar\varphi_j \otimes \psi_k)_{(j,k)\in\Lambda(m,m')}$ an orthonormal basis of $(S_m + S_{m'}, \|\cdot\|_f)$.

$$\mathbb{E}\left(\sup_{t\in B}\left|G^2(t)\right|\right) \leqslant \sum_{j,k}\mathbb{E}\left(G^2(\bar\varphi_j\otimes\psi_k)\right)$$
$$\leqslant \sum_{j,k}\frac{1}{p_n^2q_n^2}\mathbb{E}\left(\left[\sum_{l=0}^{p_n-1}\sum_{i=4lq_n+1,\ i\ \text{odd}}^{2(2l+1)q_n-1}\Gamma_i^*(\bar\varphi_j\otimes\psi_k)\right]^2\right)$$
$$\leqslant \sum_{j,k}\frac{16}{n^2}\sum_{l=0}^{p_n-1}\mathbb{E}\left(\left[\sum_{i=4lq_n+1,\ i\ \text{odd}}^{2(2l+1)q_n-1}\Gamma_i^*(\bar\varphi_j\otimes\psi_k)\right]^2\right)$$

where we used the independence of the A_l^*. Now we can replace Γ_i^* by Γ_i in the sum because A_l and A_l^* have the same distribution and we use as previously the martingale property of the Γ_i.

$$\mathbb{E}\left(\sup_{t\in B}\left|G^2(t)\right|\right) \leqslant \sum_{j,k}\frac{16}{n^2}\sum_{l=0}^{p_n-1}\mathbb{E}\left(\left[\sum_{i=4lq_n+1,\ i\ \text{odd}}^{2(2l+1)q_n-1}\Gamma_i(\bar\varphi_j\otimes\psi_k)\right]^2\right)$$
$$\leqslant \sum_{j,k}\frac{16}{n^2}\sum_{l=0}^{p_n-1}\sum_{i=4lq_n+1,\ i\ \text{odd}}^{2(2l+1)q_n-1}\mathbb{E}\left(\Gamma_i^2(\bar\varphi_j\otimes\psi_k)\right)$$
$$\leqslant \sum_{j,k}\frac{4}{n}\|\pi\|_\infty\|\bar\varphi_j\otimes\psi_k\|_f^2 \leqslant 4\|\pi\|_\infty\frac{D(m,m')}{n}.$$

Then

$$\mathbb{E}^2\left(\sup_{t\in B}\left|G(t)\right|\right) \leqslant 4\|\pi\|_\infty\frac{D(m,m')}{n} \quad \text{and} \quad H^2 = 4\|\pi\|_\infty\frac{D(m,m')}{n}.$$

According to Lemma 10, there exists $K' > 0$, $K_1 > 0$, $K_2' > 0$ such that

$$\mathbb{E}\left[\sup_{t\in B_f(m')}\left(4Z_{n,1}^*\right)^2(t) - 10H^2\right]_+ \leqslant K'\left[\frac{1}{n}\mathrm{e}^{-K_1D(m,m')} + n^{-4/3}q_n^2\mathrm{e}^{-K_2'n^{1/6}\sqrt{D(m,m')}/q_n}\right].$$

But $q_n \leqslant n^c$ with $c < \frac{1}{6}$. So

$$\sum_{m'\in\mathcal{M}_n}\mathbb{E}\Big[\sup_{t\in B_f(m')} Z_{n,1}^{*2}(t)-\frac{p(m,m')}{4}\Big]_+\leqslant\frac{K'}{n}\Big[\sum_{m'\in\mathcal{M}_n}\mathrm{e}^{-K_1D(m,m')}+n^{2c-1/3}|\mathcal{M}_n|\mathrm{e}^{-K_2'n^{1/6-c}}\Big]\leqslant\frac{A_1}{n}. \qquad (20)$$

In the same way, $\sum_{m'\in\mathcal{M}_n}\mathbb{E}[\sup_{t\in B_f(m')} Z_{n,r}^{*2}(t)-p(m,m')/4]_+\leqslant A_r/n$ for $r=2,3,4$. And then

$$\sum_{m'\in\mathcal{M}_n}\mathbb{E}\Big(\Big[\sup_{t\in B_f(m')} Z_n^2(t)-p(m,m')\Big]_+\mathbb{1}_{\Omega^*}\Big)=\sum_{m'\in\mathcal{M}_n}\mathbb{E}\Big(\Big[\sup_{t\in B_f(m')} Z_n^{*2}(t)-p(m,m')\Big]_+\mathbb{1}_{\Omega^*}\Big)\leqslant\frac{C_1}{n}.$$

7.7. *Proof of Proposition 8*

First we observe that

$$P\big(\Omega_\rho^c\cap\Omega^*\big)\leqslant P\bigg(\sup_{t\in\mathcal{B}}\big|\nu_n\big(t^2\big)\big|>1-\frac{1}{\rho}\bigg)$$

where

$$\nu_n(t)=\frac{1}{n}\sum_{i=1}^n\int\big[t\big(X_i^*,y\big)-\mathbb{E}\big(t\big(X_i^*,y\big)\big)\big]\,\mathrm{d}y\quad\text{and}\quad\mathcal{B}=\big\{t\in\mathcal{S},\ \|t\|_f=1\big\}.$$

But, if $t(x,y)=\sum_{j,k}a_{j,k}\varphi_j(x)\psi_k(y)$, then

$$\nu_n\big(t^2\big)=\sum_{j,j'}\sum_k a_{j,k}a_{j',k}\bar\nu_n(\varphi_j\varphi_{j'})$$

where

$$\bar\nu_n(u)=\frac{1}{n}\sum_{i=1}^n\big[u\big(X_i^*\big)-\mathbb{E}\big(u\big(X_i^*\big)\big)\big]. \qquad (21)$$

Let $b_j=(\sum_k a_{j,k}^2)^{1/2}$, then $|\nu_n(t^2)|\leqslant\sum_{j,j'}b_jb_{j'}|\bar\nu_n(\varphi_j\varphi_{j'})|$ and, if $t\in\mathcal{B}$, $\sum_j b_j^2=\sum_j\sum_k a_{j,k}^2=\|t\|^2\leqslant f_0^{-1}$. Thus

$$\sup_{t\in\mathcal{B}}\big|\nu_n\big(t^2\big)\big|\leqslant f_0^{-1}\sup_{\sum b_j^2=1}\sum_{j,l}b_jb_l\big|\bar\nu_n(\varphi_j\varphi_l)\big|.$$

Lemma 11. *Let $B_{j,l}=\|\varphi_j\varphi_l\|_\infty$ and $V_{j,l}=\|\varphi_j\varphi_l\|_2$. Let, for any symmetric matrix $(A_{j,l})$*

$$\bar\rho(A)=\sup_{\sum a_j^2=1}\sum_{j,l}|a_ja_l|A_{j,l}$$

and $L(\varphi)=\max\{\bar\rho^2(V),\bar\rho(B)\}$. Then, if M2 *is satisfied, $L(\varphi)\leqslant\phi_1\mathcal{D}_n^2$.*

This lemma is proved in Baraud et al. [3].

Let

$$x=\frac{f_0^2(1-1/\rho)^2}{40\|f\|_\infty L(\varphi)}\quad\text{and}\quad\Delta=\big\{\forall j\forall l\ \big|\bar\nu_n(\varphi_j\varphi_l)\big|\leqslant 4\big[B_{j,l}x+V_{j,l}\sqrt{2\|f\|_\infty x}\,\big]\big\}.$$

On Δ:

$$\begin{aligned}\sup_{t\in\mathcal{B}}\big|\nu_n\big(t^2\big)\big|&\leqslant 4f_0^{-1}\sup_{\sum b_j^2=1}\sum_{j,l}b_jb_l\big[B_{j,l}x+V_{j,l}\sqrt{2\|f\|_\infty x}\big]\\&\leqslant 4f_0^{-1}\big[\bar\rho(B)x+\bar\rho(V)\sqrt{2\|f\|_\infty x}\,\big]\\&\leqslant\Big(1-\frac{1}{\rho}\Big)\Big[\frac{f_0(1-1/\rho)}{10\|f\|_\infty}\frac{\bar\rho(B)}{L(\varphi)}+\frac{2}{\sqrt5}\Big(\frac{\bar\rho^2(V)}{L(\varphi)}\Big)^{1/2}\Big]\\&\leqslant\Big(1-\frac{1}{\rho}\Big)\Big[\frac{1}{10}+\frac{2}{\sqrt5}\Big]\leqslant\Big(1-\frac{1}{\rho}\Big).\end{aligned}$$

Then $P(\sup_{t\in\mathcal{B}}|\nu_n(t^2)|>1-1/\rho)\leqslant P(\Delta^c)$. But $\bar{\nu}_n(u)=2\bar{\nu}_{n,1}(u)+2\bar{\nu}_{n,2}(u)$ with

$$\bar{\nu}_{n,r}(u)=\frac{1}{p_n}\sum_{l=0}^{p_n-1}Y_{l,r}(u),\quad r=1,2,$$

with

$$\begin{cases} Y_{l,1}(u)=\dfrac{1}{2q_n}\displaystyle\sum_{i=4lq_n+1}^{2(2l+1)q_n}\left[u\left(X_i^*\right)-\mathbb{E}\left(u\left(X_i^*\right)\right)\right],\\ Y_{l,2}(u)=\dfrac{1}{2q_n}\displaystyle\sum_{i=2(2l+1)q_n+1}^{2(2l+2)q_n}\left[u\left(X_i^*\right)-\mathbb{E}\left(u\left(X_i^*\right)\right)\right].\end{cases}$$

To bound $P(\bar{\nu}_{n,1}(\varphi_j\varphi_l)\geqslant B_{j,l}x+V_{j,l}\sqrt{2\|f\|_\infty x})$, we will use the Bernstein inequality given in Birgé and Massart [8]. That is why we bound $\mathbb{E}|Y_{l,1}(u)|^m$:

$$\begin{aligned}\mathbb{E}\left|Y_{l,1}(u)\right|^m&\leqslant\frac{1}{4q_n^2}\left(2\|u\|_\infty\right)^{m-2}\mathbb{E}\left|\sum_{i=4lq_n+1}^{2(2l+1)q_n}\left[u\left(X_i^*\right)-\mathbb{E}\left(u\left(X_i^*\right)\right)\right]\right|^2\\&\leqslant\left(2\|u\|_\infty\right)^{m-2}\frac{1}{4q_n^2}\mathbb{E}\left|\sum_{i=4lq_n+1}^{2(2l+1)q_n}\left[u(X_i)-\mathbb{E}\left(u(X_i)\right)\right]\right|^2\\&\leqslant\left(2\|u\|_\infty\right)^{m-2}\frac{1}{2q_n}\mathbb{E}\sum_{i=4lq_n+1}^{2(2l+1)q_n}\left|u(X_1)-\mathbb{E}\left(u(X_1)\right)\right|^2\end{aligned}$$

since $X_i^*=X_i$ on Ω^* and using an elementary convex inequality. Thus

$$\begin{aligned}\mathbb{E}\left|Y_{l,1}(u)\right|^m&\leqslant\left(2\|u\|_\infty\right)^{m-2}\mathbb{E}\left|u(X_1)-\mathbb{E}\left(u(X_1)\right)\right|^2\leqslant\left(2\|u\|_\infty\right)^{m-2}\int u^2(x)f(x)\,\mathrm{d}x\\&\leqslant 2^{m-2}\left(\|u\|_\infty\right)^{m-2}\left(\sqrt{\|f\|_\infty}\,\|u\|\right)^2.\end{aligned}\tag{22}$$

With $u=\varphi_j\varphi_{j'}$, $\mathbb{E}|Y_{l,1}(\varphi_j\varphi_{j'})|^m\leqslant 2^{m-2}(B_{j,j'})^{m-2}(\sqrt{\|f\|_\infty}V_{j,j'})^2$. And then

$$P\left(\left|\bar{\nu}_{n,r}(\varphi_j\varphi_l)\right|\geqslant B_{j,l}x+V_{j,l}\sqrt{2\|f\|_\infty x}\right)\leqslant 2\mathrm{e}^{-p_nx}.$$

Given that $P(\Omega_\rho^c\cap\Omega^*)\leqslant P(\Delta^c)=\sum_{j,l}P(|\bar{\nu}_n(\varphi_j\varphi_l)|>4(B_{j,l}x+V_{j,l}\sqrt{2\|f\|_\infty x}))$,

$$\begin{aligned}P\left(\Omega_\rho^c\cap\Omega^*\right)&\leqslant 4\mathcal{D}_n^2\exp\left\{-\frac{p_nf_0^2(1-1/\rho)^2}{40\|f\|_\infty L(\varphi)}\right\}\\&\leqslant 4n^{2/3}\exp\left\{-\frac{f_0^2(1-1/\rho)^2}{160\|f\|_\infty}\frac{n}{q_nL(\varphi)}\right\}.\end{aligned}$$

But $L(\varphi)\leqslant\phi_1\mathcal{D}_n^2\leqslant\phi_1n^{2/3}$ and $q_n\leqslant n^{1/6}$ so

$$P\left(\Omega_\rho^c\cap\Omega^*\right)\leqslant 4n^{2/3}\exp\left\{-\frac{f_0^2(1-1/\rho)^2}{160\|f\|_\infty\phi_1}n^{1/6}\right\}\leqslant\frac{C}{n^{7/3}}.\tag{23}$$

Acknowledgements

I would like to thank Fabienne Comte for her constructive comments and suggestions, and Marc Hoffmann for his help regarding the lower bound.

Appendix A. Random penalty

Here we prove that Theorem 2 is valid with a penalty which does not depend on $\|\pi\|_\infty$.

Theorem 12. *We consider the following penalty*:

$$\overline{\mathrm{pen}}(m) = \overline{K_0}\|\hat{\pi}\|_\infty \frac{D_{m_1}D_{m_2}}{n}$$

where $\overline{K_0}$ is a numerical constant and $\hat{\pi} = \hat{\pi}_{m*}$ with S_{m*} a space of trigonometric polynomials such that

$$\ln n \leqslant D_{m_1*} = D_{m_2*} \leqslant n^{1/6}.$$

If the restriction of π to A belongs to $B_{2,\infty}^{(\alpha_1,\alpha_2)}(A)$ with $\alpha_1 > 3/2$ and $\alpha_2 > \max(\frac{\alpha_1}{2\alpha_1-3}, \frac{3\alpha_1}{2\alpha_1-1})$, then, under assumptions of Theorem 2, *for n large enough,*

$$\mathbb{E}\|\pi\mathbb{1}_A - \tilde{\pi}\|_n^2 \leqslant C \inf_{m\in\mathcal{M}_n}\left\{d^2(\pi\mathbb{1}_A, S_m) + \frac{D_{m_1}D_{m_2}}{n}\right\} + \frac{C'}{n}.$$

Remark 13. The condition on the regularity of π is verified for example if $\alpha_1 > 2$ and $\alpha_2 > 2$. If $\alpha_1 = \alpha_2 = \alpha$, it is equivalent to $\alpha > 2$.

Proof. We recall that $\|\pi\|_\infty$ denotes actually $\|\pi\mathbb{1}_A\|_\infty$ and we introduce the following set:

$$\Lambda = \left\{\left|\frac{\|\hat{\pi}\|_\infty}{\|\pi\mathbb{1}_A\|_\infty} - 1\right| < \frac{1}{2}\right\}.$$

As previously, we decompose the space:

$$\mathbb{E}\|\tilde{\pi} - \pi\mathbb{1}_A\|_n^2 = \mathbb{E}\big(\|\tilde{\pi} - \pi\mathbb{1}_A\|_n^2\mathbb{1}_{\Omega_\rho^*\cap\Lambda}\big) + \mathbb{E}\big(\|\tilde{\pi} - \pi\mathbb{1}_A\|_n^2\mathbb{1}_{\Omega_\rho^*\cap\Lambda^c}\big) + \mathbb{E}\big(\|\tilde{\pi} - \pi\mathbb{1}_A\|_n^2\mathbb{1}_{\Omega_\rho^{*c}}\big).$$

We have already dealt with the third term. For the first term, we can proceed as in the proof of Theorem 2 as soon as

$$\theta p(m, m') \leqslant \overline{\mathrm{pen}}(m) + \overline{\mathrm{pen}}(m')$$

with $\theta = 3\rho = 9/2$ and $p(m, m') = 10\|\pi\|_\infty D(m, m')/n$. But on Λ, $\|\pi\|_\infty < 2\|\hat{\pi}\|_\infty$ and so

$$\begin{aligned}\theta p(m, m') &= 10\theta\|\pi\|_\infty\frac{D(m, m')}{n} \leqslant 20\theta\|\hat{\pi}\|_\infty\frac{D(m, m')}{n}\\ &\leqslant 20\theta\|\hat{\pi}\|_\infty\frac{D_{m_1}D_{m_2}}{n} + 20\theta\|\hat{\pi}\|_\infty\frac{D_{m'_1}D_{m'_2}}{n}.\end{aligned}$$

It is sufficient to set $\overline{K_0} = 20\theta$.

Now, inequality (16) gives

$$\mathbb{E}\big(\|\pi\mathbb{1}_A - \hat{\pi}_{\hat{m}}\|_n^2\mathbb{1}_{\Omega_\rho^*\cap\Lambda^c}\big) \leqslant \big(\|\pi\|_\infty + 4\phi_2 n^{1/3}\big)P\big(\Omega_\rho^*\cap\Lambda^c\big).$$

It remains to prove that $P(\Omega_\rho^*\cap\Lambda^c) \leqslant Cn^{-4/3}$ for some constant C.

$$\begin{aligned}P\big(\Omega_\rho^*\cap\Lambda^c\big) &= P\big(\big|\|\hat{\pi}\|_\infty - \|\pi\mathbb{1}_A\|_\infty\big|\mathbb{1}_{\Omega_\rho^*} \geqslant \|\pi\|_\infty/2\big) \leqslant P\big(\|\hat{\pi} - \pi\mathbb{1}_A\|_\infty\mathbb{1}_{\Omega_\rho^*} \geqslant \|\pi\|_\infty/2\big)\\ &\leqslant P\big(\|\hat{\pi} - \pi_{m*}\|_\infty\mathbb{1}_{\Omega_\rho^*} \geqslant \|\pi\|_\infty/4\big) + P\big(\|\pi_{m*} - \pi\mathbb{1}_A\|_\infty \geqslant \|\pi\|_\infty/4\big)\\ &\leqslant P\left(\|\hat{\pi} - \pi_{m*}\|\mathbb{1}_{\Omega_\rho^*} \geqslant \frac{\|\pi\|_\infty}{4\phi_0\sqrt{D_{m_1*}D_{m_2*}}}\right) + P\big(\|\pi_{m*} - \pi\mathbb{1}_A\|_\infty \geqslant \|\pi\|_\infty/4\big)\end{aligned}$$

since $\|\hat{\pi} - \pi_{m*}\|_\infty \leqslant \phi_0\sqrt{D_{m*_1}D_{m*_2}}\|\hat{\pi} - \pi_{m^*}\|$.

Furthermore the inequality $\gamma_n(\hat{\pi}) \leqslant \gamma_n(\pi_{m*})$ leads to

$$\|\hat{\pi} - \pi\mathbb{1}_A\|_n^2 \leqslant \|\pi_{m*} - \pi\mathbb{1}_A\|_n^2 + \frac{1}{\theta'}\|\hat{\pi} - \pi_{m*}\|_f^2 + \theta' \sup_{t\in B_f(m*)} Z_n^2(t)$$

and then, on Ω_ρ,

$$\|\hat{\pi}-\pi_{m*}\|_f^2\left(1-\frac{2\rho}{\theta'}\right)\leqslant 4\rho\|\pi_{m*}-\pi\mathbb{1}_A\|_n^2+2\rho\theta'\sup_{t\in B_f(m*)}Z_n^2(t)$$

so

$$\begin{aligned}\|\hat{\pi}-\pi_{m*}\|^2&\leqslant\frac{4\rho\theta' f_0^{-1}}{\theta'-2\rho}\|\pi_{m*}-\pi\mathbb{1}_A\|_n^2+\frac{2\rho\theta'^2 f_0^{-1}}{\theta'-2\rho}\sup_{t\in B_f(m*)}Z_n^2(t)\\&\leqslant 12\rho f_0^{-1}|A_2|\|\pi_{m*}-\pi\mathbb{1}_A\|_\infty^2+18\rho^2 f_0^{-1}\sup_{t\in B_f(m*)}Z_n^2(t)\end{aligned}$$

with $\theta'=3\rho$ and by remarking that for t with support A, $\|t\|_n^2\leqslant|A_2|\|t\|_\infty^2$. Thus

$$\begin{aligned}P\big(\Omega_\rho^*\cap\Lambda^c\big)&\leqslant P\left(\sup_{t\in B_f(m*)}Z_n^2(t)\mathbb{1}_{\Omega_\rho^*}\geqslant\frac{\|\pi\|_\infty^2}{32\phi_0^2n^{1/3}}\frac{1}{18\rho^2 f_0^{-1}}\right)\\&\quad+P\left(\|\pi_{m*}-\pi\mathbb{1}_A\|_\infty^2\geqslant\frac{\|\pi\|_\infty^2}{32\phi_0^2D_{m_1*}D_{m_2*}}\frac{1}{12\rho f_0^{-1}|A_2|}\right)+P\big(\|\pi_{m*}-\pi\mathbb{1}_A\|_\infty\geqslant\|\pi\|_\infty/4\big)\\&\leqslant P\left(\sup_{t\in B_f(m*)}Z_n^2(t)\mathbb{1}_{\Omega^*}\geqslant\frac{a}{n^{1/3}}\right)+P\big(D_{m_1*}D_{m_2*}\|\pi_{m*}-\pi\mathbb{1}_A\|_\infty^2\geqslant b\big)\\&\quad+P\left(\|\pi_{m*}-\pi\mathbb{1}_A\|_\infty\geqslant\frac{\|\pi\|_\infty}{4}\right)\end{aligned}\tag{24}$$

with

$$a=\frac{\|\pi\|_\infty^2}{32\phi_0^2}\frac{1}{18\rho^2 f_0^{-1}}\quad\text{and}\quad b=\frac{\|\pi\|_\infty^2}{32\phi_0^2}\frac{1}{12\rho f_0^{-1}|A_2|}.$$

We will first study the two last terms in (24). Since the restriction π_A of π belongs to $B_{2,\infty}^{(\alpha_1,\alpha_2)}(A)$, the embedding theorem proved in Nikol'skiĭ [23] p. 236 implies that π_A belongs to $B_{\infty,\infty}^{(\beta_1,\beta_2)}(A)$ with $\beta_1=\alpha_1(1-1/\bar{\alpha})$ and $\beta_2=\alpha_2(1-1/\bar{\alpha})$. Then the approximation Lemma 9 (which is still valid for the trigonometric polynomial spaces with the infinite norm instead of the L^2 norm) yields to

$$\|\pi_{m*}-\pi\mathbb{1}_A\|_\infty\leqslant C\big(D_{m_1*}^{-\beta_1}+D_{m_2*}^{-\beta_2}\big).$$

And then, since $D_{m_1*}=D_{m_2*}$,

$$\begin{aligned}D_{m_1*}D_{m_2*}\|\pi_{m*}-\pi\mathbb{1}_A\|_\infty^2&\leqslant C'\big(D_{m_1*}^{2-2\beta_1}+D_{m_1*}^{2-2\beta_2}\big)\\&\leqslant C'\big((\ln n)^{2-2\beta_1}+(\ln n)^{2-2\beta_2}\big)\to 0.\end{aligned}$$

Indeed

$$\begin{cases}2-2\beta_1<0\Leftrightarrow 2\alpha_1\alpha_2-3\alpha_2-\alpha_1>0,\\2-2\beta_2<0\Leftrightarrow 2\alpha_1\alpha_2-3\alpha_1-\alpha_2>0\end{cases}$$

and this double condition is ensured when $\alpha_1>3/2$ and $\alpha_2>\max(\frac{\alpha_1}{2\alpha_1-3},\frac{3\alpha_1}{2\alpha_1-1})$. Consequently, for n large enough,

$$P\big(D_{m_1*}D_{m_2*}\|\pi_{m*}-\pi\mathbb{1}_A\|_\infty^2\geqslant b\big)+P\left(\|\pi_{m*}-\pi\|_\infty\geqslant\frac{\|\pi\|_\infty}{4}\right)=0.$$

We will now prove that

$$P\left(\sup_{t\in B_f(m*)}Z_n^2(t)\mathbb{1}_{\Omega^*}\geqslant\frac{a}{n^{1/3}}\right)\leqslant\frac{C}{n^{4/3}}$$

and then using (24), we will have $P(\Omega_\rho^* \cap \Lambda^c) \leqslant Cn^{-4/3}$. We remark that, if $(\varphi_j \otimes \psi_k)_{j,k}$ is a base of $(S_{m*}, \|\cdot\|_f)$,

$$\sup_{t\in B_f(m*)} Z_n^2(t) \leqslant \sum_{j,k} Z_n^2(\varphi_j \otimes \psi_k)$$

and we recall that, on Ω^*, $Z_n(t) = \sum_{r=1}^4 Z_{n,r}^*(t)$ (see the proof of Proposition 7). So we are interested in

$$P\left(Z_{n,1}^{*2}(\varphi_j \otimes \psi_k)\mathbb{1}_{\Omega^*} \geqslant \frac{a}{4D_{m_1*}D_{m_2*}n^{1/3}}\right).$$

Let $x = Bn^{-2/3}$ with B such that $2f_0^{-2}B^2 + 4\|\pi\|_\infty B \leqslant a/4$ (for example $B = \inf(1, a/8(f_0^{-2} + 2\|\pi\|_\infty))$). Then

$$\left(\sqrt{2\|\pi\|_\infty x} + \sqrt{D_{m_1*}D_{m_2*}}f_0^{-1}x\right)^2 \leqslant \frac{a}{4D_{m_1*}D_{m_2*}n^{1/3}}.$$

So we will now bound $P(Z_{n,1}^*(\varphi_j \otimes \psi_k)\mathbb{1}_{\Omega^*} \geqslant \sqrt{2\|\pi\|_\infty x} + \sqrt{D_{m_1*}D_{m_2*}}f_0^{-1}x)$ by using the Bernstein inequality given in [8]. That is why we bound $\mathbb{E}|\frac{1}{4q_n}\sum_{i=1,\ i \text{ odd}}^{2q_n-1}\Gamma_i^*(t)|^m$ for all integer $m \geqslant 2$,

$$\begin{aligned}
\mathbb{E}\left|\frac{1}{4q_n}\sum_{i=1,\ i\text{ odd}}^{2q_n-1}\Gamma_i^*(t)\right|^m &\leqslant \frac{(2\|t\|_\infty q_n)^{m-2}}{(4q_n)^m}\mathbb{E}\left|\sum_{i=1,\ i\text{ odd}}^{2q_n-1}\left[t\left(X_i^*, X_{i+1}^*\right) - \int t\left(X_i^*, y\right)\pi\left(X_i^*, y\right)\mathrm{d}y\right]\right|^2 \\
&\leqslant \left(\frac{\|t\|_\infty}{2}\right)^{m-2}\frac{1}{16q_n^2}\mathbb{E}\left|\sum_{i=1,\ i\text{ odd}}^{2q_n-1}\left[t(X_i, X_{i+1}) - \int t(X_i, y)\pi(X_i, y)\,\mathrm{d}y\right]\right|^2 \\
&\leqslant \left(\frac{\|t\|_\infty}{2}\right)^{m-2}\frac{1}{16}\int t^2(x, y)f(x)\pi(x, y)\,\mathrm{d}x\,\mathrm{d}y \\
&\leqslant \frac{1}{2^{m+2}}\left(\|t\|_\infty\right)^{m-2}\|\pi\|_\infty\|t\|_f^2.
\end{aligned}$$

Then

$$\mathbb{E}\left|\frac{1}{4q_n}\sum_{i=1,\ i\text{ odd}}^{2q_n-1}\Gamma_{i*}(\varphi_j \otimes \psi_k)\right|^m \leqslant \frac{1}{2^{m+2}}\left(\sqrt{D_{m_1*}D_{m_2*}}\,f_0^{-1}\right)^{m-2}\|\pi\|_\infty.$$

Thus the Bernstein inequality gives

$$P\left(\left|Z_{n,1}^*(\varphi_j \otimes \psi_k)\right| \geqslant \sqrt{D_{m_1*}D_{m_2*}}\,f_0^{-1}x + \sqrt{2\|\pi\|_\infty x}\right) \leqslant 2\mathrm{e}^{-p_n x}.$$

Hence

$$\begin{aligned}
P\left(\sup_{t\in B_f(m*)} Z_{n,1}^{*2}(t)\mathbb{1}_{\Omega^*} \geqslant \frac{a}{4n^{1/3}}\right) &\leqslant 2D_{m_1*}D_{m_2*}\exp\left\{-p_n Bn^{-2/3}\right\} \\
&\leqslant 2n^{2/3}\exp\left\{-\frac{B}{4}\frac{n^{1/3}}{q_n}\right\}.
\end{aligned}$$

But

$$2n^{2/3}\exp\left\{-\frac{B}{4}\frac{n^{1/3}}{q_n}\right\} \leqslant Cn^{-4/3}$$

since $q_n \leqslant n^{1/6}$ and so

$$P\left(\sup_{t\in B_f(m*)} Z_n^2(t)\mathbb{1}_{\Omega^*} \geqslant \frac{a}{n^{1/3}}\right) \leqslant \frac{4C}{n^{4/3}}. \qquad \square$$

References

[1] P. Ango Nzé, Critères d'ergodicité de quelques modèles à représentation markovienne, C. R. Acad. Sci. Paris Sér. I 315 (12) (1992) 1301–1304.
[2] K.B. Athreya, G.S. Atuncar, Kernel estimation for real-valued Markov chains, Sankhyā Ser. A 60 (1) (1998) 1–17.
[3] Y. Baraud, F. Comte, G. Viennet, Adaptive estimation in autoregression or β-mixing regression via model selection, Ann. Statist. 29 (3) (2001) 839–875.
[4] A. Barron, L. Birgé, P. Massart, Risk bounds for model selection via penalization, Probab. Theory Related Fields 113 (3) (1999) 301–413.
[5] A.K. Basu, D.K. Sahoo, On Berry–Esseen theorem for nonparametric density estimation in Markov sequences, Bull. Inform. Cybernet. 30 (1) (1998) 25–39.
[6] L. Birgé, Approximation dans les espaces métriques et théorie de l'estimation, Z. Wahrsch. Verw. Gebiete 65 (2) (1983) 181–237.
[7] L. Birgé, P. Massart, From model selection to adaptive estimation, in: Festschrift for Lucien Le Cam, Springer, New York, 1997, pp. 55–87.
[8] L. Birgé, P. Massart, Minimum contrast estimators on sieves: exponential bounds and rates of convergence, Bernoulli 4 (3) (1998) 329–375.
[9] D. Bosq, Sur l'estimation de la densité d'un processus stationnaire et mélangeant, C. R. Acad. Sci. Paris Sér. A-B 277 (1973) A535–A538.
[10] M. Chaleyat-Maurel, V. Genon-Catalot, Computable infinite dimensional filters with applications to discretized diffusion processes, Stochastic Process. Appl. 116 (10) (2006) 1447–1467.
[11] S. Clémençon, Adaptive estimation of the transition density of a regular Markov chain, Math. Methods Statist. 9 (4) (2000) 323–357.
[12] F. Comte, Adaptive estimation of the spectrum of a stationary Gaussian sequence, Bernoulli 7 (2) (2001) 267–298.
[13] F. Comte, Y. Rozenholc, Adaptive estimation of mean and volatility functions in (auto-)regressive models, Stochastic Process. Appl. 97 (1) (2002) 111–145.
[14] F. Comte, Y. Rozenholc, A new algorithm for fixed design regression and denoising, Ann. Inst. Statist. Math. 56 (3) (2004) 449–473.
[15] P. Doukhan, Mixing. Properties and Examples, Lecture Notes in Statistics, vol. 85, Springer-Verlag, New York, 1994.
[16] P. Doukhan, M. Ghindès, Estimation de la transition de probabilité d'une chaîne de Markov Doëblin-récurrente. Étude du cas du processus autorégressif général d'ordre 1, Stochastic Process. Appl. 15 (3) (1983) 271–293.
[17] W. Härdle, G. Kerkyacharian, P. Picard, A. Tsybakov, Wavelets, Approximation, and Statistical Applications, Lecture Notes in Statistics, vol. 129, Springer-Verlag, New York, 1998.
[18] O. Hernández-Lerma, S.O. Esparza, B.S. Duran, Recursive nonparametric estimation of nonstationary Markov processes, Bol. Soc. Mat. Mexicana (2) 33 (2) (1988) 57–69.
[19] R. Hochmuth, Wavelet characterizations for anisotropic Besov spaces, Appl. Comput. Harmon. Anal. 12 (2) (2002) 179–208.
[20] C. Lacour, Nonparametric estimation of the stationary density and the transition density of a Markov chain, Preprint MAP5 n° 2005-8: http://www.math-info.univ-paris5.fr/map5/publis/titres05.html.
[21] S.P. Meyn, R.L. Tweedie, Markov Chains and Stochastic Stability, Springer-Verlag, London, 1993.
[22] A. Mokkadem, Sur un modèle autorégressif non linéaire : ergodicité et ergodicité géométrique, J. Time Ser. Anal. 8 (2) (1987) 195–204.
[23] S.M. Nikol'skiĭ, Approximation of Functions of Several Variables and Imbedding Theorems, Springer-Verlag, New York, 1975, translated from the Russian by John M. Danskin, Jr., Die Grundlehren der Mathematischen Wissenschaften, Band 205.
[24] E. Pardoux, A.Y. Veretennikov, On the Poisson equation and diffusion approximation. I, Ann. Probab. 29 (3) (2001) 1061–1085.
[25] G.G. Roussas, Nonparametric estimation in Markov processes, Ann. Inst. Statist. Math. 21 (1969) 73–87.
[26] M. Talagrand, New concentration inequalities in product spaces, Invent. Math. 126 (3) (1996) 505–563.
[27] G. Viennet, Inequalities for absolutely regular sequences: application to density estimation, Probab. Theory Related Fields 107 (4) (1997) 467–492.