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

Denis Bosq and Delphine Blanke
(Université Pierre et Marie Curie, Paris - France)

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 Statistician. 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 \( \mathcal{F}_0 \).

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

AMS Classification
62G07, 62M.

1 Introduction

We study a data-driven projection density estimator \( \hat{f}_T \) in a general framework. 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 Statistician.

The results are, in some sense, extensions of those which were obtained in the i.i.d. case (cf Bosq 2002a, 2002b), 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 \( \left( \frac{\ln \ln T}{T} \right)^{1/2} \)-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 one does exist. 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 $\mu$ $\sigma$-finite, and such that $L^2(\mu)$ is infinite dimensional. The norm of $L^2(\mu)$ will be denoted $\| \cdot \|$. 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 $\mu$.

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

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

The class of the observable processes will be denoted $X$. 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:

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

and

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

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

We always suppose that (unless otherwise stated)

$$
k_T \to \infty, \quad \frac{k_T}{T} \to 0, \quad \gamma_T \to 0, \quad \text{as } T \to \infty.
$$
If \( \gamma_T \equiv 0 \) one obtains the projection density estimator

\[
(2.2) \quad f_T = \sum_{j=0}^{k_T} \hat{a}_j e_j
\]

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

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

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

\[
(2.3) \quad \alpha(C, D) = \sup_{C \in C, D \in D} |P(C \cap D) - P(C)P(D)|
\]

where \( C \) and \( D \) are sub-\( \sigma \)-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

\[
\alpha^{(2)}(u) = \sup_{h \in I, h+u \in I} \alpha(\sigma(Y_h), \sigma(Y_{h+u})), \quad u \geq 0 \quad \text{and} \quad \alpha_Y(u) = \sup_{h \in \mathbb{R}} \alpha(\sigma(Y_t, t \leq h, t \in I), \sigma(Y_t, t \geq h+u, t \in I)), \quad u \geq 0
\]

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

Now the main assumptions and conditions are \( H_1 \) and \( H_2 \):

\[
H_1 \begin{cases}
A_1 & : P_{(X_{s+t}, X_{s+t+h})} = P_{(X_s, X_t)}; \quad s, t, h \in \mathbb{R} \quad \text{(2-stationarity)}, \\
B_1(r) & : M_r = \sup_{j \geq 0} \|e_j(X_0)\|_r < \infty, \quad \text{where} \ 2 < r \leq \infty, \\
C_1(r) & : \int_0^\infty \left[ \alpha^{(2)}(u) \right]^{(r-2)/r} du < \infty, \\
c_1 & : \gamma_T \simeq T^{-\gamma} \quad (\gamma > 0) \ \text{and} \ k_T \simeq T^\beta \quad (0 < \beta < 1).
\end{cases}
\]
\[ H_2 \left\{ \begin{array}{l} A_2 : \text{X is strictly stationary}, \\
B_2 = B_1(\infty) : M = \sup_{j \geq 0} \|e_j(X_0)\|_\infty < \infty, \\
C_2 : \alpha_X(u) \leq a e^{-bu} (a > 0, b > 0) \\
\quad (X \text{ is geometrically strongly mixing, (GSM)}), \\
c_2 : \gamma_T = \left(\frac{\ln T \ln \ln T}{T}\right)^{1/2}. 
\end{array} \right. \]

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 \( \mu \) 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 \( \widehat{k}_T \) 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 next Proposition shows that \( \widehat{k}_T \) is actually a consistent estimator of \( K(f) \).

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

1) if \( H_1 \) holds,

\[
P(\widehat{k}_T \neq K(f)) = O(T^{\beta+2\gamma-1})
\]

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

2) If \( H_2 \) holds,

\[
P(\widehat{k}_T \neq K(f)) = o(T^{-\delta}),
\]

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

\[
\widehat{k}_{T_n} = K(f) \text{ almost surely for } n \text{ large enough.}
\]
These results show that the adaptive estimator \( \hat{f}_T \) has asymptotically the same behaviour as the pseudo-estimator

\[
(3.4) \quad g_T = \sum_{j=0}^{K(f)} \hat{a}_j e_j.
\]

The next lemma makes this fact explicit:

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

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

We now indicate the rates of \( \hat{f}_T \) on \( F_0 \), we begin with the mean integrated square error (MISE).

**Proposition 3.2** If \( f \in F_0 \), then

1) If \( H_1 \) holds we have

\[
(3.6) \quad E \left\| \hat{f}_T - f \right\|^2 = O \left( \frac{1}{T^{1-\beta}} \right)
\]

2) If \( H_2 \) holds,

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

The next statement gives an uniform result.

**Corollary 3.1**

\[
(3.8) \quad \limsup_{T \to \infty} \sup_{X \in \mathcal{X}_0(a_0,b_0,K_0)} T.E \left\| \hat{f}_T - f \right\|^2 \leq \frac{8a_0 M^2 K_0}{b_0}.
\]

Here \( \mathcal{X}_0(a_0,b_0,K_0) \) denotes the family of processes that satisfy \( H_2 \) with \( f \in F_0(K), K \leq K_0 \) and \( \alpha_X(u) \leq ae^{-bu} \) where \( a \leq a_0 \) and \( b \geq b_0 \).

We now turn to the \( \| \cdot \|_\infty \)-error:
Proposition 3.3  If $f \in F_0$ and $H_2$ holds, then
\begin{equation}
(\forall \varepsilon > 0), (\forall \delta > 0), \quad P\left( \left\| \hat{f}_T - f \right\|_{\infty} \geq \varepsilon \right) = O(T^{-\delta}),
\end{equation}
and, if $T = T_n = nh \ (h > 0), \ n \to \infty$,
\begin{equation}
\left\| \hat{f}_T - f \right\|_{\infty} = O\left( \left( \frac{\ln \ln T}{T} \right)^{1/2} \right), \text{ almost surely.}
\end{equation}
Finally the limit in distribution appears in the following statement :

Proposition 3.4  If $f \in F_0$, $H_2$ holds and $T = nh \ (h > 0)$ then
\begin{equation}
\sqrt{T} (\hat{f}_T - f) \Rightarrow N
\end{equation}
where “$\Rightarrow$” 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 $f \in F_0$. In general these rates appear if the Castellana-Leadbetter condition holds (cf 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 $F_0$.

4  Asymptotic behaviour of $\hat{f}_T$ over $F_1$

In order to study consistency of $\hat{f}_T$ when $f \in F_1$ 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 $f \in F_1$ then
1) If $H_1$ holds
\begin{equation}
P\left( \hat{k}_T < A \right) = O(T^{-1}), \ A > 0,
\end{equation}
2) If $H_2$ holds
\begin{equation}
P\left( \hat{k}_T < A \right) = O\left( \exp\left(-c_A \sqrt{T} \right) \right), \ (c_A > 0), \ A > 0.
\end{equation}
Now we specify the asymptotic comportment of $\hat{k}_T$. For this purpose we set

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

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

Now we put $q_T(\varepsilon) = q\left((1+\varepsilon)\gamma_T\right), \ \varepsilon > 0; \ q'_T(\varepsilon') = q\left((1-\varepsilon')\gamma_T\right), \ 0 < \varepsilon' < 1$ 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 $H_1$,

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

2) Under $H_2$,

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

We indicate two applications of these results :

**Example 4.1** Under $H_1$, if $|a_j| \simeq j^{-\eta} (\eta > \frac{1}{2})$ 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 $|a_j| \simeq \alpha \rho^j (\alpha > 0, \ 0 < \rho < 1)$ and $k_T > \left[1 + (2\ln 1/\rho)^{-1}\right] \ln T$ one has $q_T(\varepsilon) \simeq \frac{\ln T}{2\ln(1/\rho)}$.

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

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

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

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

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

7
Proposition 4.3 If $f \in \mathcal{F}_1$ and $q_T(\varepsilon) \leq k_T$ then

1) Under $H_1$,

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

2) Under $H_2$,

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

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

$$(4.10) \quad E \left\| \hat{f}_T - f \right\|^2 = \mathcal{O}\left(T^{-2\eta-1}\right),$$

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

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

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

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

$$(4.12) \quad E \left\| \hat{f}_T - f \right\|^2 = \mathcal{O}\left(\frac{\ln T}{T}\right).$$

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

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

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 $\mu$. More precisely we make the following assumption:

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

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

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 $\sigma$-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 $\mu_T$ defined by

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

Examples

1) If $E = \mathbb{N}$ and $\mu$ is the counting measure then $H_3$ is satisfied and

$$f_{T,0}(x) = \frac{1}{T} \int_0^T 1_{\{x\}}(X_t) \, dt, \quad x \in \mathbb{N} \tag{5.3}$$
2) If \( E = \mathbb{R} \), and \( \mu \) is Lebesgue measure, \( H_3 \) is equivalent to

\[
\liminf_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \int \int_{[0,T]^2} P(|X_t - X_s| \leq \varepsilon) \, ds \, dt < \infty, \quad T > 0
\]

(cf Geman and Horowitz, 1980).

3) If \((E, \mathcal{B}, \mu) \subseteq (E_0, \mathcal{B}_0, \mu_0)\) with \( \mu = g.\mu_0 \) and \( 0 < m \leq g \leq m' < \infty \) then if \( H_3 \) holds for \( \mu_0 \) with local time \( \ell_T^{(0)} \), it holds for \( \mu \) 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 \( L = \sum_{k \in \mathbb{Z}} \int \text{Cov} (\ell_{(1)}(x), \ell_{(k)}(x)) \, d\mu(x) \) converges.

Note that the Davydov’s inequality (DI) shows that a sufficient condition for \( H_4 \) is

\( H'_4 : X \) is strictly stationary and there exists \( r > 2 \) such that

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

Now the following statement exhibits superefficiency of \( f_{T,0} \):

**Proposition 5.1** If \( H_3 \) and \( H_4 \) hold, then

\[
(5.5) \quad T.E \|f_{T,0} - f\|^2 \rightarrow L, \quad f \in \mathcal{F}.
\]

Concerning \( \hat{f}_T \) we have
Proposition 5.2  If $H_3$ and $H_4$ hold, then

\[(5.6) \quad \mathbb{E} \left\| \hat{f}_T - f \right\|^2 = \mathcal{O} \left( \frac{1}{T} \right) + \mathbb{E} \left( \sum_{j > k_T} a_j^2 \right). \]

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

\[(5.7) \quad \mathbb{E} \left\| \hat{f}_T - \Pi^{k_T} f \right\|^2 \leq \mathbb{E} \| f_{T,0} - f \| \quad \text{and} \quad \mathbb{E} \left\| \hat{f}_T - \Pi^{k_T} f \right\|^2 = \mathcal{O} \left( \frac{1}{T} \right). \]

Consequently the efficiency of $\hat{f}_T$ depends on the “pseudo-bias” $\sum_{j > k_T} a_j^2$. Under conditions in Proposition 4.3 this pseudo-bias may be replaced by $\sum_{j > K(f)} a_j^2$ 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 $f_{T,0} = \sum_{j=0}^{\infty} \hat{a}_{j,T} e_j$, when $\hat{f}_T$ has the same asymptotic behaviour as $g_T = \sum_{j=0}^{K(f)} \hat{a}_{j,T} e_j$ and more precisely :

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

\[(5.9) \quad \liminf_{T \to \infty} T. \mathbb{E} \left\| f_{T,0} - f \right\|^2 \geq \sum_{j=0}^{\infty} \int_0^\infty \text{Cov} \left( e_j(X_0), e_j(X_u) \right) \, du \]

when

\[(5.10) \quad T. \mathbb{E} \left\| \hat{f}_T - f \right\|^2 \to \sum_{j=0}^{K(f)} \int_0^\infty \text{Cov} \left( e_j(X_0), e_j(X_u) \right) \, du. \]

It is easy to construct examples where $\int_0^\infty \text{Cov} \left( e_j(X_0), e_j(X_u) \right) \, du > 0$ 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 $B_T = \{ \exists j \ : \ 0 \leq j \leq k_T, \ |\hat{a}_{j,T}| \geq \gamma_{T} \}$, then, we have for $T$ large enough and $K = K(f)$, $B^c_T \Rightarrow |\hat{a}_{K,T}| < \gamma_T \leq |a_{K,T}| / 2 \Rightarrow |a_{K,T} - \hat{a}_{K,T}| \geq |a_{K,T}| / 2$, thus
\[ P(B_T^c) \leq \frac{4 \text{Var} \hat{a}_K T}{|a_K|^2}. \] Now, 2-stationarity yields

\[ \text{(6.1)} \quad \text{Var} \hat{a}_K T = \frac{2}{T} \int_0^T (1 - \frac{u}{T}) \text{Cov} \left( e_K(X_0), e_K(X_u) \right) \, du, \]

using Davydov’s inequality (DI), cf Bosq (1998) p. 21, one obtains

\[ \text{Var} \hat{a}_K T \leq \frac{2}{T} \int_0^T (1 - \frac{u}{T}) 2^{r-2} \frac{a^{(2)}(u)}{\alpha_X(u)} \frac{r^2}{\gamma^2} \|e_K(X_0)\|_r^2 \, du, \]

and \( H_1 \) implies

\[ \text{(6.2)} \quad \text{Var} \hat{a}_K T \leq \frac{c_r}{T}, \]

where \( c_r = \frac{4rM^22^{r-2}}{\gamma^2} \int_0^\infty \left[ \alpha_X(u) \right]^{\frac{r-2}{r}} \, du \) thus

\[ \text{(6.3)} \quad P(B_T^c) \leq \frac{4c_r}{a_K^2} \frac{1}{T}. \]

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

\[ \text{(6.4)} \quad \left\{ \hat{k}_T > K, B_T \right\} \Rightarrow \bigcup_{j=K+1}^{k_T} \left\{ |\hat{a}_j| \geq \gamma_T \right\} \]

and

\[ \text{(6.5)} \quad \left\{ \hat{k}_T < K, B_T \right\} \Rightarrow |\hat{a}_K - a_K| > |\frac{a_K}{2}| \Rightarrow |\hat{a}_K - a_K| > \gamma_T \]

thus

\[ \text{(6.6)} \quad P(\hat{k}_T \neq K, B_T) \leq \frac{1}{\gamma_T^2} \sum_{j=K}^{k_T} \text{Var} \hat{a}_K, \]

using again DI one obtains

\[ \text{(6.7)} \quad P(\hat{k}_T \neq K, B_T) = O\left(\frac{k_T + 1}{\gamma_T^2 T} \right) = O\left( T^{3\gamma^2 - 1} \right). \]
Now, since (6.3) implies

\[ P(\hat{k}_T \neq K, B^c_T) = O\left(\frac{1}{T}\right), \]

(6.8) follows. \( \square \)

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

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

\[ P\left(\frac{1}{T} \int_0^T Y_t - \mathbb{E} Y_t \, dt \geq \eta\right) \leq 4 \exp\left(-\frac{T \eta^2 / M_Y^2}{c_1 + c_2 \frac{T}{r} + c_3 M_Y^{-1} \eta r}\right) + c_4 M_Y \alpha_Y(r) \]

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

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

\[ V_T(j) = \int_{(j-1)r}^{j r} (Y_t - \mathbb{E} Y_t) \, dt \quad \text{and} \quad V_T(2[q]) = \int_{(2[q]-1)r}^{2qr} (Y_t - \mathbb{E} Y_t) \, dt. \]

So, for any \( \eta > 0 \),

\[ P\left(\frac{1}{T} \int_0^T Y_t - \mathbb{E} Y_t \, dt \geq \eta\right) \leq P\left(\sum_{j=1}^{[q]} V_T(2j) > \frac{T \eta}{2}\right) + P\left(\sum_{j=1}^{[q]} V_T(2j-1) > \frac{T \eta}{2}\right). \]

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

\[ \mathbb{E} |V'^*_T(2j) - V_T(2j)| \leq 2\|V_T(2j)\|_{\infty} \left(\sup |P(AB) - P(A)P(B)|\right) \]

13
where the supremum is taken over all sets $A$ and $B$ belonging to $\sigma$-algebras of events generated by respectively $\{V_T(2), \ldots, V_T(2j - 2)\}$ and $V_T(2j)$.

For any positive $\kappa$, one may write

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

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

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

\[
P(\{\sum_{j=1}^{[q]} V_T^*(2j) > \frac{T\eta}{2(1 + \kappa)}\} \leq 2 \exp\left(-\frac{T\eta^2/M_T^2}{c_1 + c_2 T + c_3 M_T^{-1}\eta^2}\right)
\]

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

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

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 $P(|\hat{a}_{j_T} - a_j| \geq \eta)$ for suitable $\eta$. In particular, one obtains

(6.10) \[ P(B_T^c) = O(\exp(-AN\sqrt{T})), \quad (A > 0), \]

technical details are omitted.

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

### 6.2 Proof of Lemma 3.1

It suffices to write \[ \|\hat{f}_T - g_T\|^2 = \|\hat{f}_T - g_T\|^2 \mathbf{1}_{\{\hat{k}_T \neq K\}} \leq \sum_{j=1}^{k_T} \hat{a}_{j_T}^2 \mathbf{1}_{\{\hat{k}_T \neq K\}} \]

\[ \leq M^2 k_T \mathbf{1}_{\{\hat{k}_T \neq K\}}, \]

hence (3.5) by taking expectations. ■
6.3 Proof of Proposition 3.2

First we have,

\begin{equation}
E \left\| \hat{f}_T - f \right\|^2 = E \left( \sum_{j=0}^{\hat{k}_T} (\hat{a}_{jT} - a_j)^2 \right) + E \left( \sum_{j>\hat{k}_T} a_j^2 \right)
\end{equation}

then, by DI: \(E \left( \sum_{j=0}^{\hat{k}_T} (\hat{a}_{jT} - a_j)^2 \right) \leq \sum_{j=0}^{\hat{k}_T} \text{Var} \ \hat{a}_{jT} \leq c \frac{\hat{k}_T}{T} \). On the other hand, if \(f \in \mathcal{F}_0(K)\), \(\sum_{j>\hat{k}_T} a_j^2 = \sum_{j>\hat{k}_T} a_j^2 I\{\hat{k}_T<K\}\), hence \(E \left( \sum_{j>\hat{k}_T} a_j^2 \right) \leq \|f\|^2 P(\hat{k}_T < K)\). Now from (6.5) and (6.8) it follows that \(P(\hat{k}_T < K) \leq P(|\hat{a}_{kT} - a_k| > |a_k|) + O(\frac{1}{T})\). Using DI one obtains the bound

\begin{equation}
E \left( \sum_{j>\hat{k}_T} a_j^2 \right) = O\left( \frac{1}{T} \right),
\end{equation}

and (6.11) gives (3.6). Concerning (6.7) 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 \(E \left\| \hat{f}_T - g_T \right\| = o(\frac{1}{T})\). Thus it is only necessary to study

\begin{equation}
E \left\| g_T - f \right\|^2 = \sum_{j=0}^{K} \text{Var} \ \hat{a}_{jT},
\end{equation}

but using BI one obtains

\begin{equation}
\int_0^\infty \left| \text{Cov} \ (e_j(X_0), e_j(X_u)) \right| \ du \leq 4M^2 \int_0^\infty ae^{-bu} \leq \frac{4aM^2}{b} < \infty.
\end{equation}

Now since

\begin{equation}
T \text{Var} \ \hat{a}_{jT} = 2 \int_0^T (1 - \frac{u}{T}) \text{Cov} \ (e_j(X_0), e_j(X_u)) \ du,
\end{equation}

the dominated convergence theorem gives

\begin{equation}
T \text{Var} \ \hat{a}_{jT} \rightarrow 2 \int_0^\infty \text{Cov} \ (e_j(X_0), e_j(X_u)) \ du
\end{equation}

and (6.13) yields (3.7). ■
6.4 Proof of Corollary 3.1
It suffices to apply BI in (6.15) and to verify that the other bounds are uniform over \(X_0(a_0, b_0, K_0)\); details are omitted.

6.5 Proof of Proposition 3.3
First, putting \(K(f) = K\) one has
\[
|\hat{f}_T - g_T| = |(\hat{f}_T - g_T) I_{\{\hat{k}_T \neq K\}}| \leq \sum_{j=1}^{k_T} |\hat{a}_{j_T}| I_{\{\hat{k}_T \neq K\}} \leq M^2 k_T I_{\{\hat{k}_T \neq K\}},
\]
one obtains, for all \(\epsilon > 0\) and all \(\delta > 0\),
\[
(6.17) \quad P\left(\|\hat{f}_T - g_T\|_{\infty} \geq \epsilon\right) \leq P(\hat{k}_T \neq K) = o(T^{-\delta}).
\]
Now, \(P(\|g_T - f\|_{\infty} \geq \epsilon) \leq \sum_{j=0}^{K} P(|\hat{a}_{j_T} - a_j| \geq \frac{\epsilon}{K^2M})\), then, using (3.9) for \(Y_t = e_j(X_t), 0 \leq t \leq T; 0 \leq j \leq K\), with \(r = B \ln T\) one arrives at the bound
\[
P(|\hat{a}_{j_T} - a_j| \geq \frac{\epsilon}{K^2M}) \leq 4 \exp\left(-\frac{3\epsilon/K^2MB}{T \ln T/16(1 + \kappa)(1 + o(1))}\right)
+ 64 \frac{1 + \kappa K^2M^2}{\kappa} \frac{1}{\epsilon} a \exp(-bB \ln T)
\]
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 \((Z_{ij}^{(h)} = \frac{1}{n} \int_{(i-1)h}^{ih} e_j(X_t) - a_j dt, i \geq 0)\) since these processes are bounded and geometrically strongly mixing. It follows that \(\|g_T - f\|_{\infty} = O\left((\frac{\ln \ln T}{\ln T})^{1/2}\right)\) a.s. 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 \(\sqrt{T}\left\|g_T - \hat{f}_T\right\|_{\infty} \rightarrow 0\) in probability (cf (6.17)), theorem 4.4 in Billingsley (1979) shows that it suffices to study asymptotic normality of
\[
\sqrt{T}(g_T - f) = \sum_{j=0}^{K} (\hat{a}_{j_T} - a_j)e_j.
\]
This is equivalent to asymptotic normality of the finite dimensional random vector \( \sqrt{T}(\tilde{a}_0 - a_0, \ldots, \tilde{a}_K - a_K) \) which in turn is equivalent to this of the real random variables \( \sqrt{T} \sum_{j=0}^{K} \lambda_j (\tilde{a}_j - a_j) ; \lambda_1, \ldots, \lambda_K \in \mathbb{R} \). 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

\[
\left\{ \hat{k}_T < j_0 \right\} \Rightarrow |\hat{a}_{j_0} - a_{j_0}| > \frac{|a_{j_0}|}{2}
\]

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) entails (4.2). Details are omitted.

6.8 Proof of Proposition 4.2

For \( T \) large enough we have \( |a_{q_T(\epsilon)}| > (1 + \epsilon)\gamma_T \).

1) From DI we get

\[
P(\hat{k}_T < q_T(\epsilon), B_T) \leq P(|\hat{a}_{q_T(\epsilon)} - a_{q_T(\epsilon)}| > \epsilon \gamma_T) \leq \frac{c'}{\epsilon^2 T \gamma_T^2}.
\]

Now, if \( q_T'(\epsilon') \geq k_T \) one has \( P(\hat{k}_T > q_T'(\epsilon')) = 0 \), if not, since \( |a_j| \leq (1 - \epsilon') \gamma_T \) for \( j > q_T'(\epsilon') \), we have

\[
\left\{ \hat{k}_T > q_T'(\epsilon'), B_T \right\} \Rightarrow \bigcup_{k_T \geq j > q_T'(\epsilon')} |\hat{a}_j - a_j| > \epsilon' \gamma_T
\]

thus \( P(\hat{k}_T > q_T'(\epsilon'), B_T) \leq \frac{c'(k_T+1)}{\epsilon'^2 T \gamma_T^2} \) and (4.3) follows.

2) For proving (4.5) we may and do suppose that \( q_T'(\epsilon') < k_T \). Then

\[
P(E_T^c \cap B_T) \leq P(|\hat{a}_{q_T(\epsilon)} - a_{q_T(\epsilon)}| \geq \epsilon \gamma_T) + \sum_{q_T'(\epsilon') < j \leq k_T} P(|\hat{a}_j - a_j| \geq \epsilon' \gamma_T),
\]

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

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

for some constant \( c' \) and the choice \( c > (\frac{3}{2} + \delta) b^{-1} \) entails (4.3) since \( P(B_T^c) = o(T^{-\delta}) \) for all \( \delta > 0 \).
6.9 Proof of Proposition 4.4

We start from (6.11) and write
\[ E \left( \sum_{j > k_T} a_j^2 \mathbf{1}_{E_T \cap B_T} \right) \leq \| f \|^2 P(E_T), \quad E \left( \sum_{j > k_T} a_j^2 \mathbf{1}_{E_T \cap B_T} \right) \leq \sum_{j > q_T(\varepsilon)} a_j^2, \]
\[ E \left( \sum_{j=0}^{k_T}(\hat{a}_{j_T} - a_j)^2 \mathbf{1}_{E_T \cap B_T} \right) \leq \sum_{j=0}^{k_T} \text{Var} \hat{a}_{j_T}. \]
Finally, under \( H_1 \) we write
\[ E \left( \sum_{j=0}^{k_T}(\hat{a}_{j_T} - a_j)^2 \mathbf{1}_{E_T \cap B_T} \right) \leq \sum_{j=0}^{k_T} \text{Var} \hat{a}_{j_T}, \]
\[ E \left( \sum_{j=0}^{k_T}(\hat{a}_{j_T} - a_j)^2 \mathbf{1}_{E_T \cap B_T} \right) \leq 4M^2 k_T P(E_T), \]
using the above bounds, (6.10) and (6.11) one obtains (4.8) and (4.9). \( \blacksquare \)

6.10 Proof of Proposition 4.4

Let \( \xi \) be a positive constant, for any positive \( \kappa_i, i = 1, 2 \) one obtains
\[ P(\| \hat{f}_T - f \|_\infty \geq \xi) \leq P(\| \hat{f}_T - f \|_\infty \mathbf{1}_{E_T} \geq \frac{\xi}{1 + \kappa_1}) + P(\| \hat{f}_T - f \|_\infty \mathbf{1}_{E_T^c} \geq \frac{\xi \kappa_1}{1 + \kappa_1}) \]
\[ \leq P_1 + P_2 + P_3 \]
with \( P_1 = \sum_{j=1}^{q_T(\varepsilon')} \mathbb{P}(Mq_T(\varepsilon') | \hat{a}_{j_T} - E \hat{a}_{j_T} | \geq \frac{\xi}{(1 + \kappa_1)(1 + \kappa_2)}), \quad P_3 = P(E_T^c), \quad P_3 = P(E_T^c) \)
\[ P_2 = P(M \sum_{j=q_T(\varepsilon)+1}^{\infty} |a_j| \geq \frac{\xi \kappa_2}{(1 + \kappa_1)(1 + \kappa_2)}). \]

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

Now noting that \( \sum_{j=q_T(\varepsilon)+1}^{\infty} |a_j| \leq C(\alpha, \rho) \gamma_T \), it is easy to see that for \( T_n \) large enough, \( P_2 \equiv 0 \) with previous choices of \( \gamma_T \) and \( \xi \). Moreover, Proposition 4.2 implies also \( P_3 = o(T_n^{-\delta}). \)
Finally, collecting these results, one obtains Proposition 4.3 with the help of Borel-Cantelli’s lemma since \( \sum_n \frac{\ln T}{T_n} < \infty. \) \( \blacksquare \)

6.11 Proof of Proposition 5.1

Using additivity of local time one may write
\[ \ell_T = \ell_0 + \frac{1}{T} \sum_{j=1}^{[T]} \ell(j) + \frac{\ell([T], T]}{T}. \]
Since \( E \left\| \frac{\ell_0}{T} \right\|^2 = o(\frac{1}{T}) \) and \( E \left\| \frac{\ell([T], T]}{T} \right\|^2 \leq \frac{E \left\| \ell_0 \right\|^2}{T} = o(\frac{1}{T}) \) it suffices to
study

\[(6.19) \quad nE\left|\frac{1}{n}\sum_{j=1}^{n} \ell(j) - f\right|^2 = \right.
\]
\[E\left|\ell(1) - f\right|^2 + 2\sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \int_{E} \text{Cov}(\ell(1)(x), \ell(k+1)(x)) \, d\mu(x) \]

where \(n = \lfloor T \rfloor\). 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 \(\text{sp}(e_j, 0 \leq j \leq \hat{k}_T)\), we have

\[\left\|\Pi^{\hat{k}_T}(f,0) - f\right\| \leq \left\|f,0 - f\right\| \text{ thus } E\left\|\hat{f}_T - \Pi^{\hat{k}_T}f\right\|^2 \leq E\left\|f,0 - f\right\|^2 \]

and \(5.5\) implies \(\limsup_{T \to \infty} T. E\left\|\hat{f}_T - \Pi^{\hat{k}_T}f\right\|^2 \leq L\) hence \(5.6\) from \(6.11\) and the fact that \(P(E^c_T \cup B^c_T) = o\left(\frac{1}{T}\right)\).

6.13 Proof of Proposition 5.3

\(5.10\) has been proved in Proposition 3.2. Concerning \(5.9\) first note that

\(5.1\) implies \(\frac{1}{T} \int_{0}^{T} e_j(x_t) \, dt = \frac{1}{T} \int_{E} e_j(x) \ell_T(x) \, dx\) thus \(\hat{a}_{j,T} = \int_{E} f_{T,0}(x) e_j(x) \, d\mu(x)\), \(j \geq 0\), hence \(f_{T,0} = \sum_{j=0}^{\infty} \hat{a}_{j,T} e_j\) and \(\sum \hat{a}_{j,T}^2 < \infty\) (a.s.) ; then we have

\(T E\left\|f,0 - f\right\|^2 = \sum_{j=0}^{\infty} T \text{Var} \hat{a}_{j,T}\) but \(H_2\) yields \(\int_{0}^{\infty} |\text{Cov}(e_j(X_0), e_j(X_u))| \, du < \infty\) and \(6.16\) holds. This implies \(5.9\) by using Fatou lemma for the counting measure.

References


