

JIRSS (2002)

Vol. 1, Nos. 1-2, pp 127-142

On Efficiency Criteria in Density Estimation

Denis Bosq

University Paris 6, 175, street of Chevaleret, 8th stage, building A, 75013 Paris, France. (bosq@ccr.jussieu.fr)

Abstract. We discuss the classical efficiency criteria in density estimation and propose some variants. The context is a general density estimation scheme that contains the cases of i.i.d. or dependent random variables, in discrete or continuous time. Unbiased estimation, optimality and asymptotic optimality are considered. An example of a density estimator that satisfies some suggested criteria is given.

1 Introduction

The aim of this paper is to present a general approach to density estimation, discuss the usual efficiency criteria and suggest some variants of these criteria.

The general scheme is introduced in Section 2. It contains the cases of discrete or continuous random variables, i.i.d. or correlated, and observed in discrete or continuous time.

In Section 3 we describe general construction of density estima-

Received: September 2002

Key words and phrases: Continuous time, covariance operator, density estimation, discrete time, kernel estimator, local time, locally superoptimal, maxiset, minimax, optimality, preference, superoptimal rate, truncated projection estimator.

tors : if the empirical measure is absolutely continuous, its Radon-Nikodym derivative is a natural estimator of density. If not, an appropriate smoothing leads to the classical methods (kernel, projection).

Section 4 presents the common measures of accuracy for density estimators, discusses their drawbacks and propose some remedies.

Existence and non-existence of unbiased density estimators are considered in Section 5. In discrete time the emphasis is on the case where the family of all possible densities generates a reproducing kernel space. In continuous time we give some properties of the local time estimator.

Section 6 is devoted to asymptotics. Optimal rates in minimax sense are briefly discussed and we examine possible substitutes for minimax, namely maxisets and local superoptimality.

Finally, Section 7 studies a density estimator that satisfies some of the new criteria proposed in the paper.

2 A general scheme for density estimation

Let $X = (X_i, i \in I)$ be a real stochastic process defined on a Probability space (Ω, \mathcal{A}, P) ([1]) and belonging to some family \mathcal{X} of processes. Here I is an unbounded locally compact subgroup of $(\mathbb{R}, +)$. The X_i 's take their values in (E, \mathcal{B}_E) where E is a Borel set in \mathbb{R} and \mathcal{B}_E its Borel σ -field. They are equidistributed and their common distribution μ is supposed to be absolutely continuous with respect to a σ -finite measure λ defined on (E, \mathcal{B}_E) . The density f of μ is unknown and is a member of a family \mathcal{F} of densities.

The problem is to estimate f from observed variables $(X_i, i \in J)$ where $J \subset I$ is a compact subset of \mathbb{R} . We denote by m a Haar measure over (I, \mathcal{B}_I) i.e. a measure uniform on I (cf [19]), and by m_J a version of m such that $m_J(J) = 1$.

The above general scheme contains the classical models for real density estimation. In the sequel of this paper we shall focus on three important cases :

A $I = \mathbb{Z}$, m is the counting measure over \mathbb{Z} ($m\{i\} = 1, i \in \mathbb{Z}$); $J = \{1, \dots, n\}$; E is a countable subset of \mathbb{R} and λ is the counting measure over E . \mathcal{X} is a family of sequences of i.i.d. E -valued random variables or a family of (strictly stationary) discrete time processes. Here

$$f(x) = P(X_0 = x), \quad x \in E. \quad (2.1)$$

[B] I, J, m and \mathcal{X} are as in A but $E = \mathbb{R}$ and λ is Lebesgue measure. Thus f is the usual density of μ :

$$\mu(B) = \int_B f(x) \, dx, \quad B \in \mathcal{B}_{\mathbb{R}}. \tag{2.2}$$

[C] $I = E = \mathbb{R}, m = \lambda$ is Lebesgue measure, $J = [0, T]$ ($T > 0$). \mathcal{X} is a family of, eventually strictly stationary, continuous time random processes. Then f is again characterized by (2.2) (λ a.e.).

3 Construction of density estimators

Suppose that $(i, \omega) \mapsto X_i(\omega)$ is $\mathcal{B}_I \otimes \mathcal{A}$ -measurable, then a natural nonparametric estimator of μ is the **empirical measure** defined as

$$\mu_J(B) = \int_J \mathbf{I}_B(X_i) \, dm_J(i), \quad B \in \mathcal{B}_E \tag{3.1}$$

If μ_J is absolutely continuous with respect to λ , it induces the natural density estimator (D.E.)

$$f_{0,J} = \frac{d\mu_J}{d\lambda} \tag{3.2}$$

In example A, $f_{0,J} := f_{0,n}$ is the **frequency estimator** given by

$$f_{0,n}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}_{\{X_i=x\}}, \quad x \in E \tag{3.3}$$

In example B, μ_J is not absolutely continuous when in example C existence of $f_{0,J}$ is not guaranteed. If it does exist $f_{0,J} := f_{0,T}$ is the so-called **local time (LT) estimator** characterized by the relation

$$\frac{1}{T} \int_0^T \mathbf{I}_B(X_t) \, dt = \int_B f_{0,T}(x) \, dx, \quad B \in \mathcal{B}_{\mathbb{R}} \tag{3.4}$$

The LT associated with $(X_t, 0 \leq t \leq T)$ is then

$$\ell_T(x) = T f_{0,T}(x), \quad x \in \mathbb{R} \tag{3.5}$$

Note that a more explicit expression for LT is ([17])

$$\ell_T(x) = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \lambda \{t : 0 \leq t \leq T, |x - X_t| \leq \varepsilon\}, \quad \lambda \text{ a.e.} \quad (3.6)$$

Some indications concerning existence of LT are given in Section 5.

Now, if μ_J is not absolutely continuous, one can smooth it by setting

$$f_J^{(H)}(x) = \int_E H_J(x, y) d\mu_J(y), \quad x \in E, \quad (3.7)$$

where the function H_J is regular enough.

Two choice of H_J are popular. First

$$H_J(x, y) = h_J^{-1} K(h_J^{-1}(x - y)); \quad x, y \in \mathbb{R} \quad (h_J > 0) \quad (3.8)$$

where K is bounded and such that $\int_{\mathbb{R}} K(x) dx = 1$.

This leads to the **kernel estimator** ([26]):

$$f_J^{K, h_J}(x) = \int_J h_J^{-1} K(h_J^{-1}(x - X_i)) dm_J(i), \quad x \in \mathbb{R} \quad (3.9)$$

The other choice is

$$H_J(x, y) = \sum_{\ell=0}^{d_J} e_\ell(x) e_\ell(y); \quad x, y \in E \quad (3.10)$$

where $(e_\ell, \ell \geq 0)$ is an orthonormal system in $L^2(E, \mathcal{B}_E, \lambda)$. The associated **projection estimator** ([12]) is given by

$$f_J^{e, d_J}(x) = \sum_{\ell=0}^{d_J} \hat{a}_{\ell, J} e_\ell(x), \quad x \in E \quad (3.11)$$

where

$$\hat{a}_{\ell, J} = \int_J e_\ell(X_i) dm_J(i), \quad 0 \leq \ell \leq d_J \quad (3.12)$$

Usually one chooses a classical orthonormal system (trigonometric functions, Hermite functions, wavelets ...) but it is often more convenient to choose a special system well adapted to \mathcal{F} (see Section 7).

Finally note that, if $f \in L^2(E, \mathcal{B}_E, \lambda)$, f_J^{e, d_J} is an unbiased estimator (UE) of the projection of f over the linear space $sp\{e_\ell, 0 \leq \ell \leq d_J\}$.

4 Measures of accuracy and preferences

Selection of an estimator in a class \mathcal{F}_J of measurable density estimators may be performed via a relation of preference, that is a partial preordering on \mathcal{F}_J ([27]).

One way to define such a relation is to use a loss $L(f_J, f)$ based on a norm or a semi-norm. Typical examples are

$$|f_J(x) - f(x)|^p \quad (x \in E) \tag{4.1}$$

and

$$\|f_J - f\|_p^p \tag{4.2}$$

where $1 \leq p < \infty$, $f_J \in \mathcal{F}_J$ and $\|\cdot\|_p$ is the usual norm in $L^p(\lambda)$ ([1]).

The preference associated with (4.1) is

$$\begin{aligned} f_{J,1}(x) \prec_p f_{J,2}(x) &\iff \\ (\forall X \in \mathcal{X}), \mathbb{E}_X |f_{J,1}(x) - f(x)|^p &\leq \mathbb{E}_X |f_{J,2}(x) - f(x)|^p \end{aligned} \tag{4.3}$$

where \mathbb{E}_X denotes expectation with respect to the distribution P_X of the process X , f denotes density of X_i , $f_{J,1}$ and $f_{J,2}$ belong to \mathcal{F}_J .

A similar relation corresponds to (4.2). Concerning the choice of p it should be noticed that $\|\cdot\|_1$ is more natural since it induces a distance between probability measures that is invariant under some transformations and does not depend on the dominating measure ([14]). Actually, if f_J is the density of a Probability P_J , we have ([1])

$$\sup_{B \in \mathcal{B}_E} |P_J(B) - \mu(B)| = \frac{1}{2} \|f_J - f\|_1 \tag{4.4}$$

Now, $\|\cdot\|_2$ is more convenient since it is linked to a scalar product. This allows to decompose the error as a stochastic term and a term of bias. Finally a more accurate measure of distance between functions is $\|\cdot\|_\infty$.

On the other hand, preferences induced by (4.2) have a drawback : they do not take into account geometry of the space where f lays. Thus, if for example $p = 2$, a more precise (but more intricate) preference should be :

$$\begin{aligned} f_{J,1} \prec_{[2]} f_{J,2} &\iff (\forall X \in \mathcal{X}), (\forall g \in L^2(\lambda)), \\ \mathbb{E}_X (\langle f_{J,1} - f, g \rangle^2) &\leq \mathbb{E}_X (\langle f_{J,2} - f, g \rangle^2) \end{aligned} \tag{4.5}$$

Downloaded from jirss.irstat.ir at 4:15 +0430 on Saturday August 19th 2017

where $\langle \cdot, \cdot \rangle$ is the scalar product in $L^2(\lambda)$. If $E_X \|f_{J,i}\|_2^2 < \infty$ ($i = 1, 2$), (4.5) is equivalent to

$$f_{J,1} \prec_{[2]} f_{J,2} \iff (\forall X \in \mathcal{X}), \Gamma_{f_{J,1}-f}^{(X)} \leq \Gamma_{f_{J,2}-f}^{(X)} \quad (4.6)$$

where $\Gamma_{f_{J,i}-f}^{(X)}$ is the **operator of order 2** associated with $f_{J,i} - f$, defined by

$$\Gamma_{f_{J,i}-f}^{(X)}(g) = E_X (\langle f_{J,i}, g \rangle (f_{J,i} - f)), \quad g \in L^2(\lambda)$$

and \leq is the classical ordering on the class of symmetric operators ([30]).

Despite the operator of order 2 is the natural extension of the finite dimensional matrix of order 2, it scarcely appears in literature concerning density estimation. This is really surprising even if the qualitative character of $\prec_{[2]}$ is not easy to handle. A result using Γ appears in Section 7.

Another aspect that is often neglected is the fact that a good density estimator must mimic the shape of the curve associated with f . This property is not guaranteed by a small integrated quadratic error, see Figure 1 where the empirical integrated quadratic error is small but the shape of curve is not restored. In this context $\|\cdot\|_\infty$ is interesting as well as

$$\|f - f_J\|_{2,\infty} = \|f - f_J\|_\infty + \|f' - f'_J\|_\infty + \|f'' - f''_J\|_\infty \quad (4.7)$$

Concerning **optimality** it is well known that, in many situations, an exact optimal D.E. does not exist. An outstanding exception is the case where an unbiased estimator is available (see Section 5).

Now a weak form of optimality is **minimaxity**. Recall that a D.E. $f_{M,J}$ is said to be minimax if ([27])

$$\min_{f_J \in \mathcal{F}_J} \sup_{X \in \mathcal{X}} E_X L(f_J, f) = \sup_{X \in \mathcal{X}} E_X L(f_{M,J}, f) \quad (4.8)$$

In [32], Wertz has shown existence of a minimax D.E., in the i.i.d. case, with respect to a special Banach space type loss. Unfortunately his proof uses compactness arguments, so that explicit construction of $f_{M,J}$ is not possible in general.

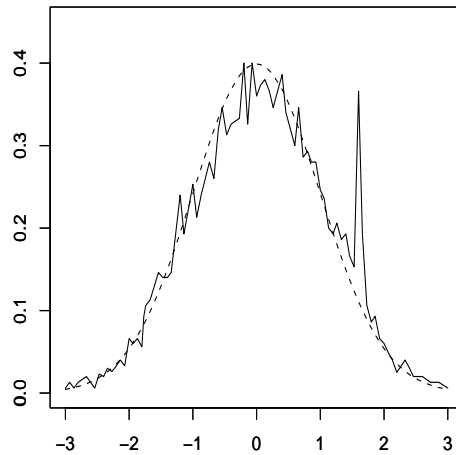


Figure 1: L.T.D.E. for Gaussian process (simulation, [25])

5 Unbiased density estimators

Unbiasedness is a property that generates optimal estimators via completeness at least in the i.i.d. case ([11]).

Recall that f_J is an unbiased density estimator (UDE) if

$$(\forall X \in \mathcal{X}), E_X f_J(x) = f(x) \quad (\lambda \text{ a.e.}) \quad (5.1)$$

If an UDE does exist, f is said to be **estimable**.

5.1 Non-existence of UDE in the general case

In his famous pioneer work ([26]), Rosenblatt showed that, in the i.i.d. case, a UDE cannot exist if \mathcal{F} is the class of all continuous real densities.

His proof is based on the following arguments : if f_n is UDE the associated measure provides a UE for the distribution function. By completeness this UE is the empirical distribution function F_n . This is a contradiction since F_n is not absolutely continuous. Thus the key of the proof is non-absolute continuity of the empirical measure with respect to Lebesgue measure.

It is also possible to explain non-existence of UDE by using the fact that $f(x)$ is a **local parameter**. We refer to [11], chapter II for

details. As an application we have the following : if \mathcal{F} is the family of densities that have regularity r in a neighbourhood V_f of x , a UE for $f(x)$ does not exist. Here we have $r \geq 0$ and ‘f has regularity r ’ means that the derivative $f^{[r]}$ does exist in V_f and satisfies a Holder condition of order $r - [r]$ (resp. is bounded and continuous if $[r] = r$).

5.2 Existence of UDE in special cases

We now turn to situations where construction of UDE is feasible.

The following statement furnishes necessary and sufficient conditions for existence of UDE if $\#J = 1$.

Proposition 5.2.1. ([11])

1) If \mathcal{F} contains the densities $[\lambda(B)]^{-1} \mathbf{1}_B$, $B \in \mathcal{B}_E$ ($0 < \lambda(B) < \infty$) then f is estimable if and only if λ is purely atomic.

2) If $\mathcal{F} \in L^2(\lambda)$ then there exists a UDE $f_1(X_1, \cdot)$ such that $f_1(\cdot, x) \in sp(\mathcal{F})$, $x \in E$, if and only if $sp(\mathcal{F})$ equipped with the scalar product of $L^2(\lambda)$ is a reproducing kernel prehilbertian space.

The first part of this Proposition characterizes model A. We give two examples of applications of the second part :

a) If $sp(\mathcal{F})$ is finite-dimensional, f is estimable and $f_1(\cdot, \cdot)$ is the reproducing kernel of $sp(\mathcal{F})$.

b) In model B, if \mathcal{F} is the family of continuous square integrable densities with support of their Fourier transform included in $[-\frac{1}{2}, \frac{1}{2}]$ then

$$f_1(x) = \frac{\sin((X_1 - x)/2)}{(X_1 - x)/2}, \quad x \in \mathbb{R} \tag{5.2}$$

This example appears in [20].

We now turn to model C. In this case, the local time estimator is UDE. More precisely it is easy to establish the following :

Lemma 5.2.1. ([7]) *Let $X = (X_t, 0 \leq t \leq T)$ be a measurable process such that $P_{X_t} = \mu$, $0 \leq t \leq T$, and that admits a local time ℓ_T .*

Then, μ is absolutely continuous with density f and $E_X \left(\frac{\ell_T(\cdot)}{T} \right)$ is a version of f .

Concerning **existence of local time**, it is ensured by the following conditions ([17]).

a) If X has absolutely continuous sample paths then

$$P(X'(t) = 0) = 0 \text{ for almost all } t \tag{5.3}$$

is necessary and sufficient for existence of LT.

b) X admits a LT satisfying $\ell_T \in L^2(\lambda \otimes P)$ if and only if

$$\liminf_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \iint_{[0,T]^2} P(|X_t - X_s| \leq \varepsilon) \, ds \, dt < \infty \tag{5.4}$$

This condition is connected with irregularity of sample paths (see [10]).

In order to study some properties of LTE we now introduce a condition slightly stronger than (5.4) :

H. *The density $f_{s,t}(y, z)$ of (X_s, X_t) is defined and measurable over $(D^c \cap [0, T]^2) \times G$ where G is an open neighbourhood of $D = \{(x, x), x \in \mathbb{R}\}$, and the function*

$$F_T(y, z) = \iint_{[0,T]^2} f_{s,t}(y, z) \, ds \, dt \tag{5.5}$$

is finite in a neighbourhood of D and continuous at each point of D .

We then have

Proposition 5.2.2. ([7]) *If H holds, X has a LT such that*

$$\lim_{h \rightarrow 0} \sup_{a \leq x \leq b} E_X \left| f_T^{K,h}(x) - f_{0,T}(x) \right| = 0, \quad a < b \tag{5.6}$$

$K \in \{K : K \geq 0, \int K = 1, K \text{ a.e. continuous, with compact support}\}$.

Moreover $E_X(f_{0,T})$ is a continuous version of f and, if $f_{s,t} = f_{|t-s|}$, $0 \leq s, t \leq T$, variance of $f_{0,T}(x)$ is

$$V_X f_{0,T}(x) = \frac{2}{T} \int_0^T \left(1 - \frac{u}{T}\right) (f_u(x, x) - f^2(x)) \, du, \quad x \in \mathbb{R} \tag{5.7}$$

Note that (5.6) signifies that the kernel estimator is an approximation of $f_{0,T}$. A similar result may be obtained for the projection estimator.

It is noteworthy that $f_{0,T}$ is accurate if sample paths are irregular. If not, (5.3) and lemma 5.2.1 ensure existence and unbiasedness of $f_{0,T}$ but in general this LTE has the non-attractive property $V_X f_{0,T}(x) = \infty$!

6 Asymptotics

Asymptotics gives information concerning selection of a ‘good’ density estimator but this point of view has some limitations. As observed by Van der Vaart ([31]) : ‘*Although asymptotics is both practically useful and of theoretical importance, it should not be taken for more than what it is : approximations*’.

In the general scheme presented in Section 2 asymptotics is expressed by $m(J) \rightarrow \infty$ where m is a fixed version of Haar measure.

In this context efficiency is evaluated by rates of convergence associated with criteria described in Section 4.

6.1 Asymptotic quadratic error

In this subsection we briefly indicate some optimal rates (in minimax sense) in mean square. Such a rate, say v_J , is defined by :

$$\lim_{m(J) \rightarrow \infty} \inf_{f_J \in \mathcal{F}_J} \sup_{X \in \mathcal{X}} v_J^{-1} E_X (f_J(x) - f(x))^2 > 0, \quad x \in \mathbb{R} \quad (6.1)$$

and

$$\sup_{X \in \mathcal{X}} E_X (f_{M,J} - f(x))^2 = O(v_J), \quad x \in \mathbb{R} \quad (6.2)$$

as $m(J) \rightarrow +\infty$ and for some $f_{M,J} \in \mathcal{F}_J$.

In case A the optimal rate is $v_J := v_n = n^{-1}$, when in case B with i.i.d. observations it is well known that $v_n = n^{-2r/(2r+1)}$ where r is regularity of f , (see e.g. [29]). This rate may be extended to strong mixing mixing processes under some conditions (see [7]). If f is analytic the rate turns to be $\frac{\log n}{n}$ ([18], [20]).

Concerning model C, $v_J := v_T$ depends on smoothness of sample paths of the observed process. If sample paths are regular enough $v_T = T^{-2r/(2r+1)}$ ([6], [7]). Now irregularity of sample paths provides additional information that allow to improve rates. In order to specify that improvement we introduce the following condition:

$$\mathbf{H}'. \quad \limsup_{T \rightarrow \infty} \frac{1}{T} \iint_{[0,T]^2} [f_{s,t}(x, x) - f^2(x)] ds dt < \infty$$

Proposition 6.1.1. ([7]) *If H and H' hold*

$$V_X f_{0,T}(x) = O(T^{-1}) \tag{6.3}$$

Note that this ‘superoptimal’ rate occurs as soon as f is continuous. Thus regularity of f does not influence rate of $f_{0,T}$.

The same exceptional rate takes place for f_T^{K,h_T} but, in order to control bias, one must choose h_T according to r (cf [5], [7]).

Minimaxity of T^{-1} is obtained in [?]. Moreover a family of intermediate minimax rates of the form $(\log T)^a T^{-b}$ ($a > 0, b > 0$) hold for appropriate families of processes (cf [3]).

Finally, it can be proved that suitable sampling schemes retain the various rate (cf [4] and [7]).

Results of the same type can be derived for $E_X \|f_J - f\|_2^2$ ([23], [28]) and $\|f_J - f\|_\infty$ ([7], [28]).

6.2 Adaptive estimators

If regularity of r is unknown one may employ procedures that, in a way, simultaneously estimate r and f (see [24]). In the i.i.d. case these ‘adaptive estimators’ are asymptotically minimax, eventually within a logarithmic term (see e.g. [15]). This kind of results is of theoretical importance and leads to nice mathematics. However their usefulness is somewhat restricted since densities that appear in practice are in general C_∞ or piecewise- C_∞ .

Now, in continuous time, the situation is a bit different: f_T^{K,h_T} and f_T^e require adaptivity (see e.g. [13]) whereas the LTE achieves the best rate as soon as f is continuous, under assumptions H and H’ (Proposition 6.1.1). Here smoothness of sample paths influences rates. An adaptive estimator with respect to that smoothness appears in [2].

6.3 Substitutes for minimax

Critics against the pessimistic ‘minimax principle’ are well known. For example, P. Hall in the discussion paper [21] claims : ‘ *If we*

organized our lives so that we performed as well as possible the worst of circumstances, but made a mess of things the rest of time, we'd not get very far. So too it is with statistical estimators'.

G. Kerkyacharian and D. Picard ([21]) have recently suggested to substitute for the minimax principle the notion of **maxiset**.

The maxiset of a given density estimator is the set of densities over which the estimator attains a specific rate of convergence.

Continuous time provides an example of maxiset: in the context of our general scheme 'the set of densities' should be replaced by 'the set of processes'. For example, the maxiset of $f_{0,T}$ is the set of processes satisfying H and H', when for f_K^{K,h_T} an additional condition appears: f must have regularity r , provided $h_T \simeq T^{-1/2r}$.

Another remedy to minimax should be selection of a privileged small subset \mathcal{F}_0 of densities. A 'good' estimator would achieve a superoptimal rate (i.e. better than the optimal rate over \mathcal{F}) over \mathcal{F}_0 dense in \mathcal{F} (**local superoptimality**) and a quasi-optimal rate over $\mathcal{F} - \mathcal{F}_0$. The example studied in the next Section illustrates this strategy.

7 An example

We now consider an example of density estimator that illustrates some ideas suggested above.

Let $(X_i, i \in \mathbb{N})$ be a sequence of i.i.d. real random variables with values in $(E, \mathcal{B}_E, \lambda)$ where $E = (a, b)$ ($-\infty \leq a < b \leq \infty$) and λ is a Probability measure.

Let $e_0 = 1, e_1, e_2, \dots$ be a bounded orthonormal basis of $L^2(\lambda)$ supposed to be infinite dimensional and separable. We suppose that f belongs to $L^2(\lambda)$ and define the estimator

$$\hat{f}_n = \sum_{j=0}^{\hat{d}_n} \hat{a}_{jn} e_j \tag{7.1}$$

where $\hat{a}_{jn} = \frac{1}{n} \sum_{i=1}^n e_j(X_i)$, $0 \leq j \leq d_n$ and

$$\hat{d}_n = \max \left\{ j : 0 \leq j \leq d_n, |\hat{a}_{jn}| \geq c \left(\frac{\log n}{n} \right)^{1/2} \right\} \tag{7.2}$$

where $c > 0$ and the integer d_n are selected by the Statistician. Slightly different estimators appear in [16] and [15].

If $d_n \rightarrow \infty$ and $\frac{d_n}{n} \rightarrow 0$ a suitable choice of c implies that \widehat{d}_n is asymptotically close to an optimal truncating index (cf [8], [9]). Thus \widehat{f}_n is an adaptive estimator.

Now set $a_j(f) = \int e_j f \, d\lambda$, $j \geq 0$, $f \in L^2(\lambda)$;

$$\mathcal{F}'_0(k) = \{f : f \in L^2(\lambda), a_k(f) \neq 0, a_j(f) = 0, j > k\}$$

and

$$\mathcal{F}'_0 = \cup_{k=0}^{\infty} \mathcal{F}'_0(k)$$

then \mathcal{F}'_0 is dense in $L^2(\lambda)$ and we have

$$nC_{\widehat{f}_n - f} \xrightarrow{\mathcal{N}} C_{(\sum_{j=0}^k e_j(X_0) e_{j-f})}, \quad f \in \mathcal{F}'_0(k), k \geq 0 \quad (7.3)$$

where \mathcal{N} stands for convergence in nuclear norm ([30]), hence

$$nE_X \left\| \widehat{f}_n - f \right\|_2^2 \rightarrow \sum_{j=0}^k \left(\int e_j^2 f - a_j^2(f) \right), \quad f \in \mathcal{F}'_0 \quad (7.4)$$

Moreover the maxiset of \widehat{f}_n associated with the superoptimal rate n^{-1} is exactly \mathcal{F}'_0 .

On the other hand

$$\left\| \widehat{f}_n - f \right\|_{\infty} = O \left(\left(\frac{\log \log n}{n} \right)^{1/2} \right) \quad \text{a.s., } f \in \mathcal{F}'_0 \quad (7.5)$$

Now it can be established that, if

$$\mathcal{F} = \{f : |a_j(f)| \leq \varphi(j), j \geq 0\}$$

where $\varphi(j) \downarrow$ and $\sum_{j=0}^{\infty} \varphi^2(j) < \infty$, then

$$\sup_{f \in \mathcal{F}} E_X \left\| \widehat{f}_n - f \right\|_2^2 = O(v_n \log n), \quad f \in \mathcal{F} - \mathcal{F}_0 \quad (7.6)$$

where $\mathcal{F}_0 = \mathcal{F} \cap \mathcal{F}'_0$ and (v_n) is the minimax rate associated with a family of linear estimators (cf [9]). A similar result holds for $\left\| \widehat{f}_n - f \right\|_{\infty}$.

Finally, according to the idea given in Section 6, one may choose a given sequence $(g_\ell, \ell \geq 0)$ of densities and construct $(e_\ell, \ell \geq 0)$ by orthonormalization. If (e_ℓ) is uniformly bounded, the associated estimator \widehat{f}_n has the rate n^{-1} over the convex set \mathcal{F}'_0 generated by $(g_\ell, \ell \geq 0)$.

References

- [1] Billingsley, P. (1979), Probability and Measure. New York: Wiley.
- [2] Blanke, D. (2002), Sample paths adaptive density estimators. Submitted, **41**.
- [3] Blanke, D. and Bosq, D. (2000), A family of minimax rates for density estimators in continuous time. *Stochastic Anal. Appl.*, **18**(6), 871-900.
- [4] Blanke, D. and Pumo, B. (2002), Optimal sampling for density estimation in continuous time. To appear in *J. Time Ser. Anal.*
- [5] Bosq, D. (1997), Parametric rates of nonparametric estimators and predictors for continuous time processes. *Ann. Statist.*, **25**(3), 982-1000.
- [6] Bosq, D. (1998), Minimax rates of density estimators for continuous time processes. *Sankhyā Ser. A*, **60**(1), 18-28.
- [7] Bosq, D. (1998), Nonparametric Statistics for Stochastic Processes. Estimation and prediction. *Lecture Notes in Statistics* (110), Second ed., New York: Springer-Verlag.
- [8] Bosq, D. (2002), Estimation localement suroptimale et adaptative de la densité. *C.R. Acad. Sci. Paris, Ser. I*, **334**, 591-595.
- [9] Bosq, D. (2002), Estimation suroptimale de la densité par projection. *Tech. Rep.*, **4**, Paris 6.
- [10] Bosq, D. and Davydov, Y. A. (1999), Local time and density estimation in continuous time. *Math. Methods Statist.*, **8**, 22-45.
- [11] Bosq, D. and Lecoutre, J. P. (1987), *Théorie de l'estimation Fonctionnelle*. Paris: Economica.
- [12] Cencov, N. N. (1962), Evaluation of an unknown distribution density from observations. *Soviet. Math.*, **3**, 1559-1562.
- [13] Comte, F., Merlevède, F. (2002), Adaptive estimation of the stationary density of discrete and continuous time mixing processes. To appear in *ESAIM Probability and Statistics*.

- [14] Devroye, L. (1987), *A Course in Density Estimation*. Boston: Birkh auser.
- [15] Donoho, D. L., Johnstone, I., Kerkyacharian, G., and Picard, D. (1996), Density estimation by wavelet thresholding. *Ann. Statist.*, **24**(2), 508-539.
- [16] Efromovich, S. (1999), *Nonparametric Curve Estimation*. New-York: Springer.
- [17] Geman, D. and Horowitz, J. (1980), Occupation densities. *Ann. Probab.*, **8**(1), 1-67.
- [18] Golubev, G. K. and Levit, B. Y. (1996), On the second order minimax estimation of distribution functions. *Maths. Methods Statist.*, **5**(1), 1-31.
- [19] Halmos, P. (1950), *Measure Theory*. Van Nostrand: Princeton.
- [20] Ibragimov, I. A. and Has'minskii, R. Z. (1990), On density estimation in the view of kolmogorov's ideas in approximation theory. *Ann. Statist.*, **18**(3), 999-1010.
- [21] Kerkyacharian, G. and Picard, D. (2000), Thresholding algorithms. *Maxisets and well-concentrated Bases. (With discussion) Test*, **9**(2), 283-344.
- [22] Kutoyants, Y. A. (1997), Some problems of nonparametric estimation by observations of ergodic diffusion processes. *Statist. Probab. Lett.*, **32**, 311-320.
- [23] Leblanc, F. (1997), Density estimation for a class of continuous time processes. *Math. Methods Statist.*, **6**(2), 171-199.
- [24] Lepskii, O. V. (1990), On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.*, **35**, 454-466.
- [25] Noubir, A. (1999), Local time density estimator. Unpublished Manusript, Mémoire Univ. Paris 6.
- [26] Rosenblatt, M. (1956), Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, **27**, 832-837.
- [27] Savage, L. (1972), *The Foundations of Statistics*. New York: Dover.

- [28] Schipper, M. (1997), Sharp asymptotics in nonparametric estimation. Ph.d. Thesis, Utrecht.
- [29] Stone, C. J. (1980), Optimal rates of convergence for nonparametric estimators. *Ann. Stat.*, **8**(6), 1348-1360.
- [30] Vakhania, N. N., Tarieladya, V. I., and Chobanyan, S. A. (1987), *Probability Distributions on Banach Spaces*. Dordrecht: Reidel.
- [31] Van der Vaart, A. W. (1998), *Asymptotic Statistics*. Cambridge Univ. Press.
- [32] Wertz, W. (1974), On the existence of density estimators. *Studia Sci. Math. Hungar.*, **9**, 45-50.