Nonparametric estimation of the derivatives of the stationary density for stationary processes

Emeline Schmisser

Nonparametric estimation of the derivatives of the stationary density for stationary processes

Emeline Schmisser
Université Paris Descartes
Laboratoire MAP5 Emeline.Schmisser@math-info.univ-paris5.fr

Abstract

In this article, our aim is to estimate the successive derivatives of the stationary density f of a strictly stationary and β-mixing process $(X_t)_{t\geq 0}$. This process is observed at discrete times $t = 0, \Delta, \dots, n\Delta$. The sampling interval Δ can be fixed or small. We use a penalized least-square approach to compute adaptive estimators. If the derivative $f^{(j)}$ belongs to the Besov space $\mathscr{B}_{2,\infty}^{\alpha}$, then our estimator converges at rate $(n\Delta)^{-\alpha/(2\alpha+2j+1)}$. Then we consider a diffusion with known diffusion coefficient. We use the particular form of the stationary density to compute an adaptive estimator of its first derivative f'. When the sampling interval Δ tends to 0, and when the diffusion coefficient is known, the convergence rate of our estimator is $(n\Delta)^{-\alpha/(2\alpha+1)}$. When the diffusion coefficient is known, we also construct a quotient estimator of the drift for low-frequency data.

Key words: derivatives of the stationary density, diffusion processes, mixing processes, nonparametric estimation, stationary processes

AMS Classification: 62G05, 60G10

1 Introduction

In this article, we consider a strictly stationary, ergodic and β-mixing process $(X_t,\ t \geq 0)$ observed at discrete times with sampling interval Δ. The jth order derivatives $f^{(j)}$ $(j \geq 0)$ of the stationary density f are estimated by model selection. Adaptive estimators of $f^{(j)}$ are constructed thanks to a penalized least-square method and the L^2 risk of these estimators is computed.

Numerous articles deal with non parametric estimation of the stationary density (or the derivatives of the stationary density) for a strictly stationary and mixing process observed in continuous time. For instance, Bosq [4] uses a kernel estimator, Comte and Merlevède [5] realize a projection estimation and Leblanc [16] utilizes wavelets. Under the Castellana and Leadbetter's conditions, when f belongs to a Besov space $\mathscr{B}_{2,\infty}^{\alpha}$, the estimator of f converges at the parametric rate $T^{-1/2}$ (where T is the time of observation). The non parametric estimation of the stationary density of a stationary and mixing process observed at discrete times $t = 0, \Delta, \dots, n\Delta$ has also been studied, especially when the sampling interval Δ is fixed. For example, Masry [19] constructs wavelets estimators, Comte and Merlevède [7] and Lerasle [17] use a penalized least-square contrast method. The L^2 rate of convergence of the estimator is in that case $n^{-\alpha/(2\alpha+1)}$. Comte and Merlevède [5] demonstrate that, if the sampling interval $\Delta \to 0$, the penalized estimator of f converges with rate $(n\Delta)^{-\alpha/(2\alpha+1)}$ and, under the conditions of Castellana and Leadbetter, the parametric rate of convergence is reached.

There are less papers about the estimation of the derivatives of the stationary density, and the main results are for independent and identically distributed random variables. For instance, Rao [22] estimates the successive derivatives $f^{(j)}$ of a multi-dimensional process by a wavelet method. He bounds the L^2 risk of his estimator and computes the rate of convergence on Sobolev spaces. This estimator converges with rate $n^{-\alpha/(2\alpha+2j+1)}$. Hosseinioun *et al.* [13] estimate the partial derivatives of the stationary density of a mixing process by a wavelet method, and their estimators converge with rate $(n\Delta)^{-\alpha/(2\alpha+1+2j)}$.

Classical examples of β-mixing processes are diffusions: if (X_t) is solution of the stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \quad \text{and} \quad X_0 = \eta,$$

then, with some classical additional conditions on b and σ, (X_t) is exponentially β-mixing. Dalalyan and Kutoyants [9] estimate the first derivative of the stationary density for a diffusion process observed at continuous time. They prove that the minimax rate of convergence is $T^{-2\alpha/(2\alpha+1)}$ where T is the time of observation. This is the same rate of convergence as for non parametric estimator of f.

A possible application is, for diffusion processes, the estimation of the drift function b by quotient. Indeed, when $\sigma = 1$, we have that $f' = 2bf$. The drift estimation is well-known when the diffusion it observed at continuous time or for high-frequency data (see Comte *et al.* [6] for instance), but it is far more difficult when Δ is fixed. Gobet *et al.* [12] build non parametric estimators of b and σ when Δ is fixed and prove that their estimators reach the minimax L^2 risk. However, their estimators are built with eigenvalues of the infinitesimal generator and are difficult to implement.

In this paper, in a first step, we consider a strictly stationary and β-mixing process $(X_t)_{t\geq 0}$ observed at discrete times $t = 0, \Delta, \dots, n\Delta$. The successive derivatives $f^{(j)}$ $(0 \leq j \leq k)$ of the stationary density f are estimated either on a compact set, or on $\mathbb{R}$ thanks to a penalized least-square method. We introduce a sequence of increasing linear subspaces (S_m) and, for each m, we construct an estimator of $f^{(j)}$ by minimising a contrast function over S_m. Then, a penalty function $pen(m)$ is introduced to select an estimator of $f^{(j)}$ in the collection. When $f^{(j)} \in \mathscr{B}^{\alpha}_{2,\infty}$, the L^2 risk of this estimator converges with rate $(n\Delta)^{-2\alpha/(2\alpha+2j+1)}$ and the procedure does not require the knowledge of α. When $j = 0$, this is the rate of convergence obtained by Comte and Merlevède [7, 5]. Moreover, when α is known, Rao [22] obtained a rate of convergence $n^{-2\alpha/(2\alpha+2j+1)}$ for independent variables.

In a second step, we assume that the process (X_t) is solution of a stochastic differential equation of known diffusion coefficient σ. Then f' can be estimated by estimating $2bf$ and f. An estimator of $2bf$ is built either on a compact set, or on $\mathbb{R}$ by a penalized least-square contrast method. It only converges when the sampling interval $\Delta \to 0$, but in this case, its rate of convergence is better than for the previous estimator: it is $(n\Delta)^{-2\alpha/(2\alpha+1)}$ when $f' \in \mathscr{B}^{\alpha}_{2,\infty}$ (and not $(n\Delta)^{-2\alpha/(2\alpha+3)}$). This is the minimax rate obtained by Dalalyan and Kutoyants [9] with continuous observations.

Then, an estimator by quotient of the drift function b is constructed. When Δ is fixed, it reaches the minimax rate obtained by Gobet *et al.* [12].

In Section 2, an adaptive estimator of the successive derivatives $f^{(j)}$ of the stationary density f of a stationary and β-mixing process is computed by a penalized least square method. In Section 3, only diffusions with known diffusion coefficients are considered. An adaptive estimator of f' (in fact, an estimator of $2bf$) is built. In Section 4, a quotient estimator of b is constructed. In Section 5, the theoretical results are illustrated via various simulations using several models. Processes (X_t) are simulated by the exact retrospective algorithm of Beskos *et al.* [3]. The proofs are given in Section 6. In the Appendix, the spaces of functions are introduced.

2 Estimation of the successive derivatives of the stationary density

2.1 Model and assumptions

In this section, a stationary process $(X_t)_{t\geq 0}$ is observed at discrete times $t = 0, \Delta, \dots, n\Delta$ and the successive derivatives $f^{(j)}$ of the stationary density $f = f^{(0)}$ are estimated for $0 \leq j \leq k$. The sampling interval Δ is fixed or tends to 0. The estimation set A is either a compact $[a_0, a_1]$, or $\mathbb{R}$. Let us consider the norms

$$\|.\|_\infty = \sup_A |.| \quad , \quad \|.\|_{L^2} = \|.\|_{L^2(A)} \quad \text{and} \quad \langle ., . \rangle = \langle ., . \rangle_{L^2(A)} \, . \tag{2.1}$$

We have the following assumptions:

Assumption M1.
The process (X_t) is ergodic, strictly stationary and arithmetically or exponentially β-mixing.

A process is arithmetically β-mixing if its β-mixing coefficient satisfies:

$$\beta_X(t) \leq \beta_0 (1+t)^{-(1+\theta)} \tag{2.2}$$

where θ and β_0 are some positive constants. A process is exponentially (or geometrically) β-mixing if there exists two positive constants β_0 and θ such that:

$$\beta_X(t) \leq \beta_0 \exp(-\theta t) \tag{2.3}$$

Assumption M2.
The stationary density f is k times differentiable and, for each $j \leq k$, its derivatives $f^{(j)}$ belong to $L^2(A) \cap L^1(A)$. Moreover, $f^{(j)}$ satisfies $\int_A x^2 \left(f^{(j)}(x)\right)^2 dx < +\infty$.

Remark 2.1. If $A = [a_0, a_1]$, Assumption **M2** is only $\forall j \leq k$, $f^{(j)} \in L^2(A)$.

Our aim is to estimate $f^{(j)}$ by model selection. Therefore an increasing sequence of finite dimensional linear subspaces (S_m) is needed. On each of these subspaces, an estimator of $f^{(j)}$ is computed, and thanks to a penalty function depending on m, the best possible estimator is chosen. Let us denote by $\mathscr{C}^l$ the space of functions l times differentiable on A and with a continuous ℓth derivative, and $\mathscr{C}^l_m$ the set of the piecewise functions $\mathscr{C}^l$. To estimate $f^{(j)}$, $0 \leq j \leq k$ on a compact set, we need a sequence of linear subspaces that satisfies the assumption:

Assumption S1 : Estimation on a compact set. *1. The subspaces S_m are increasing, of finite dimension D_m and included in $L^2(A)$.*

2. *For any m, any function $t \in S_m$ is k times differentiable (belongs to $\mathscr{C}^{k-1} \cap \mathscr{C}^k_m$) and satisfies:*

$$\forall j \leq k, \quad t^{(j)}(a_0) = t^{(j)}(a_1) = 0.$$

3. *There exists a norm connection: for any $j \leq k$, there exists a constant ψ_j such that:*

$$\forall m,\ \forall t \in S_m, \quad \left\| t^{(j)} \right\|_\infty^2 \leq \psi_j D_m^{2j+1} \|t\|_{L^2}^2 .$$

   Let us consider $(\varphi_{\lambda,m},\ \lambda \in \Lambda_m)$ an orthonormal basis of S_m with $|\Lambda_m| = D_m$. We have that $\left\|\Psi^2_{j,m}(x)\right\|_\infty \leq \psi_j D_m^{2j+1}$ where $\Psi^2_{j,m}(x) = \sum_{\lambda \in \Lambda_m} \left(\varphi^{(j)}_{\lambda,m}(x)\right)^2$.

4. *There exists a constant c such that, for any $m \in \mathbb{N}$, any function $t \in S_m$:*

$$\left\| t^{(j)} \right\|_{L^2} \leq c D_m^{2j} \|t\|_{L^2}^2 .$$

5. *For any function t belonging to the unit ball of a Besov space $\mathscr{B}^\alpha_{2,\infty}$:,*

$$\|t - t_m\|_{L^2}^2 \leq D_m^{-2} \vee D_m^{-2\alpha}$$

   where t_m is the orthogonal (L^2) projection of t over S_m.

Remark 2.2. Because of Point 2, the projection t_m converges very slowly to t on the boundaries of the compact $A = [a_0, a_1]$ and the inequality $\|t - t_m\|_{L^2}^2 \leq D_m^{-2\alpha}$ can not be satisfied for any $t \in \mathscr{B}^\alpha_{2,\infty}$.

In the Appendix, several sequences of linear subspaces satisfying this property are given. To estimate $f^{(j)}$ on $\mathbb{R}$, slightly different assumptions are needed: let us consider an increasing sequence of linear subspaces S_m generated by an orthonormal basis $\{\varphi_{\lambda,m},\ \lambda \in \mathbb{Z}\}$. We have that $\dim(S_m) = \infty$, so to build estimators, we use the restricted spaces $S_{m,N} = \mathrm{Vect}(\varphi_{\lambda,m},\ \lambda \in \Lambda_{m,N})$ with $|\Lambda_{m,N}| < +\infty$. The following assumption involves the sequences of linear subspaces (S_m) and $(S_{m,N})$.

Assumption S2 : Estimation on $\mathbb{R}$. *1. The sequence of linear subspaces (S_m) is increasing.*

2. *We have that $|\Lambda_{m,N}| := \dim(S_{m,N}) = 2^{m+1}N + 1$.*
3. *$\forall m, N \in \mathbb{N}$, $\forall t \in S_{m,N}$: $t \in \mathscr{C}^{k-1} \cap \mathscr{C}^k_m$ and $\forall j < k$, $\lim_{|x|\to\infty} t^{(j)}(x) = 0$.*
4. *$\exists \psi_j \in \mathbb{R}^+$, $\forall m \in \mathbb{N}$, $\forall t \in S_m$, $\forall j \le k$, $\left\|t^{(j)}\right\|_\infty^2 \le \psi_j 2^{(2j+1)m} \|t\|_{L^2}^2$. Particularly,*
$$\left\|\Psi_m^2(x)\right\|_\infty^2 = \left\|\sum_{\lambda\in\mathbb{Z}} \left(\varphi_{\lambda,m}^{(j)}(x)\right)^2\right\|_\infty^2 \le \psi_j 2^{(2j+1)m}.$$
5. *$\exists c$, $\forall m \in \mathbb{N}$, $\forall t \in S_m$, $\forall j \le k$: $\left\|t^{(j)}\right\|_{L^2}^2 \le c2^{2jm} \|t\|_{L^2}^2$.*
6. *For any function $t \in L^2 \cap L^1(\mathbb{R})$ such that $\int x^2 t^2(x)dx < +\infty$,*
$$\|t_m - t_{m,N}\|_{L^2}^2 \le c\frac{2^m}{N}$$
where t_m is the orthogonal (L^2) projection of t over S_m and $t_{m,N}$ its projection over $S_{m,N}$.
7. *There exists $r \ge 1$ such that, for any function t belonging to the unit ball of a Besov space $\mathscr{B}^\alpha_{2,\infty}$ (with $\alpha < r$),*
$$\|t - t_m\|_{L^2}^2 \le 2^{-2m\alpha}.$$

Proposition 2.1.
If the function φ generates a r-regular multiresolution analysis of L^2, with $r \ge k$, then the subspaces
$$S_m = Vect\{\varphi_{\lambda,m},\ \lambda \in \mathbb{Z}\} \quad \text{and} \quad S_{m,N} = Vect\{\varphi_{\lambda,m},\ \lambda \in \Lambda_{m,N}\}$$
*(where $\varphi_{\lambda,m}(x) = 2^{m/2}\varphi(2^m x - \lambda)$ and $\Lambda_{m,N} = \{\lambda \in \mathbb{Z},\ |\lambda| \le 2^m N\}$) satisfy **S2**.*

For the definition of the multi-resolution analysis, see Meyer [20], chapter 2.

2.2 Risk of the estimator for fixed m

An estimator $\hat{g}_{j,m}$ of $g_j := f^{(j)}$ is computed by minimising the contrast function
$$\gamma_{j,n}(t) = \|t\|_{L^2}^2 - \frac{2(-1)^j}{n}\sum_{k=1}^n t^{(j)}(X_{k\Delta}).$$

Under Assumptions **S1** or **S2**:
$$\mathbb{E}(\gamma_{j,n}(t)) = \|t\|_{L^2}^2 - 2(-1)^j\left\langle t^{(j)}, f\right\rangle = \|t\|_{L^2}^2 - 2\left\langle t, f^{(j)}\right\rangle = \left\|t - f^{(j)}\right\|_{L^2}^2 - C \quad \text{where} \quad C = \left\|f^{(j)}\right\|_{L^2}^2.$$

If Assumption **S1** is satisfied, let us denote
$$\hat{g}_{j,m}(t) = \arg\inf_{t\in S_m} \gamma_{j,n}(t),$$

and, under Assumption **S2**,
$$\hat{g}_{j,m,N}(t) = \arg\inf_{t\in S_{m,N}} \gamma_{j,n}(t).$$

We have the two following theorems:

Theorem 2.1 : Estimation on a compact set.
*Under Assumptions **M1-M2** and **S1**, the estimator risk satisfies, for any $j \le k$ and $m \in \mathbb{N}$:*
$$\mathbb{E}\left(\|\hat{g}_{j,m} - g_j\|_{L^2}^2\right) \le \|g_{j,m} - g_j\|_{L^2}^2 + 8c\beta_0\psi_j\frac{D_m^{2j+1}}{n}\left(1 \vee \frac{1}{\theta\Delta}\right)$$
*where $g_{j,m}$ is the orthogonal (L^2) projection of g_j over S_m. The constants β_0 and θ are defined in (2.2) or (2.3), ψ_j is defined in Assumption **S1** and c is a universal constant.*

Theorem 2.2 : Estimation on $\mathbb{R}$.
*Under Assumptions **M1-M2** and **S2**, for any $j \leq k$ and $m \in \mathbb{N}$:*

$$\mathbb{E}\left(\|\hat{g}_{j,m,N} - g_j\|_{L^2}^2\right) \leq \|g_{j,m} - g_j\|_{L^2}^2 + C\frac{2^m}{N} + 8c\beta_0\psi_j\frac{2^{(2j+1)m}}{n}\left(1 \vee \frac{1}{\theta\Delta}\right)$$

*where C depends on $\int_{-\infty}^{\infty} x^2 g^2(x)dx$ and of the chosen sequence of linear subspaces $(S_{m,N})$. According to Assumption **S2** 6., if $N \geq (n \wedge n\theta\Delta)$,*

$$\mathbb{E}\left(\|\hat{g}_{j,m,N} - g_j\|_{L^2}^2\right) \leq \|g_{j,m} - g\|_{L^2}^2 + c\beta_0\frac{2^{(2j+1)m}}{n}\left(1 \vee \frac{1}{\theta\Delta}\right).$$

If the random variables $(X_0, \dots, X_n)$ are independent, the derivatives of the density can be estimated in the same way and the two previous theorems (as well as the theorems for the adaptive risk) can be applied if we set $\theta = +\infty$.

When $\Delta = 1$, the risk bound is the same as in Hosseinioun *et al.* [13].

2.3 Optimisation of the choice of m

Under Assumption **S1** and if g_j belongs to the unit ball of a Besov space $\mathscr{B}_{2,\infty}^{\alpha}$ with $\alpha \geq 1$, then $\|g_{j,m} - g_j\|_{L^2}^2 \leq cD_m^{-2}$ and the best bias-variance compromise is obtained for $D_m \sim (n(1 \vee \theta\Delta))^{1/(2j+3)}$. In that case,

$$\mathbb{E}\left(\|\hat{g}_{j,m} - g_j\|_{L^2}^2\right) \leq (n \vee n\theta\Delta)^{-2/(2j+3)}.$$

If Assumption **S2** is satisfied and if g_j belongs to $\mathscr{B}_{2,\infty}^{\alpha}$, with $r \geq \alpha$, then $\|g_{j,m} - g_j\|_{L^2}^2 \leq c2^{-2m\alpha}$. If $N \geq n(1 \wedge \theta\Delta)$, the best bias-variance compromise is obtained for

$$m \sim \frac{1}{2j+1+2\alpha}\log_2(n(1 \vee \theta\Delta)) \quad \text{and then} \quad \mathbb{E}\left(\|\hat{g}_{j,m,N} - g_j\|_{L^2}^2\right) \leq (n \vee n\Delta)^{-2\alpha/(2\alpha+2j+1)}.$$

Rao [22] builds estimators of the successive derivatives $f^{(j)}$ for independent variables. This estimators converge with rate $n^{-2\alpha/(2\alpha+2j+1)}$.

2.4 Risk of the adaptive estimator on a compact set

An additional assumption for the process (X_t) is needed:

Assumption M3.
If the process $(X_t)_{t\geq 0}$ is arithmetically β-mixing, then the constant θ defined in (2.2) is such that $\theta > 3$.

Let us set $\mathscr{M}_{j,n} = \{m,\ D_m \leq \mathscr{D}_{j,n}\}$ where $\mathscr{D}_{j,n} \leq (n\Delta \wedge n)^{1/(2j+2)}$ is the maximal dimension. For any $m \in \mathscr{M}_{j,n}$, an estimator $\hat{g}_{j,m} \in S_m$ of $g_j = f^{(j)}$ is computed. Let us introduce a penalty function $pen_j(m)$ depending on D_m and n:

$$pen_j(m) \geq \kappa\beta_0\psi_j\frac{D_m^{2j+1}}{n}\left(1 \vee \frac{1}{\theta\Delta}\right).$$

Then we construct an adaptive estimator: choose $\hat{m}_j$ such that

$$\tilde{g}_j := \hat{g}_{j,\hat{m}_j} \quad \text{where} \quad \hat{m}_j = \arg\min_{m\in\mathscr{M}_{j,n}}\left[\gamma_{j,n}(\hat{g}_{j,m}) + pen_j(m)\right].$$

Theorem 2.3 : Adaptive estimation on a compact set.
*There exists a universal constant κ such that, if Assumptions **M1-3** and **S1** are satisfied:*

$$\mathbb{E}\left(\|\tilde{g}_j - g_j\|_{L^2}^2\right) \leq C\inf_{m\in\mathscr{M}_{j,n}}\left(\|g_{j,m} - g_j\|_{L^2}^2 + pen_j(m)\right) + \frac{c}{n}\left(1 \vee \frac{1}{\Delta}\right)$$

where C is a universal constant and c depends on ψ_j, β_0 and θ.

Remark 2.3. The adaptive estimator automatically realises the bias-variance compromise. Comte and Merlevède [5] obtain similar results when $j = 0$ and the sampling interval Δ is fixed, and their remainder term is smaller: it is $1/n$ and not $\ln^2(n)/n$.

The penalty function depends on β_0 and θ. Unfortunately, these two constants are difficult to estimate. However, the slope heuristic defined in Arlot and Massart [1] enables us to choose automatically a constant λ such that the penalty $\lambda D_m^{2j+1}/(n\Delta)$ is good. It is also possible to use the resampling penalties of Lerasle [18].

2.5 Risk of the adaptive estimator on $\mathbb{R}$

Let us denote $\mathscr{M}_{j,n} = \{m,\ 2^m \leq \mathscr{D}_{j,n}\}$ with $\mathscr{D}_{j,n}^{2j+2} \leq n\Delta \wedge n$ and fix $N = N_n = (n \wedge n\Delta)$. For any $m \in \mathscr{M}_{j,n}$, an estimator $\hat{g}_{j,m,N_n} \in S_{m,N_n}$ of g_j is computed. The best dimension $\hat{m}_j$ is chosen such that

$$\hat{m}_j = \arg\min_{m \in \mathscr{M}_{j,n}} \left[\gamma_{j,n}\left(\hat{g}_{j,m,N_n}\right) + pen_j(m)\right] \quad \text{where} \quad pen_j(m) = c\psi_j \left(\frac{2^{(2j+1)m}}{n} \vee \frac{2^{(2j+1)m}}{n\theta\Delta}\right)$$

and the resulting estimator is denoted by $\tilde{g}_j := \hat{g}_{j,\hat{m}_j,N_n}$.

Theorem 2.4 : Adaptive estimation on $\mathbb{R}$.
*Under Assumptions **M1-M3** and **S2**,*

$$\mathbb{E}\left(\|\tilde{g}_j - g_j\|_{L^2}^2\right) \leq C \inf_{m \in \mathscr{M}_{j,n}} \left(\|g_{j,m} - g_j\|_{L^2}^2 + pen_j(m)\right) + \frac{c}{n}\left(1 \vee \frac{1}{\Delta}\right)$$

where c depends on ψ_j, β_0 and θ.

3 Case of stationary diffusion processes

Let us consider the stochastic differential equation (SDE):

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = \eta, \tag{3.1}$$

where η is a random variable and $(W_t)_{t\geq 0}$ a Brownian motion independent of η. The drift function $b : \mathbb{R} \to \mathbb{R}$ is unknown and the diffusion coefficient $\sigma : \mathbb{R} \to \mathbb{R}^{+*}$ is known. The process $(X_t)_{t\geq 0}$ is assumed to be strictly stationary, ergodic and β-mixing. Obviously, we can construct estimators of the successive derivatives of the stationary density using the previous section. But in this section, we use the properties of a diffusion process to compute a new estimator of the first derivative of the stationary density. If the sampling interval Δ is small, this new estimator converge faster than the previous one.

3.1 Model and Assumptions

The process $(X_t)_{t\geq 0}$ is observed at discrete times $t = 0, \Delta, \ldots, n\Delta$.

Assumption M4.
The functions b and σ are globally Lipschitz and $\sigma \in \mathscr{C}^1$.

Assumption **M4** ensures the existence and uniqueness of a solution of the SDE (3.1).

Assumption M5.
The diffusion coefficient σ belongs to $\mathscr{C}^1$, is bounded and positive: there exist constants σ_0 and σ_1 such that:

$$\forall x \in \mathbb{R},\ 0 < \sigma_1 \leq \sigma(x) \leq \sigma_0.$$

Assumption M6.
There exist constant $r > 0$ and $1 \leq \alpha \leq 2$ such that

$$\exists M_0 \in \mathbb{R}^+, \quad \forall x, |x| \geq M_0, \quad xb(x) \leq -r\,|x|^\alpha.$$

Under Assumptions **M4-M6**, there exists a stationary density f for the SDE (3.1), and

$$f(x) \propto \sigma^{-2}(x) \exp\left(2\int_0^x b(s)\sigma^{-2}(s)ds\right). \tag{3.2}$$

Then f has moments of any orders and:

$$\int |f'(x)|^2\, dx < \infty, \quad \forall m \in \mathbb{N}, \int |x|^m\, |f'(x)|\, dx < \infty \tag{3.3}$$

$$\forall m \in \mathbb{N},\ \|x^m f(x)\|_\infty < \infty, \quad \left\|b^4(x) f(x)\right\|_\infty < \infty \quad \text{and} \quad \int \exp\left(|b(x)|\right) f(x)dx < \infty. \tag{3.4}$$

Assumption M7.
The process is stationary: $\eta \sim f$.

According to Pardoux and Veretennikov [21], Proposition 1 p.1063, under Assumptions **M5-M6**, the process (X_t) is exponentially β-mixing: there exist constants β_0 and θ such that $\beta_X(t) \leq \beta_0 e^{-\theta t}$. Moreover, Gloter [11] prove the following property:

Proposition 3.1.
Let us set $\mathscr{F}_t = \sigma\left(\eta,\, W_s, s \leq t\right)$. Under Assumptions M4 and M7, for any $k \geq 1$, there exists a constant $c(k)$ depending on b and σ such that:

$$\forall h, 0 < h \leq 1, \forall t \geq 0 \quad \mathbb{E}\left(\sup_{s\in[t,t+h]} |b(X_s) - b(X_t)|^k \middle| \mathscr{F}_t\right) \leq c\left(k\right) h^{k/2}\left(1 + |X_t|^k\right).$$

Remark 3.1. To estimate f', it is enough to have an estimator of $2bf$ and an estimator of f. Indeed, according to equation (3.2), the first derivative f' satisfies:

$$\frac{f'(x)}{f(x)} \propto \frac{2b(x)}{\sigma^2(x)} - 2\frac{\sigma'(x)}{\sigma(x)}.$$

By assumption, the diffusion coefficient σ is known. Besides, according to Assumptions **M4** and **M5**, σ' and σ^{-1} are bounded. As we have already constructed an estimator of $f = g_0$ in Section 2, it remains to estimate $2bf$.

In this section, we construct an estimator $\tilde{h}$ of $h := 2bf$ either on a compact set $[a_0, a_1]$, or on $\mathbb{R}$.

3.2 Sequence of linear subspaces

Like in the previous section, estimators $\hat{h}_m$ of h are computed on some linear subspaces S_m or $S_{m,N}$, then a penalty function $pen(m)$ is introduced to choose the best possible estimator $\tilde{h}$. If h is estimated on a compact set $A = [a_0, a_1]$, the following assumption is needed:

Assumption S3 : Estimation on a compact set. *1. The sequence of linear subspaces S_m is increasing, $D_m = \dim(S_m) < \infty$ and $\forall m$, $S_m \subseteq L^2(A)$.*

2. *There exists a norm connection: for any $m \in \mathbb{N}$, any function $t \in S_m$ satisfies*

$$\|t\|_\infty^2 \leq \phi_0 D_m \|t\|_{L^2}^2.$$

Particularly, if we note $\Phi_m(x) = \sum_{\lambda\in\Lambda_m} \left(\varphi_{\lambda,m}(x)\right)^2$ where $(\varphi_{\lambda,m},\ \lambda \in \Lambda_m)$ is an orthonormal basis of S_m, then $\left\|\Phi_m^2(x)\right\|_\infty \leq \phi_0 D_m$.

3. *There exists $r \geq 1$ such that, for any function t belonging to $\mathscr{B}_{2,\infty}^\alpha$ with $\alpha \leq r$,*

$$\|t - t_m\|_{L^2}^2 \leq D_m^{-2\alpha}$$

where t_m is the orthogonal projection of t over S_m.

In the Appendix, several examples of sequence of linear subspaces satisfying this assumption are given. To estimate h on $\mathbb{R}$, an increasing sequence of linear subspaces $S_m = \mathrm{Vect}\,(\varphi_{\lambda,m}\ \lambda \in \mathbb{Z})$ (where $\{\varphi_{\lambda,m}\}_{\lambda\in\mathbb{Z}}$ is an orthonormal basis of S_m) is considered. As the dimension of those subspaces is infinite, the truncated subspaces $S_{m,N} = \mathrm{Vect}\,(\varphi_{\lambda,m},\ \lambda \in \Lambda_{m,N})$ are used.

Assumption S4 : Estimation on $\mathbb{R}$. *1. The sequence of linear subspaces (S_m) is increasing.*

2. *The dimension of the subspace $S_{m,N}$ is $2^{m+1}N + 1$.*

3. *$\exists \phi_0$, $\forall m$, $\forall t \in S_m$, $\|t\|_\infty^2 \leq \phi_0 2^m \|t\|_{L^2}^2$. Let us set $\Phi_m(x) = \sum_{\lambda\in\mathbb{Z}} (\varphi_{\lambda,m}(x))^2$, then $\left\|\Phi_m^2(x)\right\|_\infty \leq \phi_0 2^m$ where ϕ_0 is a constant independent of N.*

4. *For any function $t \in L^2 \cap L^1(\mathbb{R})$ such that $\int x^2 t^2(x)dx < +\infty$,*
$$\|t_m - t_{m,N}\|_{L^2}^2 \leq c\frac{2^m}{N}$$
where t_m is the orthogonal (L^2) projection of t over S_m and $t_{m,N}$ its projection over $S_{m,N}$.

5. *There exists $r \geq 1$ such that for any function t belonging to the unit ball of a Besov space $\mathscr{B}_{2,\infty}^\alpha$ with $\alpha \leq r$,*
$$\|t - t_m\|_{L^2}^2 \leq c2^{-2m\alpha}.$$

Proposition 3.2.
Let us consider a function φ generating a r-regular multi-resolution analysis of L^2 with $r \geq 0$. Let us set
$$S_m = \mathit{Vect}\{\varphi_{\lambda,m},\ \lambda \in \mathbb{Z}\} \quad \text{and} \quad S_{m,N} = \mathit{Vect}\{\varphi_{\lambda,m},\ \lambda \in \Lambda_m\}$$
*where $\varphi_{\lambda,m}(x) = 2^{m/2}\varphi(2^m x - \lambda)$ and $\Lambda_m = \{\lambda \in \mathbb{Z},\ |\lambda| \leq 2^m N\}$. Then the subspaces $S_{m,N}$ satisfy Assumption **S4**.*

*Functions $\varphi(x) = \sin(x)/x$ also generate a multi-resolution of $L^2(\mathbb{R})$, but they are not even 0-regular. However, they satisfy Assumption **S4** if Sobolev spaces take the place of Besov spaces in Point 5. The definition of Sobolev spaces of regularity α is recalled here:*
$$W_\alpha = \left\{g,\ \int_{-\infty}^{\infty} |g^*(x)|^2 \left(x^2+1\right)^\alpha dx < \infty\right\}$$
where g^* is the Fourier transform of g.

3.3 Risk of the estimator with m fixed

For any $m \in \mathscr{M}_n$, where $\mathscr{M}_n = \{m,\ D_m \leq \mathscr{D}_n\}$, an estimator $\hat{h}_m$ of $h = 2bf$ is computed. The maximal dimension $\mathscr{D}_n$ is specified later. The following contrast function is considered:
$$\Gamma_n(t) = \|t\|_{L^2}^2 - \frac{4}{n\Delta}\sum_{k=1}^n \left(X_{(k+1)\Delta} - X_{k\Delta}\right) t\left(X_{k\Delta}\right).$$
As $\Delta^{-1}\left(X_{(k+1)\Delta} - X_{k\Delta}\right) = I_{k\Delta} + Z_{k\Delta} + b(X_{k\Delta})$ with
$$I_{k\Delta} = \frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta} \left(b(X_s) - b(X_{k\Delta})\right) ds \quad \text{and} \quad Z_{k\Delta} = \frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta} \sigma(X_s)dW_s, \tag{3.5}$$
we have that $\mathbb{E}\left(\Gamma_n(t)\right) = \|t\|_{L^2}^2 - 4\langle bf, t\rangle - 4\mathbb{E}\left(I_\Delta t(X_\Delta)\right)$. According to Lemma 6.4, $|\mathbb{E}\left(I_{k\Delta}t(X_{k\Delta})\right)| \leq c\Delta^{1/2}$. Moreover, $h = 2bf$, so
$$\mathbb{E}\left(\Gamma_n(t)\right) = \|t\|_{L^2}^2 - 2\langle h, t\rangle + O\left(\Delta^{1/2}\right).$$
This inequality justifies the choice of the contrast function if the sampling interval Δ is small. If Assumption **S3** is satisfied, we consider the estimator
$$\hat{h}_m = \arg\min_{t\in S_m} \Gamma_n(t)$$

and, under Assumption **S4**, we set

$$\hat{h}_{m,N} = \arg\min_{t\in S_{m,N}} \Gamma_n(t).$$

Theorem 3.1 : Estimation on a compact set.
*Under Assumptions **M4-M7** and **S3**,*

$$\mathbb{E}\left(\left\|\hat{h}_m - h\right\|_{L^2}^2\right) \leq \|h_m - h\|_{L^2}^2 + c\Delta + \left(\sigma_0^2 \|f\|_\infty + \frac{2\beta_0\phi_0}{\theta}\right)\frac{D_m}{n\Delta}$$

where h_m is the orthogonal projection of h over S_m and c a constant depending on b and σ. We remind that the β-mixing coefficient of the process (X_t) is such that $\beta_X(t) \leq \beta_0 e^{-\theta t}$.

Theorem 3.2 : Estimation on $\mathbb{R}$.
*Under Assumptions **M4-M7** and **S4***

$$\mathbb{E}\left(\left\|\hat{h}_{m,N} - h\right\|_{L^2}^2\right) \leq \|h_{m,N} - h\|_{L^2}^2 + c\frac{2^m}{N} + c\Delta + \left(\|f\|_\infty + \frac{2\beta_0\phi_0}{\theta}\right)\frac{2^m}{n\Delta}.$$

where $h_{m,N}$ is the orthogonal projection of h on the space $S_{m,N}$. If $N = N_n = n\Delta$, then

$$\mathbb{E}\left(\left\|\hat{h}_{m,N_n} - h\right\|_{L^2}^2\right) \leq \|h_m - h\|_{L^2}^2 + c\Delta + \left(\|f\|_\infty + \frac{2\beta_0\phi_0}{\theta}\right)\frac{2^m}{n\Delta}$$

where h_m is the orthogonal projection of h over S_m.

3.4 Optimisation of the choice of m

Under Assumption **S3**, if $h\mathbb{1}_A$ belongs to the unit ball of a Besov space $\mathscr{B}_{2,\infty}^\alpha$, then $\|h - h_m\|_{L^2}^2 \leq D_m^{-2\alpha}$. To minimise the bias-variance compromise, one have to choose

$$D_m \sim (n\Delta)^{1/(1+2\alpha)}$$

and in that case the estimator risk satisfies:

$$\mathbb{E}\left(\left\|\hat{h}_m - h\right\|_{L^2}^2\right) \leq C\,(n\Delta)^{-2\alpha/(1+2\alpha)} + c\Delta.$$

Under Assumption **S4**, if h belongs to $B_{2,\infty}^\alpha$, then $\|h - h_m\|_{L^2}^2 \leq 2^{-2m\alpha}$ and

$$\mathbb{E}\left(\left\|\hat{h}_{m,n\Delta} - h\right\|_{L^2}^2\right) \leq C\,(n\Delta)^{-2\alpha/(1+2\alpha)} + c\Delta.$$

Remark 3.2. Dalalyan and Kutoyants [9] estimate the first derivative of the stationary density observed at continuous time (they observe X_t for $t \in [0, T]$). In that framework, the diffusion coefficient σ^2 is known. The minimax rate of convergence of the estimator is $T^{-\alpha/(1+2\alpha)}$. It is the rate that we obtain when Δ tends to 0.

Let us set $\Delta \sim n^{-\beta}$. We obtain the following convergence table:

β	principal term of the bound	rate of convergence of the estimator
$0 < \beta \leq \frac{2\alpha}{4\alpha+1}$	Δ	$n^{-\beta}$
$\frac{2\alpha}{4\alpha+1} \leq \beta < 1$	$(n\Delta)^{-2\alpha/(1+2\alpha)}$	$n^{-2\alpha(1-\beta)/(4\alpha+1)}$

Those rates of convergence are the same as for the estimator of the drift. If $\beta \geq 1/2$, the dominating term in the risk bound is always $(n\Delta)^{-2\alpha/(1+2\alpha)}$. The rate of convergence is always smaller than $n^{-1/2}$. If (n, Δ) is fixed and if $\Delta \leq n^{-2\alpha/(4\alpha+3)}$, then the second estimator $\hat{h}_m$ converges faster than the first one $\hat{g}_{1,m}$. However, if the sampling interval Δ is larger than $n^{-2\alpha/(4\alpha+3)}$, it is the opposite.

3.5 Risk of the adaptive estimator on a compact set

For any $m \in \mathscr{M}_{n,A} = \{m,\ D_m \leq \mathscr{D}_n\}$ where the maximal dimension $\mathscr{D}_n$ is specified later, an estimator $\hat{h}_m \in S_m$ of h is computed. Let us set

$$pen(m) \geq \kappa \frac{D_m}{n\Delta}\left(1+\frac{8\beta_0}{\theta}\right) \quad \text{and} \quad \hat{m} = \inf_{m \in \mathscr{M}_{n,A}} \left\{\gamma_n\left(\hat{h}_m\right) + pen(m)\right\}.$$

The resulting estimator is denoted by $\tilde{h} := \hat{h}_{\hat{m}}$. Let us consider the asymptotic framework:

Assumption S5.

$$\frac{n\Delta}{\ln^2(n)} \to \infty \quad \textit{and} \quad \mathscr{D}_n^2 \leq \frac{n\Delta}{\ln^2(n)}.$$

Theorem 3.3 : Adaptive estimation on a compact set.
*There exists a constant κ depending only on the chosen sequence of linear subspaces (S_m) such that, under Assumptions **M4-M7** , **S3** and **S5**,*

$$\mathbb{E}\left(\left\|\tilde{h}-h\right\|_{L^2}^2\right) \leq C \inf_{m \in \mathscr{M}_{n,A}} \left\{\|h_m - h\|_{L^2}^2 + pen(m)\right\} + c\Delta + \frac{c'}{n\Delta}$$

where C is a numerical constant, c' depends on ϕ_0 and $\|f\|_\infty$ and c depends on b.

Remark 3.3. The estimator is only consistent if $\Delta \to 0$. Moreover, the adaptive estimator $\tilde{h}$ automatically realises the bias-variance compromise.

3.6 Risk of the adaptive estimator on $\mathbb{R}$

An estimator $\hat{h}_{m,n\Delta} \in S_{m,n\Delta}$ is computed for any $m \in \mathscr{M}_{n,\mathbb{R}} = \{m,\ 2^m \leq \mathscr{D}_n\}$. The following penalty function is introduced:

$$pen(m) \geq \kappa \frac{2^m}{n\Delta}\left(1+\frac{2\beta_0}{\theta}\right) \quad \text{and we set} \quad \hat{m} = \inf_{m \in \mathscr{M}_n} \left\{\gamma_n\left(\hat{h}_{m,n\Delta}\right) + pen(m)\right\}$$

Let us denote by $\tilde{h}_{n\Delta}$ the resulting estimator.

Theorem 3.4 : Adaptive estimation on $\mathbb{R}$.
*There exists a constant κ depending only on the sequence of linear subspaces (S_m) such that, if Assumptions **M4-M7** , **S4** and **S5** are satisfied:*

$$\mathbb{E}\left(\left\|\tilde{h}_{n\Delta}-h\right\|_{L^2}^2\right) \leq C \inf_{m \in \mathscr{M}_{j,n,\mathbb{R}}} \left\{\|h_m - h\|_{L^2}^2 + pen(m)\right\} + c\Delta + \frac{c'}{n\Delta}.$$

4 Drift estimation by quotient

If the process $(X_t)_{t\geq 0}$ is the solution of the stochastic differential equation (SDE)

$$dX_t = b(X_t)dt + dW_t$$

and satisfies Assumptions **M4-M7**, then

$$b = f'/2f.$$

An estimator of the drift by quotient can therefore be constructed. For high-frequency data, Comte *et al.* [6] build an adaptive drift estimator thanks to a penalized least-square method. Their estimator converges with the minimax rate $(n\Delta)^{-2\alpha/(2\alpha+1)}$ if b belongs to the Besov space $\mathscr{B}_{2,\infty}^{\alpha}$. On the contrary, there exist few results on the drift estimation where the sampling interval Δ is fixed. Gobet *et al.* [12] build a drift estimator for low-frequency data, however, their estimator is not easy to implement. In this section, a drift estimator by quotient is constructed and its risk is computed.

We estimate f and f' on $\mathbb{R}$ in order to avoid convergence problems on the boundaries of the compact. Let us consider two sequences of linear subspaces $(S_{0,m},\ m \in \mathscr{M}_{0,n})$ and $(S_{1,m},\ m \in \mathscr{M}_{1,n})$ satisfying Assumption **S2** for $k = 1$ and such that

$$\mathscr{M}_{0,n} = \left\{m_0,\ \log(n) \le 2^{m_0} \le \eta\sqrt{n\Delta}/\log(n\Delta)\right\} \text{ and } \mathscr{M}_{1,n} = \left\{m_1,\ 2^{m_1} \le (n\Delta)^{1/5}\right\}$$

where the constant η does not depend on b neither σ.

As in Section 2, adaptive estimators $\tilde{f} := \tilde{g}_{0,n\Delta}$ and $\tilde{g} := \tilde{g}_{1,n\Delta}$ of $f = g_0$ and $f' = g_1$ are computed. As b belongs to $\mathscr{B}^{\alpha}_{2,\infty}$, f and f' also belong to $\mathscr{B}^{\alpha}_{2,\infty}$ and the best bias-variance compromise for $\hat{g}_{0,m}$ is obtained for $2^{m_0} \sim (n\Delta)^{1/(1+2\alpha)}$, and for $\hat{g}_{1,m}$ it is obtained for $2^{m_1} \sim (n\Delta)^{1/(3+2\alpha)}$. If $\alpha > 1$, the restrictions on $\mathscr{M}_{0,n}$ and $\mathscr{M}_{1,n}$ do not modify the rate of convergence of ours estimators. Let us consider the estimator

$$\tilde{b} = \frac{\tilde{g}}{2\tilde{f}} \quad \text{if} \quad \tilde{g} \le 2n\Delta\tilde{f} \quad \text{and} \quad \tilde{b} = 0 \quad \text{otherwise.}$$

Theorem 4.1.
If $b \in \mathscr{B}^{\alpha}_{2,\infty}$ with $\alpha > 1$, then

$$\mathbb{E}\left(\left\|\tilde{b} - b\right\|^2_{L^2}\right) \le c\left(\mathbb{E}\left(\left\|\tilde{f} - f\right\|^2_{L^2}\right) + \mathbb{E}\left(\|\tilde{g} - g\|^2_{L^2}\right) + \frac{1}{n\Delta}\right)$$

where the constant c does not depend on n nor on Δ. Then, by Theorem 2.4,

$$\mathbb{E}\left(\left\|\tilde{b} - b\right\|^2_{L^2}\right) \le c(n\Delta)^{-2\alpha/(2\alpha+3)}$$

So $\tilde{b}$ converges towards b with the minimax rate defined by Gobet *et al.* [12].

5 Simulations

5.1 Models

Ornstein-Uhlenbeck: Let us consider the SDE $dX_t = -bX_t + dW_t$ with $b > 0$. The stationary density is a Gaussian distribution $\mathscr{N}\left(0, (2b)^{-1}\right)$ and its derivative is

$$f'(x) = -\frac{2b^{3/2}}{\sqrt{\pi}} x e^{-bx^2}.$$

Hyperbolic tangent: We consider a process (X_t) satisfying the SDE

$$dX_t = -a\tanh(aX_t)dt + dW_t.$$

The stationary density related to this SDE is

$$f(x) = \frac{a}{2\cosh^2(ax)} \quad \text{and} \quad f'(x) = -\frac{a^2\tanh(ax)}{\cosh^2(ax)}.$$

Square root: Let us consider the diffusion with parameters

$$b(x) = -\frac{ax}{\sqrt{1+x^2}} \quad \text{and} \quad \sigma = 1.$$

The stationary density is

$$f(x) = c\exp\left(-2a\sqrt{1+x^2}\right) \quad \text{and} \quad f'(x) = 2b(x)f(x)$$

Model 4: We consider the following SDE:

$$dX_t = -\frac{2aX_t}{1+X_t^2}dt + dW_t.$$

The process $(X_t)_{t\geq 0}$ does not satisfy Assumption **M6** neither the sufficient conditions to be exponentially β-mixing. If $a > 1/2$, it admits the stationary density

$$f(x) = c_a\left(1+x^2\right)^{-2a} \quad \text{and} \quad f'(x) = -\frac{4c_a ax}{(1+x^2)^{1+2a}}.$$

Sine function: Let us consider the diffusion with parameters:

$$b(x) = \sin(ax) - \frac{x}{\sqrt{1+x^2}} \quad \text{and} \quad \sigma = 1.$$

Its stationary density f satisfies:

$$f(x) = c_a \exp\left(-2a^{-1}\cos(ax) - 2\sqrt{1+x^2}\right) \quad \text{and} \quad f'(x) = 2c_a b(x) f(x)$$

5.2 Estimation of the first derivative f'

Here, we estimate the first derivative f' of the stationary density on a compact set and we compare the two estimators $\tilde{g}_1$ and $\tilde{h}$ defined in Sections 2 and 3. The subspaces S_m are generated by trigonometric polynomials: those functions are orthonormal, very regular and enable very fast computations: to compute $\hat{g}_{1,m}$ (resp $\hat{h}_m$) when $\hat{g}_{1,m-1}$ (resp $\hat{h}_{m-1}$) is known, it is only necessary to compute one or two coefficients.

Figures 1-5 show the differences between the two estimators: $\tilde{g}_1$ converges whatever the sampling interval, and $\tilde{h}$ converges only if Δ is small. In that case, $\tilde{h}$ is better than $\tilde{g}_1$: the variance term is greater for $\hat{g}_{1,m}$ (is proportional to $D_m^3/(n\Delta)$) than for $\hat{h}_m$ (is p proportional to $D_m/n\Delta$).

In Tables 1-3, for each value of n and Δ, 50 exact simulations of a diffusion process are realized using the retrospective exact algorithm of Beskos *et al.* [3] (except for the Ornstein-Uhlenbeck process which is simulated using Gaussian variables). For each path, we compute the empirical risks of the estimators $\tilde{g}_1$ and $\tilde{h}$:

$$\|\tilde{g}_1 - g_1\|_E^2 := \frac{1}{M}\sum_{k=1}^{M}(\tilde{g}_1(x_k) - g_1(x_k))^2 \quad \text{and} \quad \left\|\tilde{h} - h\right\|_E^2 := \frac{1}{M}\sum_{k=1}^{M}\left(\tilde{h}(x_k) - h(x_k)\right)^2,$$

where the points x_k are equidistributed over A. To check that the estimator is adaptive, the oracles

$$or_g = \frac{\|\tilde{g}_1 - g_1\|_E^2}{\min_{m\in\mathscr{M}_n}\|\hat{g}_{1,m} - g_1\|_E^2} \quad \text{and} \quad or_h = \frac{\left\|\tilde{h} - h\right\|_E^2}{\min_{m\in\mathscr{M}_n}\left\|\hat{h}_m - h\right\|_E^2}$$

are computed. The mean time of simulation t_{sim} of a process is measured, and for each kind of estimator, the means of the empirical risk ris_g or ris_h, of the oracles $\bar{or}_g$ or $\bar{or}_h$ and of the computation times t_g or t_h or computed.

The complexity of the retrospective exact algorithm of Beskos *et al.* [3] is proportional to $ne^{c\Delta}$ where c depends on the model. Table 3 shows that for Model 4, t_{sim} increases when n or Δ increases. For the hyperbolic tangent, the time of simulation only depends on n because the constant c is exactly equal to 0. The Ornstein-Uhlenbeck process is not simulated thanks to the retrospective algorithm, so its time of simulation does not depend on Δ. Tables 1-3 show that the first estimator $\tilde{g}_1$ is always faster to compute than the second one $\tilde{h}$. This is mainly because we have less models to test: for the first estimator, the maximal dimension $\mathscr{D}_n$ is bounded by $(n\Delta)^{1/4}$ whereas for the second estimator, $\mathscr{D}_n \leq (n\Delta)^{1/2}$.

When $\Delta = 1$, $\tilde{g}_1$ is better than $\tilde{h}$. If not, the estimators are similar and become better when $n\Delta$ increases. For the Ornstein-Uhlenbeck process and the hyperbolic tangent, the process $(X_t)_{t\geq 0}$ is exponentially β-mixing and $\tilde{g}_1$ is in general better than $\tilde{h}$. For Model 4, the process (X_t) is not exponentially β-mixing and when $\Delta < 1$, $\tilde{h}$ is (in general) better than $\tilde{g}_1$.

Figure 1: Ornstein-Uhlenbeck: estimation of f'

$n = 10^4$, $\Delta = 1$ $\qquad$ $n = 10^5$, $\Delta = 10^{-2}$

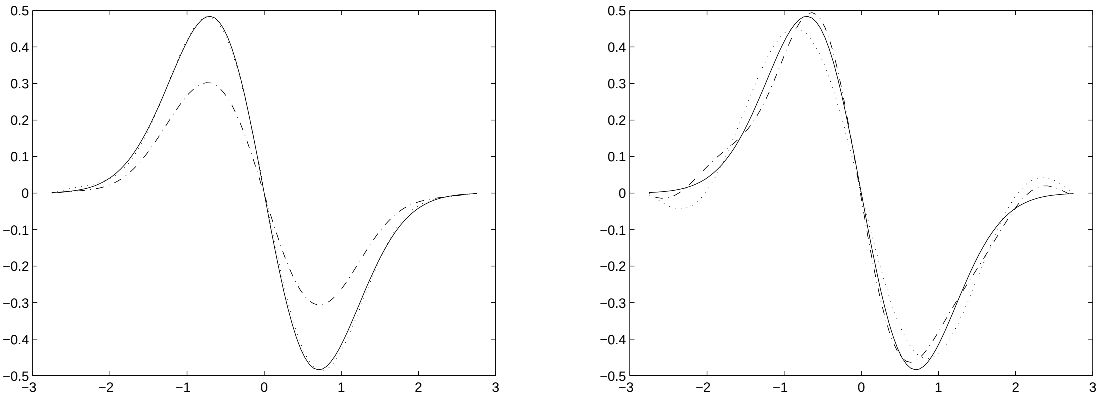


Figure 2: Hyperbolic tangent: estimation of f'

$n = 10^4$, $\Delta = 1$ $\qquad$ $n = 10^5$, $\Delta = 10^{-2}$

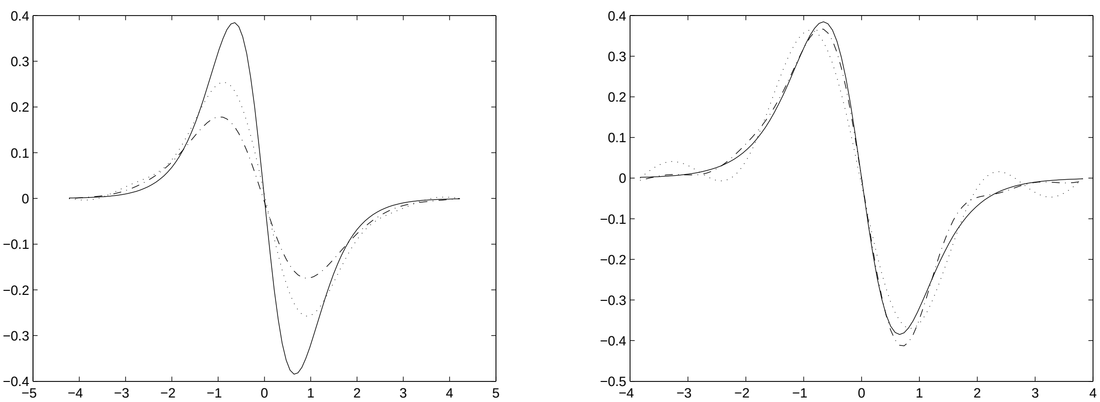
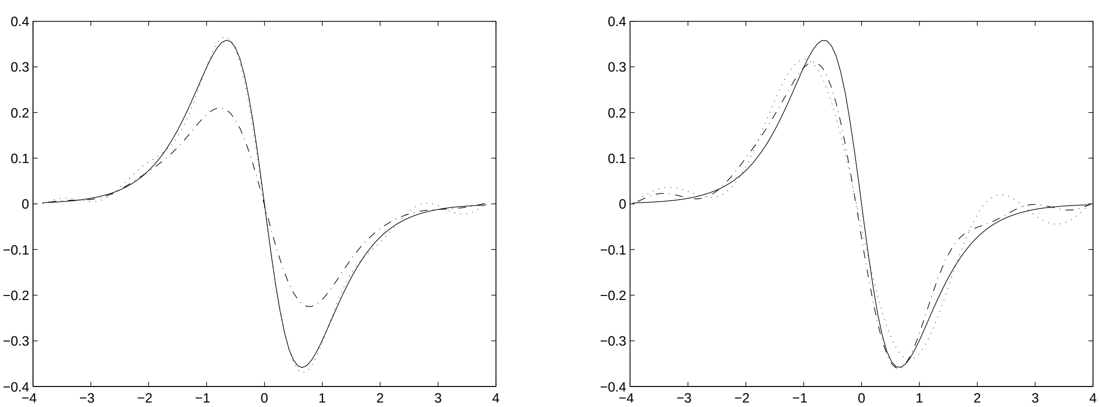


Figure 3: Square root: estimation of f'

$n = 10^4$, $\Delta = 1$ $\qquad$ $n = 10^4$, $\Delta = 10^{-1}$



- : true derivative

$\cdots$: estimator $\tilde{g}_1$ (differentiating an estimator of f)

-. : estimator $\tilde{h}$ (using to $f' = 2bf$)

Figure 4: Model 4: estimation of f'

$n = 10^4$, $\Delta = 1$ $\qquad$ $n = 10^4$, $\Delta = 10^{-1}$

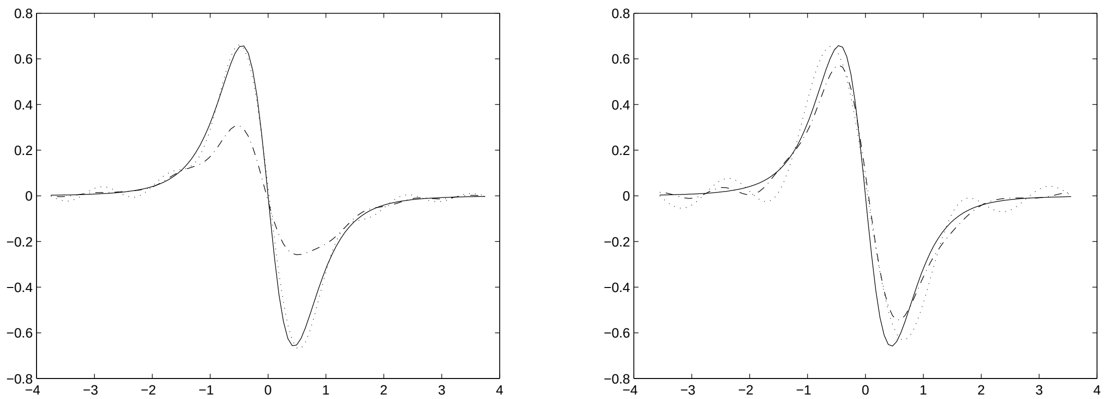
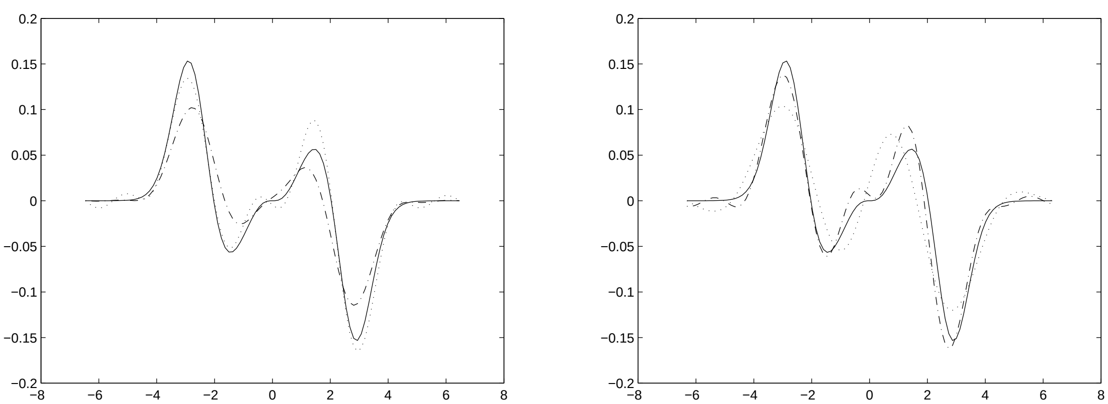


Figure 5: Sine function: estimation of f'

$n = 10^4$, $\Delta = 1$ $\qquad$ $n = 10^5$, $\Delta = 10^{-2}$



- : true derivative

$\cdots$: estimator $\tilde{g}_1$ (differentiating an estimator of f)

-. : estimator $\tilde{h}$ (using to $f' = 2bf$)

Table 1: Estimation of f' for Ornstein-Uhlenbeck

			first estimator			second estimator		
n	Δ	t_{sim}	ris_g	$\bar{or}_g$	t_g	ris_h	$\bar{or}_h$	t_h
10^4	1	0.10	0.00025	2.5	0.33	0.0090	1.0	0.73
10^4	10^{-1}	0.10	0.0010	1.8	0.17	0.00091	1.2	0.68
10^4	10^{-2}	0.099	0.0060	2.6	0.097	0.0067	2.3	0.66
10^3	1	0.0027	0.0023	4.2	0.034	0.0097	1.0	0.12
10^3	10^{-1}	0.0025	0.0058	3.0	0.020	0.0077	2.3	0.12
10^3	10^{-2}	0.0026	0.037	3.0	0.0070	0.078	4.0	0.035
10^2	1	0.00022	0.0080	2.0	0.013	0.019	1.5	0.062
10^2	10^{-1}	0.00021	0.035	2.4	0.0046	0.078	5.5	0.019
10^2	10^{-2}	0.00023	0.067	2.1	0.0048	0.11	1.4	0.0068

Table 2: Hyperbolic tangent: estimation of f'

			first estimator			second estimator		
n	Δ	t_{sim}	ris_g	$\bar{or}_g$	t_g	ris_h	$\bar{or}_h$	t_h
10^4	1	6.2	0.0027	1.1	0.33	0.0087	1.03	0.71
10^4	10^{-1}	1.2	0.0018	3.7	0.17	0.0014	1.4	0.68
10^4	10^{-2}	1.7	0.0065	2.8	0.10	0.0056	1.8	0.65
10^3	1	0.61	0.0040	1.5	0.034	0.0097	1.1	0.12
10^3	10^{-1}	0.19	0.0067	2.8	0.020	0.0087	2.1	0.12
10^3	10^{-2}	0.16	0.022	2.5	0.0068	0.036	2.6	0.03
10^2	1	0.066	0.011	1.7	0.014	0.021	1.80	0.063
10^2	10^{-1}	0.020	0.023	2.3	0.0048	0.044	3.4	0.020
10^2	10^{-2}	0.018	0.033	1.6	0.0054	0.078	1.2	0.0080

Table 3: Model 4: estimation of f'

			first estimator			second estimator		
n	Δ	t_{sim}	ris_g	$\bar{or}_g$	t_g	ris_h	$\bar{or}_h$	t_h
10^4	1	6.6	0.00073	1.8	0.33	0.020	1.0	0.71
10^4	10^{-1}	2.3	0.0032	4.2	0.17	0.0019	1.3	0.70
10^4	10^{-2}	2.1	0.016	3.8	0.10	0.0090	1.7	0.68
10^3	1	0.67	0.0049	2.4	0.035	0.022	1.1	0.12
10^3	10^{-1}	0.24	0.017	3.6	0.021	0.013	2.0	0.12
10^3	10^{-2}	0.18	0.043	2.0	0.0071	0.094	3.5	0.035
10^2	1	0.071	0.048	8.1	0.014	0.041	1.6	0.065
10^2	10^{-1}	0.022	0.046	1.91	0.0049	0.077	3.1	0.02
10^2	10^{-2}	0.019	0.070	1.4	0.005	0.12	1.1	0.0069

ris_g and ris_h: average empirical risks related for $\tilde{g}_1$ and $\tilde{h}$

$\bar{or}_g$ and $\bar{or}_h$: average oracles (empirical risks of $\tilde{g}_1$ (resp $\tilde{h}$) over the empirical risk of the best estimator $\hat{g}_{1,m}$ (resp $\hat{h}_m$))

t_g et t_h : average time of computation of $\tilde{g}_1$ and $\tilde{h}$ (times in seconds)

t_{sim}: average times of simulation of $(X_0, X_\Delta, \dots, X_{n\Delta})$ (times in seconds)

5.3 Drift estimation by quotient

Two drift estimators are compared: the estimator by quotient defined in Section 4, denoted here by $\tilde{b}_{quot}$, and a penalized least-square estimator denoted by $\tilde{b}_{pls}$. The construction of the last estimator is done in Comte *et al.* [6]. It only converges when the sampling interval Δ is small, but in that case, it reaches the minimax rate of convergence: if b belongs to a Besov space $\mathscr{B}^{\alpha}_{2,\infty}$, then the risk of the estimator $\tilde{b}_{pls}$ is bounded by

$$\mathbb{E}\left(\left\|\tilde{b}_{pls}-b\right\|^2_{L^2}\right)\leq C\left((n\Delta)^{-2\alpha/(2\alpha+1)}+\Delta\right).$$

Figures 6-10 show that, for low-frequency data, the quotient estimator $\tilde{b}_{quot}$ is better than $\tilde{b}_{pls}$. For various values of n and Δ, 50 exact simulations of $(X_0,\dots,X_{n\Delta})$ are realized and estimators $\tilde{b}_{quot}$ and $\tilde{b}_{pls}$ are computed. Table 4 and 5 give the average empirical risk for these estimators and the average computation times. The lowest risk is set in bold.

Tables 4 and 5 underline that the first estimator is always faster than the second one: to compute $\tilde{b}_{pls}$, we have to inverse a matrix $m\times m$ over each space S_m. When Δ is small and the time of observation $n\Delta$ is large, the penalized least square contrast estimator converges better than the quotient estimator. Of course, when Δ is fixed, $\tilde{b}_{quot}$ converges faster than $\tilde{b}_{pls}$.

Table 4: Ornstein-Uhlenbeck: estimation of b

		quotient estimator		least-square estimator	
n	Δ	ris_{quot}	t_{quot}	ris_{pls}	t_{pls}
10^4	1	**0.0022**	3.6	0.089	7.3
10^4	10^{-1}	0.0086	1.2	**0.0049**	1.7
10^4	10^{-2}	0.069	0.4	**0.031**	0.7
10^3	1	**0.011**	0.2	0.090	0.7
10^3	10^{-1}	0.061	0.06	**0.022**	0.3
10^3	10^{-2}	**0.31**	0.02	0.50	0.004
10^2	1	**0.073**	0.03	0.085	0.3
10^2	10^{-1}	**0.25**	0.01	0.34	0.003

Table 5: Hyperbolic tangent: estimation of b

		quotient estimator		least-square estimator	
n	Δ	ris_{quot}	t_{quot}	ris_{pls}	t_{pls}
10^4	1	**0.0023**	3.6	0.086	7.2
10^4	10^{-1}	0.019	1.2	**0.017**	1.8
10^4	10^{-2}	0.078	0.4	**0.052**	0.7
10^3	1	**0.036**	0.2	0.18	0.7
10^3	10^{-1}	0.12	0.06	**0.065**	0.3
10^3	10^{-2}	**0.17**	0.02	0.61	0.004
10^2	1	0.24	0.03	**0.10**	0.3
10^2	10^{-1}	**0.20**	0.01	0.53	0.003

ris_{quot} and ris_{pls}: average empirical risks for $\tilde{b}_{quot}$ and $\tilde{b}_{pls}$
t_{quot} and t_{pls} : average computation times of $\tilde{b}_{quot}$ and $\tilde{b}_{pls}$ (times in seconds)

6 Proofs

6.1 Important lemmas

Lemma 6.1 : Variance of β-mixing variables.
Let us set

$$A=\frac{1}{n}\sum_{k=1}^{n}g(X_{k\Delta})-\mathbb{E}\left(g(X_{k\Delta})\right).$$

Figure 6: Ornstein-Uhlenbeck: estimation of b
$n = 10^4$, $\Delta = 1$

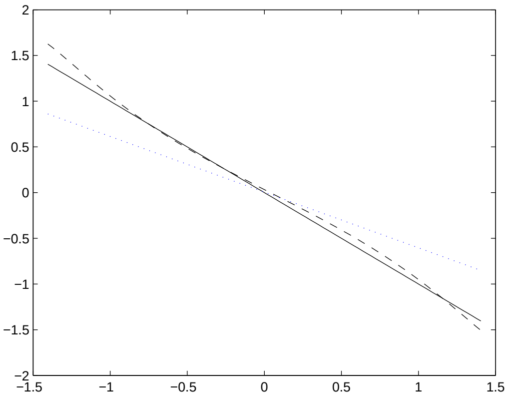


Figure 7: Hyperbolic tangent: estimation of b
$n = 10^4$, $\Delta = 1$

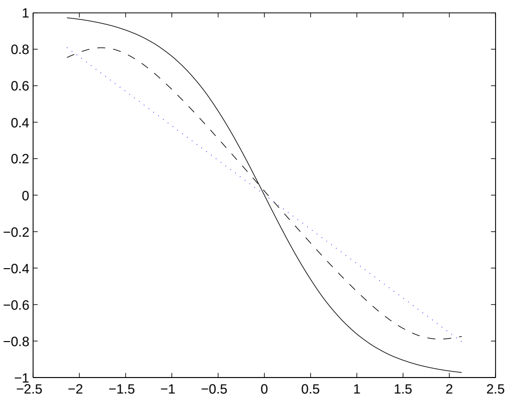
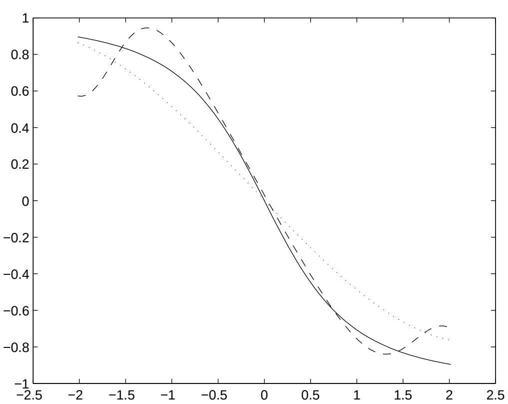


Figure 8: Square root: estimation of b
$n = 10^4$, $\Delta = 1$



- : true drift b
$--$: estimation of b by quotient: $\tilde{b}_{quot}$
.. : estimation of b like in Comte *et al.* [6]: $\tilde{b}_{pls}$

Figure 9: Model 4: estimation of b
$n = 10^4$, $\Delta = 10^{-1}$

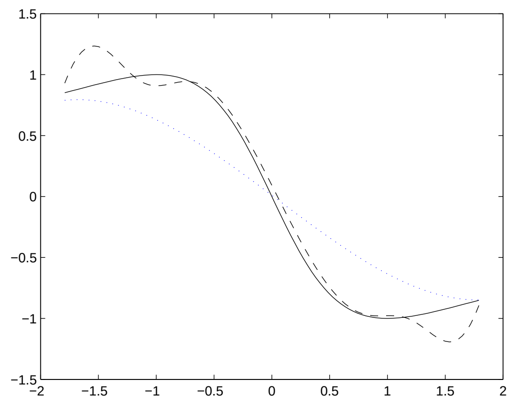
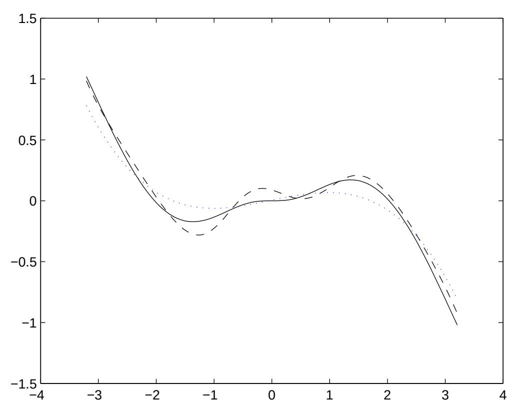


Figure 10: Sine function: estimation of b
$n = 10^4$, $\Delta = 1$



- : true drift b
$--$: estimation of b by quotient: $\tilde{b}_{quot}$
.. : estimation of b like in Comte *et al.* [6]: $\tilde{b}_{pls}$

If the random variables $(X_{k\Delta})$ are strictly stationary and β-mixing, then there exists a function B such that

$$\mathbb{E}\left(B(X_0)\right) \leq \sum_{k=1}^{+\infty} \beta_{k\Delta} \quad \textit{and} \quad \mathbb{E}\left(B^2(X_0)\right) \leq \sum_{k=1}^{+\infty} k\beta_{k\Delta}$$

and, for any function g such that $\mathbb{E}\left(g^2(X_0)\right) < +\infty$,

$$Var(A) \leq \frac{4}{n}\mathbb{E}\left(B(X_0)g^2(X_0)\right)$$

Moreover, if the β-mixing coefficients are such that $\beta_X(k) \leq \beta_0 e^{-\theta\Delta k}$ (that is if $(X_{k\Delta})$ are exponentially β-mixing), then if $\theta\Delta \geq 1$:

$$\sum_{k=1}^{+\infty} \beta_{k\Delta} \leq 2\beta_0 \quad \textit{and} \quad \sum_{k=1}^{+\infty} k\beta_{k\Delta} \leq 2\beta_0$$

and if $\Delta\theta \leq 1$ and $n\Delta \to \infty$:

$$\sum_{k=1}^{n} \beta_{k\Delta} \leq \frac{2\beta_0}{\Delta\theta} \quad \textit{and} \quad \sum_{k=1}^{n} k\beta_{k\Delta} \leq \frac{2\beta_0}{\Delta^2\theta^2}.$$

If the random variables $(X_{k\Delta})$ are arithmetically β-mixing, then:

$$\textit{if } \theta\Delta > 1, \textit{ then } \sum_{k=1}^{+\infty} \beta_{k\Delta} \leq 2\beta_0 \quad \textit{and if } \theta > 1, \quad \sum_{k=1}^{+\infty} k\beta_{k\Delta} \leq \frac{2\beta_0}{\theta - 1}$$
$$\textit{if } \theta\Delta \leq 1, \textit{ then } \sum_{k=1}^{n} \beta_{k\Delta} \leq \frac{2\beta_0}{\Delta\theta} \quad \textit{and if } \theta > 1, \quad \sum_{k=1}^{n} k\beta_{k\Delta} \leq \frac{2\beta_0}{\Delta^2(\theta - 1)}.$$

This lemma is proved in Viennet [24].

Lemma 6.2 : Coupling method for the construction of independent variables. *Let us consider a stationary and β-mixing process $(X_t)_{t\geq 0}$ observed at discrete times $t = 0, \Delta, \ldots, n\Delta$. Let us set $n = 2q_n p_n$ where $q_n = \frac{(2l+1)\ln(n)}{\theta\Delta}$ and, for $a \in \{0,1\}$, $1 \leq k \leq p_n$,*

$$U_{k,a} = \left(X_{((2(k-1)+a)q_n+1)\Delta}, \ldots, X_{(2k-1+a)q_n\Delta}\right).$$

According to Berbee's Lemma (see Viennet [24]), there exist random variables $(X^*_\Delta, \ldots, X^*_{n\Delta})$ such that the random vectors

$$U^*_{k,a} = \left(X^*_{((2(k-1)+a)q_n+1)\Delta}, \ldots, X^*_{(2k-1+a)q_n\Delta}\right) \quad \textit{where} \quad a \in \{0,1\},\ 1 \leq k \leq p_n$$

satisfy:

- *For any $a \in \{0,1\}$, vectors $U^*_{0,a}, \ldots, U^*_{(p_n-1),a}$ are independent.*
- *For any $a \in \{0,1\}$, any k, $1 \leq k \leq p_n$, $U^*_{k,a}$ and $U_{k,a}$ have the same law.*
- *For any $a \in \{0,1\}$, $1 \leq k \leq p_n$:*

$$\mathbb{P}\left(U_{k,a} \neq U^*_{k,a}\right) \leq \beta_X\left(q_n\Delta\right)$$

Let us set

$$\Omega^* = \left\{U_{k,a} = U^*_{k,a}, k = 1, \ldots, n,\ a = \{0,1\}\right\}.$$

If the process is exponentially β-mixing, then $\mathbb{P}\left(\Omega^{*c}\right) \leq 2p_n\beta_X(q_n) \leq n^{-2l}$.

Lemma 6.3 : Talagrand inequality.
Let us consider some random variables $X_1, \dots, X_n$ independent and identically distributed. Let us set $g_n : t \in \mathscr{B} \to g_n(t)$ where $\mathscr{B}$ is a countable set and

$$g_n(t) = \frac{1}{n} \sum_{k=1}^{n} F_t(X_k) - \mathbb{E}\left(F_t(X_k)\right).$$

If

$$\sup_{t \in \mathscr{B}} \|F_t\|_\infty \le M_1, \quad \mathbb{E}\left(\sup_{t \in \mathscr{B}} |g_n(t)|\right) \le H, \quad \sup_{t \in \mathscr{B}} Var(F_t(X_k)) \le V,$$

then

$$\mathbb{E}\left(\sup_{t \in \mathscr{B}} g_n^2(t) - 12H^2\right)_+ \le C\left(\frac{V}{n}\exp\left(-k_1 \frac{nH^2}{V}\right) + \frac{M_1^2}{n^2}\exp\left(-k_2 \frac{nH}{M_1}\right)\right)$$

with $k_1 = 1/6$, $k_2 = 1/(21\sqrt{2})$, and C a universal constant.There exist a constant κ independent of the process (X_t) and of the function F_t such that:

$$\mathbb{P}\left(\sup_{t \in \mathscr{B}} |g_n(t)| \ge 2H + \lambda\right) \le 3\exp\left(-\kappa n \min\left(\frac{\lambda^2}{2V}, \frac{\lambda}{7M_1}\right)\right) \tag{6.1}$$

This proof is done in Lacour [14] p156 and in Comte and Merlevède [7] p.224.

6.2 Proofs of Theorems 2.1 and 2.2

We only prove here Theorem 2.2 (the proof of Theorem 2.1 is very similar and easier). According to Pythagoras, we have

$$\|\hat{g}_{j,m,N} - g\|_{L^2}^2 = \|g_{j,m,N} - g\|_{L^2}^2 + \|\hat{g}_{j,m,N} - g_{j,m,N}\|_{L^2}^2.$$

Let us set $a_\lambda := \int_{\mathbb{R}} f^{(j)}(x)\varphi_{\lambda,m}(x)dx$. By Assumption **S2** *2.*, $a_\lambda = (-1)^j \int_{\mathbb{R}} f(x)\varphi_{\lambda,m}^{(j)}(x)dx$. Let us set $\hat{a}_\lambda = \frac{(-1)^j}{n}\sum_{k=1}^n \varphi_{\lambda,m}^{(j)}(X_{k\Delta})$. We have

$$\|\hat{g}_{j,m,N} - g_{j,m,N}\|_{L^2}^2 = \sum_{\lambda \in \Lambda_{m,N}} (\hat{a}_\lambda - a_\lambda)^2$$

and

$$\mathbb{E}\left((\hat{a}_\lambda - a_\lambda)^2\right) = \mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^{n} \varphi_{\lambda,m}^{(j)}(X_{k\Delta})\right).$$

According to Lemma 6.1,

$$\mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^{n} \varphi_{\lambda,m}^{(j)}(X_{k\Delta})\right) \le \frac{4}{n}\mathbb{E}\left(B(X_0)\left(\varphi_{\lambda,m}^{(j)}(X_0)\right)^2\right)$$

where $\mathbb{E}\left(B(X_0)\right) \le 2\beta_0\left(1 \vee \frac{1}{\theta\Delta}\right)$. So, by Assumption **S2** *3.*,

$$\mathbb{E}\left(\|\hat{g}_{j,m,N} - g_{j,m,N}\|_{L^2}^2\right) \le \frac{4}{n}\mathbb{E}\left(B(X_0)\Psi_{j,m}^2(X_0)\right) \le 8\beta_0\psi_j \frac{2^{(2j+1)m}}{n}\left(1 \vee \frac{1}{\theta\Delta}\right).$$

6.3 Proofs of Theorems 2.3 and 2.4

As previously, only Theorem 2.4 is demonstrated. Let us set

$$\nu_{j,n}(t) = \frac{1}{n}\sum_{k=1}^{n} t^{(j)}(X_{k\Delta}) - \int_{\mathbb{R}} t^{(j)}(x)f(x)dx.$$

For any m, we have

$$\gamma_{j,n}(\tilde{g}_j) + pen_j(\hat{m}_j) \leq \gamma_{j,n}(\hat{g}_{j,m,N_n}) + pen_j(m) \leq \gamma_{j,n}(g_{j,m,N_n}) + pen_j(m).$$

As, for any $t \in S_{m,N}$,

$$\gamma_{j,n}(t) = \|t-g\|_{L^2}^2 - \|g\|_{L^2}^2 + 2\nu_{j,n}(t),$$

for any $m \in \mathbb{N}$,

$$\|\tilde{g}_j - g\|_{L^2}^2 \leq \|g_{j,m,N_n} - g\|_{L^2}^2 + 2\nu_{j,n}\left(g_{j,m,N_n} - \tilde{g}_j\right) + pen_j(m) - pen_j(\hat{m}_j).$$

According to Cauchy-Schwartz, if we set $\mathscr{B}_{m,m'} = \left\{t \in S_{m,N_n} + S_{m',N_n},\ \|t\|_{L^2}^2 \leq 1\right\}$, we have:

$$\|\tilde{g}_j - g\|_{L^2}^2 \leq \|g_{j,m,N_n} - g\|_{L^2}^2 + \frac{1}{4}\|\tilde{g}_j - g_{j,m,N_n}\|_{L^2}^2 + 4\sup_{t\in\mathscr{B}_{m,\hat{m}}} \nu_{j,n}(t) + pen_j(m) - pen_j(\hat{m}_j).$$

As $\|\tilde{g}_j - g_{j,m,N_n}\|_{L^2}^2 \leq 2\|g_{j,m,N_n} - g\|_{L^2}^2 + 2\|\tilde{g}_j - g\|_{L^2}^2$:

$$\|\tilde{g}_j - g\|_{L^2}^2 \leq 3\|g_{j,m,N_n} - g\|_{L^2}^2 + 8\sup_{t\in\mathscr{B}_{\hat{m},m}} \nu_{j,n}^2(t) + pen_j(m) - pen_j(\hat{m}_j).$$

Let us consider a function $p_j(m,m')$ such that $8p_j(m,m') = pen_j(m) + pen_j(m')$. We have that

$$\begin{aligned} E: &= \mathbb{E}\left(8\sup_{t\in\mathscr{B}_{m,\hat{m}}} \nu_{j,n}^2(t) + pen_j(m) - pen_j(\hat{m}_j)\right) \\ &= 8\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,\hat{m}}} \nu_{j,n}^2(t) - p_j(m,\hat{m}_j)\right) + 2pen_j(m). \end{aligned}$$

Let us use the set Ω^* described in Lemma 6.2 where q_n is defined later. Let us set, for $a \in \{0,1\}$, $0 \leq k \leq p_n - 1$,

$$U_{k,a}^* = \frac{1}{q_n}\sum_{l=1}^{q_n} t^{(j)}\left(X^*_{((2k+a)q_n+l)\Delta}\right) \quad , \quad U_{k,a} = \frac{1}{q_n}\sum_{l=1}^{q_n} t^{(j)}\left(X_{((2k+a)q_n+l)\Delta}\right)$$

and

$$\nu_{j,n}^*(t) = \frac{1}{n}\sum_{k=1}^{n} t^{(j)}(X^*_{k\Delta}) - \mathbb{E}^*\left(t^{(j)}(X^*_{k\Delta})\right).$$

We have:

$$\sup_{t\in\mathscr{B}_{m,\hat{m}}} \nu_{j,n}^2(t) - p_j(m,\hat{m}_j) \leq \sup_{t\in\mathscr{B}_{m,\hat{m}}} \left\{\left(\nu_{j,n}^*(t)\right)^2 - p_j(m,\hat{m}_j)\right\} + \sup_{t\in\mathscr{B}_{m,\hat{m}}} \left\{\left|\nu_{j,n}^2(t) - \left(\nu_{j,n}^*(t)\right)^2\right|\right\}.$$

According to Lemma 6.2, the random variables $\left(U_{k,0}^*\right)$ are independent and identically distributed, and so are the variables $\left(U_{k,1}^*\right)$.

Bound of $\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,\hat{m}}}\left\{\left(\nu_{j,n}^*(t)\right)^2 - p_j(m,\hat{m}_j)\right\}\right)$ We have that

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,\hat{m}}} \left(\nu_{j,n}^*(t)\right)^2 - p_j(m,\hat{m}_j)\right) \leq \sum_{m'} \mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,m'}} \left(\nu_{j,n}^*(t)\right)^2 - p_j(m,m')\right). \tag{6.2}$$

Let us set, for $a \in \{0,1\}$, $0 \leq k \leq p_n - 1$,

$$\nu_{j,n,a}^*(t) = \frac{1}{2p_n}\sum_{k=1}^{p_n} U_{k,a}^* - \mathbb{E}\left(U_{k,a}^*\right).$$

We have that:

$$\nu_{j,n}^*(t) = \nu_{j,n,0}^*(t) + \nu_{j,n,1}^*(t)$$

We want to apply Lemma 6.3 to the random variables $U_{k,a}^*$. So we compute H^2, V and M_1 such that

$$\sup_{t\in\mathscr{B}_{m,m'}} \left\|U_{k,i}^*\right\|_\infty \leq M_1, \quad \mathrm{Var}\left(U_{k,j}^*\right) \leq V \quad \text{and} \quad \mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,m'}} \left(\nu_{j,n}^*(t)\right)^2\right) \leq H^2.$$

Let us denote by $\{\varphi_\lambda,\ \lambda\in\Lambda\}$ an orthonormal basis of $S_{m,N}+S_{m',N}$ and set $D=2^m+2^{m'}$. By Assumption **S2** *3.-4.*, we have

$$\sup_{t\in\mathscr{B}_{m,m'}} \left\|U_{k,a}^*\right\|_\infty \leq \left\|t^{(j)}(X_0)\right\|_\infty \leq \sqrt{\psi_j} D^{(2j+1)/2}.$$

By Lemma 6.1:

$$\begin{aligned}
\mathrm{Var}\left(U_{k,a}^*\right) &\leq \frac{4}{q_n}\mathbb{E}\left(\left(t^{(j)}(X_0)\right)^2 B(X_0)\right) \leq \frac{4}{q_n}\left\|t\right\|_\infty \left(\mathbb{E}\left(\left(t^{(j)}(X_0)\right)^2\right)\right)^{1/2}\left(\mathbb{E}\left(B^2(X_0)\right)\right)^{1/2} \\
&\leq CD^{2j+1/2}\left(\frac{1}{q_n}\vee\frac{1}{q_n\Delta}\right).
\end{aligned}$$

Besides,

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,m'}}\left(\nu_{j,n,a}^*(t)\right)^2\right) = \mathbb{E}\left(\sup_{\sum_{\lambda\in\Lambda}\alpha_\lambda^2\leq 1}\left(\sum_{\lambda\in\Lambda}\alpha_\lambda\nu_{j,n,a}^*(\varphi_\lambda)\right)^2 \mathbb{1}_{\Omega^*}\right) \leq \sum_{\lambda\in\Lambda}\mathbb{E}\left(\left(\nu_{j,n,a}^*\left(\varphi_\lambda\right)\right)^2\right)$$

and

$$\mathbb{E}\left(\left(\nu_{j,n,a}^*\left(\varphi_\lambda\right)\right)^2\right) = \mathrm{Var}\left(\frac{1}{2n}\sum_{k=1}^{p_n}\sum_{l=1}^{q_n}\varphi_\lambda^{(j)}\left(X_{((2k+a)q_n+l)\Delta}^*\right)\right).$$

The random variables $(X_{k\Delta}^*)$ are exponentially β-mixing, so according to Lemma 6.1:

$$\mathbb{E}\left(\left(\nu_{j,n,a}^*\left(\varphi_\lambda\right)\right)^2\right) \leq \frac{4}{n}\mathbb{E}\left(B(X_0)\left(\varphi_\lambda^{(j)}(X_0)\right)^2\right) \quad \text{where} \quad \mathbb{E}\left(B(X_0)\right) \leq 2\beta_0\left(\frac{1}{n}\vee\frac{1}{n\theta\Delta}\right).$$

Thus, by Assumption **S2** *3.*, we have:

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_m}\left(\nu_{j,n,a}^*(t)\right)^2\right) \leq \frac{4}{n}\mathbb{E}\left(B(X_0)\left(\Psi_{j,m}^2(X_0)+\Psi_{j,m'}^2(X_0)\right)\right) \leq 16\beta_0\psi_j\frac{D^{(2j+1)}}{n}\left(1\vee\frac{1}{\theta\Delta}\right),$$

and it follows:

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,m'}}\left(\nu_{j,n}^*(t)\right)^2\right) \leq 32\beta_0\psi_j\frac{D^{(2j+1)}}{n}\left(1\vee\frac{1}{\theta\Delta}\right).$$

Let us set

$$F := \mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,m'}}\left(\nu_{j,n}^2(t)-p_j(m,m')\right)\mathbb{1}_{\Omega^*}\right)_+.$$

We can apply Lemma 6.3 with $H^2 = 32\beta_0\psi_j D^{(2j+1)}\left(\frac{1}{n}\vee\frac{1}{n\theta\Delta}\right)$, $M_1=\sqrt{\psi_j}D^{(2j+1)/2}$ and $V = cD^{2j}$. Let us set $p_j(m,m')=12H^2$. We find:

$$F \leq C\left(\frac{D^{2j+1/2}}{n\Delta}\exp\left(-cD^{1/2}\right)+\frac{D^{2j+1}}{p_n^2}\exp\left(-c\frac{p_n}{\sqrt{n\Delta}}\right)\right).$$

where c and C are two constants independents of D, n and Δ.

As $D=2^m+2^{m'}$ and $2^{m'}\geq m'$ for any $m'\geq 0$:

$$\sum_{m'} D^{2j+1/2}\exp\left(-cD^{1/2}\right) \leq \sum_{k=1}^{\infty} k^{2j+1/2}\exp\left(-ck^{1/2}\right) \leq C.$$

Besides,

$$\sum_{m'} D^{2j+1} \leq \sum_{k=1}^{\mathscr{D}_{j,n}} k^{2j+1} \leq \mathscr{D}_{j,n}^{2j+2} \leq n\Delta$$

and if there exists $\eta > 0$ such that

$$p_n = \frac{n}{2q_n} \geq (n\Delta)^{1/2+\eta}, \tag{6.3}$$

then:

$$\mathbb{E}\left(\left(\sup_{t\in\mathscr{B}_{m,\hat{m}}} \left(\nu_{j,n}^*(t)\right)^2 - p_j(m,m')\right)\right)_+ \leq \frac{C}{n\Delta}.$$

Bound of $\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,\hat{m}}}\left\{\left|\nu_{j,n}^2(t) - \left(\nu_{j,n}^*(t)\right)^2\right|\right\}\right)$ We have that:

$$\sup_{t\in\mathscr{B}_{m,\hat{m}}}\left\{\left|\nu_{j,n}^2(t) - \left(\nu_{j,n}^*(t)\right)^2\right|\right\} \leq \sum_{m'} \sup_{t\in\mathscr{B}_{m,m'}}\left\{\left|\nu_{j,n}^2(t) - \left(\nu_{j,n}^*(t)\right)^2\right|\right\}$$

and

$$\left|\nu_{j,n}(t) - \nu_{j,n}^*(t)\right| \leq \frac{1}{2p_n}\sum_{a=0}^{1}\sum_{k=1}^{p_n}\left|U_{k,a} - U_{k,a}^*\right| \leq 2\left\|t^{(j)}\right\|_\infty \sum_{a=0}^{1}\sum_{k=1}^{p_n} \mathbb{1}_{U_{k,a}\neq U_{k,a}^*}.$$

Moreover,

$$\left|\nu_{j,n}(t) + \nu_{j,n}^*(t)\right| \leq \frac{1}{2p_n}\sum_{a=0}^{1}\sum_{k=1}^{p_n}\left|U_{k,a} + U_{k,a}^*\right| + 2\left|\mathbb{E}\left(U_{1,0}\right)\right| \leq 4\left\|t^{(j)}\right\|_\infty.$$

Lemma 6.2 and Assumption **S2** *3.* ensures that:

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,m'}}\left\{\left|\nu_{j,n}^2(t) - \left(\nu_{j,n}^*(t)\right)^2\right|\right\}\right) \leq 8\sup_{t\in\mathscr{B}_{m,m'}}\left\{\left\|t^{(j)}\right\|_\infty^2\right\}\mathbb{P}\left(U_{1,0}\neq U_{1,0}^*\right) \leq 8\psi_j\mathscr{D}_n^{2j+1}\beta_X\left(q_n\Delta\right)$$

then

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,\hat{m}}}\left\{\left|\nu_{j,n}^2(t) - \left(\nu_{j,n}^*(t)\right)^2\right|\right\}\right) \leq 8\psi_j\mathscr{D}_n^{2j+2}\beta_X(q_n\Delta).$$

As $\mathscr{D}_{j,n}^{2j+2} \simeq n\Delta$, and $\beta_X\left(q_n\Delta\right) \leq \beta_0\left(1+q_n\Delta\right)^{-(1+\theta)}$, we want that:

$$\left(1+q_n\Delta\right)^{-(1+\theta)} \leq (n\Delta)^{-2}. \tag{6.4}$$

Choice of q_n The integers q_n and $p_n = n/(2q_n)$ have to satisfy the inequalities (6.3) and (6.4). If the process is exponentially β-mixing, then $q_n = (l+1)\ln(n)/(\theta\Delta)$ with $l \in \mathbb{N}\setminus\{0\}$ fits. If the process is arithmetically β-mixing, let us set $q_n = (n\Delta)^\alpha/\Delta$. According to inequalities (6.3) and (6.4), we need:

$$\exists\eta > 0,\ \alpha \leq \frac{1}{2} - \eta \quad \text{and} \quad \alpha \geq \frac{2}{1+\theta}.$$

This condition can only be fulfilled if $\theta > 3$. In that case, we can set $\alpha = 2/(1+\theta)$.

Collecting the results, we obtain:

$$\mathbb{E}\left(\|\tilde{g}_j - g_j\|_{L^2}^2\right) \leq C\inf_{m\in\mathscr{M}_n}\left(\|g_{j,M,N_n} - g_j\|_{L^2}^2 + pen_j(m)\right) + \frac{c}{n}\left(1\vee\frac{1}{\Delta}\right).$$

6.4 Proof of Theorems 3.1 and 3.2

We only prove Theorem 3.2. We have that $\Delta^{-1}\left(X_{(k+1)\Delta}-X_{k\Delta}\right)=I_{k\Delta}+Z_{k\Delta}+b(X_{k\Delta})$ (see(3.5)). Then

$$\Gamma_n(t)-\Gamma_n(s)=\|t\|_{L^2}^2-\|s\|_{L^2}^2-\frac{4}{n}\sum_{k=1}^{n}\left(I_{k\Delta}+Z_{k\Delta}+b(X_{k\Delta})\right)\left(t(X_{k\Delta})-s(X_{k\Delta})\right).$$

Moreover,

$$\begin{aligned}\|t-h\|_{L^2}^2 &= \|t\|_{L^2}^2+\|h\|_{L^2}^2-2\int t(x)h(x)dx=\|t\|_{L^2}^2+\|h\|_{L^2}^2-4\int t(x)b(x)f(x)dx\\ &= \|t\|_{L^2}^2+\|h\|_{L^2}^2-4\mathbb{E}\left(b(X_{k\Delta})t(X_{k\Delta})\right).\end{aligned}$$

Then

$$\Gamma_n(t)-\Gamma_n(s)=\|t-h\|_{L^2}^2-\|s-h\|_{L^2}^2-2\nu_n(t-s)-2\rho_n(t-s)-2\xi_n(t-s)$$

where

$$\begin{aligned}\nu_n(t) &= \frac{2}{n}\sum_{k=1}^{n}\mathbb{E}\left(I_{k\Delta}t(X_{k\Delta})\right)\\ \rho_n(t) &= \frac{2}{n}\sum_{k=1}^{n}Z_{k\Delta}t(X_{k\Delta})\\ \xi_n(t) &= \frac{2}{n}\sum_{k=1}^{n}J_{k\Delta}t(X_{k\Delta})-\mathbb{E}\left(J_{k\Delta}t(X_{k\Delta})\right)\end{aligned}$$

and

$$J_{k\Delta}=I_{k\Delta}+b(X_{k\Delta})=\Delta^{-1}\int_{k\Delta}^{(k+1)\Delta}b(X_s)ds. \tag{6.5}$$

As

$$\Gamma_n\left(\hat{h}_{m,N}\right)\leq\Gamma_n(h_{m,N}),$$

we can write

$$\left\|\hat{h}_{m,N}-h\right\|_{L^2}^2\leq\|h_{m,N}-h\|_{L^2}^2+2\nu_n\left(\hat{h}_{m,N}-h_{m,N}\right)+2\rho_n\left(\hat{h}_{m,N}-h_{m,N}\right)+2\xi_n\left(\hat{h}_{m,N}-g_{m,N}\right).$$

According to Cauchy-Schwarz, if we set $\mathscr{B}_m=\{t\in S_{m,N},\ \|t\|_{L^2}\leq 1\}$, we have:

$$\left\|\hat{h}_{m,N}-h\right\|_{L^2}^2\leq\|h_{m,N}-h\|_{L^2}^2+\frac{1}{2}\left\|\hat{h}_{m,N}-h_{m,N}\right\|_{L^2}^2+6\sup_{t\in\mathscr{B}_m}\left(\nu_n^2(t)+\rho_n^2(t)+\xi_n^2(t)\right)$$

According to Pythagoras, $\left\|\hat{h}_{m,N}-h_{m,N}\right\|_{L^2}^2=\left\|\hat{h}_{m,N}-h\right\|_{L^2}^2-\|h_{m,N}-h\|_{L^2}^2$, so

$$\left\|\hat{h}_{m,N}-h\right\|_{L^2}^2\leq\|h_{m,N}-h\|_{L^2}^2+12\sup_{t\in\mathscr{B}_m}\left(\nu_n^2(t)+\rho_n^2(t)+\zeta_n^2(t)\right).$$

The following lemma is very useful and is proved later.

Lemma 6.4.
We have that

1. $\mathbb{E}\left[I_{k\Delta}^2\middle|\mathscr{F}_{k\Delta}\right]=c\Delta\left(1+X_{k\Delta}^2\right)$ *and* $\mathbb{E}\left[I_{k\Delta}^4\middle|\mathscr{F}_{k\Delta}\right]\leq c\Delta^2\left(1+X_{k\Delta}^4\right)$.
2. $\mathbb{E}\left[Z_{k\Delta}\middle|\mathscr{F}_{k\Delta}\right]=0$, $\mathbb{E}\left[Z_{k\Delta}^2\middle|\mathscr{F}_{k\Delta}\right]\leq\frac{\sigma_0^2}{\Delta}$ *and* $\mathbb{E}\left[Z_{k\Delta}^4\middle|\mathscr{F}_{k\Delta}\right]\leq\frac{\sigma_0^4}{\Delta^2}$.
3. $\mathbb{E}\left[t^4(X_{k\Delta})b^4(X_{k\Delta})\right]\leq c\|t\|_\infty^2\|t\|_{L^2}^2$.
4. $\mathbb{E}\left(J_{k\Delta}^2\right)\leq c$, $\mathbb{E}\left(J_{k\Delta}^4\right)\leq c$ *and* $Var\left(J_{k\Delta}t(X_{k\Delta})\right)\leq c\|t\|_{L^2}^2$.

where the filtration $\mathscr{F}_t = \sigma\left(\eta,\ (W_s)_{0\leq s\leq t}\right)$ *is defined in Proposition 3.1 and the constant* c *depends on* b *and* σ.

Then

$$\begin{aligned}
\sup_{t\in\mathscr{B}_m} \nu_n^2(t) &= \sup_{t\in\mathscr{B}_m}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(I_{k\Delta}t(X_{k\Delta})\right)\right)^2 \\
&\leq \frac{1}{n}\sum_{k=1}^n \mathbb{E}\left(t^2(X_{k\Delta})\mathbb{E}\left(\left. I_{k\Delta}^2\right|\mathscr{F}_{k\Delta}\right)\right) \\
&\leq \frac{c\Delta}{n}\sum_{k=1}^n \mathbb{E}\left(t^2(X_{k\Delta})\left(1+X_{k\Delta}^2\right)\right) = c\Delta\int_{-\infty}^{+\infty}\left(1+x^2\right)f(x)t^2(x)dx
\end{aligned}$$

where the constant c depends on b. By (3.3), $\left\|\left(1+x^2\right)f(x)\right\|_\infty \leq c$ and we have that

$$\sup_{t\in\mathscr{B}_m} \nu_n^2(t) \leq c\Delta\left\|t\right\|_{L^2}^2.$$

As $(\varphi_{\lambda,m})_{\lambda\in\Lambda_m}$ is an orthonormal basis of S_m for the L^2-norm,

$$\sup_{t\in\mathscr{B}_m}\rho_n^2(t) \leq \sum_{\lambda\in\Lambda_m}\rho_n^2(\varphi_{\lambda,m}).$$

Besides,

$$\mathbb{E}\left(\rho_n^2(\varphi_{\lambda,m})\right) \leq \frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\left(\varphi_{\lambda,m}^2(X_{k\Delta})\mathbb{E}\left(\left. Z_{k\Delta}^2\right|\mathscr{F}_{k\Delta}\right)\right) \leq \frac{\sigma^2}{n\Delta}\mathbb{E}\left(\varphi_{\lambda,m}^2(X_0)\right).$$

So, by Assumption S(4) *2.*,

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_m}\rho_n^2(t)\right) \leq \frac{\sigma^2}{n\Delta}\mathbb{E}\left(\Phi_m^2(X_0)\right) \leq \frac{\phi_0\sigma^2 D_m}{n\Delta}.$$

We know that

$$\sup_{t\in\mathscr{B}_m}\xi_n^2(t) \leq \sum_{\lambda\in\Lambda_m}\xi_n^2(\varphi_{\lambda,m})$$

As

$$J_{k\Delta} = \frac{1}{\Delta}\int_{k\Delta}^{(k+1)\Delta} b(X_s)ds,$$

the random sequence $(J_{k\Delta}, X_{k\Delta})$ is stationary and β-mixing such that $\beta_{J,X}(n) \leq \beta_X(n\Delta)$. According to Lemma 6.1, we have that

$$\mathbb{E}\left(\xi_n^2\left(\varphi_{\lambda,m}\right)\right) \leq \frac{4}{n}\mathbb{E}\left(B(J_0,X_0)J_0^2\varphi_{\lambda,m}^2(X_0)\right).$$

Then, as $\mathbb{E}\left(J_0^4\right) \leq C$ and $\mathbb{E}\left(B^2(J_0,X_0)\right) \leq c/(\theta^2\Delta^2)$,

$$\begin{aligned}
\mathbb{E}\left(\sup_{t\in\mathscr{B}_m}\xi_n^2(t)\right) &\leq \frac{4}{n}\mathbb{E}\left(B(J_0,X_0)J_0^2\Phi_m^2(X_0)\right) \leq \frac{4\phi_0 D_m}{n}\mathbb{E}\left(B(J_0,X_0)J_0^2\right) \\
&\leq \frac{4\phi_0 D_m}{n}\left(\mathbb{E}\left(B^2(J_0,X_0)\right)\right)^{1/2}\left(\mathbb{E}\left(J_0^4\right)\right)^{1/2} \leq \frac{cD_m}{n\theta\Delta}.
\end{aligned}$$

So

$$\mathbb{E}\left(\left\|\hat{h}_{m,N}-h\right\|_{L^2}^2\right) \leq \left\|h_{m,N}-h\right\|_{L^2}^2 + c\Delta + c\frac{D_m}{n\Delta}\left(\frac{1}{\theta}+\sigma_0^2\right).$$

Proof of Lemma 6.4 According to Proposition 3.1,

$$\mathbb{E}\left(\sup_{s\in[0,\Delta]}\left(b(X_{k\Delta+s})-b(X_{k\Delta})\right)^{2l}\Big|\mathscr{F}_{k\Delta}\right)\leq c\Delta^l\left(1+X_{k\Delta}^{2l}\right),$$

which proves *(1)*. Points *(2)* and *(3)* are obvious, thus we only prove *(4)*. We know that

$$\mathrm{Var}\left(J_{k\Delta}t(X_{k\Delta})\right)\leq 2\mathbb{E}\left(I_{k\Delta}^2t^2(X_{k\Delta})\right)+2\mathrm{Var}(b(X_{k\Delta})t(X_{k\Delta}))$$

and

$$\mathrm{Var}(b(X_{k\Delta})t(X_{k\Delta}))\leq\int_A b^2(x)t^2(x)f(x)dx\leq\left\|b_A\right\|_\infty^2\left\|f\right\|_\infty\left\|t\right\|_{L^2}^2.$$

According to Proposition 3.1, we have that

$$\begin{aligned}\mathbb{E}\left(t^2(X_{k\Delta})\mathbb{E}\left(I_{k\Delta}^2\big|\mathscr{F}_{k\Delta}\right)\right) &\leq c\Delta\mathbb{E}\left(\left(1+X_{k\Delta}^2\right)t^2(X_{k\Delta})\right)\\ &\leq c\Delta\int_{-\infty}^{+\infty}\left(1+x^2\right)f(x)t^2(x)dx.\end{aligned}$$

By (3.3):

$$\int_{-\infty}^{+\infty}\left(1+x^2\right)f(x)t^2(x)dx\leq c\left\|t\right\|_{L^2}^2, \tag{6.6}$$

which ends the proof.

6.5 Proofs of Theorems 3.3 and 3.4

As previously, we only demonstrate Theorem 3.4. We have:

$$\left\|\tilde{h}_{N_n}-h\right\|_{L^2}^2\leq\inf_{m\in\mathscr{M}_n}\left\|h_{m,N_n}-h\right\|_{L^2}^2+12\sup_{t\in\mathscr{B}_{\hat{m},m}}\left(\nu_n^2(t)+\rho_n^2(t)+\xi_n^2(t)+pen(m)-pen(\hat{m})\right)$$

where $\mathscr{B}_{m,m'}=\{t\in S_{m,N_n}+S_{m',N_n},\ \left\|t\right\|_{L^2}\leq 1\}$. Let us consider a function $p(m,m')$ such that $12p(m,m')=pen(m)+pen(m')$. We have that

$$\left\|\tilde{h}-h\right\|_{L^2}^2\leq\inf_{m\in\mathscr{M}_n}\left\|h_{m,N_n}-h\right\|_{L^2}^2+2pen(m)+12\sup_{t\in\mathscr{B}_{\hat{m},m}}\left(\nu_n^2(t)+\rho_n^2(t)+\xi_n^2(t)-p(m,\hat{m})\right).$$

We already prove that $\sup_{t\in\mathscr{B}_{\hat{m},m}}\nu_n^2(t)\leq c\Delta$. Moreover,

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{\hat{m},m}}\rho_n^2(t)-p(m,\hat{m})\right)\leq\sum_{m'\in\mathscr{M}_n}\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m',m}}\rho_n^2(t)-p(m,m')\right)$$

and

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{\hat{m},m}}\xi_n^2(t)-p(m,\hat{m})\right)\leq\sum_{m'\in\mathscr{M}_n}\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m',m}}\xi_n^2(t)-p(m,m')\right)$$

The triplet $(X_{k\Delta},Z_{k\Delta},J_{k\Delta})$ is β-mixing and its β-mixing coefficient is smaller than $\beta_0e^{-\theta t}$. So we can construct a set Ω^* like in Lemma 6.2 with

$$q_n=\frac{(2l+3)\ln(n)}{\theta\Delta}.$$

Let us set, for $a=0,1$ and $0\leq k\leq p_n-1$:

$$U_{k,a}^*=\frac{1}{q_n}\sum_{l=1}^{q_n}J_{((2k+a)q_n+l)\Delta}^*t\left(X_{((2k+a)q_n+l)\Delta}^*\right)\quad\text{and}\quad V_{k,a}^*(t)=\frac{1}{q_n}\sum_{l=1}^{q_n}Z_{((2k+a)q_n+l)\Delta}^*t\left(X_{((2k+a)q_n+l)\Delta}^*\right).$$

Let us set:

$$\left\|t\right\|_{k,a}^2=\frac{1}{q_n}\sum_{l=1}^{q_n}t^2\left(X_{((2k+a)q_n+l)\Delta}^*\right) \tag{6.7}$$

As for the proof of Theorem 2.4, we denote $D = 2^m + 2^{m'}$ and we consider $(\varphi_\lambda,\ \lambda \in \Lambda)$ a basis of $S_m + S_{m'}$. Let us consider the spaces

$$\Omega_{Z,\Lambda} = \left\{\omega,\ \forall k,\ \forall a \in \{0,1\},\ \forall \lambda \in \Lambda,\ \left(V^*_{k,a}(\varphi^*_\lambda)\right)^2 \le 2\sigma_0^2\theta \left\|\varphi_\lambda\right\|^2_{k,1}\right\},$$
$$\Omega_J = \{\omega,\ \forall k,\ |J^*_{k\Delta}| \le (2l+1)\ln(n)\} \quad \text{and} \quad \mathcal{O} = \Omega^* \cap \Omega_{Z,\Lambda} \cap \Omega_J. \quad (6.8)$$

Risk bound on $\mathcal{O}$ We apply Lemma 6.3 to the variables $U^*_{k,a}$ and $V^*_{k,a}$. We have that

$$\rho_n(t) = \rho_{n,0}(t) + \rho_{n,1}(t) \quad \text{with} \quad \rho_{n,a}(t) = \frac{1}{2p_n}\sum_{k=1}^{p_n} V^*_{k,a} - \mathbb{E}\left(V^*_{k,a}\right)$$

and

$$\xi_n(t) = \xi_{n,0}(t) + \xi_{n,1}(t) \quad \text{with} \quad \xi_{n,a}(t) = \frac{1}{2p_n}\sum_{k=1}^{p_n} U^*_{k,a} - \mathbb{E}\left(U^*_{k,a}\right).$$

Applying Lemma 6.3 to the variables $V^*_{k,a}$. We have that

$$\mathrm{Var}\left(V^*_{k,a}\mathbb{1}_{\mathcal{O}}\right) \le \frac{1}{q_n}\mathbb{E}\left(Z_0^2 t^2(X_0)\right) = \frac{1}{q_n}\mathbb{E}\left(t^2(X_0)\mathbb{E}\left(Z_0^2\middle|\mathscr{F}_0\right)\right) \le \frac{\sigma_0^2}{q_n\Delta}.$$

Let us set $\mathscr{B} := \left\{t \in S_m + S_{m'},\ \|t\|^2_{L^2} \le 1\right\}$. By (6.8), we have that

$$\sup_{t\in\mathscr{B}}\left(V^*_{k,a}(t)\mathbb{1}_{\mathcal{O}}\right)^2 = \sup_{\sum_{\lambda\in\Lambda} a_\lambda^2 \le 1}\left(\sum_{\lambda\in\Lambda} a_\lambda V^*_{k,a}(\varphi_\lambda)\mathbb{1}_{\mathcal{O}}\right)^2 \le \sum_{\lambda\in\Lambda}\left(V^*_{k,a}(\varphi_\lambda)\mathbb{1}_{\mathcal{O}}\right)^2 \le 2\sigma_0^2\theta\sum_{\lambda\in\Lambda}\|\varphi_\lambda\|^2_{k,a}$$

where the semi-norm $\|.\|_{k,a}$ is defined by (6.7). So by Assumption **S4** *2*,

$$\sup_{t\in\mathscr{B}}\left(V^*_{k,a}(t)\mathbb{1}_{\mathcal{O}}\right)^2 \le 2\sigma_0^2\phi_0\theta D \quad \text{where} \quad D = 2^m + 2^{m'}.$$

Moreover, in the previous section it is demonstrated that

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,m'}}\rho_n^2(t)\mathbb{1}_{\mathcal{O}}\right) \le \frac{\phi_0 D}{n\Delta}.$$

Lemma 6.3 can be applied with $H^2 = \phi_0\sigma_0^2 D/(n\Delta)$, $V = \sigma_0^2 q_n^{-1}\Delta^{-1}$ and $M_1^2 = 2\sigma_0^2\phi_0\theta D$. We find:

$$\mathbb{E}\left(\left(\sup_{t\in\mathscr{B}_{m,m'}}\rho_n^2(t) - 12\frac{\phi_0 D}{n\Delta}\right)\mathbb{1}_{\mathcal{O}}\right)_+ \le C\left(\frac{1}{n\Delta}\exp\left(-cD\right) + \frac{D\ln^2(n)}{n^2\Delta^2}\exp\left(-\frac{c}{\ln(n)}\right)\right).$$

We know that $\sum_{m'}\exp\left(-cD\right) = \sum_{m'}\exp\left(-c\left(2^m + 2^{m'}\right)\right) \le C$ where the constant C does not depend on m nor on m'. Besides, $\sum_{m'} D \le \mathscr{D}_n^2$. As

$$\mathscr{D}_n^2 \le \frac{n\Delta}{\ln^2(n)},$$

we have

$$\sum_{m'}\mathbb{E}\left(\left(\sup_{t\in\mathscr{B}_{m,m'}}\rho_n^2(t) - 12\frac{\phi_0 D}{n\Delta}\right)\mathbb{1}_{\mathcal{O}}\right)_+ \le \frac{C}{n\Delta}.$$

Applying Lemma 6.3 to the variables $U^*_{k,a}$. According to Lemma 6.1, we have that

$$\mathrm{Var}\left(U^*_{k,a}\mathbb{1}_{\mathcal{O}}\right) \leq \frac{4}{q_n}\mathbb{E}\left(J_0^2 t^2\left(X_0\right)B\left(X_0\right)\right) \leq \frac{4}{q_n}\left(\mathbb{E}\left(J_0^4 t^4\left(X_0\right)\right)\right)^{1/2}\left(\mathbb{E}\left(B^2\left(X_0\right)\right)\right)^{1/2}$$

where $\mathbb{E}\left(B^2\left(X_0\right)\right) \leq 2\beta_0/\left(\theta\Delta\right)$. Moreover, as $J_0 = I_0 + b(X_0)$, we have, by Lemma 6.4:

$$\begin{aligned}\mathbb{E}\left(J_0^4 t^4\left(X_0\right)\right) &\leq c\mathbb{E}\left[t^4\left(X_0\right)\left(b^4\left(X_0\right)+\mathbb{E}\left(I_0^4\middle|\mathscr{F}_0\right)\right)\right]\\ &\leq c\left\|t\right\|_\infty^2\mathbb{E}\left(\left[\Delta^2\left(1+X_{k\Delta}^4\right)+b^4\left(X_0\right)\right]t^2\left(X_0\right)\right).\end{aligned}$$

By Equation (3.3):

$$\mathbb{E}\left(J_0^4 t^4\left(X_0\right)\right) \leq c\left\|t\right\|_\infty^2\int_{\mathbb{R}}\Delta^2\left(1+x^4\right)f(x)t^2(x)+b^4(x)f(x)t^2(x)dx$$

and

$$\mathbb{E}\left(J_0^4 t^4\left(X_0\right)\right) \leq cD.$$

Collecting terms, we obtain:

$$\mathrm{Var}\left(U^*_{k,a}\mathbb{1}_{\mathcal{O}}\right) \leq \frac{cD^{1/2}}{q_n\theta\Delta} = c\frac{D^{1/2}}{\ln(n)}.$$

Moreover,

$$\left\|U^*_{k,a}\mathbb{1}_{\mathcal{O}}\right\|_\infty \leq \left\|J_0 t(X_0)\mathbb{1}_{\mathcal{O}}\right\|_\infty \leq (2l+1)D^{1/2}\ln(n)$$

and we have proved in the previous section that

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,m'}}\xi_n^2(t)\mathbb{1}_{\mathcal{O}}\right) \leq 8\beta_0\phi_0\frac{D}{n\theta\Delta}.$$

We can apply Lemma 6.3 with $M_1 = CD^{1/2}\ln(n)$, $V = C'D^{1/2}/\ln(n)$ and $H^2 = 8\beta_0\phi_0 D/(n\theta\Delta)$. We find that

$$\mathbb{E}\left(\left(\sup_{t\in\mathscr{B}_{m,m'}}\nu_n^2(t) - 84\beta_0\phi_0\frac{D}{n\theta\Delta}\right)\mathbb{1}_{\mathcal{O}}\right)_+ \leq C\left(\frac{D^{1/2}}{n\theta\Delta}\exp\left(-cD^{1/2}\right)+\frac{D\ln^4(n)}{n^2\Delta^2}\exp\left(-c\frac{\sqrt{n\Delta}}{\ln^2(n)}\right)\right)$$

where the constant c is independent of D, n and Δ. We have that $\sum_{m'}D^{1/2}\exp\left(-cD^{1/2}\right) \leq \sum_{k=1}^{\infty}k^{1/2}\exp\left(-ck^{1/2}\right) < +\infty$. So, if

$$\mathscr{D}_{j,n} \leq \frac{n\Delta}{\ln^3(n)},$$

we have that

$$\sum_{m'\in\mathscr{M}_n}\mathbb{E}\left(\left(\sup_{t\in\mathscr{B}_{m,m'}}\nu_n^2(t) - 84\beta_0\phi_0\frac{D}{n\theta\Delta}\right)\mathbb{1}_{\mathcal{O}}\right)_+ \leq \frac{C}{n\Delta}.$$

Risk bound on $\mathcal{O}^c$ We know that

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m',m}}\left(\rho_n^2(t)+\xi_n^2(t)\right)\mathbb{1}_{\mathcal{O}^c}\right) \leq 2\sqrt{\mathbb{P}\left(\mathcal{O}^c\right)}\left(\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m',m}}\left(\rho_n^2(t)+\xi_n^2(t)\right)^2\right)\right)^{1/2}$$

and

$$\mathbb{P}\left(\mathcal{O}^c\right) \leq \mathbb{P}\left(\Omega^{*c}\right)+\mathbb{P}\left(\Omega^c_{Z,\Lambda}\right)+\mathbb{P}\left(\Omega^c_J\right).$$

According to Lemma 6.2,

$$\mathbb{P}\left(\Omega^{*c}\right) \leq n^{-2l}. \tag{6.9}$$

The following lemma is proved later:

Lemma 6.5.

$$\mathbb{P}\left(\Omega_{Z,\Lambda}^c\right) \leq \frac{c}{n^{2l}} \quad \textit{and} \quad \mathbb{P}(\Omega_J^c) \leq \frac{c}{n^{2l}}.$$

We have that

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m',m}} \left(\rho_n^2(t)+\xi_n^2(t)\right)^2\right) \leq \mathbb{E}\left(\left(\sum_{\lambda\in\Lambda}\rho_n^2(\varphi_\lambda)+\xi_n^2(\varphi_\lambda)\right)^2\right).$$

Besides,

$$\begin{aligned}\rho_n^2(\varphi_\lambda)+\xi_n^2\left(\varphi_\lambda\right) &= \left(\frac{1}{n}\sum_{k=1}^n \varphi_\lambda(X_{k\Delta})\left(Z_{k\Delta}+J_{k\Delta}\right)-\mathbb{E}\left(J_0\varphi_\lambda(J_0)\right)\right)^2\\ &\leq \frac{3}{n}\sum_{k=1}^n \varphi_\lambda^2(X_{k\Delta})\left(Z_{k\Delta}^2+J_{k\Delta}^2\right)+\mathbb{E}\left(\varphi_\lambda^2(X_0)\right)\mathbb{E}\left(J_0^2\right)\end{aligned}$$

According Assumption $\mathbf{S}(4)$ Point *2*, we know that $\sup_x \sum_{\lambda\in\Lambda}\varphi_\lambda^2(x)\leq \phi_0$, so:

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m',m}} \left(\rho_n^2(t)+\xi_n^2(t)\right)^2\right) \leq 27\phi_0^2\frac{1}{n}\sum_{k=1}^n\left[\mathbb{E}\left(Z_{k\Delta}^4+J_{k\Delta}^4\right)+\left(\mathbb{E}\left(J_{k\Delta}^2\right)\right)^2\right].$$

By Lemma 6.4, we obtain that:

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m',m}} \left(\rho_n^2(t)+\xi_n^2(t)\right)^2\right) \leq c\left(1+\frac{1}{\Delta^2}\right).$$

where c does not not depend on m,m', n, nor on Δ. So, by (6.9) and Lemma 6.5,

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{\hat{m}',m}} \left(\rho_n^2(t)+\xi_n^2(t)\right)\mathbb{1}_{\mathcal{O}^c}\right) \leq c\sum_{m'}\frac{1}{n^l\Delta} \leq \frac{\mathscr{D}_n}{n^l\Delta}.$$

As $\mathscr{D}_n \leq n\Delta$, as soon as $l\geq 2$:

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_{\hat{m}',m}} \left(\rho_n^2(t)+\xi_n^2(t)\right)\mathbb{1}_{\mathcal{O}^c}\right) \leq \frac{c}{n}.$$

Proof of Lemma 6.5

Bound of $\mathbb{P}\left(\Omega_J^c\right)$: We have that

$$\mathbb{P}\left(\Omega_J^c\right) = \mathbb{P}\left(\exists k,\ |J_{k\Delta}|\geq (2l+3)\ln(n)\right) \leq n\mathbb{P}\left(|J_0|\geq (2l+3)\ln(n)\right).$$

It is known that

$$\mathbb{P}\left(|J_0|\geq (2l+3)\ln(n)\right) \leq n^{-(2l+3)}\mathbb{E}\left(\exp\left(|J_0|\right)\right).$$

For any m, by stationarity,

$$\mathbb{E}\left(|J_0|^m\right) \leq \mathbb{E}\left(|b(X_0)|^m\right) \leq \int |b(x)|^m f(x)dx.$$

By (3.3),

$$\mathbb{E}\left(\exp\left(|J_0|\right)\right) \leq \int_{-\infty}^{\infty}\exp\left(|b(x)|\right)f(x)dx < c$$

and

$$\mathbb{P}\left(\Omega_J^c\right) \leq n^{-2l}.$$

Bound of $\mathbb{P}\left(\Omega_{Z,\Lambda}^c\right)$: According to Lemma 2 p.533 of Comte *et al.* [6], we have that

$$\mathbb{P}\left(\left(V_{k,a}^*(\varphi_\lambda)\right)^2 \geq 2\sigma_0^2\theta \left\|\varphi_\lambda\right\|_{k,a}^2\right) \leq 2\exp\left(-q_n\Delta\theta\right).$$

As $q_n = (2l+3)/(\theta\Delta)$, we obtain:

$$\mathbb{P}\left(\left(V_{k,a}^*(\varphi_\lambda)\right)^2 \geq 2\sigma_0^2\theta \left\|\varphi_\lambda\right\|_{k,a}^2\right) \leq 2n^{-(2l+3)}.$$

So we can write:

$$\mathbb{P}\left(\Omega_Z^c\right) = \mathbb{P}\left(\exists a,\ \exists k,\ \exists\lambda,\ \left(V_{k,a}^*(\varphi_\lambda)\right)^2 \geq 2\sigma_0^2\theta \left\|\varphi_\lambda\right\|_{k,a}^2\right) \leq |\Lambda|\, n\mathbb{P}\left(\left(V_{k,a}^*(\varphi_\lambda)\right)^2 \geq 2\sigma_0^2\theta \left\|\varphi_\lambda\right\|_{k,a}^2\right).$$

As$|\Lambda| \leq D.K_n$ with $K_n = n\Delta$, we have:

$$\mathbb{P}\left(\Omega_Z^c\right) \leq \mathscr{D}_n(n\Delta)n^{-2l-2} \leq (n\Delta)^{3/2}\, n^{2l-2} \leq n^{2l}.$$

6.6 Proof of Theorem 4.1

This proof follows the lines of Lacour [15], section 6.8. Let us set $\mathcal{E} = \left\{\omega,\ \left\|f - \tilde{f}\right\|_\infty \leq f_0/2\right\}$.

Risk bound on $\mathcal{E}$. On $\mathcal{E}$, $\tilde{f} \geq f_0/2$. We know that

$$\tilde{g}(x) = \sum_{\lambda\in\Lambda_{\hat{m}_1}} \left(\frac{1}{n}\sum_{k=1}^n \varphi_\lambda'(X_{k\Delta})\right)\varphi_\lambda(x),$$

so

$$\|\tilde{g}\|_{L^2}^2 = \sum_{\lambda\in\Lambda_{\hat{m}_1}} \left(\frac{1}{n}\sum_{k=1}^n \varphi_\lambda'(X_{k\Delta})\right)^2 \leq \left\|\sum_{\lambda\in\Lambda_{\hat{m}_1}} \left(\varphi_\lambda'\right)^2\right\|_\infty \leq \psi_1 2^{3\hat{m}_1}.$$

As $\|\tilde{g}\|_\infty^2 \leq \psi_0 2^{\hat{m}_1}\|\tilde{g}\|_{L^2}^2$ and $2^{5\hat{m}_1} \leq n\Delta$,

$$\|\tilde{g}\|_\infty^2 \leq \psi_0\psi_1 2^{4\hat{m}_1} \leq \psi_0\psi_1\,(n\Delta)^{4/5}$$

and for $n\Delta$ large enough, $\|\tilde{g}\|_\infty^2 \leq n\Delta f_0/2 \leq n\Delta \min_{x\in A}\tilde{f}(x)$. So, on $\mathcal{E}$, $\tilde{b} = \tilde{g}/(2\tilde{f})$ and:

$$\tilde{b} = b_A + \left(\frac{\tilde{g}-g}{2\hat{f}} + \frac{g}{2}\left(\frac{1}{\tilde{f}} - \frac{1}{f}\right)\right).$$

Therefore

$$\mathbb{E}\left(\left\|\tilde{b} - b_A\right\|_{L^2}^2 \mathbb{1}_{\mathcal{E}}\right) \leq f_0^{-2}\mathbb{E}\left(\|\tilde{g}-g\|_{L^2}^2\right) + f_0^{-4}\left\|g_A^2\right\|_\infty \mathbb{E}\left(\left\|\tilde{f} - f\right\|_{L^2}^2\right).$$

Risk bound on $\mathcal{E}^c$. As $\left\|\hat{b}\right\|_\infty \leq n\Delta$, we have that

$$\mathbb{E}\left(\left\|\tilde{b} - b_A\right\|_{L^2}^2 \mathbb{1}_{\mathcal{E}^c}\right) \leq \left((n\Delta)^2 + \|b_A\|_\infty^2\right)\mathbb{P}\left(\mathcal{E}^c\right)$$

It is known that:

$$\left\|f - \tilde{f}\right\|_\infty \leq \inf_{m_0\in\mathscr{M}_{0,n}}\left(\|f - f_{m_0}\|_\infty + \left\|f_{m_0} - \tilde{f}\right\|_\infty\right).$$

As $f \in \mathscr{B}_{2,\infty}^\alpha$, by DeVore and Lorentz [10] p182 and Barron *et al.* [2] (Lemma 12):

$$\|f - f_{m_0}\|_\infty \leq C2^{m_0(-\alpha+1/2)} \leq C\ln(n\Delta)^{-\alpha+1/2}.$$

So $\|f-f_{m_0}\|_\infty \leq f_0/4$ for n large enough, and $\mathcal{E}^c \subseteq \left\{\left\|f_{m_0}-\tilde{f}\right\|_\infty \geq f_0/4\right\}$. As f_{m_0} and $\tilde{f}$ belongs to the linear space $S_{\hat{m}_0}+S_{m_0}$ which satisfies Assumption **S2**, we have that

$$\left\|f_{m_0}-\tilde{f}\right\|_\infty^2 \leq \psi_0 \sup_{m_0'\in\mathscr{M}_{0,n}} 2^{m_0'\vee m_0}\left\|f_{m_0}-\hat{f}_{m_0'}\right\|_{L^2}^2.$$

We know:

$$\left\|f_{m_0}-\hat{f}_{m_0'}\right\|_{L^2}^2 = \sup_{t\in\mathscr{B}_{m_0,m_0'}} \nu_{0,n}^2(t) \quad \text{where} \quad \nu_{0,n}(t)=\frac{1}{n}\sum_{k=1}^n t(X_{k\Delta})-\int_{\mathbb{R}} t(x)f(x)dx.$$

Then

$$\mathbb{P}\left(\mathcal{E}^c\right) \leq \sup_{m_0'\in\mathscr{M}_{0,n}} \mathbb{P}\left(\sup_{t\in\mathscr{B}_{m_0,m_0'}} \nu_{0,n}^2(t) \geq 2^{-m_0'\vee m_0}\frac{f_0^2}{16\psi_0}\right).$$

As in Subsection 6.3, we use the set Ω^*. We have that

$$\mathbb{P}\left(\Omega^{*c}\right) \leq \frac{1}{n^{2l}}$$

so $\mathbb{P}\left(\mathcal{E}^c\right) \leq \mathbb{P}\left(\mathcal{E}^c\cap\Omega^{*c}\right)+\frac{1}{n^{2l}}$. Let us consider the random variables

$$U_{k,1}^* = \frac{1}{q_n}\sum_{l=1}^{q_n} t\left(X^*_{(2(k-1)q_n+l)\Delta}\right) \quad \text{and} \quad U_{k,2}^* = \frac{1}{q_n}\sum_{l=1}^{q_n} t\left(X^*_{((2k-1)q_n+l)\Delta}\right).$$

The random variables $\left(U_{k,a}^*\right)_{1\leq k\leq p_n}$ are independent and identically distributed. It is demonstrated in Subsection 6.3 that

$$\sup_{t\in\mathscr{B}_{m_0,m_0'}} \left\|U_{k,i}^*\right\|_\infty \leq \sqrt{\psi_0}D^{3/2} \quad , \quad \operatorname{Var}\left(U_{k,j}^*\right)\leq c \quad \text{and} \quad H^2 := \mathbb{E}\left(\sup_{t\in\mathscr{B}_{m_0,m_0'}} \nu_{0,n}^2(t)\right) \leq C\frac{D}{n\Delta}$$

where $D=2^{m_0}+2^{m_0'}$. As, by assumption, $D^2\leq n\Delta/\log^2(n\Delta)$ for n large enough, we have that $H^2 = CD/(n\Delta) \leq f_0^2/64\psi_0 D$. then

$$\mathbb{P}\left(\mathcal{E}^c\cap\Omega^*\right) \leq \sup_{m_0'\in\mathscr{M}_{0,n}} \mathbb{P}\left(\sup_{t\in\mathscr{B}_{m_0,m_0'}} \nu_{0,n}^2(t) \geq 2H^2+\frac{f_0^2}{64\psi_0 D}\right).$$

According to (6.1), we have that

$$\mathbb{P}\left(\mathcal{E}^c\cap\Omega^*\right) \leq \sup_{m_0'\in\mathscr{M}_{n,0}} \exp\left(-\frac{cn\Delta}{\ln(n)D^2}\right)$$

where the constant c is independent of n and D_m.By assumption, $D^2\leq \eta^2 n\Delta/\ln^2(n\Delta)$, so

$$\mathbb{P}\left(\mathcal{E}^c\cap\Omega^*\right) \leq (n\Delta)^{-c\eta^2}.$$

If η^2 is large enough, $\mathbb{P}\left(\mathcal{E}^c\cap\Omega^*\right)\leq (n\Delta)^{-3}$ and if $l\geq 2$, we have that

$$\mathbb{E}\left(\left\|\tilde{b}-b_A\right\|_{L^2}^2 \mathbb{1}_{\mathcal{E}^c}\right) \leq \frac{1}{n\Delta}$$

which ends the proof.

A Linear subspaces

A.1 Linear subspaces satisfying Assumptions S1 or S3

To use simple notations, we set in this section $A=[0,1]$.

Trigonometric polynomials

The trigonometric polynomial linear subspaces $V_m = \text{Vect}\left\{1, \cos(\pi\lambda x)\right\}_{1\leq\lambda\leq 2m}$ satisfy Assumption **S3**. The linear subspaces $S_m = \left\{\sin(\pi\lambda x)\right\}_{1\leq\lambda\leq 2m}$ satisfy Assumption **S1** for $k = 0, 1$.

Proof. DeVore and Lorentz [10] (Corollary 2.5 p205) and Barron *et al.* [2] (p120) prove that Assumption **S3** is satisfied by subspaces V_m.

Points *1.* et *2.* of Assumption **S1** are fulfilled by the subspaces (S_m). Moreover, for any $t \in S_m$,

$$\|t\|_\infty^2 \leq \|t\|_{L^2}^2 \left\| \sum_{\lambda=1}^m \sin^2(\lambda x) \right\|_\infty \leq D_m \|t\|_{L^2}^2 .$$

We have that

$$\left\|\Psi_m^2(x)\right\|_\infty = \left\| \sum_{\lambda=1}^m \lambda^2 \cos^2(\lambda x) \right\|_\infty \leq m^3 = D_m^3.$$

Besides, any function $t \in S_m$ can be written $\sqrt{(2/\pi)} \sum_{\lambda=1}^m a_\lambda \sin(\lambda x)$, so

$$\|t'\|_{L^2} = \frac{2}{\pi} \sum_{\lambda=1}^m a_\lambda^2 \lambda^2 \left\|\cos^2(\lambda x)\right\|_{L^2}^2 = \sum_{\lambda=1}^m a_\lambda^2 \lambda^2 \leq m^2 \|t\|_{L^2}^2 .$$

Points *3.* and *4.* of Assumption **S1** are satisfied.

□

Piecewise polynomials

Let us set

$$g_0(x) = \mathbb{1}_{[0,1]}(x),\ g_1(x) = x\mathbb{1}_{[0,1]}(x), \ldots,\ g_r(x) = x^r \mathbb{1}_{[0,1]}(x)$$

and $\varphi_{a,\lambda,m} = 2^{m/2} g_a\left(2^m x - \lambda\right)$. The linear subspaces

$$V_m = \text{Vect}\left(\varphi_{a,\lambda,m},\ 0 \leq a \leq r,\ 0 \leq \lambda \leq 2^m - 1\right)$$

satisfy Assumption **S3**. The linear subspaces

$$S_m = \text{Vect}\left(\left\{\varphi_{a,\lambda,m}\right\}_{0\leq a\leq r,\ 1\leq\lambda\leq 2^m-1} \cup \left\{\varphi_{a,\lambda,m}\right\}_{l\leq a\leq r,\ \lambda\in\{0,2^m\}}\right)$$

satisfy Assumption **S1** for $k \leq r$.

Proof. DeVore and Lorentz [10] (Theorem 3.4 p362) and Barron *et al.* [2] (p120) prove that (V_m) satisfy Assumption **S3**.

The linear subspaces (S_m) satisfy Points *1.* and *2.* of Assumption **S1**. Moreover, the functions $\varphi_{a,\lambda,m}$ have disjoint supports if $\lambda \neq \lambda'$, and for any a, $\|g_a\|_\infty \leq 1$. So

$$\|t\|_\infty^2 \leq \|t\|_{L^2}^2 \left\| \sum_{\lambda\in\Lambda_m} \sum_{a=0}^r \left(\varphi_{a,\lambda,m}\right)^2 \right\|_\infty = \|t\|_{L^2}^2 \left\| \sum_{a=0}^r \left(\varphi_{a,\lambda,m}\right)^2 \right\|_\infty \leq (r+1) 2^m \|t\|_{L^2}^2 .$$

In the same way, we obtain:

$$\left\|\Psi_m^2(x)\right\|_\infty = \left\| \sum_{\lambda\in\Lambda_m} \sum_{a=0}^r \left(\varphi'_{a,\lambda,m}\right)^2 \right\|_\infty = \left\| \sum_{a=0}^r \left(\varphi'_{a,\lambda,m}\right)^2 \right\|_\infty = \left\| \sum_{a=0}^r 2^{3m} \left(g'_r\left(2^m x - \lambda\right)\right)^2 \right\|_\infty \leq (r+1) 2^{3m}.$$

For any function $t \in S_m$,

$$\|t'\|_{L^2}^2 = \left\| \sum_{\lambda\in\Lambda_m} \sum_{a=0}^r \left(\varphi'_{a,\lambda,m}\right) \right\|_{L^2}^2 = \sum_{\lambda\in\Lambda_m} 2^m \left\| \sum_{a=0}^r 2^m g'_a\left(2^m x - \lambda\right) \right\|_{L^2}^2 = 2^{2m} \left\| \sum_{a=0}^r g'_a(x) \right\|_{L^2}^2 \leq r(r+1) 2^{2m}.$$

Points *2.*, *3.*, and *4.* are proved.

□

Spline functions restricted to $[0,1]$

Spline functions g_r, where g_r is the $r+1$ time convolution of the indicator function of $[0,1]$, generates a r-regular multi-resolution analysis of $L^2(\mathbb{R})$. Their supports are included in $[0,r+1]$ and they belong to $\mathscr{C}_p^r \cap \mathscr{C}^{r-1}$. Let us set $\varphi_{\lambda,m} = 2^m g_r\left(2^m x - \lambda\right) \mathbb{1}_{[0,1]}(x)$. Then

$$V_m = \mathrm{Vect}\left(\varphi_{\lambda,m},\ \lambda = -r+1, \ldots, 2^m\right)$$

satisfies Assumption **S3** for $k \leq r$ and

$$S_m = \mathrm{Vect}\left(\varphi_{\lambda,m},\ \lambda = 0, \ldots, 2^m - r\right)$$

satisfies Assumption **S2.1** for $k \leq r$.

Proof. Schmisser [23] proved that the linear subspaces (V_m) satisfy Assumption **S3.1**. The functions g_r have a compact support: to prove that the subspaces (S_m) fulfil Assumption **S1**, we use the same arguments as in the previous paragraph. □

A.2 Restricted spaces of wavelets

The properties of wavelets are defined in Meyer [20] p21-22 (Definitions 1 and 2).

Definition A.1.
Let us consider

$$S_m = \left\{\varphi_{\lambda,m} := 2^{m/2}\varphi(2^m. - \lambda),\ \lambda \in \mathbb{Z}\right\}$$

a multi-resolution analysis of $L^2(\mathbb{R})$ *such that* $\left(\varphi_{\lambda,m}\right)_{\lambda\in\mathbb{Z}}$ *is an orthonormal basis of* S_m*. Let us set*

$$S_{m,N} = \left\{\varphi_{\lambda,m} := 2^{m/2}\varphi(2^m. - \lambda),\ |\lambda| \leq 2^m N\right\}$$

and denote, for any function $t \in L^2(\mathbb{R})$, t_m *(resp* $t_{m,N}$*) its orthogonal projection over* S_m *(resp* $S_{m,N}$*).*

Lemma A.1.
If

$$\int x^2 t^2(x)dx < +\infty \quad , \quad t \in L_1 \quad \textit{and} \quad \sup_{x\in\mathbb{R}}\left(|x\varphi(x)|\right) < +\infty,$$

then

$$\|t_{m,N} - t_m\|_{L^2}^2 \leq \frac{c}{N}$$

where the constant c *is independent of* m *and* N*.*

The proof is done in Comte *et al.* [8].

References

[1] Arlot, S. and Massart, P. (2009) Data-driven calbration of penalties for least-squares regression. *Journal of Machine Learning Research*, 10 pp. 245–279.

[2] Barron, A., Birgé, L. and Massart, P. (1999) Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113 (3) pp. 301–413.

[3] Beskos, A., Papaspiliopoulos, O. and Roberts, G.O. (2006) Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12 (6) pp. 1077–1098.

[4] Bosq, D. (1997) Parametric rates of nonparametric estimators and predictors for continuous time processes. *Ann. Statist.*, 25 (3) pp. 982–1000.

[5] Comte, F. and Merlevède, F. (2005) Super optimal rates for nonparametric density estimation via projection estimators. *Stochastic Process. Appl.*, 115 (5) pp. 797–826.

[6] Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2007) Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, 13 (2) pp. 514–543.

[7] Comte, F. and Merlevède, F. (2002) Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM Probab. Statist.*, 6 pp. 211–238 (electronic). New directions in time series analysis (Luminy, 2001).

[8] Comte, F., Rozenholc, Y. and Taupin, M.L. (2006) Penalized contrast estimator for adaptive density deconvolution. *Canad. J. Statist.*, 34 (3) pp. 431–452.

[9] Dalalyan, A.S. and Kutoyants, Y.A. (2003) Asymptotically efficient estimation of the derivative of the invariant density. *Stat. Inference Stoch. Process.*, 6 (1) pp. 89–107.

[10] DeVore, R.A. and Lorentz, G.G. (1993) *Constructive approximation*, *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, volume 303. Springer-Verlag, Berlin.

[11] Gloter, A. (2000) Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM Probab. Statist.*, 4 pp. 205–227.

[12] Gobet, E., Hoffmann, M. and Reiß, M. (2004) Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.*, 32 (5) pp. 2223–2253.

[13] Hosseinioun, N., Doosti, H. and Niroumand, H.A. (2009) Wavelet-based estimators of the integrated squared density derivatives for mixing sequences. *Pakistan J. Statist.*, 25 (3) pp. 341–350.

[14] Lacour, C. (2007) *Estimation non paramétrique adaptative pour les chaînes de Markov et les chaînes de Markov cachées*. Ph.D. thesis, Université Paris Descartes.

[15] Lacour, C. (2008) Nonparametric estimation of the stationary density and the transition density of a Markov chain. *Stochastic Process. Appl.*, 118 (2) pp. 232–260.

[16] Leblanc, F. (1997) Density estimation for a class of continuous time processes. *Math. Methods Statist.*, 6 (2) pp. 171–199.

[17] Lerasle, M. (2009) Adaptive density estimation of stationary β-mixing and τ-mixing processes. *Math. Methods Statist.*, 18 (1) pp. 59–83.

[18] Lerasle, M. (2010) Optimal model selection for stationary data under various mixing conditions.

[19] Masry, E. (1994) Probability density estimation from dependent observations using wavelets orthonormal bases. *Statist. Probab. Lett.*, 21 (3) pp. 181–194. URL `http://dx.doi.org/10.1016/0167-7152(94)90114-7`.

[20] Meyer, Y. (1990) *Ondelettes et opérateurs. I*. Actualités Mathématiques. [Current Mathematical Topics]. Hermann, Paris. Ondelettes. [Wavelets].

[21] Pardoux, E. and Veretennikov, A.Y. (2001) On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, 29 (3) pp. 1061–1085.

[22] Rao, B.L.S.P. (1996) Nonparametric estimation of the derivatives of a density by the method of wavelets. *Bull. Inform. Cybernet.*, 28 (1) pp. 91–100.

[23] Schmisser, E. (2009b) Nonparametric drift estimation from noisy data. Preprint 2009-21, MAP5, Université Paris Descartes.

[24] Viennet, G. (1997) Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Related Fields*.