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 \( \pi \) 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 \( \pi \) under mild conditions. The resulting estimate is adaptive with an optimal rate of convergence over a large range of anisotropic Besov spaces \( B^{(a_1,a_2)}_{2,\infty} \). Some simulations are also presented.

© 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 \( \pi(x,y) \, dy = P(X_{i+1} \in dy \mid X_i = x) \). 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).

All these authors have estimated $\pi$ 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 $\pi$ 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 $\hat{\pi}$ of $\pi$ from the observations $X_1, \ldots, X_{n+1}$ such that the order of the $L^2$ risk depends only on the regularity of $\pi$ 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 $\beta$-mixing Markov chains (the notion of “regular” Markov chains in [11] is equivalent to $\Phi_1$-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 $\mu$ (for more details, we refer to [21]). We assume that the distribution $\mu$ 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 $\pi$. Since the number of observations is finite, $\pi$ 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 $\mu$, thus the chain is (strictly) stationary.
A3. The transition density $\pi$ 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 $A_1$, $f(x) \geq f_0$.
A5. The chain is geometrically $\beta$-mixing ($\beta_\gamma \leq e^{-\gamma d}$), or arithmetically $\beta$-mixing ($\beta_\gamma \leq q^{-\gamma}$).

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

$$\beta_\gamma = \int \|P^\gamma(x, \cdot) - \mu\|_{TV} f(x) \, dx$$

where $\|\cdot\|_{TV}$ 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 \((X_i \Delta)_{1 \leq i \leq n}\) where \(\Delta > 0\) is the observation step and \((X_t)_{t \geq 0}\) is defined by

\[
dX_t = b(X_t) \, dt + \sigma(X_t) \, dW_t
\]

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

1. There exist \(\lambda_- > 0\) such that for all \(x \neq 0\), \(0 < \lambda_- < \sigma^2(x) < \lambda_+\).
2. There exists \(M_0 \geq 0\), \(\alpha > -1\) and \(r > 0\) such that for all \(|x| \geq M_0\), \(xb(x) \leq -r|x|^{\alpha+1}\).

Then, if \(X_0\) follows the stationary distribution, the discretized process \((X_i \Delta)_{1 \leq i \leq n}\) satisfies Assumptions A1–A5.

Note that the mixing is geometrical as soon as \(\alpha \geq 0\). The continuity of the transition density ensures that Assumption A3 holds. Moreover, we can write

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

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

\[
\|f\|_\infty \leq \frac{1}{M\lambda_-} \exp \left[ 2 \sup_{x \in A_1} \int_0^x b(u) \, du \right]
\]

and

\[
f_0 \geq \frac{1}{M\lambda_+} \exp \left[ 2 \inf_{x \in A_1} \int_0^x b(u) \, du \right].
\]

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 \(\varepsilon_{X_{n-1},n}\) 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 \(\sup_{|x| \leq M} |\varphi(x)| < \infty\).
2. There exist \(l_0 > 0\), \(l_1 > 0\) such that for all \(x\), \(y\), \(l_0 \leq l(x, y) \leq l_1\).

Then Mokkadem [22] proves that the chain is Harris recurrent and geometrically ergodic. It implies that Assumptions A1 and A5 are satisfied. Moreover \(\pi(x, y) = lx(y - \varphi(x))\) and \(f(y) = \int f(x)\pi(x, y) \, dx\) and then Assumptions A3, A4 hold with \(f_0 \geq l_0\) and \(\|f\|_\infty \leq \|\pi\|_\infty \leq l_1\).

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 \((\xi_n), (Z_n)\) 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 \(h : \mathbb{R} \to \mathbb{R}^+\) such that \(\mathbb{E}h(Z_n) < \infty\) and positive constants \(M, c\) such that

\[
\forall (u, v) > M \quad |F(u, v)| < \rho |u| + h(v) - c \quad \text{and} \quad \sup_{|x| \leq M} |F(x)| < \infty.
\]
Then Doukhan [15] proves (p. 102) that \((X_n)\) is a geometrically \(\beta\)-mixing process. We can write
\[
\pi(x, y) = \int l(z) f_\xi(y - F(x, z)) \, dz
\]
where \(f_\xi\) is the density of \(\xi_n\). So, if we assume furthermore that there exist \(a_0, a_1 > 0\) such that \(a_0 \leq f_\xi \leq a_1\), then Assumptions A3, A4 are verified with \(f_0 \geq a_0\) and \(\|f\|_\infty \leq \|\pi\|_\infty \leq a_1\).

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 \(\varepsilon_n\) has a positive density \(l\) with respect to the Lebesgue measure and that there exists \(s \geq 1\) such that \(E|\varepsilon_n|^s < \infty\). The chain \((X_n)\) satisfies Assumptions A1 and A5 if (see [1]):
\[
\lim sup_{|x| \to \infty} \frac{|F(x)| + |G(x)||E|\varepsilon_n|^s}{{|x|}^{1/s}} < 1.
\]
(2)
In addition, we assume that \(\forall x, l_0 \leq l(x) \leq l_1\). Then Assumption A3 is verified with \(\pi \leq l_1/\inf_{x \in A_1} G(x)\). And Assumption A4 holds with \(f_0 \geq l_0 \int f G^{-1}\) and \(\|f\|_\infty \leq l_1 \int f G^{-1}\).

2.3. Assumptions on the models

In order to estimate \(\pi\), 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 \((\psi^m_j)_{j \in J_m}\) such that, for all \(j\), the support of \(\psi^m_j\) 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 \(\psi^m_k\) 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
\[
S_m = F_{m_1} \otimes H_{m_2} = \left\{ t, t(x, y) = \sum_{j \in J_m} \sum_{k \in K_m} a^m_{j,k} \psi^m_j(x) \psi^m_k(y) \right\}.
\]

The assumptions on the models are the following:

M1. For all \(m_2\), \(D_{m_2} \leq n^{1/3}\) and \(D_n := \max_{m \in \mathcal{M}_n} D_{m_1} \leq n^{1/3}\).

M2. There exist positive reals \(\phi_1, \phi_2\) such that, for all \(u \in F_{m_1}\), \(\|u\|_\infty \leq \phi_1 D_{m_1} \int u^2\), and for all \(v \in H_{m_2}\), \(\sup_{x \in A_2} |v(x)|^2 \leq \phi_2 D_{m_2} \int v^2\). By letting \(\phi_0 = \sqrt{\phi_1 \phi_2}\), that leads to
\[
\forall t \in S_m \quad \|t\|_\infty \leq \phi_0 \sqrt{D_{m_1} D_{m_2}} \|t\|\quad (3)
\]
where \(\|t\|^2 = \int_{\mathbb{R}^2} t^2(x, y) \, dx \, dy\).

M3. \(D_{m_1} \leq D_{m_1} \Rightarrow F_{m_1} \subset F_{m_1'}\) and \(D_{m_2} \leq D_{m_2} \Rightarrow H_{m_2} \subset H_{m_2'}\).

The first assumption guarantees that \(\dim S_m = D_{m_1} D_{m_2} \leq n^{2/3} \leq n\) 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 \(D_{m''} = \max(D_{m_1}, D_{m_1'})\) and \(D_{m''_2} = \max(D_{m_2}, D_{m_2'})\)). We denote by \(S\) the space with maximal dimension among the \((S_m)_{m \in \mathcal{M}_n}\). Thus for all \(m\) in \(\mathcal{M}_n\), \(S_m \subset S\).

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 $A = [0, 1]$, we chose $\langle \varphi_0, \ldots, \varphi_{m_1 - 1} \rangle$ with $\varphi_0 = \mathbb{1}_{[0,1]}$, $\varphi_{2^j} = \sqrt{2} \sin(2\pi j x) \mathbb{1}_{[0,1]}(x)$, $\varphi_{2^j - 1} = \sqrt{2} \sin(2\pi j x) \mathbb{1}_{(0,1)}(x)$ for $j \geq 1$. For this model $D_{m_1} = m_1$ and $\phi_1 = 2$ hold.
• Histogram basis: for $A = [0, 1]$, we chose $\langle \varphi_1, \ldots, \varphi_{2m_1} \rangle$ with $\varphi_j = 2^{m_1/2} \mathbb{1}_{[j(1-1)/2^{m_1}, j/2^{m_1}]}$ for $j = 1, \ldots, 2^{m_1}$. Here $D_{m_1} = 2^{m_1}$, $\phi_1 = 1$.
• Regular piecewise polynomial basis: for $A = [0, 1]$, polynomials of degree $0, \ldots, r$ (where $r$ is fixed) on each interval $[(l-1)/2^D, l/2^D], l = 1, \ldots, 2^D$. In this case, $m_1 = (D, r)$, $J_m = \{ j = (l, d), 1 \leq l \leq 2^D, 0 \leq d \leq r \}$, $D_{m_1} = (r+1)2^D$. We can put $\phi_1 = \sqrt{r+1}$.
• Regular wavelet basis: $\langle \Psi_{lk}, l = -1, \ldots, m_1, k \in \Lambda(l) \rangle$ where $\Psi_{-1,k}$ points out the translates of the father wavelet and $\Psi_{lk}(x) = 2^{l/2} \psi(2^l x - k)$ where $\psi$ is the mother wavelet. We assume that the support of the wavelets is included in $A_1$ and that $\Psi_{-1}$ belongs to the Sobolev space $W^2_2$.

3. Estimation procedure

3.1. Definition of the contrast

To estimate the function $\pi$, we define the contrast

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

(4)

We choose this contrast because

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

where

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

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 $\pi$ by minimizing this contrast on $S_m$. Let $t(x, y) = \sum_{j \in J_m} \sum_{k \in \mathbb{K}_m} a_{j,k} \psi_j^m(x) \psi_k^m(y)$ a function in $S_m$. Then, if $A_m$ denotes the matrix $(a_{j,k})_{j \in J_m, k \in \mathbb{K}_m}$,

$$\forall j_0 \forall k_0 \quad \frac{\partial \gamma_n(t)}{\partial a_{j_0,k_0}} = 0 \iff G_m A_m = Z_m,$$

where

$$G_m = \left\{ \frac{1}{n} \sum_{i=1}^{n} \psi_j^m(X_i) \psi_k^m(X_i) \right\}_{j \in J_m, k \in \mathbb{K}_m},$$

$$Z_m = \left\{ \frac{1}{n} \sum_{i=1}^{n} \psi_j^m(X_i) \psi_k^m(X_{i+1}) \right\}_{j \in J_m, k \in \mathbb{K}_m}.$$

Indeed,

$$\frac{\partial \gamma_n(t)}{\partial a_{j_0,k_0}} = 0 \iff \sum_{j \in J_m} a_{j_0,k_0} \frac{1}{n} \sum_{i=1}^{n} \psi_j^m(X_i) \psi_j^m(X_i) = \frac{1}{n} \sum_{i=1}^{n} \psi_j^m(X_i) \psi_j^m(X_{i+1}).$$

(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 $\varphi_{j_0}$ ($G_m$ 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.
\[ \forall j, k \quad \frac{\partial \gamma_n(t)}{\partial a_{j_0, k_0}} = 0 \iff \forall y \quad (t(X_i, y))_{1 \leq i \leq n} = PW \left( \left( \sum_k \psi_k^m(X_{i+1}) \psi_k^m(y) \right)_{1 \leq i \leq n} \right) \]

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

Thus the minimization of \( \gamma_n(t) \) leads to a unique vector \((\hat{\pi}_m(X_i, y))_{1 \leq i \leq n}\) defined as the projection of \((\sum_k \psi_k(X_{i+1}) \psi_k(y))_{1 \leq i \leq n}\) 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 \((\hat{\pi}_m(X_i, y))_{1 \leq i \leq n}\). This remark leads to consider the risk defined with the empirical norm
\[ \| t \|_n = \left( \frac{1}{n} \sum_{i=1}^{n} \int t^2(X_i, y) \, dy \right)^{1/2}. \]

This norm is the natural distance in this problem and we can notice that if \( t \) is deterministic with support included in \( A_1 \times \mathbb{R} \)
\[ f_0^n \leq \mathbb{E} \| t \|_n^2 \leq \| f \|_\infty \| t \|_n^2 \]
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
\[ Y_{i,k} = \psi_k^m(X_{i+1}) \quad \text{for} \ i \in \{1, \ldots, n\}, \]
\[ t_k(x) = \int t(x, y) \psi_k^m(y) \, dy \quad \text{for all} \ t \in L^2(\mathbb{R}^2). \]

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 \( \psi_k^m \) by \( \psi_k \) and \( \psi_j^m \) by \( \psi_j \). Then, if \( t \) belongs to \( S_m \),
\[ t(x, y) = \sum_{j \in J_m} \sum_{k \in K_m} \left( \int t(x', y') \psi_j(x') \psi_k(y') \, dx' \, dy' \right) \psi_j(x) \psi_k(y) \]
\[ = \sum_{k \in K_m} \left( \int t_k(x') \psi_k(x') \, dx' \right) \psi_j(x) \psi_k(y) = \sum_{k \in K_m} t_k(x) \psi_k(y) \]

and then, by replacing this expression of \( t \) in \( \gamma_n(t) \), we obtain
\[ \gamma_n(t) = \frac{1}{n} \sum_{i=1}^{n} \left[ \int \sum_{k,k'} \left( \int t_k(X_i) t_k(X_i) \psi_k(y) \psi_k'(y) \, dy - 2 \sum_k t_k(X_i) \psi_k(X_{i+1}) \right) \right] \]
\[ = \frac{1}{n} \sum_{i=1}^{n} \sum_{k \in K_m} \left[ t_k^2(X_i) - 2 t_k(X_i) Y_{i,k} \right] = \frac{1}{n} \sum_{i=1}^{n} \sum_{k \in K_m} \left[ t_k(X_i) - Y_{i,k} \right]^2 - Y_{i,k}^2. \]

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 $\pi_k = \int \pi(\cdot, y)\psi_k(y)\,dy$ which verifies

$$Y_{i,k} = \pi_k(X_i) + \epsilon_{i,k}$$

where

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

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 $(\hat{\pi}_k(X_i))_{1 \leq i \leq n}$ 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 $\pi$ on some $S_m$ by estimating first each $\pi_k$, 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 $\pi$ for all $S_m$. Let now

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

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

4. Calculation of the risk

For a function $h$ and a subspace $S$, let

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

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 $\hat{\pi}$ the estimator of the transition density $\pi$ described in Section 3 with models verifying Assumptions M1–M3 and the following penalty:

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

where $K_0$ is a numerical constant. Then

$$\mathbb{E}[\|\pi \mathbb{1}_A - \hat{\pi}\|^2_n] \leq C \inf_{m \in \mathcal{M}_n} \left\{ d^2(\pi \mathbb{1}_A, S_m) + \text{pen}(m) \right\} + \frac{C'}{n}$$

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 $\pi$. 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 $\pi$ 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 $\pi$ restricted to $A$ belongs to the anisotropic Besov space on $A$ with regularity $\alpha = (\alpha_1, \alpha_2)$. Note that if $\pi$ belongs to $B^\alpha_{2,\infty}(\mathbb{R}^2)$, then $\pi$ restricted to
A belongs to $B_{2,\infty}^{\alpha}(A)$. Let us recall the definition of $B_{2,\infty}^{\alpha}(A)$. Let $e_1$ and $e_2$ be the canonical basis vectors in $\mathbb{R}^2$ and for $i = 1, 2$, $A'_{h,i} = \{ x \in \mathbb{R}^2; x, x + he_i, \ldots, x + rhe_i \in A \}$. Next, for $x$ in $A'_{h,i}$, let

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

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

$$\omega_{r,i}(g,t) = \sup_{|h| \leq t} \left( \int_{A'_{h,i}} |\Delta^r_{h,i}g(x)|^2 \, dx \right)^{1/2}.$$

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

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

for $r_i$ integers larger than $\alpha_i$. The transition density $\pi$ 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 $D_{m_2} = D_{m_1}^{\beta_2/\beta_1} < D_{m_1}$ for the second direction, where $\beta$ is more regular. We can thus write the following corollary.

**Corollary 3.** We suppose that $\pi$ restricted to $A$ belongs to the anisotropic Besov space $B_{2,\infty}^{\alpha}(A)$ with regularity $\alpha = (\alpha_1, \alpha_2)$ 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 $\alpha_1 - 1$). Then, under the assumptions of Theorem 2,

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

where $\bar{\alpha}$ is the harmonic mean of $\alpha_1$ and $\alpha_2$.

The harmonic mean of $\alpha_1$ and $\alpha_2$ 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 $\alpha_1 \geq 1$ and the condition $\alpha_2 - 2\alpha_1 + 2\alpha_1\alpha_2 > 0$ as soon as $\alpha_2 \geq 1$.

Thus we obtain the rate of convergence $n^{-\frac{2\bar{\alpha}}{\bar{\alpha} + 2}}$, 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:

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

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 $\pi$ described in Section 5.1. Then

$$\mathbb{E} \| \tilde{\pi}^* - \pi 1_A \|^2 \leq C \inf_{m \in \mathcal{M}_a} \left\{ \mathbb{E}(\tilde{\pi} 1_A, S_m) + \text{pen}(m) \right\} + C' \frac{n}{n}$$
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 $\pi$ is regular, we can state the following corollary:

**Corollary 5.** We suppose that the restriction of $\pi$ to $A$ belongs to the anisotropic Besov space $B^{\alpha}_{2,\infty}(A)$ with regularity $\alpha = (\alpha_1, \alpha_2)$ 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 $\alpha_1 - 1$). Then, under the assumptions of Theorem 4,

$$E \|\pi - \hat{\pi}^*\|^2 = O(n^{-\frac{2\alpha}{m+2}}).$$

where $\hat{\alpha}$ is the harmonic mean of $\alpha_1$ and $\alpha_2$.

The same rate of convergence is then achieved with the $L^2$ norm instead of the empirical norm. The procedure allows to adapt automatically the two dimensions of the projection spaces to the regularities $\alpha_1$ and $\alpha_2$ of the transition density $\pi$. If $\alpha_1 = 1$ we recognize the rate $n^{-\frac{4\alpha}{m+2}}$ 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 $\hat{\alpha} = \alpha$ and our result is thus an improvement of the one of Clémençon [11], whose procedure achieves only the rate $(\log(n)/n)^{\frac{2\alpha}{m+2}}$ 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 $\pi$ is regular enough.

Actually, in the case $\alpha_1 = \alpha_2$, an estimation with isotropic spaces ($D_{m1} = D_{m2}$) is preferable. Indeed, in this framework, the models are nested and so we can consider spaces with larger dimension ($D_m^2 \leq n$ instead of $D_m^2 \leq n^{2/3}$). 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

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

and $E_\pi$ the expectation corresponding to the distribution of $X_1, \ldots, X_n$ if the true transition density of the Markov chain is $\pi$ and the initial distribution is the stationary distribution.

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

$$\inf \sup_{\hat{\pi}_n} E_\pi \|\hat{\pi}_n - \pi\|_A^2 \geq Cn^{-\frac{2\alpha}{m+2}}$$

where the infimum is taken over all estimators $\hat{\pi}_n$ of $\pi$ 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 $\pi$, without needing to know $\alpha$.

### 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 $\gamma_n(\hat{\pi}_m) = \text{Tr}(A_m G_m A_m - 2^T Z_m A_m)$,
- find $\hat{m}$ such that $\gamma_n(\hat{\pi}_m) + \text{pen}(m)$ 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 $\text{pen}(m) = 0.5 D_m D_m^2 / n$.

We consider three Markov chains:

- An autoregressive process defined by $X_{n+1} = a X_n + b + \varepsilon_{n+1}$, where the $\varepsilon_n$ are i.i.d. centered Gaussian random variables with variance $\sigma^2$. 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 $(\xi_1, \xi_2, \xi_3)$ whose components are i.i.d. processes satisfying, for $j = 1, 2, 3, \xi_{n+1}^j = a \xi_n^j + \beta \varepsilon_n^j$ where $\varepsilon_n^j$ are i.i.d. standard Gaussian. This process is studied in detail in [10]. Its transition density is $\pi(x,y) = \varphi_{1/2}(y - \sqrt{ax^2/\beta^2}) I_{1/2}(\frac{ax}{\beta^2} \sqrt{\frac{y}{ax}})$ where $I_{1/2}$ is the Bessel function with index $1/2$. The stationary density of this chain is $f(x) = \varphi_{1/2}(\xi_n + \beta \sqrt{\rho^2/(1-a^2)})$, 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 $\varepsilon_{n+1}$ are i.i.d. standard Gaussian. We verify that the condition (2) is satisfied. Here the transition density is $\pi(x,y) = \varphi_{1/2}(y - \sin(x) \cos(x) + 3) \frac{1}{\cos(x) + 3}$ 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 $z = \hat{\pi}(x,y)$. 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 $z = \hat{\pi}(4.6, y)$ and the curves $z = \pi(x, 5)$ versus $z = \hat{\pi}(x, 5)$. 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, $n = 1000$.](image-url)
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 $\text{pen}(m) = 0.5 D_{m_1} D_{m_2} / n$, averaged over $N = 200$ samples

<table>
<thead>
<tr>
<th>Law</th>
<th>$n$</th>
<th>50</th>
<th>100</th>
<th>250</th>
<th>500</th>
<th>1000</th>
<th>basis</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td></td>
<td>0.067</td>
<td>0.055</td>
<td>0.043</td>
<td>0.038</td>
<td>0.033</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.096</td>
<td>0.081</td>
<td>0.063</td>
<td>0.054</td>
<td>0.045</td>
<td>T</td>
</tr>
<tr>
<td>$\sqrt{\text{CTR}}$</td>
<td></td>
<td>0.026</td>
<td>0.023</td>
<td>0.019</td>
<td>0.016</td>
<td>0.014</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.019</td>
<td>0.015</td>
<td>0.009</td>
<td>0.007</td>
<td>0.006</td>
<td>T</td>
</tr>
<tr>
<td>ARCH</td>
<td></td>
<td>0.031</td>
<td>0.027</td>
<td>0.016</td>
<td>0.015</td>
<td>0.014</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.020</td>
<td>0.012</td>
<td>0.008</td>
<td>0.007</td>
<td>0.007</td>
<td>T</td>
</tr>
</tbody>
</table>

H: histogram basis, T: trigonometric basis.

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

<table>
<thead>
<tr>
<th>Law</th>
<th>$n$</th>
<th>50</th>
<th>100</th>
<th>250</th>
<th>500</th>
<th>1000</th>
<th>basis</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td></td>
<td>0.242</td>
<td>0.189</td>
<td>0.132</td>
<td>0.109</td>
<td>0.085</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.438</td>
<td>0.357</td>
<td>0.253</td>
<td>0.213</td>
<td>0.180</td>
<td>T</td>
</tr>
<tr>
<td>$\sqrt{\text{CTR}}$</td>
<td></td>
<td>0.152</td>
<td>0.130</td>
<td>0.094</td>
<td>0.066</td>
<td>0.054</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.152</td>
<td>0.123</td>
<td>0.072</td>
<td>0.052</td>
<td>0.046</td>
<td>T</td>
</tr>
<tr>
<td>ARCH</td>
<td></td>
<td>0.367</td>
<td>0.303</td>
<td>0.168</td>
<td>0.156</td>
<td>0.144</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.249</td>
<td>0.137</td>
<td>0.096</td>
<td>0.092</td>
<td>0.090</td>
<td>T</td>
</tr>
</tbody>
</table>

H: histogram basis, T: trigonometric basis.

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

<table>
<thead>
<tr>
<th>Law</th>
<th>$n$</th>
<th>50</th>
<th>100</th>
<th>250</th>
<th>500</th>
<th>1000</th>
<th>basis</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td></td>
<td>0.052</td>
<td>0.038</td>
<td>0.026</td>
<td>0.020</td>
<td>0.015</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.081</td>
<td>0.069</td>
<td>0.046</td>
<td>0.037</td>
<td>0.031</td>
<td>T</td>
</tr>
<tr>
<td>$\sqrt{\text{CTR}}$</td>
<td></td>
<td>0.016</td>
<td>0.014</td>
<td>0.010</td>
<td>0.006</td>
<td>0.004</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.018</td>
<td>0.012</td>
<td>0.008</td>
<td>0.006</td>
<td>0.004</td>
<td>T</td>
</tr>
</tbody>
</table>

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 $\pi$ 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 $\psi_{k_0}^m(y)$,

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

Then, we sum over $k_0$ in $K_m$:

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

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

$$\sum_{i=1}^n \left[ t(X_i, y) - \sum_{k_0 \in K_m} \psi_{k_0}^m(X_{i+1}) \psi_{k_0}^m(y) \right] \sum_{k \in K_m} \sum_{j_0 \in J_m} a_{j_0,k}^i \psi_k^m(X_i) \psi_{j_0}^m(y) = 0$$

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 $(t(X_i, y) - \sum_{k \in K_m} \psi_k^m(X_{i+1}) \psi_k^m(y))_{1 \leq i \leq n}$ 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 $\rho$ a real larger than 1, let

$$\Omega_\rho = \left\{ \forall t \in S \mid \| t \|^2_2 \leq \rho \| t \|^2_n \right\}.$$

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

$$\begin{cases} 0 < c < \frac{1}{6}, \\ \gamma c > \frac{7}{3}. \end{cases}$$

We set in this case $q_n = \frac{1}{12} \lfloor n^c \rfloor$. In the case of a geometrical mixing, we set $q_n = \frac{1}{2} [c \ln(n)]$ 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 $i = 1, \ldots, n/2$, $U_i = (X_{2i-1}, X_{2i})$.

Let

$$\begin{cases} A_l = (U_{2lq_n+1}, \ldots, U_{(2l+1)q_n}), \quad l = 0, \ldots, p_n - 1, \\ B_l = (U_{(2l+1)q_n+1}, \ldots, U_{(2l+2)q_n}), \quad l = 0, \ldots, p_n - 1. \end{cases}$$
We use now the mixing assumption A5. As in Viennet [27] we can build a sequence \((A_r^t)\) such that
\[
\begin{align*}
A_t & \text{ and } A_t^\circ \text{ have the same distribution}, \\
A_t^r & \text{ and } A_t^r \text{ are independent if } l \neq l', \\
P(A_t \neq A_t^l) & \leq \beta_{2q_n}.
\end{align*}
\]

In the same way, we build \((B_r^t)\) and we define for any \(l \in \{0, \ldots, p_n - 1\},
\]
\[
A_t^r = (U_{2q_n + 1}^*, \ldots, U_{(2l + 1)q_n}^*), \quad B_t^r = (U_{(2l + 1)q_n + 1}^*, \ldots, U_{(2l + 2)q_n}^*)
\]
so that the sequence \((U_{1}^*, \ldots, U_{n/2}^*)\) and then the sequence \((X_{1}^*, \ldots, X_n^*)\) are well defined.

Let now \(V_i = (X_{2i}, X_{2i + 1})\) for \(i = 1, \ldots, n/2\) and
\[
\begin{align*}
C_l & = (V_{2lq_n + 1}, \ldots, V_{(2l + 1)q_n}), \quad l = 0, \ldots, p_n - 1, \\
D_l & = (V_{(2l + 1)q_n + 1}, \ldots, V_{(2l + 2)q_n}), \quad l = 0, \ldots, p_n - 1.
\end{align*}
\]
We can build \((V_{1}^{**}, \ldots, V_{n/2}^{**})\) and then \((X_{2}^{**}, \ldots, X_{n + 1}^{**})\) such that
\[
\begin{align*}
C_l & \text{ and } C_l^{**} \text{ have the same distribution}, \\
C_l^{**} & \text{ and } C_l^{**} \text{ are independent if } l \neq l', \\
P(C_l \neq C_l^{**}) & \leq \beta_{2q_n}.
\end{align*}
\]

We put \(X_{n + 1}^{*} = X_{n + 1}\) and \(X_{1}^{**} = X_1\). Now let
\[
\Omega^* = \{\forall i \; X_i = X_i^{**} = X_i^*\} \quad \text{and} \quad \Omega^*_\rho = \Omega^*_\rho \cap \Omega^*.
\]
We denote by \(\pi_m\) the orthogonal projection of \(\pi\) on \(S_m\). Now,
\[
\mathbb{E}\|\tilde{\pi} - \pi 1_A\|^2_{H^n} = \mathbb{E}\left(\|\tilde{\pi} - \pi 1_A\|^2_{H^n} + \Omega^*_{\rho}\right) + \mathbb{E}\left(\|\tilde{\pi} - \pi 1_A\|^2_{H^n} + \Omega^*_{\rho}\right)
\]
To bound the first term, we observe that for all \(s, t\)
\[
\gamma_n(t) - \gamma_n(s) = |t - s| \leq \|\pi\|^2_{H^n} - \|\pi\|^2_{H^n} - 2Z_n(t - s)
\]
where
\[
Z_n(t) = \frac{1}{n} \sum_{i=1}^{n} \left\{ t(X_i, X_{i+1}) - \int_{\mathbb{R}} t(X_i, y)\pi(X_i, y)\,dy \right\}.
\]
Since \(\|t - \pi\|^2_{H^n} = \|t - \pi 1_A\|^2_{H^n} + \|\pi 1_{A^c}\|^2_{H^n}\), we can write
\[
\gamma_n(t) - \gamma_n(s) = \|t - \pi 1_A\|^2_{H^n} - \|s - \pi 1_A\|^2_{H^n} - 2Z_n(t - s).
\]
The definition of \(\tilde{m}\) gives, for some fixed \(m \in \mathcal{M}_n\),
\[
\gamma_n(\tilde{\pi}) + \text{pen}(\tilde{m}) \leq \gamma_n(\pi_m) + \text{pen}(m).
\]
And then
\[
\|\tilde{\pi} - \pi 1_A\|_{H^n}^2 \leq \|\pi_m - \pi 1_A\|_{H^n}^2 + 2Z_n(\tilde{\pi} - \pi_m) + \text{pen}(m) - \text{pen}(\tilde{m})
\]
where, for all \(m', B_f(m') = \{t \in S_m + S_{m'} : \|t\|_f = 1\}\). Let \(\theta\) a real larger than \(2\rho\) and \(p(\cdot, \cdot)\) a function such that
\[
\|\pi_m - \pi_m\|_{H^n}^2 \leq \|\pi_m - \pi_m\|_{H^n}^2 + \frac{1}{\theta}\|\tilde{\pi} - \pi_m\|_{H^n}^2 + 2\text{pen}(m)
\]
\[
+ \theta \sum_{m' \in \mathcal{M}_n} \left[ \sup_{t \in B_f(m')} Z_n^2(t) - p(m, m') \right].
\]
But \(\|\tilde{\pi} - \pi_m\|_{H^n}^2, \|\tilde{\pi} - \pi_m\|_{H^n}^2, \|\tilde{\pi} - \pi_m\|_{H^n}^2 \leq 2\rho\|\pi 1_A\|_{H^n}^2 + 2\rho\|\pi 1_A - \pi_m\|_{H^n}^2\).
Then, inequality (12) becomes
\[ \|\tilde{\pi} - \pi \mathbb{1}_A \|_n^2 \leq \frac{1}{n \Omega_n^2} \left( 1 - \frac{2\rho}{\theta} \right) \|\pi_m - \pi \mathbb{1}_A\|_n^2 + 2\text{pen}(m) \]
\[ + \theta \sum_{m' \in M_n} \left[ \sup_{t \in B_f(m')} Z_n^2(t) - p(m, m') \right]_{+} \Omega_n^2 \]
so
\[ \mathbb{E}(\|\tilde{\pi} - \pi \mathbb{1}_A \|_n^2) \leq \frac{\theta + 2\rho}{\theta - 2\rho} \mathbb{E}(\|\pi \mathbb{1}_A - \pi_m\|_n^2) + \frac{2\theta}{\theta - 2\rho} \text{pen}(m) \]
\[ + \frac{\theta^2}{\theta - 2\rho} \sum_{m' \in M_n} \mathbb{E}(\left[ \sup_{t \in B_f(m')} Z_n^2(t) - p(m, m') \right]_{+} \Omega_n^2). \]

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
\[ \sum_{m' \in M_n} \mathbb{E}(\left[ \sup_{t \in B_f(m')} Z_n^2(t) - p(m, m') \right]_{+}) \leq C_1 \frac{1}{n}. \]
(14)

Then, with \( \theta = 3\rho \), inequalities (13) and (14) yield
\[ \mathbb{E}(\|\tilde{\pi} - \pi \mathbb{1}_A \|_n^2) \leq 5\|\pi \|_{\infty} \|\pi_m - \pi \mathbb{1}_A\|_n^2 + 6\text{pen}(m) + \frac{9\rho C_1}{n}. \]
(15)

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

To bound the second term in (11), we recall (see Section 3) that \( (\tilde{\pi}_m(X_i, y))_{1 \leq i \leq n} \) is the orthogonal projection of \( \{\epsilon_k(X_{i+1}) \psi_k(y)\}_{1 \leq i \leq n} \) on \( W = \{(t(X_i, y))_{1 \leq i \leq n} : t \in S_m\} \)
where \( \psi_k = \psi_k^\hat{m} \). Thus, since \( P_W \) denotes the orthogonal projection on \( W \), using (7), (8)
\[ (\tilde{\pi}_m(X_i, y))_{1 \leq i \leq n} = P_W \left( \left( \sum_k \psi_k(X_{i+1}) \psi_k(y) \right)_{1 \leq i \leq n} \right) \]
\[ = P_W \left( \left( \sum_k \pi_k(X_i) \psi_k(y) \right)_{1 \leq i \leq n} \right) + P_W \left( \left( \sum_k \epsilon_{i,k} \psi_k(y) \right)_{1 \leq i \leq n} \right) \]
\[ = P_W (\pi \mathbb{1}_A(X_i, y))_{1 \leq i \leq n} + P_W \left( \left( \sum_k \epsilon_{i,k} \psi_k(y) \right)_{1 \leq i \leq n} \right). \]

We denote by \( \| \cdot \|_{\mathbb{R}^n} \) the Euclidean norm in \( \mathbb{R}^n \), by \( X \) the vector \( (X_i)_{1 \leq i \leq n} \) and by \( \epsilon_k \) the vector \( (\epsilon_{i,k})_{1 \leq i \leq n} \). Thus
\[ \|\pi \mathbb{1}_A - \tilde{\pi}_m\|_n^2 = \frac{1}{n} \int \|\pi \mathbb{1}_A(X, y) - P_W (\pi \mathbb{1}_A(X, y)) - P_W \left( \sum_k \epsilon_k \psi_k(y) \right) \|_{\mathbb{R}^n}^2 dy \]
\[ = \frac{1}{n} \int \|\pi \mathbb{1}_A(X, y) - P_W (\pi \mathbb{1}_A(X, y))\|_{\mathbb{R}^n}^2 dy + \frac{1}{n} \int P_W \left( \sum_k \epsilon_k \psi_k(y) \right) \|_{\mathbb{R}^n}^2 dy \]
Lemma 9.

7.3 Proof of Corollary 3

But Assumption M2 implies \( \| \sum_{k \in K_m} \psi_k^2 \| \leq \phi_2 D_{\hat{m}_2} \). So, using (8),

\[
\epsilon_{i,k}^2 \leq 2 \psi_k^2 (X_{i+1}) + 2 E[\psi_k (X_{i+1})|X_i]^2
\]

and

\[
\sum_k \epsilon_{i,k}^2 \leq 2 \sum_k \psi_k^2 (X_{i+1}) + 2 E \left[ \sum_k \psi_k^2 (X_{i+1})|X_i \right] \leq 4 \phi_2 D_{\hat{m}_2}.
\]

Thus we obtain

\[
\| \pi \mathbb{1}_A - \hat{\pi}_m \|^2 \leq \| \pi \|_\infty + 4 \phi_2 D_{\hat{m}_2} \leq \| \pi \|_\infty + 4 \phi_2 n^{1/3}
\]

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

We now remark that \( P(\Omega_{\hat{m}}^c) = P(\Omega_{\hat{m}}^c) + P(\Omega_{\hat{m}}^c \cap \Omega^*) \). In the geometric case \( \beta_{2q_n} \leq e^{-c n \log(n)} \leq n^{-c} \) and in the other case \( \beta_{2q_n} \leq (2q_n)^{-c} \). Then

\[
P(\Omega_{\hat{m}}^c) \leq 4 \rho_n \beta_{2q_n} \leq n^{-1-cy}.
\]

But we have choose \( c \) such that \( cy > 7/3 \) and so \( P(\Omega_{\hat{m}}^c) \leq n^{-4/3} \). 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

\[
P(\Omega_{\hat{m}}^c \cap \Omega^*) \leq C_2 n^{-7/3}.
\]

This proposition implies that \( E(\| \pi \mathbb{1}_A - \hat{\pi}_m \|^2 \mathbb{1}_{\Omega_{\hat{m}}^c}) \leq C_3/n \).

Now we use (15) and we observe that this inequality holds for all \( m \in M_n \), so

\[
E\| \pi - \pi \mathbb{1}_A \|_n^2 \leq C \inf_{m \in M_n} (\| \pi \mathbb{1}_A - \pi_m \|^2 + \text{pen}(m)) + \frac{C_4}{n}
\]

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 \( \pi_A \) belong to \( B_{2,\infty}^A \). We consider that \( S_{m_1}^i \) is one of the following spaces on A:

- a space of piecewise polynomials of degrees bounded by \( s_j > \alpha_i - 1 \) \( (i = 1, 2) \) based on a partition with rectangles of vertices \( 1/D_{m_1} \) and \( 1/D_{m_2} \),
- a linear span of \( \{ \phi_k \psi_k \lambda, \lambda \in \bigcup_{0}^{m_1} A(j), \mu \in \bigcup_{m_2} M(k) \} \) where \( \{ \phi_k \} \) and \( \{ \psi_k \} \) are orthonormal wavelet bases of respective regularities \( s_j > \alpha_i - 1 \) and \( s_2 > \alpha_2 - 1 \) (here \( D_{m_1} = 2^{m_1}, i = 1, 2 \)),
- 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 \( \pi_m' \) be the orthogonal projection of \( \pi_A \) on \( S_m' \). Then, there exists a positive constant \( C_0 \) such that
\[
\left( \int_A |\pi_A - \pi_m'|^2 \right)^{1/2} \leq C_0 \left[ D_{m_1}^{-\alpha_1} + D_{m_2}^{-\alpha_2} \right].
\]

**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
\[
\left( \int_A |\pi_A - \pi_m'|^2 \right)^{1/2} \leq C \left[ \omega_{31+1.1}(\pi, D_{m_1}) + \omega_{32+1.2}(\pi, D_{m_2}) \right].
\]
The definition of \( B_{2,\infty}^* \) (A) implies \( \left( \int_A |\pi_A - \pi_m'|^2 \right)^{1/2} \leq C_0 \left[ D_{m_1}^{-\alpha_1} + D_{m_2}^{-\alpha_2} \right] \). \( \square \)

If we choose for \( S_m' \) the set of the restrictions to \( A \) of the functions of \( S_m \) and \( \pi_A \) the restriction of \( \pi \) to \( A \), we can apply Lemma 9. But \( \pi_m' \) is also the restriction to \( A \) of \( \pi_m \) so that
\[
\| \pi\mathbb{1}_A - \pi_m\| \leq C_0 \left[ D_{m_1}^{-\alpha_1} + D_{m_2}^{-\alpha_2} \right].
\]

According to Theorem 2
\[
\mathbb{E}\|\tilde{\pi} - \pi\mathbb{1}_A\|^2 \leq C\inf_{\mathbb{M}_A} \left\{ D_{m_1}^{-2\alpha_1} + D_{m_2}^{-2\alpha_2} + \frac{D_{m_1}D_{m_2}}{n} \right\}.
\]

In particular, if \( m^* \) is such that \( D_{m_1}^* = \lfloor n^{\alpha_2/2(\alpha_1 + \alpha_2)} \rfloor \) and \( D_{m_2}^* = \lfloor (D_{m_1}^*)^{\alpha_1/\alpha_2} \rfloor \) then
\[
\mathbb{E}\|\tilde{\pi} - \pi\mathbb{1}_A\|^2 \leq C m^* \left\{ D_{m_1}^{-2\alpha_1} + \frac{D_{m_2}^{1+\alpha_1/\alpha_2}}{n} \right\} = O(n^{-\frac{2(\alpha_1 + \alpha_2)}{\alpha_2(\alpha_1 + \alpha_2)}}).
\]

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

The condition \( D_{m_1} \leq n^{1/3} \) allows this choice of \( m \) only if \( \frac{\alpha_2}{\alpha_1 + 2\alpha_2} < \frac{1}{3} \) 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
\[
\begin{align*}
B_1 &= \mathbb{E}(\|\tilde{\pi}^* - \pi\mathbb{1}_A\|^2)_{\|\tilde{\pi}^*\| \leq k_1} , \\
B_2 &= \mathbb{E}(\|\tilde{\pi}^* - \pi\mathbb{1}_A\|^2)_{\|\tilde{\pi}^*\| > k_1} , \\
B_3 &= \mathbb{E}(\|\tilde{\pi}^* - \pi\mathbb{1}_A\|^2) .
\end{align*}
\]

To bound the first term, we observe that for all \( m \in \mathbb{M}_A \), on \( \Omega^*_\rho \), \( \|\tilde{\pi} - \pi_m\|^2 \leq f_0^{-1}\rho \|\tilde{\pi} - \pi_m\|^2 \). Then
\[
\|\tilde{\pi} - \pi\mathbb{1}_A\|^2 \leq 2\|\tilde{\pi} - \pi_m\|^2 + 2\|\pi_m - \pi\mathbb{1}_A\|^2 \leq 2f_0^{-1}\rho \|\tilde{\pi} - \pi_m\|^2 + 2\|\pi_m - \pi\mathbb{1}_A\|^2 \leq 2f_0^{-1}\rho \left\{ 2\|\tilde{\pi} - \pi\mathbb{1}_A\|^2 + 2\|\pi_m - \pi\mathbb{1}_A\|^2 \right\} + 2\|\pi_m - \pi\mathbb{1}_A\|^2 .
\]

Thus
\[
B_1 \leq \mathbb{E}(\|\tilde{\pi} - \pi\mathbb{1}_A\|^2) \leq 4f_0^{-1}\rho \mathbb{E}(\|\tilde{\pi} - \pi\mathbb{1}_A\|^2) + (4f_0^{-1}\rho + 2)\|\pi_m - \pi\mathbb{1}_A\|^2 .
\]
But, using (15), we obtain
\[ B_1 \leq (24f_0^{-1}\rho \|f\|_\infty + 2)\|\pi_m - \pi 1_A\|^2 + 24f_0^{-1}\rho \text{pen}(m) + 36f_0^{-1}\rho^2 C_1/n. \]
Since \( \rho = 3/2 \), by setting \( C = \max(36f_0^{-1}\|f\|_\infty + 1, 36f_0^{-1}) \),
\[ B_1 \leq C\left\{ \|\pi_m - \pi 1_A\|^2 + \text{pen}(m) \right\} + \frac{81f_0^{-1}C_1}{n} \]
for all \( m \in \mathcal{M}_n \).

Next, the definition of \( \tilde{\pi}^* \) and the Markov inequality provide
\[ B_2 \leq \mathbb{E}\left(\|\pi 1_A\|^2 1_{\Omega_{\rho}^* 1_{|\tilde{\pi}| > k_n}}\right) \leq \pi \pi^2 \frac{\mathbb{E}(\|\tilde{\pi}\|^2 1_{\Omega_{\rho}^*})}{k_n^2}. \] (17)
But \( \|\tilde{\pi}\|^2 1_{\Omega_{\rho}^*} \leq \rho f_0^{-1}\|\tilde{\pi}\|^2 \leq 2\rho f_0^{-1}(\|\tilde{\pi} - \pi 1_A\|^2 + \|\pi 1_A\|^2) \). Now we use (16) to state
\[ \|\tilde{\pi}\|^2 1_{\Omega_{\rho}^*} \leq 2\rho f_0^{-1}(\|\pi\|_\infty + 4\phi_2 n^{1/3} + \|\pi 1_A\|^2) \leq 2\rho f_0^{-1}(2\|\pi\|_\infty + 4\phi_2 n^{1/3}) \]
Then, since \( k_n = n^{2/3} \), (17) becomes
\[ B_2 \leq \pi \frac{2\rho f_0^{-1}(2\|\pi\|_\infty + 4\phi_2 n^{1/3})}{k_n^2} \leq 4\rho f_0^{-1}\|\pi\|^2 \left(\frac{\|\pi\|_\infty}{n^{4/3}} + \frac{2\phi_2}{n}\right). \]
Lastly
\[ B_3 \leq \mathbb{E}(2(\|\pi^*\|^2 + \|\pi 1_A\|^2) 1_{\Omega_{\rho}^*}) \leq 2(k_n^2 + \|\pi\|^2) P(\Omega_{\rho}^{sc}). \]
We now remark that \( P(\Omega_{\rho}^{sc}) = P(\Omega_{\rho}^{sc}) + P(\Omega_{\rho}^{c} \cap \Omega^*). \) In the geometric case \( \beta_{2q_n} \leq e^{-\gamma c \ln(n)} \leq n^{-\gamma c} \) and in the other case \( \beta_{2q_n} \leq (2q_n)^{-\gamma} \leq n^{-\gamma c}. \) Then
\[ P(\Omega_{\rho}^{sc}) \leq 4p_n \beta_{2q_n} \leq n^{-1-c\gamma}. \]
But, if \( \gamma > 20 \) in the arithmetic case, we can choose \( c \) such that \( c\gamma > 10/3 \) and so \( P(\Omega_{\rho}^{sc}) \leq n^{-7/3}. \) Then, using Proposition 8,
\[ B_3 \leq 2(n^{4/3} + \|\pi\|^2) \frac{1 + C_2}{n^{7/3}} \leq \frac{2(C_2 + 1)(1 + \|\pi\|^2)}{n}. \]

7.5. Proof of Theorem 6

Let \( \psi \) 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
\[ \psi_{J,K}(x, y) = 2^{(j_1+j_2)/2}\psi(2^{j_1}x - k_1)\psi(2^{j_2}y - k_2). \]
Let \( \pi_0(x, y) = c_0 1_B(y) \) with \( B \) a compact set such that \( A \subset B \times B \) and \( |B| \geq 2|A|^{1/2}/L \), and \( c_0 = |B|^{-1} \). So \( \pi_0 \) is a transition density with \( \|\pi_0\|_{L^\infty(A)} \leq L/2 \). Now we set \( R_J \) the maximal subset of \( \mathbb{Z}^2 \) such that
\[ \text{Supp}(\psi_{J,K}) \subset A \quad \forall K \in R_J, \quad \text{Supp}(\psi_{J,K}) \cap \text{Supp}(\psi_{J,K'}) = \emptyset \quad \text{if} \ K \neq K'. \]
The cardinal of \( R_J \) is \( |R_J| = c\pi_{j_1+j_2} \), with \( c \) a positive constant which depends only on \( A \) and the support of \( \psi \). Let, for all \( \epsilon = (\epsilon_k) \in (\epsilon)_{|R_J|} = (1, 1)^{|R_J|} \),
\[ \pi_\epsilon = \pi_0 + \frac{1}{\sqrt{n}} \sum_{K \in R_J} \epsilon_k \psi_{J,K}. \]
Let us denote by $G$ the set of all such $\pi_\epsilon$. Since $\int \psi = 0$ and $\pi_0$ is a transition density, for all $x$ in $\mathbb{R}$,
\[
\int \pi_\epsilon(x, y) \, dy = 1.
\]
Additionally $\pi_\epsilon(x, y) = \pi_0(x, y) \geq 0$ if $(x, y) \notin A$, and if $(x, y) \in A$: $\pi_\epsilon \geq c_0 - 2^{(j_1 + j_2)/2} \|\psi\|_\infty^2 / \sqrt{n}$ and then $\pi_\epsilon(x, y) \geq c_0 / 2 > 0$ as soon as
\[
\left( \frac{2^{j_1 + j_2} \sqrt{n}}{c_0} \right)^{1/2} \leq \frac{c_0}{2 \|\psi\|_\infty^2}.
\]
Thus, if (18) holds, $\pi_\epsilon(x, y) \geq (c_0 / 2) \mathbb{1}_B(y)$ 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 $\pi_\epsilon \in \mathcal{B}$, we still have to compute $\|\pi_\epsilon\|_{B_{2, \infty}^0(A)}$. Hochmuth [19] proves that for $\psi$ smooth enough $\|\sum_{K \in R_j} \varepsilon K \psi_{JK} \|_{B_{2, \infty}^0(A)} \leq (2^{j_1 \alpha_1} + 2^{j_2 \alpha_2}) \|\sum_{K \in R_j} \varepsilon K \psi_{JK}\|_A$. Since
\[
\left\| \sum_{K \in R_j} \varepsilon K \psi_{JK} \right\|^2_A = \sum_{K \in R_j} |\varepsilon K|^2 = c_2^{j_1 + j_2},
\]
then
\[
\|\pi_\epsilon\|_{B_{2, \infty}^0(A)} \leq \frac{L}{2} + \frac{2^{j_1 \alpha_1} + 2^{j_2 \alpha_2}}{\sqrt{n}} c_2^{j_1 + j_2}/2.
\]
From now on, we suppose that Condition C is verified where
\[
\text{Condition C: } \frac{(2^{j_1 \alpha_1} + 2^{j_2 \alpha_2})2^{(j_1 + j_2)/2}}{\sqrt{n}} \leq \frac{L}{2c_2^{1/2}}.
\]
It entails in particular that (18) holds if $j_1$ and $j_2$ are great enough. Then for all $\epsilon$, $\pi_\epsilon \in \mathcal{B}$. We now use Lemma 10.2 p. 160 in Härdle et al. [17]. The likelihood ratio can be written
\[
A_n(\pi_{\epsilon_k}, \pi_\epsilon) = \prod_{i=1}^n \pi_{\epsilon_k}(X_i, X_{i+1}) / \pi_\epsilon(X_i, X_{i+1}).
\]
Note that $\pi_\epsilon(X_i, X_{i+1}) > 0$ $P_{\pi_\epsilon}$- and $P_{\pi_{\epsilon_k}}$-almost surely (actually the chain “lives” on $B$). Then
\[
\log\left( A_n(\pi_{\epsilon_k}, \pi_\epsilon) \right) = \sum_{i=1}^n \log \left( 1 - \frac{2}{\sqrt{n}} \frac{\varepsilon K \psi_{JK}(X_i, X_{i+1})}{\pi_\epsilon(X_i, X_{i+1})} \right).
\]
We set $U_{JK}(X_i, X_{i+1}) = -\varepsilon K \psi_{JK}(X_i, X_{i+1})/\pi_\epsilon(X_i, X_{i+1})$ so that
\[
\log\left( A_n(\pi_{\epsilon_k}, \pi_\epsilon) \right) = \sum_{i=1}^n \log \left( 1 + \frac{2}{\sqrt{n}} U_{JK}(X_i, X_{i+1}) \right)
\]
\[
= \sum_{i=1}^n \left\{ \theta \left( \frac{2}{\sqrt{n}} U_{JK}(X_i, X_{i+1}) \right) + \frac{2}{\sqrt{n}} U_{JK}(X_i, X_{i+1}) - \frac{2}{n} U_{JK}^2(X_i, X_{i+1}) \right\}
\]
\[
= u_n + v_n - w_n
\]
with $\theta$ the function defined by $\theta(u) = \log(1 + u) - u + u^2/2$. Now we prove the three following assertions

$1^o$ $E_{\pi_\epsilon}(|u_n|) = E_{\pi_\epsilon}(\sum_{i=1}^n \theta\left( \frac{2}{\sqrt{n}} U_{JK}(X_i, X_{i+1}) \right)) \to 0$,

$2^o$ $E_{\pi_\epsilon}(v_n) = E_{\pi_\epsilon}\left( \frac{2}{n} \sum_{i=1}^n U_{JK}^2(X_i, X_{i+1}) \right) \leq 4$,

$3^o$ $E_{\pi_\epsilon}(w_n^2) = E_{\pi_\epsilon}\left( \frac{4}{n} \sum_{i=1}^n U_{JK}^2(X_i, X_{i+1})^2 \right) \leq 8$. 
1°: First we observe that
\[
\left\| \frac{2}{\sqrt{n}} U_{JK} \right\|_{\infty} \leq \frac{2}{\sqrt{n}} \frac{2(j_1 + j_2)^2}{c_0} \| \psi \|_{\infty}^2 = O \left( \frac{2(j_1 + j_2)^2}{\sqrt{n}} \right)
\]
and \(2(j_1 + j_2)/n \to 0\) since Condition C holds. So there exists some integer \(n_0\) such that \(\forall n \geq n_0, \forall x, y,
\[
\left| \theta(2U_{JK}(x, y)/\sqrt{n}) \right| \leq \left| 2U_{JK}(x, y)/\sqrt{n} \right|^3.
\]
But
\[
\left\| \frac{2U_{JK}(x, y)}{\sqrt{n}} \right\|^3 \leq \frac{8}{n} \int f(x) \psi(x, y) \, dx \, dy \leq \frac{8}{n} \frac{2(j_1 + j_2)^2}{c_0} \frac{\| \psi \|_{\infty}^2}{(c_0/2)^2} \int \psi_{JK}(x, y)^2 \, dx \, dy \leq \frac{32 \| \psi \|_{\infty}^2}{c_0 n} \left( \frac{2(j_1 + j_2)}{n} \right)^{1/2}.
\]
Then
\[
E_{\pi_e}[u_n] = \sum_{i=1}^{n} \frac{32 \| \psi \|_{\infty}^2}{c_0 n} \left( \frac{2(j_1 + j_2)}{n} \right)^{1/2} \to 0.
\]

2°: We bound the expectation of \(U_{JK}(X_i, X_{i+1})^2\):
\[
E_{\pi_e}(U_{JK}(X_i, X_{i+1})^2) = \left\| \frac{\psi_{JK}^2(x, y)}{\psi(x, y)} \right\|_{\infty} \leq c_0 \left\| \frac{\psi_{JK}(x, y)}{c_0} \right\|_{\infty} \leq 2.
\]
And then \(E_{\pi_e}(w_n) = E_{\pi_e}(2/n) \sum_{i=1}^{n} U_{JK}(X_i, X_{i+1})^2 \leq 4.

3°: We observe that \(E_{\pi_e}(U_{JK}(X_i, X_{i+1}) | X_i, \ldots, X_i) = 0\) and thus \(\sum_{i=1}^{n} U_{JK}(X_i, X_{i+1})\) is a martingale. A classic property of square integrable martingales involves
\[
E_{\pi_e}(\left[ \sum_{i=1}^{n} U_{JK}(X_i, X_{i+1}) \right]^2) = \sum_{i=1}^{n} E_{\pi_e}[U_{JK}(X_i, X_{i+1})^2].
\]
Thus, using (19), \(E_{\pi_e}(v_n^2) = (4/n) \sum_{i=1}^{n} E_{\pi_e}[U_{JK}(X_i, X_{i+1})^2] \leq 8.

We deduce easily from the three previous assertions 1°, 2° and 3° that there exists \(\lambda > 0\) and \(p_0\) such that \(P_{\pi_e}(\lambda n, \pi_{e\lambda}, \pi_e) > e^{-\lambda} \geq p_0\). Thus, according to Lemma 10.2 in [17],
\[
\max_{\pi \in \mathcal{G}} E_{\pi_e}[\| \hat{\pi}_n - \pi \|_{\Lambda}^2] \geq \frac{\| R |}{\delta} e^{-\lambda} p_0
\]
where \(\delta = \inf_{\pi \neq \pi'} \| \pi - \pi' \|_{\Lambda} / 2 = \| \pi \|_{\Lambda} / 2 = 1 / \sqrt{n}.

Now for all \(n\) we choose \(J = J(n) = (j_1(n), j_2(n)) \) such that
\[
c_1/2 \leq 2 \frac{2}{j} n^{-\frac{a_1 + a_2}{2}} \leq c_1 \quad \text{and} \quad c_2/2 \leq 2 \frac{2}{j} n^{-\frac{a_1 + a_2}{2}} \leq c_2
\]
with \(c_1\) and \(c_2\) such that \(c_1 + c_2 \leq \sqrt{c_1} \leq L/(2c_1^{1/2})\) so that Condition C is satisfied. Moreover, we have
\[
\| R |^2 \geq \frac{cc_1 c_2}{4} n^{-\frac{a_1 + a_2}{2}} - 1 \geq \frac{cc_1 c_2}{4} n^{-\frac{a_1 + a_2}{2}}.
\]
Thus
\[
\max_{\pi \in \mathcal{G}} E_{\pi_e}[\| \hat{\pi}_n - \pi \|_{\Lambda}^2] \geq \frac{cc_1 c_2}{8} \frac{p_0 c_1}{n} \frac{n^{-\frac{a_1 + a_2}{2}}}{n^{-\frac{a_1 + a_2}{2}}}
\]
And then for all estimator
\[
\sup_{\pi \in \mathcal{B}} E_{\pi}[\| \hat{\pi}_n - \pi \|_{\Lambda}^2] \geq C n^{-\frac{a_1 + a_2}{2}}
\]
with \(C = cc_1 c_2 / 8\).
Let
\[
\begin{align*}
\Gamma(t) &= t(X_i, X_{i+1}) - \int t(X_i, y)\pi(X_i, y) \, dy, \\
\Gamma_1^*(t) &= t(X_i^*, X_{i+1}^*) - \int t(X_i^*, y)\pi(X_i^*, y) \, dy, \\
\Gamma_2^*(t) &= t(X_i^*, X_{i+1}^*) - \int t(X_i^*, y)\pi(X_i^*, y) \, dy.
\end{align*}
\]

We now define \( Z_n^*(t) \):
\[
Z_n^*(t) = \frac{1}{n} \sum_{i \text{ odd}} I_i^*(t) + \frac{1}{n} \sum_{i \text{ even}} I_i^*(t).
\]

Let us remark that \( Z_n^*(t) \mathbb{1}_{\Omega^*} = Z_n(t) \mathbb{1}_{\Omega^*} \). Next we split each of these terms:
\[
\begin{align*}
Z_n^*_{1}(t) &= \frac{1}{n} \sum_{l=0}^{p_n-1} 2(2l+1)q_n^{-1} \Gamma_l^*(t), \\
Z_n^*_{2}(t) &= \frac{1}{n} \sum_{l=0}^{p_n-1} 2(2l+2)q_n^{-1} \sum_{i \text{ odd}} \Gamma_i^*(t), \\
Z_n^*_{3}(t) &= \frac{1}{n} \sum_{l=0}^{p_n-1} 2(2l+1)q_n \sum_{i \text{ even}} \Gamma_i^*(t), \\
Z_n^*_{4}(t) &= \frac{1}{n} \sum_{l=0}^{p_n-1} 2(2l+2)q_n \sum_{i \text{ even}} \Gamma_i^*(t).
\end{align*}
\]

We use the following lemma:

**Lemma 10.** (Talagrand [26]) Let \( \mathcal{U}_0, \ldots, \mathcal{U}_{N-1} \) i.i.d. variables and \( (\xi_t)_{t \in B} \) a set of functions.

Let \( G(t) = \frac{1}{N} \sum_{i=0}^{N-1} \xi_t(\mathcal{U}_i) \). We suppose that
\[
\begin{align*}
(1) & \sup_{t \in B} \| \xi_t \|_\infty \leq M_1, \\
(2) & \mathbb{E} \left( \sup_{t \in B} |G(t)| \right) \leq H, \\
(3) & \text{Var} \left[ \xi_t(\mathcal{U}_0) \right] \leq v.
\end{align*}
\]

Then, there exists \( K > 0, K_1 > 0, K_2 > 0 \) such that
\[
\mathbb{E} \left[ \sup_{t \in B} G^2(t) - 10H^2 \right] \leq K \left[ \frac{v}{N} e^{-K_1 \frac{vH^2}{2}} + \frac{M_1^2}{N^2} e^{-K_2 \frac{N \mathbb{E} \xi_t^2}{4}} \right].
\]

Here \( N = p_n, B = B(f(m')) \) and for \( l \in \{0, \ldots, p_n-1\} \), \( \mathcal{U}_l = (X_{4l}^*, \ldots, X_{2(2l+1)}^*) \),
\[
\xi_t(x_1, \ldots, x_{2q_n}) = \frac{1}{q_n} \sum_{i=1}^{2q_n-1} t(x_i, x_{i+1}) - \int t(x_i, y)\pi(x_i, y) \, dy.
\]

Then
\[
G(t) = \frac{1}{p_n} \sum_{l=0}^{p_n-1} \frac{1}{q_n} \sum_{i=4lq_n+1}^{(2l+1)q_n} \Gamma_i^*(t) = 4Z_n^*_{1}(t).
\]

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 \( \max(D_{m_1}, D_{m'_1}) \max(D_{m_2}, D_{m'_2}) \).

\[
\sup_{t \in B} \| \xi_t \|_\infty \leq \sup_{t \in B} \| t \|_\infty \frac{1}{q_n} \sum_{i=1}^{2q_n-1} \left( 1 + \int \pi(x_i, y) \, dy \right)
\leq 2\phi_0 \sqrt{\max(D_{m_1}, D_{m'_1}) \max(D_{m_2}, D_{m'_2})} \| t \| \leq \frac{2\phi_0}{\sqrt{f_0}} n^{1/3}.
\]
Then we set \( M_1 = (2\phi_0/f_0)n^{1/3} \).

(2) Since \( A_0 \) and \( A_0^* \) have the same distribution, \( \zeta_t(U_0) = \frac{1}{q_n} \sum_{i=1, i \text{ odd}}^{2q_n-1} \Gamma_1^*(t) \) has the same distribution than \( \frac{1}{q_n} \sum_{i=1, i \text{ odd}}^{2q_n-1} \Gamma_1^*(t) \). We observe that \( \mathbb{E}(\Gamma_1(t)|X_i) = 0 \) and then for all set \( I \)

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

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

In particular

\[
\text{Var}[\zeta_t(U_0)] = \mathbb{E}\left( \left( \frac{1}{q_n} \sum_{i=1, i \text{ odd}}^{2q_n-1} \Gamma_i(t) \right)^2 \right) = \frac{1}{q_n^2} \sum_{i=1, i \text{ odd}}^{2q_n-1} \mathbb{E}[\Gamma_i^2(t)]
\]

\[
\leq \frac{1}{q_n^2} \sum_{i=1, i \text{ odd}}^{2q_n-1} \mathbb{E}[\Gamma_i^2(X_i, X_{i+1})] \leq \frac{1}{q_n} \|\pi\|_{\infty} \|t\|_j^2.
\]

Then \( v = \|\pi\|_{\infty}/q_n \).

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

\[
\mathbb{E}\left( \sup_{t \in B} G^2(t) \right) \leq \sum_{j,k} \mathbb{E}\left( G^2(\varphi_j \otimes \psi_k) \right)
\]

\[
\leq \sum_{j,k} \frac{1}{p_n^2q_n^2} \mathbb{E}\left( \left( \sum_{l=0}^{p_n-1} \sum_{i=4ql+1, i \text{ odd}}^{2(2l+1)q_n-1} \Gamma_1^*(\varphi_j \otimes \psi_k) \right)^2 \right)
\]

\[
\leq \sum_{j,k} \frac{16}{n^2} \sum_{l=0}^{p_n-1} \mathbb{E}\left( \left( \sum_{i=4ql+1, i \text{ odd}}^{2(2l+1)q_n-1} \Gamma_1^*(\varphi_j \otimes \psi_k) \right)^2 \right)
\]

where we used the independence of the \( A_1^* \). Now we can replace \( \Gamma_1^* \) by \( \Gamma_1 \) in the sum because \( A_1 \) and \( A_1^* \) have the same distribution and we use as previously the martingale property of the \( \Gamma_i \).

\[
\mathbb{E}\left( \sup_{t \in B} G^2(t) \right) \leq \sum_{j,k} \frac{16}{n^2} \sum_{l=0}^{p_n-1} \mathbb{E}\left( \left( \sum_{i=4ql+1, i \text{ odd}}^{2(2l+1)q_n-1} \Gamma_1(\varphi_j \otimes \psi_k) \right)^2 \right)
\]

\[
\leq \sum_{j,k} \frac{16}{n^2} \sum_{l=0}^{p_n-1} \mathbb{E}\left( \Gamma_1^2(\varphi_j \otimes \psi_k) \right)
\]

\[
\leq 4\|\pi\|_{\infty} \|\varphi_j \otimes \psi_k\|_j^2 \leq 4\|\pi\|_{\infty} \frac{D(m,m')}{n}.
\]

Then

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

According to Lemma 10, there exists \( K' > 0, K_1 > 0, K_2' > 0 \) such that

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

But \( q_n \leq n^c \) with \( c < \frac{1}{6} \). So
7.7. Proof of Proposition 8

First we observe that

$$P(\Omega^c \cap \Omega^*) \leq P \left( \sup_{t \in B} |v_n(t^2)| > 1 - \frac{1}{\rho} \right)$$

where

$$v_n(t) = \frac{1}{n} \sum_{i=1}^{n} \int \left[ t(X^*_i, y) - E(t(X^*_i, y)) \right] dy$$

and $B = \{ t \in S, \|t\|_f = 1 \}$.

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

$$v_n(t^2) = \sum_{j,j'} \sum_k a_{j,k} a_{j',k} \tilde{v}_n(\varphi_j \varphi_{j'})$$

where

$$\tilde{v}_n(u) = \frac{1}{n} \sum_{i=1}^{n} \left[ u(X^*_i) - E(u(X^*_i)) \right].$$

Let $b_j = (\sum_k a_{j,k}^2)^{1/2}$, then $|v_n(t^2)| \leq \sum_{j,j'} b_j b_{j'} |\tilde{v}_n(\varphi_j \varphi_{j'})|$ and, if $t \in B$, $\sum_j b_j^2 = \sum_j \sum_k a_{j,k}^2 = \|t\|^2 \leq f_0^{-1}$.

Thus

$$\sup_{t \in B} |v_n(t^2)| \leq f_0^{-1} \sup_{j,l} \sum_{j,l} b_j b_l |\tilde{v}_n(\varphi_j \varphi_{l})|.$$

**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 $(A_{j,l})$

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

and $L(\varphi) = \max[\bar{\rho}(V), \bar{\rho}(B)]$. Then, if M2 is satisfied, $L(\varphi) \leq \rho \Delta^2 n$.

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

Let

$$x = f_0^2 (1 - 1/\rho)^2 \frac{40}{40 \|f\|_\infty L(\varphi)}$$

and $\Delta = \{ \forall j \forall l \left| \tilde{v}_n(\varphi_j \varphi_l) \right| \leq 4 \left[ B_{j,l} x + V_{j,l} \sqrt{2 \|f\|_\infty x} \right] \}$.

On $\Delta$:

$$\sup_{t \in B} |v_n(t^2)| \leq 4 f_0^{-1} \sup_{\sum b_j^2 = 1} \sum_{j,l} b_j b_l \left[ B_{j,l} x + V_{j,l} \sqrt{2 \|f\|_\infty x} \right]$$

$$\leq 4 f_0^{-1} \left[ \bar{\rho}(B) x + \bar{\rho}(V) \sqrt{2 \|f\|_\infty x} \right]$$

$$\leq \left( 1 - \frac{1}{\rho} \right) \left[ \frac{f_0 (1 - 1/\rho) \bar{\rho}(B)}{10 \|f\|_\infty L(\varphi)} + \frac{2}{\sqrt{5}} \left( \frac{\bar{\rho}(V)^2}{L(\varphi)} \right)^{1/2} \right]$$

$$\leq \left( 1 - \frac{1}{\rho} \right) \left[ \frac{1}{10} + \frac{2}{\sqrt{5}} \right] \leq \left( 1 - \frac{1}{\rho} \right).$$
Then $P(\sup_{t \in \mathbb{B}} |v_n(t^2)| > 1 - 1/\rho) \leq P(\Delta^c)$. But $\tilde{v}_n(u) = 2\tilde{v}_{n,1}(u) + 2\tilde{v}_{n,2}(u)$ with

$$\tilde{v}_{n,r}(u) = \frac{1}{p_n} \sum_{i=0}^{p_n-1} Y_{i,r}(u), \quad r = 1, 2,$$

with

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

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

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

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

$$\begin{align*}
\mathbb{E}|Y_{i,1}(u)|^m &\leq (2\|u\|_{\infty})^{m-2} \mathbb{E}|u(X_1) - \mathbb{E}(u(X_1))|^2 \leq (2\|u\|_{\infty})^{m-2} \int u^2(x) f(x) \, dx \\
&\leq 2^{m-2} (\|u\|_{\infty})^{m-2} \left( \sqrt{\|f\|_{\infty}} \|u\| \right)^2.
\end{align*}$$

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

$$P(|\tilde{v}_{n,r}(\varphi_j \varphi_l)| \geq B_{j,l}x + V_{j,l}\sqrt{2\|f\|_{\infty}x}) \leq 2e^{-pnx}.$$

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

$$P(\Omega^c_\rho \cap \Omega^*) \leq 4D_n^2 \exp \left\{ -\frac{p_n f_0^2 (1 - 1/\rho)^2}{40\|f\|_{\infty} L(\varphi)} \right\} \leq 4n^{2/3} \exp \left\{ -\frac{f_0^2 (1 - 1/\rho)^2}{160\|f\|_{\infty}} \frac{n}{q_n L(\varphi)} \right\}.$$

But $L(\varphi) \leq \phi_1 D_n^2 \leq \phi_1 n^{2/3}$ and $q_n \leq n^{1/6}$ so

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

**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:

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

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

$$\ln n \leq D_{m^*} = D_{m_2^*} \leq n^{1/6}.$$

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

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

**Remark 13.** The condition on the regularity of $\pi$ 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 $\|\pi 1_A\|_{\infty}$ and we introduce the following set:

$$A = \left\{ \left| \frac{\|\hat{\pi}\|_{\infty}}{\|\pi 1_A\|_{\infty}} - 1 \right| \leq \frac{1}{2} \right\}.$$

As previously, we decompose the space:

$$\mathbb{E}\|\pi - \pi 1_A\|_{n}^2 = \mathbb{E}\left( \|\pi - \pi 1_A\|_{n}^2 \mathbb{1}_{\Omega^c} \cap A \right) + \mathbb{E}\left( \|\pi - \pi 1_A\|_{n}^2 \mathbb{1}_{\Omega^c} \cap \Lambda^c \right) + \mathbb{E}\left( \|\pi - \pi 1_A\|_{n}^2 \mathbb{1}_{\Omega^c} \cap \Lambda^c \right).$$

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

$$\theta p(m, m') \leq \overline{\text{pen}}(m) + \overline{\text{pen}}(m')$$

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

$$\theta p(m, m') = 10\theta \|\pi\|_{\infty} \frac{D(m, m')}{n} \leq 20\theta \|\hat{\pi}\|_{\infty} \frac{D(m, m')}{n} \leq 20\theta \|\hat{\pi}\|_{\infty} \frac{D_{m_1} D_{m_2}}{n} + 20\theta \|\hat{\pi}\|_{\infty} \frac{D_{m_1} D_{m_2}}{n}.$$

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

Now, inequality (16) gives

$$\mathbb{E}\left( \|\pi 1_A - \pi_{m^*}\|_{n}^2 \mathbb{1}_{\Omega^c} \cap \Lambda^c \right) \leq \left( \|\pi\|_{\infty} + 4\phi_2 n^{1/3} \right) P(\Omega_{\rho}^c \cap \Lambda^c).$$

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

$$P(\Omega_{\rho}^c \cap \Lambda^c) = P\left( \|\pi\|_{\infty} \geq \frac{\|\pi 1_A\|_{\infty} + \|\Lambda^c\|_{\infty}}{2} \right) \leq P\left( \|\pi - \pi 1_A\|_{\infty} \geq \frac{\|\pi\|_{\infty}}{2} \right) \leq P\left( \|\pi - \pi 1_A\|_{\infty} \geq \frac{\|\pi\|_{\infty}}{4} \right) + P\left( \|\pi_{m^*} - \pi 1_A\|_{\infty} \geq \frac{\|\pi\|_{\infty}}{4} \right) \leq P\left( \|\pi - \pi_{m^*}\|_{\infty} \geq \frac{\|\pi\|_{\infty}}{4\phi_0 \sqrt{D_{m_1} D_{m_2}}} \right) + P\left( \|\pi_{m^*} - \pi 1_A\|_{\infty} \geq \frac{\|\pi\|_{\infty}}{4} \right)$$

since $\|\pi - \pi_{m^*}\|_{\infty} \leq \phi_0 \sqrt{D_{m_1} D_{m_2}} \|\pi - \pi_{m^*}\|_1$.

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

$$\|\pi - \pi 1_A\|_{n}^2 \leq \|\pi_{m^*} - \pi 1_A\|_{n}^2 + \frac{1}{\theta'} \|\pi - \pi_{m^*}\|_1^2 + \theta' \sup_{t \in B_f(m^*)} Z_n^2(t)$$

where $B_f(m^*)$ is a ball of radius $m^*$.
and then, on $\Omega_\rho$,
\[
\|\hat{p} - \pi_{m*}\|^2_f \left(1 - \frac{2\rho}{\theta'}\right) \leq 4\rho \|\pi_{m*} - \pi \|_n^2 + 2\rho \theta' \sup_{t \in B_{f}(m*)} Z_n^2(t)
\]
so
\[
\|\hat{p} - \pi_{m*}\|^2 \leq \frac{4\rho \theta' f_0^{-1}}{\theta' - 2\rho} \|\pi_{m*} - \pi \|_n^2 + \frac{2\rho \theta'^2 f_0^{-1}}{\theta' - 2\rho} \sup_{t \in B_{f}(m*)} Z_n^2(t)
\]
\[
\leq 12\rho f_0^{-1} |A_2| \|\pi_{m*} - \pi \|_\infty^2 + 18\rho^2 f_0^{-1} \sup_{t \in B_{f}(m*)} Z_n^2(t)
\]
with $\theta' = 3\rho$ and by remarking that for $t$ with support $A$, $\|t\|_n^2 \leq |A_2| \|t\|_\infty^2$. Thus
\[
P(\Omega_\rho^* \cap A^c) \leq P \left( \sup_{t \in B_{f}(m_*)} Z_n^2(t) \|_{\Omega_\rho^*} \geq \frac{\|\pi\|_\infty^2}{32\phi_0^2 n^{1/3}} \frac{1}{18\rho^2 f_0^{-1}} \right) + P \left( \|\pi_{m*} - \pi \|_\infty^2 \geq \frac{\|\pi\|_\infty}{4} \right)
\]
with
\[
a = \frac{\|\pi\|_\infty^2}{32\phi_0^2} \frac{1}{18\rho^2 f_0^{-1}} \quad \text{and} \quad b = \frac{\|\pi\|_\infty^2}{12\rho f_0^{-1} |A_2|}.
\]

We will first study the two last terms in (24). Since the restriction $\pi_A$ of $\pi$ belongs to $B_{2,\infty}(\alpha_1,\alpha_2)$, the embedding theorem proved in Nikol’skii [23] p. 236 implies that $\pi_A$ belongs to $B_{\infty,\infty}(\beta_1,\beta_2)$ with $\beta_1 = \alpha_1(1 - 1/\alpha)$ and $\beta_2 = \alpha_2(1 - 1/\alpha)$. 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
\[
\|\pi_{m*} - \pi \|_\infty \leq C \left(D_{m_{1*}}^{-\beta_1} + D_{m_{2*}}^{-\beta_2}\right).
\]
And then, since $D_{m_{1*}} = D_{m_{2*}}$,
\[
D_{m_{1*}}D_{m_{2*}} \|\pi_{m*} - \pi \|_\infty^2 \leq C' \left(D_{m_{1*}}^{-2\beta_1} + D_{m_{1*}}^{-2\beta_2}\right) \leq C' \left((\ln n)^{2-2\beta_1} + (\ln n)^{2-2\beta_2}\right) \to 0.
\]
Indeed
\[
\begin{align*}
2 - 2\beta_1 &< 0 \Leftrightarrow 2\alpha_1\alpha_2 - 3\alpha_2 - \alpha_1 > 0, \\
2 - 2\beta_2 &< 0 \Leftrightarrow 2\alpha_1\alpha_2 - 3\alpha_1 - \alpha_2 > 0
\end{align*}
\]
and this double condition is ensured when $\alpha_1 > 3/2$ and $\alpha_2 > \max\left\{\frac{\alpha_1}{2\alpha_1 - 3}, \frac{3\alpha_1}{2\alpha_1 - 1}\right\}$. Consequently, for $n$ large enough,
\[
P \left( D_{m_{1*}}D_{m_{2*}} \|\pi_{m*} - \pi \|_\infty^2 \geq b \right) + P \left( \|\pi_{m*} - \pi \|_\infty \geq \frac{\|\pi\|_\infty}{4} \right) = 0.
\]
We will now prove that
\[
P \left( \sup_{t \in B_{f}(m_*)} Z_n^2(t) \|_{\Omega^*} \geq \frac{a}{n^{1/3}} \right) \leq \frac{C}{n^{4/3}}
\]
and then using (24), we will have $P(\Omega^*_p \cap A^c) \leq Cn^{-4/3}$. We remark that, if $(\varphi_j \otimes \psi_k)_{j,k}$ is a base of $(S_{m*}, \| \cdot \|)$, 

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

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

$$P \left( \sup_{t \in B_f(m*)} Z^2_n(\varphi_j \otimes \psi_k) \leq \frac{a}{4D_{m1s}D_{m2s}n^{1/3}} \right).$$

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

$$\left( \sqrt{2\|\pi\|_\infty x} + \sqrt{D_{m1s}D_{m2s}f_0^{-1}x} \right)^2 \leq \frac{a}{4D_{m1s}D_{m2s}n^{1/3}}.$$

So we will now bound $P(\Omega^* \cap A^c)$ by using the Bernstein inequality given in [8]. That is why we bound $E\left| \frac{1}{4q_n} \sum_{i=1, i \text{ odd}}^{2q_n-1} \Gamma_i^*(t) \right|^m$ for all integer $m \geq 2$,

$$E \left| \frac{1}{4q_n} \sum_{i=1, i \text{ odd}}^{2q_n-1} \Gamma_i^*(t) \right|^m \leq \left( \frac{\|t\|_\infty^{m-2}}{2} \right)^m \left( \frac{\|t\|_\infty}{2} \right)^m \frac{1}{16q_n^2} \sqrt{D_{m1s}D_{m2s}f_0^{-1}} \left( \int t(X_i, X_i + 1) - \int t(X_i, y) \pi(X_i, y) dy \right)^2$$

$$\leq \left( \frac{\|t\|_\infty}{2} \right)^m \frac{1}{16} \int t^2(x, y) f(x) \pi(x, y) dx dy$$

$$\leq \frac{1}{2m+2} \left( \frac{\|t\|_\infty}{2} \right)^m \frac{\|t\|_\infty}{2} \|t\|_{L^2}^2.$$

Then

$$E \left| \frac{1}{4q_n} \sum_{i=1, i \text{ odd}}^{2q_n-1} \Gamma_i^*(\varphi_j \otimes \psi_k) \right|^m \leq \frac{1}{2m+2} \left( \sqrt{D_{m1s}D_{m2s}f_0^{-1}} \right)^{m-2} \|\pi\|_\infty.$$

Thus the Bernstein inequality gives

$$P \left( \Omega^* \cap A^c \right) \leq \sqrt{D_{m1s}D_{m2s}f_0^{-1}}x + \sqrt{2\|\pi\|_\infty x} \leq 2e^{-p_n x}.$$

Hence

$$P \left( \sup_{t \in B_f(m*)} Z^2_{n1}(t) \leq \frac{a}{4n^{1/3}} \right) \leq 2D_{m1s}D_{m2s} \exp \left\{ -p_n Bn^{-2/3} \right\}$$

$$\leq 2n^{2/3} \exp \left\{ -\frac{Bn^{1/3}}{4q_n} \right\}.$$

But

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

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

$$P \left( \sup_{t \in B_f(m*)} Z^2_n(t) \leq \frac{a}{n^{1/3}} \right) \leq \frac{4C}{n^{4/3}}. \quad \Box$$
References


