Modelization and Nonparametric estimation for a dynamical system with noise

Dominique Guegan, D. Bosq, Delphine Blanke

Modelization and Nonparametric Estimation for a Dynamical System with Noise

D. Blanke[1] D. Bosq[2] and D. Guégan[3]

Abstract: We examine the effect of two specific noises (either known or small ones) on a dynamical system. We obtain consistent estimates with their rates of convergence for the invariant density in that context.

Keywords: Invariant Measure, Nonparametric Estimation, Dynamical System, Small Noise, Deconvolution.

1 Introduction

In this paper, we focus on estimation of the invariant measure of a dynamical system disturbed by noise. Let us recall that a discrete time dynamical system is usually defined via a measurable mapping $\varphi : E \to E$ where E is a closed subset of $\mathbb{R}^d$ such that the state of the system at time t is given by:

$$x_t = \varphi^{(t)}(x_0) \tag{1.1}$$

where $\varphi^{(t)} = \varphi \, o \, \varphi \cdots o \, \varphi$ (t times) and $x_0 \in E$ is the state of the system at time $t = 0$. Note that (1.1) implies

$$x_t = \varphi(x_{t-1}), \qquad t \geq 1. \tag{1.2}$$

Nevertheless a purely theoretical system like (1.2) is quite unrealistic since observations x_t are in general corrupted by some noise.

We assume therefore that we observe a "noisy" trajectory $y_1, y_2, \cdots, y_n$ which leads to the more natural model:

$$y_t = \psi(y_{t-1}, \delta_t) \;, \qquad t \geq 1 \tag{1.3}$$

where ψ is a measurable function: $E \times F \to E$, ($F \in \mathcal{B}_{\mathbb{R}^d}$) and where $(\delta_t, t \geq 1)$ is the noise which pollutes the system. In the following, typical kinds of noise will be introduced

[1]University of Le Havre, 25 rue Philippe Lebon, BP 540, 76058 Le Havre Cedex, France, email: Delphine.Blanke@univ-lehavre.fr

[2]University Paris 6, I.S.U.P., 4 place Jussieu, 75252 Paris Cedex 05, France

[3]UPRESA 6056, University of Reims, CREST-ENSAE-Paris, 3 av. P. Larousse, Timbre J120, 92245 Malakoff Cedex, France, email: guegan@ensae.fr

and studied.

In this paper, we are interested in ergodic dynamical system, thus we consider a probability space $(E, \mathcal{B}(E), \mu)$ where E is a Borel set of $\mathbb{R}^d$ and $\mathcal{B}(E)$ is its Borel σ-field. We thus re-write (1.2) as:

$$X_t = \varphi(X_{t-1}), \qquad t \geq 1 \tag{1.4}$$

where X_0 is a given E-valued random variable.
In the same way, (1.3) becomes

$$Y_t = \psi(Y_{t-1}, \Delta_t), \qquad t \geq 1 \tag{1.5}$$

where (Y_t) and (Δ_t) are sequences of random variables.

Now, we intend to estimate nonparametrically the density f of the (possible) invariant measure μ associated to models such as (1.4) when observations are given by (1.5). Only two particular but important cases of (1.5) (see (2.1) and (2.3)) will be considered in the sequel.

Many works have been devoted to dynamical systems, we refer to Lasota-MacKey (1994) and the references therein for further reading. Note that in this paper, we have rather followed a statistical approach of such dynamical systems.
The plan of our paper is the following: in Section 2 we specify the noises which will be studied. In Section 3 we give conditions for existence and uniqueness of an invariant measure. Section 4 and 5 deal with estimation of f for two specific systems: known and small noises. Proofs appear in Section 6.

2 Classification of Noises

One may encounter many kind of noises in experimental systems, we now specify those which are studied in our paper.

- First suppose that we observe a dynamical system such (1.4) but with errors-in-variables. Thus, we get the following model:

$$\begin{cases} Y_t &= X_t + \varepsilon_t, \qquad t \geq 0 \\ X_t &= \varphi(X_{t-1}), \qquad t \geq 1. \end{cases} \tag{2.1}$$

  Note that (2.1) gives one possible model which is a particular case of (1.5) since one may write $Y_t = \psi(Y_{t-1}, \Delta_t)$ with

$$\begin{cases} \psi(y, \delta) &= \varphi(y) + \delta \ , \\ \Delta_t &= \varphi(X_{t-1}) - \varphi(Y_{t-1}) + \varepsilon_t \ , \qquad t \geq 1 \ , \end{cases} \tag{2.2}$$

  provided that $Y_t \in E$ i.e. ε_t satisfies the condition $X_t + \varepsilon_t \in E$ for $t \geq 1$.

This latter condition is somewhat restricting, but it can be relaxed by supposing that $E = \mathbb{R}^d$, even if the system lives in some subset of $\mathbb{R}^d$. In the particular case where ε_t is bounded ($\|\varepsilon_t\| \le \varepsilon$) it is only necessary to suppose that φ has a natural extension to $E^\varepsilon = \{x : \inf_{y \in E} \|x - y\| \le \varepsilon\}$ and then to work over this set.

System (2.1) corresponds to model with measurement errors. This case appears, for instance, when one wants to simulate systems, in particular aperiodic ones (like chaotic systems, see Guégan and Mercier, 1998). Now, if in (2.1), ε_t is independent of X_t and ε_t has a known invariant measure ν, $t \ge 1$, existence of an invariant measure, say μ for (X_t) implies the same property for (Y_t) and the invariant measure associated with (Y_t) is $\mu * \nu$ where $*$ denotes convolution product. If the characteristic function of ν does not vanish, then $\mu * \nu$ determines μ: in that case one deals with a deconvolution problem.

- Now, we consider a general model which corresponds to propagation of errors. We take model (1.5) with $\psi(y, \delta) = \varphi(y) + \delta$ and $E = \mathbb{R}$:

$$Y_t = \varphi(Y_{t-1}) + \Delta_t. \tag{2.3}$$

Assume Y_0 is observed, we get:

$$Y_1 = \varphi(Y_0) + \Delta_1 \ ,$$

thus

$$Y_2 = \varphi[\varphi(Y_0) + \Delta_1] + \Delta_2 \ .$$

For a general t, the relation between Y_t and Y_0 is intricate. However, this representation may be simplified by using successive approximations (see appendix for details):

$$Y_t = \varphi^{(t)}(Y_0) + \xi_t \ , \qquad t \ge 1 \tag{2.4}$$

where

$$\xi_t = \Delta_t + \Gamma_{t-1}\Delta_{t-1} + \Gamma_{t-1}\Gamma_{t-2}\Delta_{t-2} + \cdots + \Gamma_{t-1} \cdots \Gamma_1 \Delta_1 \ ,$$

with

$$\Gamma_t = (\varphi' o \varphi^{(t)})(Y_0) \ ,$$

where φ', the derivative of φ, is supposed to exist except in a countable set of points.

Notice that if φ is linear, the models given by (2.3) and (2.4) coincide. Models such as (2.4) can be easily found in experimental situations, for example we refer to the Couette-Taylor fluid flow experiment described in Brandstäter and Swinney (1987), other examples of deterministic noise amplifiers can be found in Deissler and Farmer (1992). In these experimental systems, the smallness of the noise is fundamental. Thus, in order to modelize the smallness of Δ we consider a sequence of observed r.v.'s $(Y_{1n}, \cdots, Y_{nn})$ associated with the noise $(\Delta_{1n}, \cdots, \Delta_{nn})$ where the r.v.'s Δ_{in} are i.i.d., zero-mean, with variance σ_n^2 and such that (Δ_{in}) is independent of Y_0 for any i.

Now, in order to control the noise in (2.4) we can make a classical assumption upon φ' (which is satisfied in usual cases) given by $1 \le \|\varphi'\|_\infty < \infty$ where $\|\cdot\|_\infty$ denotes the essential supremum.
Then the conditional variance of $\xi_t = \xi_{tn}$, (introduced in (2.4)) with respect to Y_0 is:

$$\mathrm{Var}\,(\xi_{tn}|Y_0) = \sigma_n^2(1+\Gamma_{t-1}^2+\cdots+\Gamma_{t-1}^2\cdots\Gamma_1^2) \le \sigma_n^2 n\cdot\|\varphi'\|_\infty^{2(n-1)},$$

and since $\mathrm{Var}\,(\xi_{tn}) = \mathrm{E}\ (\mathrm{Var}\,(\xi_{tn}/Y_0)) + \mathrm{Var}\ (\mathrm{E}\ (\xi_{tn}/Y_0))$, we get

$$\mathrm{Var}\,(\xi_{tn}) \le \sigma_n^2 n\cdot\|\varphi'\|_\infty^{2(n-1)}.$$

Hence we have, $\mathrm{Var}\ (\xi_{tn}) \to 0$ as $n \to +\infty$, provided $\lim\limits_{n\to+\infty} n\sigma_n^2\|\varphi'\|_\infty^{2(n-1)} = 0$.

3 Existence and uniqueness of invariant measure

In this section we give a result about existence and uniqueness of invariant measure for the dynamical system (1.5). We use the so-called FOAIS operator associated to ψ (defined in (1.5)), see Lasota-Mackey (1994, p. 414). Here, we relax somewhat assumptions concerning this operator in order to extend Theorem 12.5.1 in Lasota-Mackey.

Let us suppose that $(\Delta_t, t \ge 1)$ is a sequence of i.i.d. random variables with common distribution ν, that for each $t \ge 1$, (Y_t) admits a distribution measure μ_t, and that (Δ_t) and (Y_t) are independent. Note that the model (2.2) does not satisfy such assumptions but as seen before, the invariant measure will exist as soon as X_t has an invariant measure and (ε_t) are i.i.d. variables independent of (X_t).

Let Π be the FOAIS operator defined over $\mathcal{P}(E)$ (the space of probability measures over $(E, \mathcal{B}_E)$) such that:

$$\begin{aligned}(\Pi\mu)(B) &= \int_E d\mu(x)\int_F 1_B[\psi(x,y)]d\nu(y)\\ &= \int (1_B o\psi)d(\mu\otimes\nu), \qquad B\in\mathcal{B}_E.\end{aligned}$$

Therefore

$$\Pi\mu_t = \mu_{t+1}\ , \qquad t \ge 0\ .$$

We now specify our main assumptions:

Assumptions A3.1
(i) There exists μ_0 belonging to $\mathcal{P}(E)$ such that for all $\eta > 0$, there exists a bounded B in $\mathcal{B}_E$ and such that $(\Pi^t\mu_0)(B) \ge 1-\eta$, for any $t \ge 0$.
(ii) : Π is continuous over $\mathcal{P}(E)$ with respect to the weak topology.

These assumptions are slightly more general than those in Lasota-Mackey (1994, p. 417), since ψ is not assumed to be continuous with respect to the first argument. We have:

Proposition 3.1

If assumptions A3.1 hold, then Π has an unique invariant measure.

4 Model with errors-in-variables

Suppose that we have n observations $Y_0, \ldots, Y_{n-1}$ from the model (2.1):

$$Y_t \;=\; X_t + \varepsilon_t, \qquad t = 0, 1, \ldots, n-1 \tag{4.1}$$

with

$$X_t \;=\; \varphi(X_{t-1}) \qquad t = 1, \ldots, n-1. \tag{4.2}$$

We make the following general assumptions about (4.1).

Assumptions A4.1

(i) $\{X_t,\, t \in \boldsymbol{Z}\}$ is an E-valued process.
(ii) $\{\varepsilon_t,\, t \in \boldsymbol{Z}\}$ are i.i.d. random variables with known density ξ_ε and with independent components.
(iii) $\{X_t,\, t \in \boldsymbol{Z}\}$ and $\{\varepsilon_t,\, t \in \boldsymbol{Z}\}$ are independent.

If we denote by μ the invariant measure of (4.2), our goal is then to estimate its density f (if it exists) with respect to Lebesgue measure over E, when only observations $Y_0, \ldots, Y_{n-1}$ are available and the law of the noise is known. Note that this latter condition is somewhat strong but it permits to ensure the identifiability of our problem. Moreover this condition will be satisfied when for example, one may preliminarily calibrate the measuring instrument. We now make the following assumptions upon the dynamical system $\{X_t,\, t \in \boldsymbol{Z}\}$:

Assumptions A4.2

(i) φ preserves μ and μ has a density f with respect to Lebesgue measure over E.
(ii) X_0 has density f_0 and for any $t \geq 1$, X_t has density $\Pi^t f_0$ where Π is the so-called Frobenius operator (see Lasota-MacKey p. 202) such that

$$\sum_{t=0}^{+\infty} \|\Pi^t f_0 - f\|_\infty < +\infty.$$

(iii) There exists a sequence $c_n \to +\infty$ defined by

$$c_n = \iint_{E\times E} \left| \sum_{\substack{0\le i,j\le n-1\\ i\neq j}} \Big(dP_{(X_i,X_j)}(u,v) - d\mu(u)d\mu(v) \Big) \right|$$

and such that $c_n = o(n^2)$ as $n \to +\infty$.

Note that these assumptions are quite similar to those used in the no-noise case. We refer to Smili (1990), Bosq (1995, 1998) and Bosq-Guégan (1995) for examples of systems satisfying assumptions A4.2. Furthermore if c_n is of same order as n^2, we get a degenerate case.

We finally remark that if f_t, g_t are the respective densities of X_t and Y_t, assumptions $A4.1$ imply that $g_t = f_t * \xi_\varepsilon$ for any t, so we deal here with the classical deconvolution problem but in a non-stationary context.

4.1 Definition of the estimator

Deconvolution kernel estimators have been widely studied for i.i.d. observed variables see e.g. Carrol and Hall (1988), Liu and Taylor (1989), Stefanski and Carrol (1990), Fan (1991), or for mixing and stationary processes: Masry (1991)-(1993), Fan and Masry (1993), as well as in a continuous time context, see Blanke (1995, 1996). These works are related both with density and regression estimation. The nonparametric kernel-type density estimator is defined by:

$$\widehat{f}_n(x) \quad = \quad \frac{1}{nh_n^d}\sum_{i=0}^{n-1} W_h\left(\frac{x-Y_i}{h_n}\right), \quad x \in \mathbb{R}^d \tag{4.3}$$

with

$$W_h(x) = \left(\frac{1}{2\pi}\right)^d \int_{\mathbb{R}^d} e^{-i<t,x>} \frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)}\, dt$$

where ϕ_K is the Fourier transform of a kernel K, ϕ_ε is the characteristic function of the noise ε and $< .,. >$ denotes the scalar product over $\mathbb{R}^d$.

In the following, we will choose K as a kernel product function $K = \widetilde{K} \otimes \ldots \otimes \widetilde{K}$ where $\widetilde{K}$ is a real symetric bounded density such that $\lim\limits_{u\to 0} |u|\widetilde{K}(u) = 0$ and $\int_{\mathbb{R}} u^2\widetilde{K}(u)\, du < +\infty$.

Then, we may write:

$$W_h(x) = \prod_{j=1}^{d} \widetilde{W_h}(x_j) \text{ with } \widetilde{W_h}(x_j) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx_j} \frac{\widetilde{\phi}_K(t)}{\widetilde{\phi}_\varepsilon(t/h_n)}\, dt.$$

Furthermore we will suppose in the sequel that $|\widetilde{\phi}_\varepsilon(t)| \neq 0$ and that $\left|\frac{\widetilde{\phi}_K(.)}{\widetilde{\phi}_\varepsilon(./h)}\right| \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$ for all real t and positive h. Note that the first condition is close to the "minimal" condition (i.e. $|\widetilde{\phi}_\varepsilon(t)| \neq 0$ for almost all t) which ensures identifiability of our problem (see Devroye, 1989). Note that this assumption excludes characteristic functions with compact support such for instance the Beta distribution $\beta(1/2, 1/2)$ defined on $[-1, 1]$.

4.2 Properties of kernel W_h

We consider two general classes of noises specified below by (4.4) and (4.6). As usual, noises satisfying (4.4) and (4.6) will be respectively refered in the sequel as "ordinary smooth noise distributions" and "supersmooth noise distributions".

4.2.1 Ordinary smooth noise distributions

We make the following assumptions on the characteristic functions $\widetilde{\phi}_\varepsilon(t)$ and $\widetilde{\phi}_K(t)$ (see Masry 1991, 1993).

Assumptions A4.3

$\widetilde{\phi}_K(.)$ and $\widetilde{\phi}_\varepsilon(.)$ are twice continuously differentiable with bounded derivatives such that :

(i) $|\widetilde{\phi}_\varepsilon(t)| > 0 \; \forall \, t \in \mathbb{R}$,

(ii) $\exists \, \beta \geq 1$ *such that*

$$t^\beta \, \widetilde{\phi}_\varepsilon(t) \quad \rightarrow \quad A_1 \;\; \text{as } |t| \rightarrow +\infty \;\; \text{with } |A_1| > 0, \tag{4.4}$$

(iii) $\int_{\mathbb{R}} |t|^{\beta-2} \, |\widetilde{\phi}_K(t)| \, dt < +\infty$ *if* $\beta > 1$, $\int_{\mathbb{R}} |t|^{2\beta} \, |\widetilde{\phi}_K(t)|^2 \, dt < +\infty$,

(iv) $\int_{\mathbb{R}} |t|^{\beta-1} \, |\widetilde{\phi}'_K(t)| \, dt < +\infty$, $\int_{\mathbb{R}} |t|^\beta \, |\widetilde{\phi}''_K(t)| \, dt < +\infty$.

Condition (ii) specifies the asymptotic behaviour of the noise, it includes, in particular Laplacian densities ($\beta = 2$) and also the Gamma ones $\Gamma_{(\lambda,t)}$ ($\beta = t$). Assumptions (iii) and (iv) are technical and are satisfied by e.g. Gaussian kernels $\widetilde{K}$.

Under such assumptions Masry (1991, 1993a) gives the following useful properties of W_h:

Lemma 4.1

Under assumptions A4.3 and for any $1 \leq q \leq +\infty$ *we get*

$$\|\widetilde{W}_h\|_q \quad = \quad O\left(h_n^{-\beta}\right) \qquad n \rightarrow +\infty. \tag{4.5}$$

4.2.2 Supersmooth noise distributions

We now consider noises satisfying the following assumptions.

Assumptions A4.4

(i) For all real t, $|\widetilde{\phi}_\varepsilon(t)| > 0$, furthermore there exist positive constants B_1, a, β and a real constant β_0 such that

$$|\widetilde{\phi}_\varepsilon(t)| \quad \geq \quad B_1 |t|^{\beta_0} \exp\left(-a|t|^\beta\right) \;\; \text{as } t \rightarrow +\infty, \tag{4.6}$$

(ii) $\widetilde{\phi}_K(.)$ has a compact support $]-\tau, \tau[$,

(iii) $\widetilde{\phi}_K$ is an even, real, decreasing and bounded function over $[0, +\infty[$ with $\widetilde{\phi}_K(0) = 1$, $\widetilde{\phi}_K$ admits $(p+1)$ bounded derivatives such that $\widetilde{\phi}_K(\tau) = \ldots = \widetilde{\phi}_K^{(p-1)}(\tau) = 0$ and $\widetilde{\phi}_K^{(p)}(\tau) \neq 0$.

Assumption (i) specifies the asymptotic behaviour of the noise and will be satisfied for e.g. Gaussian noises (with $\beta_0 = 0$ and $\beta = 2$) or those following a Cauchy law ($\beta_0 = 0$ and $\beta = 1$). Conditions (ii) and (iii) are technical and will be fulfilled for $\widetilde{K}$ with Fourier transform such that $\widetilde{\phi}_K(t) = (\tau^2 - t^2)^2 \, \mathbf{I}_{[-\tau,\tau]}(t)$.

Properties of kernel W_h for such noises are given by Stefanski (1990) and extended to any dimension d in the following lemma.

Lemma 4.2
Under assumptions A4.4, as $h_n \to 0$ we get

$$\|W_h\|_\infty = O\left((n\Lambda_n)^{1/2} h_n^d\right) \tag{4.7}$$

$$\|W_h\|_q = O\left((n\Lambda_n)^{1/2} h_n^{(1-\beta/q)d}\right) \quad \text{for } 2 \le q < \infty, \tag{4.8}$$

where Λ_n is defined by

$$\Lambda_n = n^{-1} h_n^{2d[(r+1)\beta+\beta_0-1]} e^{2ad(\tau/h_n)^\beta}. \tag{4.9}$$

4.3 Asymptotic results of convergence for $\widehat{f}_n$

A surprising result concerns the asymptotic bias of the estimator $\widehat{f}_n$ defined by (4.3) which does not depend on the noise distribution. We denote by $C_{2,d}(b)$ the space of twice continuously differentiable real valued functions f , defined on $\mathbb{R}^d$, and such that $\|f\|_\infty \le b$ and $\|f^{(2)}\|_\infty \le b$ where $f^{(2)}$ denotes any partial derivative of order 2 for f.

Theorem 4.1
If $f \in C_{2,d}(b)$ and if $h_n \to 0$ such that $nh_n^2 \to +\infty$, then assumptions A4.1 and A4.2(i)-(ii) imply that

$$\mathrm{E}\, \widehat{f}_n(x) - f(x) = O(h_n^2), \quad n \to +\infty.$$

In order to study the asymptotic variance of our estimator, we now have to treat separately the two classes of noise distributions introduced above.

4.3.1 Ordinary smooth noise distributions

First we deal with ordinary smooth noise distributions of order β satisfying relation (4.4).

Theorem 4.2
Under assumptions A4.1, A4.2, A4.3 and if ξ_ε is bounded, then for any $h_n \to 0$ such that $nh_n^{(2\beta+1)d} \to +\infty$, we get

$$\mathrm{Var}\, \widehat{f}_n(x) = O\left(\frac{1}{nh_n^{(2\beta+1)d}}\right) + O\left(\frac{c_n}{n^2 h_n^{2\beta d}}\right), \quad n \to +\infty.$$

Note that the covariance term depends on c_n: $c_n = O(n)$ is a sufficient condition for the variance of our estimator to tend to zero. Moreover, the smallest is c_n, more rates will be close to the optimal ones of the i.i.d. case (see Fan, 1991). Theorems 4.1 and 4.2 imply the following corollary which gives the asymptotic quadratic error of the estimator.

Corollary 4.1
Under assumptions of theorem 4.1 and theorem 4.2, as $n \to \infty$ we get
(1) *for* $c_n = O\left(n^{\frac{4+d(2\beta+2)}{4+d(2\beta+1)}}\right)$ *and* $h_n = n^{-\frac{1}{4+d(2\beta+1)}}$,

$$\mathrm{E}\left(\widehat{f}_n(x) - f(x)\right)^2 = O\left(n^{-\frac{4}{4+d(2\beta+1)}}\right);$$

(2) *if* $n^{-\frac{4+2d(\beta+1)}{4+d(2\beta+1)}} c_n \to +\infty$ *then for* $h_n = c_n^{\frac{1}{4+2\beta d}} n^{-\frac{1}{2+\beta d}}$

$$\mathrm{E}\left(\widehat{f}_n(x) - f(x)\right)^2 = O\left(\left(\frac{c_n}{n^2}\right)^{\frac{2}{2+\beta d}}\right).$$

Remark that if condition (1) is satisfied in corollary 4.1, then we find again the optimal rates (i.e. independent of the choice of the estimator) of the i.i.d. case with $d = 1$ (see Fan, 1991). Such assumption will be satisfied for e.g. the r-adic function defined by $\varphi_r(x) = rx \pmod 1$ since it can be shown that c_n does not exceed n (see Smili, 1990). Furthermore, the no-noise case may be deduced from corollary 4.1 by putting $\beta = 0$, thus, under asumptions of theorems 4.1 and 4.2 we get that

(i) $c_n = O\left(n^{\frac{4+2d}{4+d}}\right)$ implies a mean-square error of order $O\left(n^{-\frac{4}{4+d}}\right)$,
(ii) $n^{-\frac{4+2d}{4+d}} c_n \to +\infty$ implies a mean-square error of order $O\left(n^{-2}c_n\right)$.

4.3.2 Super smooth noise distributions

We now establish the asymptotic mean square error of our estimator when the noise has a super smooth distribution of order β given by (4.6).

Theorem 4.3
Under assumptions A4.1 A4.2 and A4.4, as $n \to +\infty$ we have

$$\mathrm{Var}\,\widehat{f}_n(x) = O(\Lambda_n) + O\left(\frac{\Lambda_n h_n^{(1-\beta)d}}{n}\right) + O\left(\frac{1}{nh_n^d}\right) + O\left(\frac{\Lambda_n c_n}{n}\right),$$

where Λ_n is given by (4.9).

In order to get convergence of the variance to zero, we now have to choose h_n decreasing to zero logarithmically, thus the term of bias becomes dominant and then the mean-square error is given by

Corollary 4.2

Under assumptions of theorems 4.1 *and* 4.3, *for* $c_n = O\left(n^\theta\right)$ *and* $h_n = \tau\left(\frac{2ad}{\theta'}\right)^{1/\beta}(\ln n)^{-1/\beta}$ *with* θ, θ' *such that* $0 < \theta < 2$ *and* $0 < \theta' < 2 - \theta$, *we have*

$$\mathrm{E}\left(\widehat{f}_n(x) - f(x)\right)^2 = O\left(\left(\frac{1}{\ln n}\right)^{4/\beta}\right).$$

We obtain the same rates as Fan (1991) who has shown in the i.i.d. case with $d = 1$ that these rates are optimal in presence of such kind of noise. Remark that condition upon c_n is still satisfied for the r-adic function.

5 Models with small noise

We have seen in the previous section that rates of convergence for our estimator are very sensitive with respect to the law of the noise. In particular for a Gaussian noise, rates seem too poor (corollary 4.2) for ensuring good estimation. However frequently, one may suppose that the noise has low level and then it is negligible. In this section, we intend to fix the maximum level of noise for which usual nonparametric estimation remains consistent with classical rates.

In this section, we consider models (2.1) and (2.3) with respectively "small" ε_t and Δ_t. For sake of simplicity, only real-valued processes will be considered but our results can be extended to any dimension d.

As we no more follow a deconvolution approach, we will use for both models the classical kernel density estimator (see Rosenblatt, 1956)

$$f_n(x) = \frac{1}{nh_n}\sum_{i=0}^{n-1} K\left(\frac{x - Y_i}{h_n}\right) \tag{5.10}$$

where the Y_i's ($i = 0, \ldots, n$) are real-valued observations arising respectively from models (2.1) and (2.3) and where K is a real positive lipschitzian kernel satisfying the same conditions as kernel $\widetilde{K}$ introduced in section 4.

5.1 Model with errors-in-variables

Fan (1992) shows that for i.i.d. variables and small Gaussian noise, the deconvolution kernel estimator reaches the optimal i.i.d. rates. We now generalize this result in several ways: we consider a large class of densities for the noise (including the Gaussian one) and then, we study the asymptotic behaviour of the usual kernel estimator (whose use is easier than the deconvolution one) for systems given by (4.1) and (4.2). Note that for stationary, mixing or irregularly observed continuous-time processes, similar results may be found in Blanke (1997).

In concrete terms, we suppose that we observe $Y_0, \ldots, Y_{n-1}$ given by

$$Y_i \;=\; X_i + \sigma\varepsilon_i. \tag{5.11}$$

with (X_t) satisfying relation (4.2). This model is similar to (2.1).

Furthemore we consider the following assumptions:

Assumptions A5.1

(i) $\{X_t,\, t \in \boldsymbol{Z}\}$ *is a real process.*
(ii) $\{\varepsilon_t,\, t \in \boldsymbol{Z}\}$ *are i.i.d. real random variables with unknown density* ξ_ε.
(iii) $\{X_t,\, t \in \boldsymbol{Z}\}$ *and* $\{\varepsilon_t,\, t \in \boldsymbol{Z}\}$ *are independent.*

We now show that if σ is small enough ($\sigma = \beta_n \to 0$ as $n \to +\infty$), we may consistently estimate the invariant density f when only observations $Y_0, \ldots, Y_{n-1}$ are given. Note that such models may be associated with instruments which are more and more accurate.

It is important to note that in assumptions $A5.1$, the law of the noise is no more supposed to be known. Assumptions will only be made upon its two first moments.

Assumptions A5.2

(i) $\int_{\mathbb{R}} u\xi_\varepsilon(u)\, du = 0$,
(ii) $\int_{\mathbb{R}} u^2\xi_\varepsilon(u)\, du < +\infty$,
(iii) $f \in C_{2,1}(b)$.

The following theorem gives the level of noise under which optimal rates still remain.

Theorem 5.1
Under assumptions $A4.2$, $A5.1$ *and* $A5.2$, *if furthermore* $\beta_n = O\left(n^{-4/5}\right)$ *then the choice* $h_n = n^{-1/5}$ *implies that*

$$\mathrm{E}\,\left(f_n(x) - f(x)\right)^2 \;=\; O\left(n^{-\frac{4}{5}}\right) + O\left(\frac{c_n}{n^2}\right) \quad \text{as } n \to +\infty.$$

For higher order level noise (i.e. when $n^{4/5}\beta_n \to +\infty$), optimal rates may still be obtained but under stronger conditions upon c_n.

Theorem 5.2
Under assumptions $A4.2$, $A5.1$ *and* $A5.2$,

1) *if* c_n *is such that* $c_n = O\left(n^{4/5}\right)$ *and* $h_n = n^{-1/5}$, *then for* $\beta_n = O(h_n)$ *we get*

$$\mathrm{E}\,\left(f_n(x) - f(x)\right)^2 \;=\; O\left(n^{-\frac{4}{5}}\right) \quad n \to +\infty;$$

2) *if* $n^{-4/5}c_n \to +\infty$, *then for* $h_n = \left(c_n/n^2\right)^{1/6}$ *and* $\beta_n = O(h_n)$ *we get*

$$\mathrm{E}\,\left(f_n(x) - f(x)\right)^2 \;=\; O\left(\left(\frac{c_n}{n^2}\right)^{2/3}\right) \quad n \to +\infty.$$

Note that if assumptions of theorem 5.1 (with $c_n = O\left(n^{6/5}\right)$) or theorem 5.2-(1) hold then we find the optimal rates of convergence for nonparametric density estimation in the i.i.d. case (see e.g. Farrel, 1972). Furthermore for processes with c_n of order n (such as the r-adic ones) the worst asymptotic rates are of order $n^{-4/5}$ for small β_n: $\beta_n = O\left(n^{-4/5}\right)$ and of order $n^{-2/3}$ for larger β_n given by $\beta_n = O\left(n^{-1/6}\right)$.

5.2 Model with propagation of errors

We now consider the model close to (2.3) given by:

$$Y_t = \varphi(Y_{t-1}) + \sigma\varepsilon_t. \tag{5.12}$$

We denote by f the invariant density (when it exists) associated to the no-noise model $Y_t = \varphi(Y_{t-1})$. Our goal is to estimate f when observations are given by (5.12) with $\sigma = \beta_n \to 0$ as $n \to +\infty$. Assumptions over the noise are still the same as in section 5.1. Furthermore we will suppose that φ is a a.e. differentiable function. For such models, we give some levels of noise under which optimal rates of convergence are reached.

Theorem 5.3
Suppose that assumptions A4.2 and A5.2 hold, if furthermore
(i) $\|\varphi'\|_\infty = 1$ *and* $\beta_n = O\left(n^{-13/10}\right)$
or
(ii) $1 < \|\varphi'\|_\infty < +\infty$ *and* $\beta_n = O\left(e^{-an}\right)$ *with* $a > \ln\|\varphi'\|_\infty$,
then the choice $h_n = n^{-1/5}$ *implies that*

$$\mathrm{E}\,\left(f_n(x) - f(x)\right)^2 \;=\; O\left(n^{-\frac{4}{5}}\right) + O\left(\frac{c_n}{n^2}\right) \quad \text{as } n \to +\infty.$$

Remark that in usual cases, we have $1 < \|\varphi'\|_\infty < +\infty$. Theorem 5.3 indicates that for such models, the noise should have very low level in order to reach optimal rates of convergence.

6 Proofs

6.1 Proof of Proposition 3.1

We use a similar of the one developed by Lasota-Mackey (1994, p. 417-419). Let us set:

$$M_n = \frac{1}{n}\sum_{i=0}^{n-1} \Pi^i(\mu_0) = \frac{1}{n}\sum_{i=1}^{n} \mu_i$$

and let (h_k) be a sequence which is dense in the space $C_0(E)$ of continuous bounded functions over E, then

$$|h_k(M_n)| = \left| \int h_k dM_n \right| \leq \|h_k\|_\infty \ , \qquad k \geq 1 \ .$$

Then, for each k, there exists $(M_{n_k}) \subset (M_n)$ such that the sequence $(h_k(M_{n_k}))$ converges. By diagonalization we may claim that there exists (M_{n_n}) such that $(h_k(M_{n_n}))$ converges for each $k \geq 1$.

Therefore there exists an unique measure $\mu_* \in \mathcal{P}(E)$ such that we have the weak convergence of M_{n_n} to μ_*, (we refer to Theorem 12.2.2 and Remark 12.2.2 in Lasota-Mackey, 1994).

Let us now verify that μ_* is invariant for Π. Since we may set that the set B is compact, $E - B$ is then open and using a well-known property (see Billingsley, 1969), we get:

$$\begin{aligned} \mu_*(E-B) &\leq \lim\ \inf\ M_{n_n}(E-B) \\ &\leq 1 - \inf_n\ \mu_n(B) \leq 1-(1-\varepsilon) = \varepsilon. \end{aligned}$$

Now we can write:

$$M_{nn} = \frac{1}{k_n} \sum_{i=0}^{k_{n-1}} \Pi^i \mu_0, \qquad k_n \nearrow \infty \ .$$

Thus,

$$\Pi M_{n_n} - M_{n_n} = \frac{1}{k_n}(\Pi^{k_n}\mu_0 - \mu_0) \ ,$$

and if $h \in C_0(E)$

$$\begin{aligned} |\Pi M_{n_n}(h) - M_{n_n}(h)| &= \left| \frac{1}{k_n}(\Pi^{k_n}\mu_0(h) - \mu_0(h)) \right| \\ &\leq \frac{2\|h\|_\infty}{k_n} \ . \end{aligned}$$

Hence, when $h > 0$, using (A_2) and the convergence of M_{n_n} towards μ_*, when $k_n \to \infty$, we get:

$$(\Pi\mu_*)(h) = \mu_*(h) \ , \qquad h \in C_0 \ ,$$

thus

$$\Pi\mu_* = \mu_* \ .\blacksquare$$

6.2 Proof of theorem 4.1

We have to show that the bias is independent of the noise distribution.

$$\mathrm{E}\,\widehat{f}_n(x) \ = \ \frac{1}{nh_n^d} \sum_{j=0}^{n-1} \mathrm{E}\, W_h\left(\frac{x-Y_j}{h_n}\right)$$

$$= \frac{1}{n}\left(\frac{1}{2\pi h_n}\right)^d \sum_{j=0}^{n-1} \mathrm{E}\left(\int_{\mathbb{R}^d} e^{-i<t,\frac{x-Y_j}{h_n}>}\frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)}\,dt\right).$$

For characteristic functions such that $\widetilde{\phi}_K(t)/\widetilde{\phi}_\varepsilon(t/h) \in L^1(\mathbb{R})$ for all h, we get by Fubini's theorem

$$\mathrm{E}\,\widehat{f}_n(x) = \frac{1}{n}\left(\frac{1}{2\pi h_n}\right)^d \sum_{j=0}^{n-1} \int_{\mathbb{R}^d} e^{-i<t,\frac{x}{h_n}>}\phi_{Y_j}\left(t/h_n\right)\frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)}\,dt.$$

Independence implies that $\phi_{Y_j}(t/h_n) = \phi_{X_j}(t/h_n)\phi_\varepsilon(t/h_n)$ so if we set $t/h_n = s$, we get

$$\begin{aligned}
\mathrm{E}\,\widehat{f}_n(x) &= \frac{1}{n}\left(\frac{1}{2\pi}\right)^d \sum_{j=0}^{n-1} \int_{\mathbb{R}^d} e^{-i<s,x>}\phi_{X_j}(s)\phi_K(sh_n)\,ds \\
&= \frac{1}{n}\left(\frac{1}{2\pi}\right)^d \sum_{j=0}^{n-1} \int_{\mathbb{R}^d} e^{-i<s,x>}\phi_{K_h*\Pi^j f_0}(s)\,ds \\
&= \frac{1}{n}\sum_{j=0}^{n-1} \int_{\mathbb{R}^d} K_h\left(x-y\right)\Pi^j f_0(y)\,dy \\
&= \frac{1}{nh_n^d}\sum_{j=0}^{n-1} \mathrm{E}\,K\left(\frac{x-X_j}{h_n}\right). \qquad (6.13)
\end{aligned}$$

1) Assume first that X_0 has the invariant density f, i.e. $f_0 = f$, this implies that $f_j = \Pi^j f_0 = f$ for all j. Then we get

$$\frac{1}{h_n^d}\mathrm{E}\,W_h\left(\frac{x-Y_j}{h_n}\right) = \frac{1}{h_n^d}\int_{\mathbb{R}^d} K\left(\frac{x-y}{h_n}\right) f(y)\,dy.$$

For f belonging to $C_{2,d}(b)$ we get by Taylor's formula

$$\left(\mathrm{E}_f\,\widehat{f}_n(x) - f(x)\right)^2 = O\left(h_n^4\right) \qquad (6.14)$$

where E_f denotes expectation under f.

2) Let E_{f_t} be the expectation under f_t, then for $f_0 \neq f$ we get

$$\begin{aligned}
\mathrm{E}_{f_t}\,\widehat{f}_n(x) - \mathrm{E}_f\,\widehat{f}_n(x) &= \frac{1}{nh_n^d}\sum_{j=0}^{n-1}\int_{\mathbb{R}^d} K\left(\frac{x-y}{h_n}\right)\left[\Pi^j f_0(y) - f(y)\right]\,dy \\
&\le \frac{1}{n}\sum_{j=0}^{n-1}\|\Pi^j f_0 - f\|_\infty.
\end{aligned}$$

Assumption $A4.2(ii)$ implies that $\sum_{j=0}^{\infty}\|\Pi^j f_0 - f\|_\infty < +\infty$ so

$$\left(\mathrm{E}_{f_t}\,\widehat{f}_n(x) - \mathrm{E}_f\,\widehat{f}_n(x)\right)^2 = O\left(\frac{1}{n^2}\right)$$

which becomes neglictible as soon as $nh_n^2 \to +\infty$, then theorem 4.1 is proved with majoration

$$\left(\mathrm{E}_{f_t}\,\widehat{f}_n(x) - f(x)\right)^2 \le 2\left(\mathrm{E}_{f_t}\,\widehat{f}_n(x) - \mathrm{E}_f\,\widehat{f}_n(x)\right)^2 + 2\left(\mathrm{E}_f\,\widehat{f}_n(x) - f(x)\right)^2. \quad \blacksquare$$

6.3 Proof of theorem 4.2

The variance of our estimator can be decomposed into two terms V_n and C_n given by

$$\operatorname{Var} \widehat{f}_n(x) \quad = \quad V_n + C_n \tag{6.15}$$

where

$$V_n \quad = \quad \frac{1}{n^2 h_n^{2d}} \sum_{i=0}^{n-1} \operatorname{Var} W_h \left(\frac{x - Y_i}{h_n} \right), \tag{6.16}$$

$$C_n \quad = \quad \frac{1}{n^2 h_n^{2d}} \sum_{i \neq j} \operatorname{Cov} \left(W_h \left(\frac{x - Y_i}{h_n} \right), W_h \left(\frac{x - Y_j}{h_n} \right) \right). \tag{6.17}$$

6.3.1 Study of V_n

1) Assume firstly that $f_0 = f$ (this implies $g_t = g_0$ for all t). We get

$$\begin{aligned}
\frac{1}{n^2 h_n^{2d}} \sum_{i=0}^{n-1} \operatorname{Var} W_h \left(\frac{x - Y_i}{h_n} \right) \quad &= \quad \frac{1}{n h_n^{2d}} \operatorname{Var} W_h \left(\frac{x - Y_0}{h_n} \right) \\
&\leq \quad \frac{1}{n h_n^{2d}} \int_{\mathbb{R}^d} W_h^2 \left(\frac{x - y}{h_n} \right) g_0(y)\, dy \\
&\leq \quad \|\xi_\varepsilon\|_\infty \frac{\|W_h\|_2^2}{n h_n^d} \\
&= \quad O \left(\frac{1}{n h_n^{(2\beta+1)d}} \right)
\end{aligned}$$

since the density of the noise is bounded and $\|W_h\|_2$ is given by lemma 4.1.

2) Assume now that $f_0 \neq f$, we have $g_i = f_i * \xi_\varepsilon$. Let Var_{g_i} and Var_g be the variances under densities g_i and $g = f * \xi_\varepsilon$. We now write

$$\frac{1}{n^2 h_n^{2d}} \sum_{i=0}^{n-1} \left(\operatorname{Var}_{g_i} W_h \left(\frac{x - Y_i}{h_n} \right) - \operatorname{Var}_g W_h \left(\frac{x - Y_i}{h_n} \right) \right) \quad := \quad A + B$$

where A and B are given by

$$A \quad = \quad \frac{1}{n^2 h_n^{2d}} \sum_{i=0}^{n-1} \left[\int_{\mathbb{R}^d} W_h^2 \left(\frac{x - y}{h_n} \right) g_i(y)\, dy - \int_{\mathbb{R}^d} W_h^2 \left(\frac{x - y}{h_n} \right) g(y)\, dy \right] \tag{6.18}$$

$$B \quad = \quad \frac{1}{n^2 h_n^{2d}} \sum_{i=0}^{n-1} \left[\left(\int_{\mathbb{R}^d} K \left(\frac{x - y}{h_n} \right) f(y)\, dy \right)^2 - \left(\int_{\mathbb{R}^d} K \left(\frac{x - y}{h_n} \right) f_i(y)\, dy \right)^2 \right]. \tag{6.19}$$

Study of term A

$$A \quad = \quad \frac{1}{n^2 h_n^{2d}} \sum_{i=0}^{n-1} \int_{\mathbb{R}^d} W_h^2 \left(\frac{x - y}{h_n} \right) (g_i(y) - g(y))\, dy$$

$$
\begin{aligned}
&= \frac{1}{n^2h_n^{2d}}\sum_{i=0}^{n-1}\int_{\mathbb{R}^d} W_h^2\left(\frac{x-y}{h_n}\right)\left(\int_{\mathbb{R}^d}\xi_\varepsilon(u)\left[f_i(y-u)-f(y-u)\right]\,du\right)dy\\
&\leq \frac{1}{n^2h_n^{2d}}\sum_{i=0}^{n-1}\int_{\mathbb{R}^d} W_h^2\left(\frac{x-y}{h_n}\right)\|f_i-f\|_\infty\,dy\\
&\leq \sum_{i=0}^{+\infty}\|\Pi^i f-f\|_\infty\frac{\|W_h\|_2^2}{n^2h_n^d}\\
&= O\left(\frac{1}{n^2h_n^{(2\beta+1)d}}\right)=o\left(\frac{1}{nh_n^{(2\beta+1)d}}\right)
\end{aligned}
$$

by assumption $A4.2(ii)$ and lemma 4.1.

Study of term B

$$
\begin{aligned}
B &= \frac{1}{n^2h_n^{2d}}\sum_{i=0}^{n-1}\left(\int_{\mathbb{R}^d}K\left(\frac{x-y}{h_n}\right)(f(y)+f_i(y))\,dy\right)\left(\int_{\mathbb{R}^d}K\left(\frac{x-y}{h_n}\right)(f(y)-f_i(y))\,dy\right)\\
&\leq \frac{\|K\|_\infty}{n^2h_n^d}\sum_{i=0}^{\infty}\|f_i-f\|_\infty\\
&= O\left(\frac{1}{n^2h_n^d}\right)=o\left(\frac{1}{nh_n^{(2\beta+1)d}}\right)
\end{aligned}
$$

by assumption $A4.2(ii)$.

Finally assumptions $A4.1$, $A4.2(i)$-(ii) and $A4.3$ imply since $n\to+\infty$

$$
\frac{1}{n^2h_n^{2d}}\sum_{i=0}^{n-1}\operatorname{Var}W_h\left(\frac{x-Y_i}{h_n}\right) \quad = \quad O\left(\frac{1}{nh_n^{(2\beta+1)d}}\right). \tag{6.20}
$$

6.3.2 Study of C_n

We have

$$
\begin{aligned}
C_n &= \frac{2}{n^2h_n^{2d}}\sum_{i=0<j\leq n-1}\mathrm{E}_{g_{i,j}}W_h\left(\frac{x-Y_i}{h_n}\right)W_h\left(\frac{x-Y_j}{h_n}\right)-\mathrm{E}_{g_i}W_h\left(\frac{x-Y_i}{h_n}\right)\mathrm{E}_{g_j}W_h\left(\frac{x-Y_j}{h_n}\right)\\
&\leq I+J
\end{aligned} \tag{6.21}
$$

where

$$
I=\frac{2}{n^2h_n^{2d}}\left|\sum_{i=0<j\leq n-1}\mathrm{E}_{g_{i,j}}W_h\left(\frac{x-Y_i}{h_n}\right)W_h\left(\frac{x-Y_j}{h_n}\right)-\mathrm{E}_gW_h\left(\frac{x-Y_i}{h_n}\right)\mathrm{E}_gW_h\left(\frac{x-Y_j}{h_n}\right)\right| \tag{6.22}
$$

$$
J=\frac{2}{n^2h_n^{2d}}\left|\sum_{i=0<j\leq n-1}\mathrm{E}_gW_h\left(\frac{x-Y_i}{h_n}\right)\mathrm{E}_gW_h\left(\frac{x-Y_j}{h_n}\right)-\mathrm{E}_{g_i}W_h\left(\frac{x-Y_i}{h_n}\right)\mathrm{E}_{g_j}W_h\left(\frac{x-Y_j}{h_n}\right)\right| \tag{6.23}
$$

Study of term *I*

$$I = \frac{2}{n^2 h_n^{2d}} \left| \sum_{i=0<j\le n-1} \iint_{\mathbb{R}^{2d}} W_h\left(\frac{x-y}{h_n}\right) W_h\left(\frac{x-z}{h_n}\right) \left(g_{(Y_i,Y_j)}(y,z) - g(y)g(z)\right)\, dydz \right|. \quad (6.24)$$

The sum being finite, we get by independence of the ε_i's

$$\begin{aligned}
I &= \frac{2}{n^2 h_n^{2d}} \left| \iint_{\mathbb{R}^{2d}} W_h\left(\frac{x-y}{h_n}\right) W_h\left(\frac{x-z}{h_n}\right) \right. \\
&\qquad \left. \left(\iint_{\mathbb{R}^{2d}} \xi_\varepsilon(y-s)\xi_\varepsilon(z-t) \sum_{i=0<j\le n-1} \left(dP_{(X_i,X_j)}(s,t) - d\mu(s)d\mu(t)\right)\right) dydz \right| \\
&\le \frac{2}{n^2} \|\xi_\varepsilon\|_\infty^2 \|W_h\|_1^2 \iint_{E\times E} \left| \sum_{i=0<j\le n-1} \left(dP_{(X_i,X_j)}(s,t) - d\mu(s)d\mu(t)\right)\right| \\
&= O\left(\frac{c_n}{n^2 h_n^{2\beta d}}\right) \qquad (6.25)
\end{aligned}$$

with assumption $A4.2(iii)$ and lemma 4.1 since the density of the noise is supposed to be bounded.

Study of term *J*

The bias is independent of the noise distribution, so (6.13) implies that

$$\begin{aligned}
J &\le \frac{2}{n^2 h_n^{2d}} \sum_{i=0<j\le n-1} \left| \mathrm{E}_g W_h\left(\frac{x-Y_i}{h_n}\right) \mathrm{E}_g W_h\left(\frac{x-Y_j}{h_n}\right) - \mathrm{E}_{g_i} W_h\left(\frac{x-Y_i}{h_n}\right) \mathrm{E}_{g_j} W_h\left(\frac{x-Y_j}{h_n}\right)\right| \\
&= \frac{2}{n^2} \sum_{i=0<j\le n-1} \left| ((K_h * f)(x))^2 - ((K_h * f_i)(x))\,((K_h * f_j)(x))\right|,
\end{aligned}$$

where we have set $K_h(.) = h^{-d}K(./h)$.

By using inequality $|a_1a_2 - b_1b_2| \le |a_2||a_1 - b_1| + |b_1||a_2 - b_2|$, we get

$$\begin{aligned}
J &\le \frac{2}{n^2} \sum_{i=0<j\le n-1} \left\{ \left| \int_{\mathbb{R}^d} K_h(x-z) f(z)\, dz \right| \left| \int_{\mathbb{R}^d} K_h(x-y)\,(f(y) - f_i(y))\, dy \right| \right. \\
&\qquad \left. + \left| \int_{\mathbb{R}^d} K_h(x-y) f_i(y)\, dy \right| \left| \int_{\mathbb{R}^d} K_h(x-z)\,(f(z) - f_j(z))\, dz \right| \right\} \\
&\le \frac{4\|K\|_\infty}{n h_n^d} \sum_{i=0}^{+\infty} \|f - \Pi^i f_0\|_\infty \\
&= O\left(\frac{1}{n h_n^d}\right) = o\left(\frac{1}{n h_n^{(2\beta+1)d}}\right), \qquad (6.26)
\end{aligned}$$

with assumption $A4.2(ii)$. Equations (6.20), (6.21)-(6.26) now yield to theorem 4.2. ■

6.4 Proof of theorem 4.3

We make use of same notations as in the proof of theorem 4.2.

6.4.1 Study of term V_n

1) Assume that $f_0 = f$, then we have

$$\begin{aligned}\frac{1}{n^2 h_n^{2d}} \sum_{i=0}^{n-1} \operatorname{Var}_g W_h \left(\frac{x - Y_i}{h_n} \right) &\leq \frac{1}{n h_n^{2d}} \int W_h^2 \left(\frac{x - y}{h_n} \right) g(y)\, dy \\ &\leq \frac{\|W_h\|_\infty^2}{n h_n^{2d}} = O\left(\Lambda_n \right)\end{aligned}$$

by using lemma 4.2.

2) If $f_0 \neq f$, the decomposition $A + B$ given by (6.18) and (6.19) implies

$$A = O\left(\frac{\Lambda_n h_n^{(1-\beta)d}}{n} \right)$$

with lemma 4.2 and

$$B = O\left(\frac{1}{n^2 h_n^d} \right).$$

Finally under assumptions $A4.2(ii)$ and $A4.4$ we get when $n \to +\infty$

$$\frac{1}{n^2 h_n^{2d}} \sum_{i=0}^{n-1} \operatorname{Var} W_h \left(\frac{x - Y_i}{h_n} \right) = O\left(\Lambda_n \right) + O\left(\frac{\Lambda_n h_n^{(1-\beta)d}}{n} \right) + O\left(\frac{1}{n^2 h_n^d} \right). \quad (6.27)$$

6.4.2 Study of term C_n

Term J in (6.23) is still the same since it does not depend on the noise distribution, so we get

$$J = O\left(\frac{1}{n h_n^d} \right). \quad (6.28)$$

For term I given by (6.22), Fubini's theorem implies that

$$\begin{aligned} I &\leq \frac{2\|W_h\|_\infty^2}{n^2 h_n^{2d}} \iint_{\mathbb{R}^{2d}} \left| \sum_{i<j} \left(dP_{(X_i, X_j)}(t, s) - d\mu(t) d\mu(s) \right) \right| \\ &= O\left(\frac{\Lambda_n c_n}{n} \right) \quad (6.29)\end{aligned}$$

under assumption $A4.2(iii)$ and lemma 4.2. Theorem 4.3 is then proved by using relations (6.21) and (6.27) to (6.29). ■

6.5 Proof of theorem 5.1

Suppose that we have observed $Z_0, \dots, Z_{n-1}$ given by the "ideal" model:

$$\begin{cases} Z_t &= \varphi(Z_{t-1}) \\ Z_0 &= X_0. \end{cases}$$

One may construct the associated estimator

$$\widetilde{f}_n(x) = \frac{1}{nh_n} \sum_{i=0}^{n-1} K\left(\frac{x - Z_i}{h_n}\right).$$

Now the mean-square error of our estimator $\widehat{f}_n$ is bounded by

$$\mathrm{E}\left(\widehat{f}_n(x) - f(x)\right)^2 \le 2\mathrm{E}\left(\widehat{f}_n(x) - \widetilde{f}_n(x)\right)^2 + 2\mathrm{E}\left(\widetilde{f}_n(x) - f(x)\right)^2.$$

It's easy to show (see the remark following corollary 4.1) that under assumptions $A4.2$ and for $h_n = n^{-1/5}$ we have

$$\mathrm{E}\left(\widetilde{f}_n(x) - f(x)\right)^2 = O\left(n^{-4/5}\right) + O\left(\frac{c_n}{n^2}\right).$$

Thus we have only to consider the term $\mathrm{E}\left(\widehat{f}_n(x) - \widetilde{f}_n(x)\right)^2$. Since K is supposed to be Lipschitzian, Cauchy-Schwarz's inequality implies

$$\mathrm{E}\left(\widehat{f}_n(x) - \widetilde{f}_n(x)\right)^2 \le \frac{C}{nh_n^4} \sum_{i=0}^{n-1} \mathrm{E}\left(Y_i - Z_i\right)^2,$$

where C is a positive generic constant. We have $Y_i = \varphi^{(i)}(X_0) + \beta_n \varepsilon_i$ and $Z_i = \varphi^{(i)}(Z_0) = \varphi^{(i)}(X_0)$, so

$$\mathrm{E}\left(\widehat{f}_n(x) - \widetilde{f}_n(x)\right)^2 \le \frac{C\beta_n^2}{h_n^4}.$$

Thus the choice $h_n = n^{-1/5}$ and $\beta_n = O\left(n^{-4/5}\right)$ implies

$$\mathrm{E}\left(\widehat{f}_n(x) - \widetilde{f}_n(x)\right)^2 = O\left(n^{-4/5}\right)$$

and finally

$$\mathrm{E}\left(\widehat{f}_n(x) - f(x)\right)^2 = O\left(n^{-4/5}\right) + O\left(\frac{c_n}{n^2}\right). \quad \blacksquare$$

6.6 Proof of theorem 5.2

6.6.1 Study of the bias term

(1) When $f_0 = f$, the bias of the estimator is given by

$$\begin{aligned} \mathrm{E}\, f_n(x) - f(x) &= \frac{1}{h_n} \int_{\mathbb{R}} K\left(\frac{x-y}{h_n}\right) \left[\int_{\mathbb{R}} f(y-u) \frac{1}{\beta_n} \xi_\varepsilon\left(\frac{u}{\beta_n}\right) du - f(x)\right] dy \\ &= \int_{\mathbb{R}} K(v) \left[\int_{\mathbb{R}} f(x - vh_n - u) \frac{1}{\beta_n} \xi_\varepsilon\left(\frac{u}{\beta_n}\right) du - f(x)\right] dv. \qquad (6.30) \end{aligned}$$

Assumptions $A5.2$ and Taylor's formula imply for $0 < \theta < 1$,

$$\begin{aligned}
&\int_{\mathbb{R}} f(x - vh_n - t\beta_n)\xi_\varepsilon(t)\ dt - f(x) \\
&\quad = \int_{\mathbb{R}} [f(x - vh_n - t\beta_n) - f(x)]\ \xi_\varepsilon(t)\, dt \\
&\quad = \int_{\mathbb{R}} \left[(-vh_n - t\beta_n) f^{'}(x) + \frac{(vh_n + t\beta_n)^2}{2} f^{''}(x - \theta(vh_n + t\beta_n))\right] \xi_\varepsilon(t)\, dt \\
&\quad = J_1(v) + J_2(v) + J_3(v) + J_4(v) + J_5(v) \qquad (6.31)
\end{aligned}$$

where the $J_i(v)$ are respectively given by

$$\begin{aligned}
J_1(v) &= -vh_n f^{'}(x), \\
J_2(v) &= -\beta_n f^{'}(x) \int_{\mathbb{R}} t\xi_\varepsilon(t)\, dt, \\
J_3(v) &= \int_{\mathbb{R}} \frac{v^2 h_n^2}{2} f^{''}(x - \theta(vh_n + t\beta_n))\, \xi_\varepsilon(t)\, dt, \\
J_4(v) &= \frac{\beta_n^2}{2} \int_{\mathbb{R}} t^2 f^{''}(x - \theta(vh_n + t\beta_n))\, \xi_\varepsilon(t)\, dt, \\
J_5(v) &= h_n\beta_n \int_{\mathbb{R}} vtf^{''}(x - \theta(vh_n + t\beta_n))\, \xi_\varepsilon(t)\, dt.
\end{aligned}$$

By assumptions $A5.2$ and since the kernel K is symetric we get:

$$\int_{\mathbb{R}} K(v) J_1(v)\, dv = 0 \qquad (6.32)$$

$$\int_{\mathbb{R}} K(v) J_2(v)\, dv = 0. \qquad (6.33)$$

Furthermore, we have successively,

$$\begin{aligned}
\int_{\mathbb{R}} K(v) J_3(v)\, dv &= \frac{h_n^2}{2} \iint_{\mathbb{R}^2} v^2 K(v) f^{''}\left(x - \theta\left(v\, h_n + t\, \beta_n\right)\right) \xi_\varepsilon(t)\, dtdv, \\
\int_{\mathbb{R}} K(v) J_4(v)\, dv &= \frac{\beta_n^2}{2} \iint_{\mathbb{R}^2} t^2 f^{''}\left(x - \theta\left(h_n\, v + \beta_n\, t\right)\right) \xi_\varepsilon(t) K(v)\, dtdv, \\
\int_{\mathbb{R}} K(v) J_5(v)\, dv &= h_n\beta_n \iint_{\mathbb{R}^2} tvK(v) f^{''}\left(x - \theta\left(h_n\, v + \beta_n\, t\right)\right) \xi_\varepsilon(t)\, dtdv.
\end{aligned}$$

Since $f^{''}$ is supposed to be continuous and bounded, Lebesgue's theorem implies when $n \to +\infty$

$$\lim_{n\to+\infty} h_n^{-2} \int_{\mathbb{R}} K(v) J_3(v)\, dv = \frac{1}{2} f^{''}(x) \int_{\mathbb{R}} v^2 K(v)\, dv, \qquad (6.34)$$

$$\lim_{n\to+\infty} \beta_n^{-2} \int_{\mathbb{R}} K(v) J_4(v)\, dv = \frac{1}{2} f^{''}(x) \int_{\mathbb{R}} t^2 \xi_\varepsilon(t)\, dt, \qquad (6.35)$$

$$\lim_{n\to+\infty} \beta_n^{-1} h_n^{-1} \int_{\mathbb{R}} K(v) J_5(v)\, dv = 0. \qquad (6.36)$$

We finally get by relations (6.32)-(6.36):

(i) When $\beta_n = h_n$,

$$\lim_{n\to+\infty} h_n^{-2} \left(\mathrm{E}\, f_n(x) - f(x)\right) = \frac{1}{2} f''(x) \left(\int_{\mathbb{R}} v^2 K(v)\, dv + \int_{\mathbb{R}} t^2 \xi_\varepsilon(t)\, dt \right).$$

(ii) When $\beta_n = o(h_n)$,

$$\lim_{n\to+\infty} h_n^{-2} \left(\mathrm{E}\, f_n(x) - f(x)\right) = \frac{1}{2} f''(x) \int_{\mathbb{R}} v^2 K(v)\, dv.$$

2) Assume now that $f_0 \neq f$, we have

$$\begin{aligned} \mathrm{E}_{f_t}\, f_n(x) - \mathrm{E}_f\, f_n(x) &= \frac{1}{nh_n} \sum_{j=0}^{n-1} \int_{\mathbb{R}} K\left(\frac{x-y}{h_n}\right) \left[\Pi^j f_0(y) - f(y)\right] dy \\ &\leq \frac{1}{n} \sum_{j=0}^{n-1} \|\Pi^j f_o - f\|_\infty \end{aligned}$$

which with assumption $A4.2(ii)$ implies

$$\mathrm{E}_{f_t}\, f_n(x) - \mathrm{E}_f\, f_n(x) = O\left(\frac{1}{n}\right)$$

which is neglictible as soon as $nh_n^2 \to +\infty$.

6.6.2 Study of variance term

The proof is quite similar to those of theorem 4.2. We make use of same notations and only main changes are outlined. Note that kernel K is now use instead of W_h and d equals to 1. We consider decomposition (6.15) with C_n and V_n respectively given by (6.16) and (6.17).

Study of V_n

1) Assume firstly that $f_0 = f$, we get

$$\frac{1}{n^2 h_n^2} \sum_{i=0}^{n-1} \mathrm{Var}\, K\left(\frac{x - Y_i}{h_n}\right) \leq \frac{1}{nh_n^2} \int_{\mathbb{R}^2} K^2\left(\frac{x-y}{h_n}\right) g(y)\, dy$$

Since f is supposed to be continuous, the same property holds for g and Bochner's lemma implies that

$$\limsup_{n\to+\infty} \frac{nh_n}{n^2 h_n^2} \sum_{i=0}^{n-1} \mathrm{Var}\, K\left(\frac{x - Y_i}{h_n}\right) \leq g(x) \int_{\mathbb{R}} K^2(v)\, dv \tag{6.37}$$

2) Assume now that $f_0 \neq f$, then we get $g_i = f_i * \frac{1}{\beta_n}\xi_\varepsilon(./\beta_n)$ and terms A and B are bounded by

$$\begin{aligned} A &\leq \frac{\|K\|_2^2}{n^2 h_n} \sum_{i=0}^{n-1} \|\Pi^i f - f\|_\infty \\ &= O\left(\frac{1}{n^2 h_n}\right) = o\left(\frac{1}{nh_n}\right) \end{aligned}$$

with assumption $A4.2(ii)$. In the same way,

$$\begin{aligned} B &\leq \frac{\|K\|_\infty}{n^2 h_n} \sum_{i=0}^{\infty} \|f_i - f\|_\infty \\ &= o\left(\frac{1}{nh_n}\right). \end{aligned}$$

Finally assumptions $A4.1$, $A4.2(i)$-(ii) and $A5.2$ imply since $n \to +\infty$

$$\limsup_{n\to+\infty} \frac{nh_n}{n^2 h_n^2} \sum_{i=0}^{n-1} \operatorname{Var} K\left(\frac{x - Y_i}{h_n}\right) \leq g(x) \int_{\mathbb{R}} K^2(v)\, dv. \tag{6.38}$$

Study of C_n

For the term I given by relation (6.24), we obtain by Fubini's theorem and boundedness of K

$$\begin{aligned} I &\leq \frac{2}{n^2 h_n^2} \|K\|_\infty^2 \iint_{E\times E} \left| \sum_{i=0}^{n-1} \left(dP_{(X_i,X_j)}(s,t) - d\mu(s)d\mu(t) \right) \right| \\ &= O\left(\frac{c_n}{n^2 h_n^2}\right) \end{aligned} \tag{6.39}$$

by assumption $A4.2(iii)$.

For term J of relation (6.23), we have

$$\begin{aligned} J &\leq \frac{4\|K\|_\infty}{nh_n} \sum_{i=0}^{n-1} \left| \int_{\mathbb{R}} K_h(x-y) \int_{\mathbb{R}} \frac{1}{\beta_n} \xi_\varepsilon(t/\beta_n) \left[f(y-t) - \Pi^i f_0(y-t) \right] dtdy \right| \\ &\leq \frac{4\|K\|_\infty}{nh_n} \sum_{i=0}^{\infty} \|f - \Pi^i f_0\|_\infty \\ &= O\left(\frac{1}{nh_n}\right). \end{aligned} \tag{6.40}$$

Finally (6.15)-(6.17), (6.21) and (6.38)-(6.40) imply that

$$\operatorname{Var} f_n(x) = O\left(\frac{1}{nh_n}\right) + O\left(\frac{c_n}{n^2 h_n^2}\right) \quad \text{for } n \to +\infty. \ \blacksquare$$

6.7 Proof of theorem 5.3

The proof is quite similar to those of theorem 5.1. We consider the "ideal" model (i.e. without noise):

$$\begin{cases} Z_t &= \varphi(Z_{t-1}) \\ Z_0 &= Y_0. \end{cases}$$

We suppose that one has n observations $Z_0, \dots, Z_{n-1}$ and we construct the estimator of the invariant density associated to this model:

$$\widetilde{f}_n(x) = \frac{1}{nh_n} \sum_{i=0}^{n-1} K\left(\frac{x - Z_i}{h_n}\right).$$

Using the same decomposition as in the proof of theorem 5.1 we get

$$\mathrm{E}\left(\widehat{f}_n(x) - f(x)\right)^2 \leq 2\mathrm{E}\left(\widehat{f}_n(x) - \widetilde{f}_n(x)\right)^2 + 2\mathrm{E}\left(\widetilde{f}_n(x) - f(x)\right)^2.$$

For the second term, we have under assumptions $A4.2$ and with $h_n = n^{-1/5}$ that

$$\mathrm{E}\left(\widetilde{f}_n(x) - f(x)\right)^2 = O\left(n^{-4/5}\right) + O\left(\frac{c_n}{n^2}\right).$$

For the first term, since K is supposed to be Lipschitzian, Cauchy-Schwarz's inequality implies

$$\mathrm{E}\left(\widehat{f}_n(x) - \widetilde{f}_n(x)\right)^2 \leq \frac{C}{nh_n^4} \sum_{i=0}^{n-1} \mathrm{E}\left(Y_i - Z_i\right)^2.$$

With the help of Taylor's formula, it can be established by recurrence that:

$$Y_i = \varphi^{(i)}(Y_0) + \beta_n \theta_i \Delta_i$$

where Δ_i is given by:

$$\begin{cases} \Delta_i &= \varepsilon_i + \Delta_{i-1}\varphi'(\varphi^{(i-1)}(Y_0) + \beta_n \theta_{i-1}\Delta_{i-1}) \\ \Delta_0 &= 0 \end{cases}$$

with $0 < \theta_i < 1$ for any i.
Note that for any i, Δ_i is independent of ε_{i+1}, thus one obtains successively

$$\begin{aligned} \mathrm{E}\,\Delta_i^2 &\leq \mathrm{E}\,\varepsilon_i^2 + \|\varphi'\|_\infty^2 \mathrm{E}\,\Delta_{i-1}^2 \\ \mathrm{E}\left(Y_i - Z_i\right)^2 &\leq \beta_n^2\,\mathrm{E}\,\Delta_i^2 \\ \mathrm{E}\left(Y_i - Z_i\right)^2 &\leq \beta_n^2\,\mathrm{E}\left(\varepsilon^2\right) \sum_{j=0}^{i-1} \|\varphi'\|_\infty^{2j}. \end{aligned}$$

These results imply that

$$\mathrm{E}\left(\widehat{f}_n(x) - \widetilde{f}_n(x)\right)^2 \leq \frac{C\beta_n^2 \mathrm{E}\left(\varepsilon^2\right)}{nh_n^4} \sum_{i=1}^{n-1}\sum_{j=0}^{i-1} \|\varphi'\|_\infty^{2j}.$$

The behaviour of $\|\varphi'\|_\infty^2$ now yields to theorem 5.3. ■

Appendix

1) Consider the linear autoregressive process

$$X_t = \rho\, X_{t-1} + \varepsilon_t,\ t \in \boldsymbol{Z},\ |\rho| < 1.$$

By reccurence, one obtains

$$X_t = \rho^t\, X_0 + \delta_t,\ t \geq 0$$

where

$$\delta_t = \varepsilon_t + \rho\ \varepsilon_{t-1} + \ldots + \rho^{t-1}\, \varepsilon_1,\ t \geq 1$$

and $\delta_0 = 0$. For "small" δ_t one gets the approximation

$$Y_t = \rho^t\, Y_0,\ t \geq 0$$

where $Y_0 = X_0$. Then it's easy to see that $\lim\limits_{t \to +\infty} Y_t = 0$.

2) Now let us consider a (possibly) non linear autoregressive process such as

$$X_t = \varphi(X_{t-1}) + \varepsilon_t,\ t \geq 1$$

where φ is a measurable function and X_0 is a given random variable.
We intend to approximate this model with a dynamical system such as $Y_t = \varphi^{(t)}(Y_0) + \delta_t$ where $\varphi^{(t)} = \varphi\, o \ldots o\, \varphi$, t times.

We can write

$$\begin{aligned} X_2 &= \varphi(X_1) + \varepsilon_2 \\ &= \varphi\left(\varphi(X_0) + \varepsilon_1\right) + \varepsilon_2. \end{aligned}$$

Let assume that ε_1 is small and that φ is an a.e. continuously differentiable function. We assume furthermore that X_1 has a density. Then, we have $\mathrm{P}\left(X_1 \in \bar{D_\varphi}\right) = 0$ where $\bar{D_\varphi}$ is the set of x such that $\varphi'(x)$ is not defined.

Thus,

$$\begin{aligned} \varphi\left(\varphi(X_0) + \varepsilon_1\right) &= \varphi\left(\varphi(X_0)\right) + \varepsilon_1 \varphi'\left(\varphi(X_0) + \theta\varepsilon_1\right) \\ &\sim \varphi^{(2)}(X_0) + \varepsilon_1 \varphi'\left(\varphi(X_0)\right) \end{aligned}$$

so we get approximatively

$$X_2 \sim \varphi^{(2)}(X_0) + \varepsilon_2 + \varepsilon_1 \varphi'\left(\varphi(X_0)\right).$$

Now setting $\delta_2 = \varepsilon_2 + \varepsilon_1 \varphi'\left(\varphi(X_0)\right)$, the next step gives

$$\begin{aligned} X_3 &= \varphi(X_2) + \varepsilon_3 \\ &= \varphi\left(\varphi(X_1) + \varepsilon_2\right) + \varepsilon_3 \\ &\sim \varphi\left(\varphi^{(2)}(X_0) + \delta_2\right) + \varepsilon_3 \\ &\sim \varphi^{(3)}(X_0) + \delta_2 \varphi'\left(\varphi^{(2)}(X_0)\right) + \varepsilon_3. \end{aligned} \tag{0.41}$$

Note that without approximations, one obtains:

$$\begin{aligned} X_3 &= \varphi(X_2) + \varepsilon_3 \\ &= \varphi\left(\varphi(X_1) + \varepsilon_2\right) + \varepsilon_3 \\ &= \varphi\left[\varphi\left(\varphi(X_0) + \varepsilon_1\right) + \varepsilon_2\right] + \varepsilon_3 \\ &= \varphi\left[\varphi\left(\varphi(X_0)\right) + \varepsilon_1 \varphi'\left(\varphi(X_0) + \theta\varepsilon_1\right) + \varepsilon_2\right] + \varepsilon_3 \\ &= \varphi^{(3)}(X_0) + \gamma_1 \varphi'\left[\varphi\left(\varphi(X_0)\right) + \theta_1 \gamma_1\right] + \varepsilon_3 \end{aligned} \tag{0.42}$$

where we have set $\gamma_1 = \varepsilon_1 \varphi'\left(\varphi(X_0) + \theta\varepsilon_1\right) + \varepsilon_2$ and $\theta = \theta(X_0, \varepsilon_1)$, $\theta_1 = \theta_1(X_0, \varepsilon_1, \varepsilon_2)$ with θ, $\theta_1 \in [0, 1]$.
Now (0.41) can be deduced from (0.42) by setting θ and θ_1 equal to 0.

More generally, one gets the following approximation:

$$\begin{aligned} X_k \sim\ & \varphi^{(k)}(X_0) + \varepsilon_k + \varepsilon_{k-1} \varphi'\left(\varphi^{(k-1)}(X_0)\right) + \ldots \\ & + \varepsilon_j \varphi'\left(\varphi^{(k-1)}(X_0)\right) \varphi'\left(\varphi^{(k-2)}(X_0)\right) \cdot \cdot \varphi'\left(\varphi^{(j)}(X_0)\right) + \ldots \\ & + \varepsilon_1 \varphi'\left(\varphi^{(k-1)}(X_0)\right) \ldots \varphi'\left(\varphi(X_0)\right). \end{aligned}$$

Finally, if we set $\Gamma_k = \varphi'\left(\varphi^{(k)}(X_0)\right)$, we have

$$X_k \sim \varphi^{(k)}(X_0) + \varepsilon_k + \Gamma_{k-1}\varepsilon_{k-1} + \Gamma_{k-1}\Gamma_{k-2}\varepsilon_{k-2} + \ldots + \Gamma_{k-1}\Gamma_{k-2} \ldots \Gamma_1 \varepsilon_1.$$

References

[1] Billingsley, P. (1969), *Convergence of Probability Measures*, Wiley, New-York.

[2] Blanke, D. (1995), Estimation de la densité pour des processus bruités à temps continu, *C. R. Acad. Sci. Paris*, t. 321, Série I, 623-626.

[3] Blanke, D. (1996), Estimation de la densité pour des trajectoires non directement observables, *Publ. Inst. Stat. Univ. Paris*, XXXX, fasc. 2, 21-36.

[4] Blanke, D. (1997), Ph. D. thesis, University Paris 6.

[5] Bosq, D. (1995), Optimal asymptotic quadratic error of density estimators for strong mixing or chaotic data, *Stat. and Prob. Letters*, 22, 339-347.

[6] Bosq, D. (1997), Une loi du zéro-l'infini. Application à l'estimation d'un filtre non linéaire faiblement bruité, *Romanian J. Pure and Appl. Math.* , 42, 7-8, 513-524.

[7] Bosq, D. (1998), *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*, Lecture Notes in Statistics 110, 2nd ed., New-York: Springer-Verlag.

[8] Bosq, D. and Guégan, D. (1995), Nonparametric estimation of the chaotic function and the invariant measure of a dynamical system, *Stat. Proba. Letters*, 25, 201-212.

[9] Brandstäter and Swinney (1987), *Phys. Rev. A.*, 35, 2207-.

[10] Carroll, R.J. and Hall, P. (1988), Optimal rates of convergence for deconvolving a density, *J. of the Amer. Stat. Assoc.*, Vol. 83, 1084-1088.

[11] Deissler and Farmer (1992), Deterministic noise amplifiers, *Physica D.*, 55, 155-165.

[12] Devroye, L. (1989), Consistent deconvolution in density estimation, *Canad. Journ. of Stat.*, Vol. 17, 2, 235-239.

[13] Fan, J. (1991a), On the optimal rates of convergence for nonparametric deconvolution problems, *Ann. of Stat.*, 19, 2, 1257-1272.

[14] Fan, J. (1991b), Asymptotic normality for deconvolution kernel density estimators, *Sankhya Ser. A*, 53, 97-110.

[15] Fan, J. (1992), Deconvolution with supersmooth distributions, *Canad. Journ. of Stat.*, Vol. 20, 2, 155-169.

[16] Fan, J. and Masry, E (1992), Multivariate regression estimation with errors-in- variables: asymptotic normality for mixing processes, *J. of Mult. Anal.*, Vol. 43, 2, 237-271.

[17] Guégan, D. and Mercier, L. (1998), Nonparametric forecasting techniques for mixing chaotic time series, *in Signal Analysis and Prediction*, ed. A. Procházka, J. Uhlíř, P.J.W. Razner, N. Kingsbury, Birkhäuser Boston.

[18] Farrel, R. (1972), On the best obtainable asymptotic rates of convergence in estimation of a density function at a point, *Ann. Math. Stat.*, 43, 170-180.

[19] Lasota, A. and Mac Key, M. (1994), *Chaos, Fractals and Noise - Stochastic Aspects of Dynamics*, 2nd ed., Springer-Verlag, New-York.

[20] Liu, M.C. and Taylor, R.L. (1989), A consistent nonparametric density estimator for the deconvolution problem, *Canad. Journ. of Stat.*, Vol. 17, 4, 427-438.

[21] Mäes, J. (1999), Ph. D. thesis, University Paris 6.

[22] Masry, E. (1991), Multivariate probability density deconvolution for stationary processes, *IEEE Trans. Inform. Theory*, 37, 1105-1115.

[23] Masry, E. (1993a), Strong consistency and rates for deconvolution estimators of multivariate densities of stationary processes, *Stoch. Proc. and Applic.*, 47, 53-74.

[24] Masry, E. (1993b), Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes, *Jour. of Mult. Anal.*, 44, 1, 47-68.

[25] Masry, E. (1993c), Multivariate regression estimation with errors in variables for stationary processes, *Nonparametric Statist.*, 3, 13-36.

[26] Neumann, M.H. (1997), On the effect of estimating the error density in nonparametric deconvolution, *Nonparametric Statist.*, 7, 307-330.

[27] Rosenblatt, M. (1956), Remarks on some nonparametric estimates of a density function, *Ann. Math. Statist.*, 27, 832-837.

[28] Smili, D. (1990), Ph. D. Thesis, University Paris 6.

[29] Stefanski, L.A. (1990), Rates of convergence of some estimators in a class of deconvolution problems, *Stat. and Proba. Letters*, 9, 229-235.

[30] Stefanski, L. and Carroll, R.J. (1990), Deconvoluting kernel density estimators, *Statistics*, 21, 2, 169-184.