SORT 28 (1) January-June 2004, 37-54

Local superefficiency of data-driven projection density estimators in continuous time

Denis Bosq and Delphine Blanke*

Université Pierre et Marie Curie, Paris

Abstract

We construct a data-driven projection density estimator for continuous time processes. This estimator reaches superoptimal rates over a class $\mathcal{F}_0$ of densities that is dense in the family of all possible densities, and a «reasonable» rate elsewhere. The class $\mathcal{F}_0$ may be chosen previously by the analyst. Results apply to $\mathbb{R}^d$-valued processes and to $\mathbb{N}$-valued processes. In the particular case where square-integrable local time does exist, it is shown that our estimator is strictly better than the local time estimator over $\mathcal{F}_0$.

MSC: 62G07, 62M

Keywords: Density estimation, data-driven, superefficiency, continuous time processes

1 Introduction

We study a data-driven projection density estimator $\hat{f}_T$ in a general framework where data are in continuous time. The purpose is to reach a superoptimal rate on a class $\mathcal{F}_0$ of densities that is dense in $\mathcal{F}$, the family of all possible densities, and a «reasonable» rate elsewhere. The class $\mathcal{F}_0$ can be previously chosen by the analyst.

The results are, in some sense, extensions of those which where obtained in the i.i.d. case (cf Bosq 2002*a*, 2002*b*), but in this new context the methods are often different.

Section 2 contains notation and assumptions. In Section 3 we study the estimator over $\mathcal{F}_0$. We obtain a $\frac{1}{T}$-rate with respect to the mean integrated square error, a $(\frac{\ln\ln T}{T})^{1/2}$-

* *Address for correspondence:* (Université Pierre et Marie Curie, Paris-France)
Received: October 2003
Accepted: January 2004

rate with respect to uniform error, and a Gaussian limit in distribution with coefficient of normalization $\sqrt{T}$. Results concerning the asymptotic behaviour of $\hat{f}_T$ over $\mathcal{F} - \mathcal{F}_0$ appear in Section 4. Finally, Section 5 is devoted to comparison of $\hat{f}_T$ with the local time estimator $f_{T,0}$ when this estimator exists. It is shown that, in a special case, $\hat{f}_T$ is strictly better than $f_{T,0}$. The proofs are postponed until Section 6.

2 Notation and assumptions

Let $(E, \mathcal{B}, \mu)$ be a measure space, with μ σ-finite, and such that $L^2(\mu)$ is infinite dimensional. The norm of $L^2(\mu)$ will be denoted $\|.\|$. Let $(e_j,\ j \geq 0)$ be an orthonormal system in $L^2(\mu)$.

We consider a stochastic process $X = (X_t,\ t \in \mathbb{R})$ defined on a probability space $(\Omega, \mathcal{A}, P)$ and with values in $(E, \mathcal{B})$. X is supposed to be measurable and such that the X_t's are identically distributed with density f with respect to μ.

Denote $\mathcal{F}$ the family of densities f such that

$$f = \sum_{j=0}^{\infty} a_j e_j, \quad \sum_{j=0}^{\infty} a_j^2 < \infty. \tag{2.1}$$

The class of the observable processes will be denoted $\mathcal{X}$. Note that two different processes may have the same f. In order to estimate f from the data $(X_t,\ 0 \leq t \leq T)$ $(T > 0)$ we use a data-driven projection estimator :

$$\hat{f}_T = \sum_{j=0}^{\hat{k}_T} \widehat{a}_{j_T} e_j \text{ with } \widehat{a}_{j_T} = \frac{1}{T} \int_0^T e_j(X_t)\,\mathrm{d}t, \quad j \geq 0$$

and

$$\hat{k}_T = \max\left\{ j \ :\ 0 \leq j \leq k_T,\ \left|\widehat{a}_{j_T}\right| \geq \gamma_T \right\}$$

where γ_T and the integer k_T are chosen by the analyst. If $\{\ldots\} = \emptyset$ one sets $\hat{k}_T = k_T$.

We always suppose that (unless otherwise stated)

$$k_T \to \infty, \quad \frac{k_T}{T} \to 0, \quad \gamma_T \to 0, \text{ as } T \to \infty.$$

If $\gamma_T = 0$ one obtains the projection density estimator

$$f_T = \sum_{j=0}^{k_T} \widehat{a}_{j_T} e_j \tag{2.2}$$

Now $\mathcal{F}_0(K)$ will denote the class of $f \in \mathcal{F}$ such that

$$f = \sum_{j=0}^{K} a_j e_j, \quad a_K \neq 0, \text{ and } \mathcal{F}_0 = \bigcup_{K=0}^{\infty} \mathcal{F}_0(K),$$

and finally we put

$$\mathcal{F}_1 = \mathcal{F} - \mathcal{F}_0.$$

In order to study the rates of convergence of $\hat{f}_T$ over $\mathcal{F}_0$ and $\mathcal{F}_1$ we shall use strong mixing coefficients of the form

$$\alpha(\mathcal{C}, \mathcal{D}) = \sup_{C \in \mathcal{C}, D \in \mathcal{D}} |P(C \cap D) - P(C)P(D)| \tag{2.3}$$

where $\mathcal{C}$ and $\mathcal{D}$ are sub-σ-algebra of $\mathcal{A}$.

For a given process $Y = (Y_t,\ t \in I)$, where $I \subseteq \mathbb{R}$, one defines its strong mixing functions as

$$\alpha_Y^{(2)}(u) = \sup_{h \in I, h+u \in I} \alpha(\sigma(Y_h),\ \sigma(Y_{h+u})), \quad u \geq 0 \quad \text{and}$$

$$\alpha_Y(u) = \sup_{h \in \mathbb{R}} \alpha(\sigma(Y_t, t \leq h, t \in I), \quad \sigma(Y_t, t \geq h+u, t \in I)), \quad u \geq 0$$

with the convention $\alpha(.,.) = 0$ if one of the two sub-σ-algebras is not defined. These two classical coefficients will be used in the sequel.

Now the main assumptions and conditions are H_1 and H_2 :

$$H_1 \begin{cases} A_1 & : \ P_{(X_{s+h}, X_{t+h})} = P_{(X_s, X_t)}; \quad s, t, h \in \mathbb{R} \ \textit{(2-stationarity)}, \\ B_1(r) & : \ M_r = \sup_{j \geq 0} \left\| e_j(X_0) \right\|_r < \infty, \ \textit{where } 2 < r \leq \infty, \\ C_1(r) & : \ \displaystyle\int_0^\infty \left[\alpha_X^{(2)}(u)\right]^{(r-2)/r} \mathrm{d}u < \infty, \\ c_1 & : \ \gamma_T \simeq T^{-\gamma} \ (\gamma > 0) \textit{ and } k_T \simeq T^\beta \ (0 < \beta < 1). \end{cases}$$

$$H_2 \begin{cases} A_2 & : \ X \textit{ is strictly stationary}, \\ B_2 = B_1(\infty) & : \ M = \sup_{j \geq 0} \left\| e_j(X_0) \right\|_\infty < \infty, \\ C_2 & : \ \alpha_X(u) \leq a\, e^{-bu} \ (a > 0, b > 0) \\ & \quad (X \textit{ is geometrically strongly mixing, (GSM)}). \\ c_2 & : \ \gamma_T = \left(\dfrac{\ln T \ln \ln T}{T}\right)^{1/2}. \end{cases}$$

Note that A_2 and C_2 are satisfied as soon as X is an enough regular stationary diffusion process (cf Doukhan, 1994). Note also in some situations, one may choose $\gamma_T = c(\frac{\ln T}{T})^{1/2}$ with constant c large enough.

Concerning B_2, it is satisfied in many classical cases, for example if (e_j) is a trigonometric system on a compact interval or the Hermite functions over $\mathbb{R}$. In the particular case where $E = \mathbb{N}$ and μ is the counting measure, the natural system $(1_{\{j\}},\ j \geq 0)$ is, of course, uniformly bounded.

Finally some special assumptions concerning local time will appear in Section 5.

3 Rates of $\hat{f}_T$ over $\mathcal{F}_0$

If $f \in \mathcal{F}_0$ we shall denote $K(f)$ the only integer K such that $f \in \mathcal{F}_0(K)$. The following proposition shows that $\hat{k}_T$ is actually a consistent estimator of $K(f)$.

Proposition 3.1 *If $f \in \mathcal{F}_0$, then*

1) if H_1 holds,

$$P(\hat{k}_T \neq K(f)) = \mathcal{O}(T^{\beta+2\gamma-1}) \tag{3.1}$$

thus, if $\beta + 2\gamma < 1$, $\hat{k}_T \to K(f)$ in probability.

2) If H_2 holds,

$$P(\hat{k}_T \neq K(f)) = o(T^{-\delta}), \tag{3.2}$$

for each $\delta > 0$, in particular, if $T = T_n \uparrow \infty$ with $\sum_n T_n^{-\delta} < \infty$, for some $\delta > 0$, then

$$\hat{k}_{T_n} = K(f) \text{ almost surely for } n \text{ large enough.} \tag{3.3}$$

These results show that the adaptive estimator $\hat{f}_T$ has asymptotically the same behaviour as the pseudo-estimator

$$g_T = \sum_{j=0}^{K(f)} \widehat{a}_{j_T} e_j. \tag{3.4}$$

The following lemma makes this fact explicit :

Lemma 3.1 *If $M = \sup_{j \geq 0} \left\| e_j(X_0) \right\|_\infty < \infty$, one has*

$$\mathrm{E} \left\| \hat{f}_T - g_T \right\|^2 \leq M^2 k_T P(\hat{k}_T \neq K(f)). \tag{3.5}$$

We now indicate the rates of $\hat{f}_T$ on $\mathcal{F}_0$, we begin with the mean integrated square error (MISE).

Proposition 3.2 *If $f \in \mathcal{F}_0$, then*

1) If H_1 holds, we have

$$\mathrm{E} \left\| \hat{f}_T - f \right\|^2 = \mathcal{O}\left(\frac{1}{T^{1-\beta}}\right) \tag{3.6}$$

2) If H_2 holds,

$$T.\mathrm{E} \left\| \hat{f}_T - f \right\|^2 \xrightarrow[T\to\infty]{} 2 \sum_{j=0}^{K(f)} \int_0^\infty \mathrm{Cov}\left(e_j(X_0), e_j(X_u)\right) \mathrm{d}u. \tag{3.7}$$

The next statement gives a uniform result.

Corollary 3.1

$$\limsup_{T\to\infty} \sup_{X\in\mathcal{X}_0(a_0,b_0,K_0)} T.\mathrm{E}\,\big\|\hat{f}_T - f\big\|^2 \le \frac{8a_0M^2K_0}{b_0}. \tag{3.8}$$

Here $\mathcal{X}_0(a_0, b_0, K_0)$ denotes the family of processes that satisfy H_2 with $f \in \mathcal{F}_0(K)$, $K \le K_0$ and $\alpha_X(u) \le ae^{-bu}$ where $a \le a_0$ and $b \ge b_0$.

We now turn to the $\|.\|_\infty$-error :

Proposition 3.3 *If $f \in \mathcal{F}_0$ and H_2 holds, then*

$$(\forall\,\varepsilon > 0), (\forall\,\delta > 0), \quad P(\big\|\hat{f}_T - f\big\|_\infty \ge \varepsilon) = \mathcal{O}(T^{-\delta}), \tag{3.9}$$

and if $T = T_n = nh$ $(h > 0)$, $n \to \infty$,

$$\big\|\hat{f}_T - f\big\|_\infty = \mathcal{O}\Big(\big(\frac{\ln\ln T}{T}\big)^{1/2}\Big), \text{ almost surely.} \tag{3.10}$$

Finally the limit in distribution appears in the following statement:

Proposition 3.4 *If $f \in \mathcal{F}_0$, H_2 holds and $T = nh$ $(h > 0)$ then*

$$\sqrt{T}(\hat{f}_T - f) \Rightarrow N \tag{3.11}$$

where «⇒» means weak convergence in $L^2(\mu)$ and N is a zero-mean Gaussian $L^2(\mu)$-valued random variable with $K(f) + 1$-dimensional support.

Proposition 3.2(2), 3.3 and 3.4 exhibit superoptimal rates if $f \in \mathcal{F}_0$. In general these rates appear if the Castellana-Leadbetter condition holds (see Castellana and Leadbetter (1986), Bosq (1998)). Here this condition is *not* needed; this means that local irregularity of the sample paths is not necessary for obtaining these parametric rates over $\mathcal{F}_0$.

4 Asymptotic behaviour of $\hat{f}_T$ over $\mathcal{F}_1$

In order to study consistency of $\hat{f}_T$ when $f \in \mathcal{F}_1$ we need results concerning the behaviour of the truncation index $\hat{k}_T$ as T tends to infinity.

Below the first statement expresses the fact that $\hat{k}_T \to \infty$ in some sense when the second one shows that $\hat{k}_T$ is not far from an «optimal k_T».

Proposition 4.1 *If $f \in \mathcal{F}_1$ then*

1) If H_1 holds

$$P(\hat{k}_T < A) = \mathcal{O}(T^{-1}), \quad A > 0, \tag{4.1}$$

2) If H_2 holds

$$P(\hat{k}_T < A) = \mathcal{O}(\exp(-c_A\sqrt{T})), \quad (c_A > 0),\ A > 0. \tag{4.2}$$

Now we specify the asymptotic behaviour of $\hat{k}_T$. For this purpose we set

$$q(\eta) = \min\left\{q \in \mathbb{N},\ \left|a_j\right| \leq \eta \text{ for all } j > q\right\},\ \eta > 0. \tag{4.3}$$

Note that $q(\eta)$ does exist since $a_j \to 0$, and that, if $q(\eta) > 0$, then $\left|a_{q(\eta)}\right| > \eta$. On the other hand $\eta < \eta'$ implies $q(\eta') \leq q(\eta)$.

We put $q_T(\varepsilon) = q((1+\varepsilon)\gamma_T)$, $\varepsilon > 0$; $q'_T(\varepsilon') = q((1-\varepsilon')\gamma_T)$, $0 < \varepsilon' < 1$ and we consider the event

$$E_T := \left\{q_T(\varepsilon) \leq \hat{k}_T \leq q'_T(\varepsilon')\right\}.$$

Then:

Proposition 4.2 *If $f \in \mathcal{F}_1$ and $q_T(\varepsilon) \leq k_T$, we have*

1) Under H_1,

$$P(E_T^c) = \mathcal{O}(T^{\beta+2\gamma-1}), \tag{4.4}$$

2) Under H_2,

$$P(E_T^c) = o(T^{-\delta}) \textit{ for all } \delta > 0. \tag{4.5}$$

We indicate two applications of these results:

Example 4.1 Under H_1, if $\left|a_j\right| \simeq j^{-\eta}$ ($\eta > \frac{1}{2}$) one has $q_T(\varepsilon) \simeq T^{\gamma/\eta}$, then $2\gamma \leq \beta$ ensures $q_T(\varepsilon) \leq k_T$ for T large enough and $\beta < \frac{1}{2}$ yields $P(E_T^c) \to 0$.

Example 4.2 Under H_2, if $\left|a_j\right| \simeq \alpha\rho^j$ ($\alpha > 0,\ 0 < \rho < 1$) and $k_T > \left[1+(2\ln 1/\rho)^{-1}\right]\ln T$, one has $q_T(\varepsilon) \simeq \frac{\ln T}{2\ln(1/\rho)}$,

$$P(|\frac{\hat{k}_T}{\ln T} - (2\ln 1/\rho)^{-1}| > \xi) = o(T^{-\delta}),\ \xi > 0,\ \delta > 0. \tag{4.6}$$

In particular, if $T = T_n$ with $\sum_n T_n^{-\delta} < \infty$ for some $\delta > 0$, then

$$\frac{\hat{k}_{T_n}}{\ln T_n} \to (2\ln 1/\rho)^{-1} \text{ almost surely.} \tag{4.7}$$

Note that, from (4.7), one may deduce an estimator of ρ, namely $\widehat{\rho_T} = T^{-\frac{1}{2\hat{k}_T+1}}$ which converges almost surely.

We now may state results concerning the MISE of $\hat{f}_T$.

Proposition 4.3 *If $f \in \mathcal{F}_1$ and $q_T(\varepsilon) \leq k_T$ then*

1) Under H_1,

$$\mathrm{E}\left\|\hat{f}_T - f\right\|^2 = \mathcal{O}(T^{-(1-\beta-2\gamma)}) + \sum_{j>q_T(\varepsilon)} a_j^2. \tag{4.8}$$

2) Under H_2,

$$\mathrm{E}\left\|\hat{f}_T - f\right\|^2 = \mathcal{O}(\frac{q'_T(\varepsilon')}{T}) + \sum_{j>q_T(\varepsilon)} a_j^2. \tag{4.9}$$

Thus if H_1 and conditions in Example 4.1 hold then, taking $\beta = \frac{1}{2\eta}$, yields

$$\mathrm{E}\,\|f_T - f\|^2 = \mathcal{O}(T^{-\frac{2\eta-1}{2\eta}}), \tag{4.10}$$

when $\mathrm{E}\,\left\|\hat{f}_T - f\right\|^2 = \mathcal{O}(T^{-\frac{2\eta-1}{2\eta}+2\gamma})$.

Suppose now that conditions in Example 4.2 and H_2 hold. Then, if $\ln T = \mathcal{O}(k_T)$, we have

$$\mathrm{E}\,\left\|\hat{f}_T - f\right\|^2 = \mathcal{O}\Big(\frac{\ln T \ln\ln T}{T}\Big), \tag{4.11}$$

when, if $k_T \simeq a \ln T$ with $a \geq (2\ln 1/\rho)^{-1}$,

$$\mathrm{E}\,\|f_T - f\|^2 = \mathcal{O}\Big(\frac{\ln T}{T}\Big). \tag{4.12}$$

In some special cases one may construct a process for which the rates (4.10) and (4.12) are the true rates for f_T. For example, if (e_j) is the trigonometric basis over $L^2[0,1]$, one may consider the process

$$X_t = Y_{[t]}, \quad t \in \mathbb{R}$$

where $(Y_n,\ n \in \mathbb{Z})$ is a sequence of independent $[0,1]$-valued random variables with common density f. For this process the rates are $T^{-(2\eta-1)/2\eta}$ and $\frac{\ln T}{T}$ respectively. This trick has been used previously in Blanke and Bosq (2000) and Bosq (1998) for the kernel density estimator.

Finally, at least in this special case, the loss of rate for $\hat{f}_T$ is a logarithm. Thus $\hat{f}_T$ has a $1/T$-rate on $\mathcal{F}_0$ and a «good» rate on $\mathcal{F}_1$.

We now turn to uniform rate. We have the following proposition:

Proposition 4.4 *Under H_2, if $\left|a_j\right| \simeq \alpha\rho^j$ ($\alpha > 0$, $0 < \rho < 1$), $j \geq 0$ and $k_T \gg \ln T$, if $T = T_n$ where $\sum \frac{\ln T_n}{T_n^\delta} < \infty$ for some $\delta > 0$ then for $f \in \mathcal{F}_1$:*

$$\limsup_{T_n\to\infty} \frac{\sqrt{T}_n}{(\ln T_n)^{3/2}} \left\|\hat{f}_{T_n} - f\right\|_\infty \leq 2\sqrt{\frac{2a\delta}{b}\,\frac{M^2}{\ln(1/\rho)}} \quad \textit{(almost surely).} \tag{4.13}$$

Note that the rate in (4.13) is almost optimal since the law of the iterated logarithm shows that the rate cannot be better than $(\frac{\ln\ln T}{T})^{1/2}$.

5 Comparison with the local time estimator

We now suppose that X admits an *occupation density* (or *local time*) with respect to μ. More precisely we make the following assumption:

H_3 : $\forall\, T \geq 0,\ \exists\, \ell_T \in L^2(\mu \otimes P)$:

$$\int_0^T \varphi(X_t)\,\mathrm{d}t = \int_E \varphi(x)\ell_T(x)\,\mathrm{d}\mu(x), \quad \varphi \in \mathcal{M}(E, \mathbb{R}^+), \tag{5.1}$$

where $\mathcal{M}(E, \mathbb{R}^+)$ is the family of $\mathcal{B}$-$\mathcal{B}_{\mathbb{R}}$ measurable positive real functions defined on E ($\mathcal{B}_{\mathbb{R}}$ is the Borel σ-algebra on $\mathbb{R}$).

In such a situation one defines the local time density estimator as

$$f_{T,0} = \frac{\ell_T}{T}, \quad T > 0 \tag{5.2}$$

$f_{T,0}$ is then the density of the empirical measure μ_T defined by

$$\mu_T(B) = \frac{1}{T}\int_0^T \mathbf{1}_B(X_t)\,\mathrm{d}t, \quad B \in \mathcal{B}.$$

Example 5.1 *If $E = \mathbb{N}$ and μ is the counting measure then H_3 is satisfied and*

$$f_{T,0}(x) = \frac{1}{T}\int_0^T \mathbf{1}_{\{x\}}(X_t)\,\mathrm{d}t, \quad x \in \mathbb{N} \tag{5.3}$$

Example 5.2 *If $E = \mathbb{R}$, and μ is Lebesgue measure, H_3 is equivalent to*

$$\liminf_{\varepsilon\downarrow 0} \frac{1}{\varepsilon}\int_{[0,T]^2} P\left(|X_t - X_s| \le \varepsilon\right)\,\mathrm{d}s\,\mathrm{d}t < \infty,\ T > 0 \tag{5.4}$$

(cf Geman and Horowitz, 1980).

Example 5.3 *If $(E, \mathcal{B}, \mu) \subseteq (E_0, B_0, \mu_0)$ with $\mu = g.\mu_0$ and $0 < m \le g \le m' < \infty$ then if H_3 holds for μ_0 with local time $\ell_T^{(0)}$, it holds for μ with local time $\ell_T = \ell_T^{(0)}/g$.*

Note that, if $E = \mathbb{R}$, the Castellana-Leadbetter condition, 1986 (cf also Bosq, 1998) implies H_3 under mild regularity conditions, if X is strictly stationary.

Results and references concerning the local time estimator appear in Bosq and Davydov (1999) and Bosq (1998). Note that, in particular, $f_{T,0}$ is an unbiased estimator of f : E $f_{T,0} = f$ (a.e.).

Now we need a result concerning the MISE of $f_{T,0}$. For this purpose we denote $\ell_{(k)}$ the local time of X on $]k-1, k]$, $k \in \mathbb{Z}$ and make the following assumption :

H_4 : *X is strictly stationary and the series $L = \sum_{k\ge1}\int_E \mathrm{Cov}\,(\ell_{(1)}(x), \ell_{(k)}(x))d\mu(x)$ converges.*

Note that the Davydov's inequality shows that a sufficient condition for H_4 is

H_4' : *X is strictly stationary and there exists $r > 2$ such that*

$$\int_E \left[E\ell_{(1)}^r(x)\right]^{2/r}\,\mathrm{d}\mu(x) < \infty \text{ and } \sum_{k\ge1}\left[\alpha_X(k)\right]^{(r-2)/r} < \infty.$$

Now the following statement exhibits superefficiency of $f_{T,0}$:

Proposition 5.1 *If H_3 and H_4 hold, then*

$$T.\mathrm{E}\left\|f_{T,0} - f\right\|^2 \to L, \quad f \in \mathcal{F}. \tag{5.5}$$

Concerning $\hat{f}_T$ we have

Proposition 5.2 *If H_3 and H_4 hold, then*

$$\mathrm{E}\left\|\hat{f}_T - f\right\|^2 = \mathcal{O}(\frac{1}{T}) + \mathrm{E}\,(\sum_{j>\hat{k}_T} a_j^2). \qquad (5.6)$$

Note that the key of the proof of Proposition 5.2 is the fact that $\hat{f}_T = \Pi^{\hat{k}_T} f_{T,0}$ where $\Pi^{\hat{k}_T}$ is the orthogonal projector of $\mathrm{sp}(e_j,\ 0 \le j \le \hat{k}_T)$. A similar property for f_T has been noticed in Frenay (2001). Thus

$$\left\|\hat{f}_T - \Pi^{\hat{k}_T} f\right\| \le \left\|f_{T,0} - f\right\| \quad \text{and} \qquad (5.7)$$

$$\mathrm{E}\left\|\hat{f}_T - \Pi^{\hat{k}_T} f\right\|^2 \le \mathrm{E}\left\|f_{T,0} - f\right\|^2 = \mathcal{O}(\frac{1}{T}). \qquad (5.8)$$

Consequently the efficiency of $\hat{f}_T$ depends on the «pseudo-bias» $\mathrm{E}\sum_{j>\hat{k}_T} a_j^2$. Under conditions in Proposition 4.3 this pseudo-bias may be replaced by $\sum_{j>q_T(\varepsilon)} a_j^2$ and the rates (4.11) and (4.12) do not change. However, $\hat{f}_T$ is better than $f_{T,0}$ over $\mathcal{F}_0$ because $f_{T,0} = \sum_{j=0}^{\infty} \widehat{a}_{j,T} e_j$, when $\hat{f}_T$ has the same asymptotic behaviour as $g_T = \sum_{j=0}^{K(f)} \widehat{a}_{j_T} e_j$ and more precisely :

Proposition 5.3 *If $f \in \mathcal{F}_0$ and H_2, H_3, H_4 hold then*

$$\liminf_{T\to\infty} T.\mathrm{E}\left\|f_{T,0} - f\right\|^2 \ge 2\sum_{j=0}^{\infty}\int_0^{\infty} \mathrm{Cov}\,(e_j(X_0), e_j(X_u))\,\mathrm{d}u \qquad (5.9)$$

when

$$T.\mathrm{E}\left\|\hat{f}_T - f\right\|^2 \to 2\sum_{j=0}^{K(f)}\int_0^{\infty} \mathrm{Cov}\,(e_j(X_0), e_j(X_u))\,\mathrm{d}u. \qquad (5.10)$$

It is easy to construct examples where $\int_0^{\infty} \mathrm{Cov}\,(e_j(X_0), e_j(X_u))\,\mathrm{d}u > 0$ for some $j > K(f)$; in that case $\hat{f}_T$ is strictly better than $f_{T,0}$ on $\mathcal{F}_0$.

6 Proofs

6.1 Proof of Proposition 3.1

Set $B_T = \left\{\exists\, j\ :\ 0 \le j \le k_T,\ \left|\widehat{a}_{j_T}\right| \ge \gamma_T\right\}$, then, we have for T large enough and $K = K(f)$, $B_T^c \Rightarrow \left|\widehat{a}_{KT}\right| < \gamma_T \le \frac{|a_K|}{2} \Rightarrow \left|a_{KT} - \widehat{a}_{KT}\right| \ge \frac{|a_K|}{2}$ thus $P(B_T^c) \le \dfrac{4\mathrm{Var}\,\widehat{a}_{KT}}{|a_K|^2}$. Now, 2-stationarity yields

$$\operatorname{Var}\widehat{a_{KT}} = \frac{2}{T}\int_0^T (1-\frac{u}{T})\operatorname{Cov}(e_K(X_0), e_K(X_u))\,\mathrm{d}u, \tag{6.1}$$

using Davydov's inequality, see Bosq (1998, p. 21), one obtains

$$\operatorname{Var}\widehat{a_{KT}} \leq \frac{2}{T}\int_0^T (1-\frac{u}{T})\frac{2r}{r-2}2^{\frac{r-2}{r}}[\alpha_X^{(2)}(u)]^{\frac{r-2}{r}}\,\|e_K(X_0)\|_r^2\,\mathrm{d}u$$

and H_1 implies

$$\operatorname{Var}\widehat{a_{KT}} \leq \frac{c_r}{T}, \tag{6.2}$$

where $c_r = \frac{4rM_r^2}{r-2}2^{\frac{r-2}{r}}\int_0^\infty [\alpha_X^{(2)}(u)]^{\frac{r-2}{r}}\,\mathrm{d}u$ thus

$$P(B_T^c) \leq \frac{4c_r}{a_K^2}\frac{1}{T}. \tag{6.3}$$

Now, as soon as $k_T > K$ and $\gamma_T \leq \frac{|a_K|}{2}$,

$$\left\{\hat{k}_T > K, B_T\right\} \Rightarrow \bigcup_{j=K+1}^{k_T}\left\{\left|\widehat{a_{j_T}}\right| \geq \gamma_T\right\} \tag{6.4}$$

and

$$\left\{\hat{k}_T < K, B_T\right\} \Rightarrow \left|\widehat{a_{KT}} - a_K\right| > \frac{|a_K|}{2} \Rightarrow \left|\widehat{a_{KT}} - a_K\right| > \gamma_T \tag{6.5}$$

thus

$$P(\hat{k}_T \neq K, B_T) \leq \frac{1}{\gamma_T^2}\sum_{j=K}^{k_T}\operatorname{Var}\widehat{a_{KT}}, \tag{6.6}$$

again using Davydov's inequality one obtains

$$P(\hat{k}_T \neq K, B_T) = O(\frac{k_T+1}{\gamma_T^2 T}) = O(T^{\beta+2\gamma-1}). \tag{6.7}$$

Now, since (6.3) implies

$$P(\hat{k}_T \neq K, B_T^c) = O(\frac{1}{T}), \tag{6.8}$$

(3.1) follows. □

The proof of (3.2) is similar. It uses the following exponential inequality:

Lemma 6.1 *Let $Y = (Y_t,\ 0 \le t \le T)$ be a real measurable stationary strong mixing process such that $\int_0^\infty \alpha_Y(u)\,\mathrm{d}u < \infty$ and $M_Y = \sup_{0\le t\le T} \|Y_t\|_\infty < \infty$. Then for all $r \in [1, \frac{T}{2}]$ and all positive constants η, κ one has*

$$P\Big(\Big|\frac{1}{T}\int_0^T Y_t - \mathrm{E}\,Y_t\,\mathrm{d}t\Big| \ \ge\ \eta\Big) \ \le\ 4\exp\Big(-\frac{T\eta^2/M_Y^2}{c_1 + c_2\frac{r}{T} + c_3 M_Y^{-1}\eta r}\Big) + \frac{c_4}{\eta} M_Y \alpha_Y(r) \quad (6.9)$$

with $c_1 = 32(1+\kappa)^2 \int_0^\infty \alpha_Y(u)\,\mathrm{d}u$, $c_2 = 4c_1$, $c_3 = \frac{16}{3}(1+\kappa)$, $c_4 = 16\frac{(1+\kappa)}{\kappa}$.

Proof of Lemma 6.1. For q, r such that $2qr = T$, we consider blocks of variables $V_T(j),\ j = 1, \dots, 2[q]-1$, defined by

$$V_T(j) = \int_{(j-1)r}^{jr} (Y_t - \mathrm{E}\,Y_t)\,\mathrm{d}t \text{ and } V_T(2[q]) = \int_{(2[q]-1)r}^{2qr} (Y_t - \mathrm{E}\,Y_t)\,\mathrm{d}t.$$

So, for any $\eta > 0$,

$$P\Big(\Big|\frac{1}{T}\int_0^T Y_t - \mathrm{E}\,Y_t\,\mathrm{d}t\Big| \ge \eta\Big) \le P\Big(\Big|\sum_{j=1}^{[q]} V_T(2j)\Big| > \frac{T\eta}{2}\Big) + P\Big(\Big|\sum_{j=1}^{[q]} V_T(2j-1)\Big| > \frac{T\eta}{2}\Big).$$

The two terms may be handled similarly. Consider the first one, for example: we use Rio's (2000) coupling result recursively to approximate $V_T(2), \dots, V_T(2[q])$ by independent variables. For any $j \ge 1$, there exists a random variable $V_T^*(2j)$, measurable function of $V_T(2), \dots, V_T(2j)$ such that $V_T^*(2j)$ is independent of $V_T(2), \dots, V_T(2j-2)$ and with same law as $V_T(2j)$. Moreover :

$$\mathrm{E}\,\Big|V_T^*(2j) - V_T(2j)\Big| \le 2\|V_T(2j)\|_\infty \Big(\sup |P(AB) - P(A)P(B)|\Big)$$

where the supremum is taken over all sets A and B belonging to σ-algebras of events generated by respectively $\{V_T(2), \dots, V_T(2j-2)\}$ and $V_T(2j)$.

For any positive κ, one may write

$$P\Big(\Big|\sum_{j=1}^{[q]} V_T(2j)\Big| > \frac{T\eta}{2}\Big) \le P\Big(\Big|\sum_{j=1}^{[q]} V_T^*(2j)\Big| > \frac{T\eta}{2(1+\kappa)}\Big) + P\Big(\Big|\sum_{j=1}^{[q]} V_T(2j) - V_T^*(2j)\Big| > \frac{T\eta\kappa}{2(1+\kappa)}\Big)$$

Since the $V_T^*(2j)$ are independent, Bernstein's inequality (written as in Pollard (1984)) implies

$$P\Big(\Big|\sum_{j=1}^{[q]} V_T^*(2j)\Big| > \frac{T\eta}{2(1+\kappa)}\Big) \le 2\exp\Big(-\frac{T\eta^2/M_Y^2}{c_1 + c_2\frac{r}{T} + c_3 M_Y^{-1}\eta r}\Big)$$

with the help of Billingsley's inequality (1979), and constants c_i as stated as in Lemma 6.1. Moreover, Markov's inequality yields

$$P(|\sum_{j=1}^{[q]} V_T(2j) - V_T^*(2j)| > \frac{T\eta\kappa}{2(1+\kappa)}) \leq \frac{2(1+\kappa)}{T\eta\kappa} \sum_{j=1}^{[q]} \mathrm{E}\,|V_T(2j) - V_T^*(2j)|$$

and the result follows from Rio's coupling result. □

Now the proof of (3.2) consists in applying (6.9) to the processes $(e_j(X_t)-a_j,\ 0 \leq t \leq T)$ for $j = K, \ldots, k_T$. This allows to bound the quantities $P(\left|\widehat{a}_{j_T} - a_j\right| \geq \eta)$ for suitable η. In particular, one obtains

$$P(B_T^c) = \mathcal{O}(\exp(-A\sqrt{T})), \quad (A > 0) \tag{6.10}$$

Technical details are omitted.

Finally (3.3) comes from Borel-Cantelli lemma. □

6.2 Proof of Lemma 3.1

It suffices to write $\left\|\hat{f}_T - g_T\right\|^2 = \left\|\hat{f}_T - g_T\right\|^2 \mathbf{1}_{\{\hat{k}_T \neq K\}} \leq (\sum_{j=1}^{\hat{k}_T} \widehat{a}_{j_T}^2)\mathbf{1}_{\{\hat{k}_T \neq K\}}$ $\leq M^2 k_T \mathbf{1}_{\{\hat{k}_T \neq K\}}$, hence (3.5) by taking expectations. □

6.3 Proof of Proposition 3.2

First we have,

$$\mathrm{E}\left\|\hat{f}_T - f\right\|^2 = \mathrm{E}\,(\sum_{j=0}^{\hat{k}_T}(\widehat{a}_{j_T} - a_j)^2) + \mathrm{E}\,(\sum_{j>\hat{k}_T} a_j^2) \tag{6.11}$$

then, by Davydov's inequality: $\mathrm{E}\,(\sum_{j=0}^{\hat{k}_T}(\widehat{a}_{j_T} - a_j)^2) \leq \sum_{j=0}^{k_T} \mathrm{Var}\,\widehat{a}_{j_T} \leq c_r \frac{k_T}{T}$. On the other hand, if $f \in \mathcal{F}_0(K)$, $\sum_{j>\hat{k}_T} a_j^2 = \sum_{j>\hat{k}_T} a_j^2 \mathbf{1}_{\{\hat{k}_T<K\}}$, hence $\mathrm{E}\,(\sum_{j>\hat{k}_T} a_j^2) \leq \|f\|^2\, P(\hat{k}_T < K)$. Now from (6.5) and (6.8) it follows that $P(\hat{k}_T < K) \leq P(\left|\widehat{a}_{KT} - a_K\right| > \frac{|a_K|}{2}) + \mathcal{O}(\frac{1}{T})$. Using Davydov's inequality one obtains the bound

$$\mathrm{E}\,(\sum_{j>\hat{k}_T} a_j^2) = \mathcal{O}(\frac{1}{T}), \tag{6.12}$$

and (6.11) gives (3.6). Concerning (3.7) first note that, if $f \in \mathcal{F}_0(K)$, $P(\hat{k}_T \neq K) = o(T^{-\delta})$ for each $\delta > 0$ (cf (3.2)), thus Lemma 3.1 entails $\mathrm{E}\left\|\hat{f}_T - g_T\right\| = o(T^{-1})$. Thus it is only necessary to study

$$\mathrm{E}\,\|g_T - f\|^2 = \sum_{j=0}^{K} \mathrm{Var}\,\widehat{a}_{j_T}, \tag{6.13}$$

but using Billingsley's inequality one obtains

$$\int_0^\infty \left|\mathrm{Cov}\,(e_j(X_0), e_j(X_u))\right| \,\mathrm{d}u \le 4M^2 \int_0^\infty ae^{-bu} \le \frac{4aM^2}{b} < \infty. \tag{6.14}$$

Now since

$$T.\mathrm{Var}\,\widehat{a}_{j_T} = 2\int_0^T (1-\frac{u}{T})\mathrm{Cov}\,(e_j(X_0), e_j(X_u))\,\mathrm{d}u, \tag{6.15}$$

the dominated convergence theorem gives

$$T.\mathrm{Var}\,\widehat{a}_{j_T} \to 2\int_0^\infty \mathrm{Cov}\,(e_j(X_0), e_j(X_u))\,\mathrm{d}u \tag{6.16}$$

and (6.13) yields (3.7). □

6.4 Proof of Corollary 3.1

It suffices to apply Billingsley's inequality in (6.15) and to verify that the other bounds are uniform over $\mathcal{X}_0(a_0, b_0, K_0)$; details are omitted. □

6.5 Proof of Proposition 3.3

First, putting $K(f) = K$ one has
$|\hat{f}_T - g_T| = |(\hat{f}_T - g_T)\mathbf{1}_{\{\hat{k}_T \neq K\}}| \le \sum_{j=1}^{k_T} |\widehat{a}_{j_T}|\,|e_j|\mathbf{1}_{\{\hat{k}_T \neq K\}} \le M^2 k_T \mathbf{1}_{\{\hat{k}_T \neq K\}}$, one obtains, for all $\varepsilon > 0$ and all $\delta > 0$,

$$P(\left\|\hat{f}_T - g_T\right\|_\infty \ge \varepsilon) \le P(\hat{k}_T \neq K) = o(T^{-\delta}). \tag{6.17}$$

Now, $P(\|g_T - f\|_\infty \ge \varepsilon) \le \sum_{j=0}^{K} P(\left|\widehat{a}_{j_T} - a_j\right| \ge \frac{\varepsilon}{KM})$, then, using (6.9) for $Y_t = e_j(X_t)$, $0 \le t \le T$; $0 \le j \le K$, with $r = B\ln T$ one arrives at the bound

$$P(\left|\widehat{a}_{j_T} - a_j\right| \ge \frac{\varepsilon}{KM}) \le 4\exp(-\frac{T}{\ln T}\frac{3\varepsilon/KM^2B}{16(1+\kappa)(1+o(1))}) + 64\frac{1+\kappa}{\kappa}\frac{KM^2}{\varepsilon}a\exp(-bB\ln T)$$

For a given $\delta > 0$ and choosing $B = \delta b^{-1}$ one obtains (3.9).

Concerning (3.10), note that $(e_j(X_t),\ t \in \mathbb{R})$ satisfies the law of the iterated logarithm (LIL) : actually using the LIL for strongly mixing discrete time processes (cf Rio, 2000) one obtains the LIL for the processes $(Z_{ij}^{(h)} = \frac{1}{h}\int_{(i-1)h}^{ih}(e_j(X_t) - a_j)\,\mathrm{d}t,\ i \ge 0)$ since these processes are bounded and geometrically strongly mixing. It follows that $\|g_T - f\|_\infty = O\Big((\frac{\ln\ln T}{\ln T})^{1/2}\Big)$ almost surely, hence (3.10) by using (6.17) for $T = nh$. □

6.6 Proof of Proposition 3.4

Since $\sqrt{T}(\hat{f}_T - f) = \sqrt{T}(\hat{f}_T - g_T) + \sqrt{T}(g_T - f)$ and $\sqrt{T}\,\left\|g_T - \hat{f}_T\right\|_\infty \to 0$ in probability (see (6.17)), Theorem 4.4 in Billingsley (1979) shows that it suffices to study asymptotic normality of

$$\sqrt{T}(g_T - f) = \sum_{j=0}^{K}(\widehat{a}_{j_T} - a_j)e_j.$$

This is equivalent to asymptotic normality of the finite dimensional random vector $\sqrt{T}(\widehat{a_0}_T - a_0, \ldots, \widehat{a_K}_T - a_K)$ which in turn is equivalent to this of the real random variables $\sqrt{T}\sum_{j=0}^{K}\lambda_j(\widehat{a}_{j_T} - a_j)$; $\lambda_1, \ldots, \lambda_K \in \mathbb{R}$. Finally using the processes $(Z_{ij}^{(h)},\ i \geq 0), 0 \leq j \leq K$ and Rio (2000), the desired result follows. □

6.7 Proof of Proposition 4.1

1) Let j_0 such that $a_{j_0} \neq 0$, similarly as in the proof of Proposition 3.1 one obtains

$$\left\{\hat{k}_T < j_0\right\} \Rightarrow \left|\widehat{a_{j_0}}_T - a_{j_0}\right| > \frac{\left|a_{j_0}\right|}{2} \tag{6.18}$$

as soon as $k_T \geq j_0$, hence $P(\hat{k}_T < j_0) = \mathcal{O}(T^{-1})$. Since $f \in \mathcal{F}_1$, j_0 may be taken arbitrarily large, hence (4.1).

2) (6.18) and the exponential inequality (6.9) lead to (4.2). Details are omitted. □

6.8 Proof of Proposition 4.2

For T large enough we have $\left|a_{q_T(\varepsilon)}\right| > (1+\varepsilon)\gamma_T$.

1) From Davydov's inequality we get $P(\hat{k}_T < q_T(\varepsilon), B_T) \leq P(\left|\widehat{a}_{q_T(\varepsilon),T} - a_{q_T(\varepsilon)}\right| > \varepsilon\gamma_T) \leq \frac{c_r}{\varepsilon^2 T\gamma_T^2}$.

Now, if $q'_T(\varepsilon') \geq k_T$ one has $P(\hat{k}_T > q'_T(\varepsilon')) = 0$, if not, since $\left|a_j\right| \leq (1-\varepsilon')\gamma_T$ for $j > q'_T(\varepsilon')$, we have $\left\{\hat{k}_T > q'_T(\varepsilon'), B_T\right\} \Rightarrow \bigcup_{k_T \geq j > q'_T(\varepsilon')}\left|\widehat{a}_{j_T} - a_j\right| > \varepsilon'\gamma_T$ thus $P(\hat{k}_T > q'_T(\varepsilon'), B_T) \leq \frac{c_r(k_T+1)}{\varepsilon'^2 T\gamma_T^2}$ and (4.4) follows.

2) For proving (4.5) we may and do suppose that $q'_T(\varepsilon') < k_T$. Then

$$P(E_T^c \cap B_T) \leq P(|\widehat{a}_{q_T(\varepsilon),T} - a_{q_T(\varepsilon)}| \geq \varepsilon\gamma_T) + \sum_{q'_T(\varepsilon') < j \leq k_T} P(|\widehat{a}_{j_T} - a_j| \geq \varepsilon'\gamma_T),$$

Choosing $r = c\ln T$ in (6.9) one arrives at

$$P(E_T^c \cap B_T) = \mathcal{O}(k_T T^{-(c'\ln\ln T)}) + \mathcal{O}(k_T\gamma_T^{-1}T^{-cb})$$

for some constant c' and the choice $c > (\frac{3}{2} + \delta)b^{-1}$ leads to (4.5) since $P(B_T^c) = o(T^{-\delta})$ for all $\delta > 0$. □

6.9 Proof of Proposition 4.3

We start from (6.11) and write

$\mathrm{E}\left(\sum_{j>\hat{k}_T} a_j^2 \mathbf{1}_{E_T^c \cap B_T}\right) \leq \|f\|^2 P(E_T^c)$, $\mathrm{E}\left(\sum_{j>\hat{k}_T} a_j^2 \mathbf{1}_{E_T \cap B_T}\right) \leq \sum_{j>q_T(\varepsilon)} a_j^2$, $\mathrm{E}\left(\sum_{j=0}^{\hat{k}_T} (\widehat{a}_{j_T} - a_j)^2 \mathbf{1}_{E_T \cap B_T}\right) \leq \sum_{j \leq q_T'(\varepsilon')} \mathrm{Var}\, \widehat{a}_{j_T}$. Finally, under H_1 we write

$\mathrm{E}\left(\sum_{j=0}^{\hat{k}_T} (\widehat{a}_{j_T} - a_j)^2 \mathbf{1}_{E_T^c \cap B_T}\right) \leq \sum_{j=0}^{k_T} \mathrm{Var}\, \widehat{a}_{j_T}$, when under H_2,

$\mathrm{E}\left(\sum_{j=0}^{\hat{k}_T} (\widehat{a}_{j_T} - a_j)^2 \mathbf{1}_{E_T^c \cap B_T}\right) \leq 4M^2 k_T P(E_T^c)$, using the above bounds, (6.10) and (6.11) one obtains (4.8) and (4.9). □

6.10 Proof of Proposition 4.4

Let ξ be a positive constant, for any positive κ_i, $i = 1, 2$ one obtains

$$\begin{aligned} P(\|\hat{f}_T - f\|_\infty \geq \xi) &\leq P(\|\hat{f}_T - f\|_\infty \mathbf{1}_{E_T} \geq \frac{\xi}{1+\kappa_1}) + P(\|\hat{f}_T - f\|_\infty \mathbf{1}_{E_T^c} \geq \frac{\xi\kappa_1}{1+\kappa_1}) \\ &\leq P_1 + P_2 + P_3 \end{aligned}$$

with $P_1 = \sum_{j=1}^{q_T'(\varepsilon')} P(Mq_T'(\varepsilon') \left|\widehat{a}_{j_T} - \mathrm{E}\,\widehat{a}_{j_T}\right| \geq \frac{\xi}{(1+\kappa_1)(1+\kappa_2)})$, $P_3 = P(E_T^c)$ and $P_2 = P(M \sum_{j=q_T(\varepsilon)+1}^{\infty} \left|a_j\right| \geq \frac{\xi\kappa_2}{(1+\kappa_1)(1+\kappa_2)})$.

Concerning P_1, the assumptions imply in particular that $q_T'(\varepsilon')$ is of the same order as $\ln T/(2\ln(1/\rho))$. Now (6.9) and the choices $Y_t = e_j(X_t)$, $M_Y = M$, $r = R \ln T$, $\eta = \frac{2\ln(1/\rho)\xi}{M(1+\kappa_1)(1+\kappa_2)\ln T}$ with $\xi^2 = c\frac{(\ln T)^3}{T}$ and $T = T_n$ yield $\sum_n P_1 = O(\frac{\ln T_n}{T_n^\delta})$ as soon as $R = (\frac{1}{2} + \delta)b^{-1}$ and $c = \frac{8M^4(1+\kappa_1)^2(1+\kappa_2)^2(1+\kappa)^2 a\delta}{b \ln^2(1/\rho)}$.

Now noting that $\sum_{j=q_T(\varepsilon)+1}^{\infty} \left|a_j\right| \leq C(\alpha, \rho)\gamma_T$, it is easy to see that for T_n large enough, $P_2 = 0$ with previous choices of γ_T and ξ. Moreover, Proposition 4.2 implies also $P_3 = o(T_n^{-\delta})$. Finally, collecting these results, one obtains Proposition 4.4 with the help of Borel-Cantelli's lemma since $\sum_n \frac{\ln T_n}{T_n^\delta} < \infty$. □

6.11 Proof of Proposition 5.1

Using additivity of local time one may write $\frac{\ell_T}{T} = \frac{\ell_{\{0\}}}{T} + \frac{1}{T}\sum_{j=1}^{[T]} \ell_{(j)} + \frac{\ell_{][T],T]}}{T}$. Since $\mathrm{E}\left\|\frac{\ell_{\{0\}}}{T}\right\|^2 = o(\frac{1}{T})$ and $\mathrm{E}\left\|\frac{\ell_{][T],T]}}{T}\right\|^2 \leq \frac{\mathrm{E}\|\ell_{(1)}\|^2}{T} = o(\frac{1}{T})$ it suffices to study

$$n\mathrm{E}\,\|\frac{1}{n}\sum_{j=1}^{n} \ell_{(j)} - f\|^2 =$$

$$\mathrm{E}\,\|\ell_{(1)} - f\|^2 + 2\sum_{k=1}^{n-1}(1 - \frac{k}{n})\int_E \mathrm{Cov}\,(\ell_{(1)}(x), \ell_{(k+1)}(x))\,\mathrm{d}\mu(x) \quad (6.19)$$

where $n = [T]$. A classical trick allows to prove that the second member of (6.19) tends to L, hence (5.5). □

6.12 Proof of Proposition 5.2

Let $\Pi^{\hat{k}_T}$ be the orthogonal projector of $\mathrm{sp}(e_j,\ 0 \leq j \leq \hat{k}_T)$, we have

$\left\|\Pi^{\hat{k}_T}(f_{T,0} - f)\right\| \leq \left\|f_{T,0} - f\right\|$ thus $\mathrm{E}\left\|\hat{f}_T - \Pi^{\hat{k}_T} f\right\|^2 \leq \mathrm{E}\left\|f_{T,0} - f\right\|^2$ and (5.5) implies $\limsup_{T\to\infty} T.\mathrm{E}\left\|\hat{f}_T - \Pi^{\hat{k}_T} f\right\|^2 \leq L$ hence (5.6) from (6.11) and the fact that $P(E_T^c \cup B_T^c) = o(\frac{1}{T})$. □

6.13 Proof of Proposition 5.2

This is clear from (6.11), (5.7) and (5.8). □

6.14 Proof of Proposition 5.3

(5.10) has been proved in Proposition 3.2. Concerning (5.9) first note that (5.1) implies $\frac{1}{T}\int_0^T e_j(X_t)\,\mathrm{d}t = \frac{1}{T}\int_E e_j(x)\ell_T(x)\,\mathrm{d}x$ thus $\widehat{a}_{j_T} = \int_E f_{T,0}(x)e_j(x)\,\mathrm{d}\mu(x),\ j \geq 0$, hence $f_{T,0} = \sum_{j=0}^{\infty} \widehat{a}_{j_T} e_j$ and $\sum \widehat{a}_{j_T}^2 < \infty$ (almost surely); then we have $T\mathrm{E}\left\|f_{T,0} - f\right\|^2 = \sum_{j=0}^{\infty} T\mathrm{Var}\,\widehat{a}_{j_T}$ but H_2 yields $\int_0^{\infty}\left|\mathrm{Cov}\,(e_j(X_0), e_j(X_u))\right|\mathrm{d}u < \infty$ and (6.16) holds. This implies (5.9) by using Fatou lemma for the counting measure. □

References

Billingsley, P. (1979). *Probability and Measure*, Wiley, New-York.

Blanke, D. and Bosq, D. (2000). A family of minimax rates for density estimators in continuous time, *Stochastic Anal. Appl.*, 18, 871-900.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*, Lecture Notes in Statistics, 110, Second edn, Springer-Verlag, New-York.

Bosq, D. (2002*a*). Estimation localement suroptimale et adaptative de la densité, *C.R. Acad. Sci. Paris Ser. I*, 334, 591-595.

Bosq, D. (2002*b*). Estimation suroptimale de la densité par projection. Submitted for publication.

Bosq, D. and Davydov, Yu. (1999). Local time and density estimation in continuous time, *Math. Methods Statist.*, 8, 22-45.

Castellana, J.V. and Leadbetter, M.R. (1986). On smoothed probability density estimation for stationary processes, *Stochastic Process. Appl.*, 21, 179-193.

Doukhan, P. (1994). *Mixing : Properties and Examples*, Lecture Notes in Statistics, 85, Springer-Verlag.

Frenay, A. (2001). Sur l'estimation de la densité marginale d'un processus à temps continu par projection orthogonale, *Ann. I.S.U.P.*, 45, 55-92.

Geman, D. and Horowitz, J. (1980). Occupation densities, *Ann. Probab.*, 8, 1-67.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer-Verlag, New-York.

Rio, E. (2000). *Théorie Asymptotique des Processus Aléatoires Faiblements Dépendants*, Mathématiques & Applications, (31), Springer-Verlag.

Resum

Construim un estimador de projecció conduïda per les dades per a processos en temps continu. Aquest estimador assoleix taxes super-òptimes sobre una classe $\mathcal{F}_0$ de densitats que és densa en la família de totes les densitats, i assoleix, a la vegada, taxes "raonables". La classe $\mathcal{F}_0$ pot ésser escollida prèviament per l'Estadístic.
Els resultats s'apliquen a processos a valors $\mathbb{R}^d$ i a valors N. En el cas particular on existeix un temps local de quadrat integrable, es demostra que el nostre estimador és estrictament millor que l'estimador temps local sobre $\mathcal{F}_0$.

MSC: 62G07, 62M

Paraules clau: estimació de densitats, conduït per les dades, processos a temps continu