Adaptive sampling schemes for density estimation

Delphine Blanke

Université Paris 6, L.S.T.A., 175 rue du Chevaleret, 8ème étage, Bat. A, 75013 Paris, France

Abstract

In this paper, first we propose sharp sampling schemes for nonparametric density estimation for discretely observed continuous time processes. Next, since such samplings depend on two coefficients: $\gamma_0, r_0$ linked with regularity of the underlying sample paths and the density respectively, we give and study the pointwise asymptotic behaviour of an adaptive kernel estimator in the case of known $\gamma_0$ and unknown $r_0$. Finally we propose a procedure in the case of both coefficients unknown.

Key words: continuous time, density estimation, sampling scheme, adaptive estimation, pointwise estimation, kernel estimator.

1991 MSC: 62G07, 62H12, 62M.

1 Introduction

In functional estimation for continuous time processes it is of great theoretical interest to examine the case where a whole sample path is observed. On the contrary considering data collected in discrete time is more appropriate for applications. In this context various deterministic or random samplings have been studied by various authors: see e.g. the work of Wu (1997) and references therein. In this paper, we investigate high rate sampling: we suppose that observations $X_{t_{i,n}}$ are delivered at instants $t_{i,n}$, $i = 1, \ldots, n$ such that $t_{i,n} = i\delta_n$, $\delta_n \to 0$, $n\delta_n \to \infty$ : in other words the process is observed frequently and during a long time. What is the meaning of such a modelization? Actually, we want to adjust the sampling scheme to the feature of the underlying process.

1 I am grateful toward the referees whose remarks and comments enabled me to improve the presentation of this work.

Email address: dblanke@ccr.jussieu.fr (Delphine Blanke).
Clearly a very irregular sample path will indicate that the successive observed variables $X_{\delta_n}, X_{2\delta_n}, \ldots$, are not highly correlated when a very regular one, should imply a great local dependance. A natural consequence is that more the sample path is regular more sampled observations should be chosen far away from each others, in order to obtain a good behaviour of the kernel estimator. The practical interest appears also in the following two situations: first if the total time of observation is a given, large enough, $T_n$ the relation $T_n = n\delta_n$ shows that a minimal $\delta_n$ will indicate a maximal number $n$ of points in $[0, T_n]$ playing a part in estimation. On the other hand, if a maximal (large enough) sample size $n$ is given (e.g. one wants to minimize costs for taking observations from the underlying process), a suitable choice of $\delta_n$ makes possible to reduce the total time $T_n = n\delta_n$ of the experiment.

Our work should be seen in the continuity of earlier papers about density estimation in continuous time. First of all Castellana and Leadbetter (1986) (‘C.L.’) pointed out that if $\{X_t, t \in [0, T]\}$ is observed then for some family of processes, one may construct estimators with a mean-square (‘m.s.’) parametric rate of convergence $1/T$. Such a phenomenon was next studied and developed by various authors. We refer e.g. to the book by Bosq (1998) and references within as well as to Kutoyants (1997, 1999, 2003) for works concerning ergodic diffusion processes. It is shown in Blanke and Bosq (1997, 2000) that all a family of rates may be expected: they are minimax and depend on a positive coefficient, $\gamma_0$, linked with regularity of sample paths. Strongly speaking, more local irregularity is present, more $\gamma_0$ is small and more the rate of convergence is good. For example, still for the ‘m.s.’ criterion, $\gamma_0 < 1$ (e.g. real diffusion processes) gives the rate $1/T$, $\gamma_0 = 1$ (e.g. real ‘m.s.’ differentiable ones) gives $\ln T/T$, and $\gamma_0 > 1$ furnishes weaker rates. Related results were also independently obtained by Sköld and Hössjer (1999), Sköld (2001a, 2001b) under slightly different conditions.

Concerning the sampled case, Bosq (1995, 1998) proposed and studied under ‘C.L.’s condition (namely $\gamma_0 < 1$), an ‘admissible sampling scheme’ that is: observations separated by $\delta_n, \delta_n \to 0$ and $T_n := n\delta_n \to \infty$, where $\delta_n$ is chosen in order to reach the full rate $1/T_n$. Next, this work was extended to regression estimation, see Bosq and Cheze-Payaud (1999), wavelet estimators, see Leblanc (1995). For general values of $\gamma_0$ we refer to Blanke and Pumo (2003) where sampling schemes are proposed and studied for the m.s. criterion.

In Section 2, first we give the main assumptions and notation used along the paper. Next the pointwise convergence of the kernel estimator is established. We also present some specific classes of processes meeting our conditions. Since sampling schemes are governed by $r_0$, usually unknown, we propose in Section 3 an adaptive estimator with respect to (w.r.t.) such parameter. The number of publications in the area of adaptive estimation is growing very fast and there exist many approaches. Here we follow in spirit that initiated by Lepskii (1990). Further literature concerning adaptive density estimation includes
among others: Birgé and Massart (1997), Donoho, Johnstone, Kerkyacharian and Picard (1996), Efromovich (1996), Tribouley and Viennet (1998). Finally we refer to Lepski, Mammen and Spokoiny (1997) for a comprehensive discussion of different approaches to nonparametric adaptive estimation. More closer to our results in continuous time, adaptive estimators have also been studied by Comte and Merlevède (2002a, 2002b) but in a different framework since their estimator is a data-driven projection one (built via minimization of a penalized contrast) and the obtained results stand only under the ‘C.L.’s framework ($\gamma_0 < 1$). Here, the originality of our paper is that we consider more general classes of processes (including e.g. real differentiable ones) and moreover in Section 4, we propose an adaptive procedure with respect to both $r_0$ and $\gamma_0$. We show that rates of Section 2 ($r_0, \gamma_0$ known) are reached in these adaptive contexts. Proofs appear in the final Section.

2 Almost sure convergence with known $(r_0, \gamma_0)$

2.1 Assumptions and Notation

Let $\{X_t, t \in \mathbb{R}\}$ be a $\mathbb{R}^d$-valued measurable process, defined on the probability space $(\Omega, \mathcal{A}, P)$ and suppose that the $X_t$’s have a common distribution $\mu$ admitting a density $f$ w.r.t. Lebesgue measure $\lambda$ over $\mathbb{R}^d$. In this section, we study strong consistency of the density estimator $\hat{f}_n(x)$, $x \in \mathbb{R}^d$, built on a discretized sample path $\{X_{t_1}, \ldots, X_{t_n}\}$. We use the standard kernel density estimator defined as

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_{t_i}}{h_n}\right)$$

where $h_n$ is the bandwidth, $h_n \to 0$, $nh_n^d \to \infty$ as $n \to +\infty$, and $K$ belongs to $\mathcal{K}_E$ for some compact set $E \subset \mathbb{R}^d$:

$$\mathcal{K}_E = \left\{ K : \mathbb{R}^d \mapsto \mathbb{R}, \int_E K(u) \, du = 1, K(u) = 0 \text{ for } u \notin E, \|K\|_\infty < \infty, \|K\|_1 < \infty \right\}.$$

Concerning the pointwise behaviour of $\hat{f}_n(x) - E\hat{f}_n(x)$, we need the following assumptions.

Assumption 2.1 (A2.1)

(i) $f$ is bounded and continuous at $x$,

(ii) for $s \neq t$, the bivariate density of $(X_s, X_t)$, say $f_{(X_s,X_t)}$, does exist and satisfies $f_{(X_s,X_t)} = f_{(X_0,X_{t-s})}$ for $t > s$;

(iii) $\{X_t, t \in \mathbb{R}\}$ is a geometrically strongly mixing (G.S.M.) process :

$$\exists 0 < \rho < 1, \alpha_0 > 0 : \forall u \in [0, +\infty[, \alpha_x(u) \leq \alpha_0 \rho^u$$
where \( \alpha_x(.) \) is the classical strong mixing coefficient of \( \{X_t, t \in \mathbb{R}\} \) defined by
\[
\alpha_x(u) = \sup_{t \in \mathbb{R}} \sup_{A \subset \sigma(X_{s}, s \leq t)} |P(A \cap B) - P(A)P(B)|, \quad u > 0;
\]
(iv) let \( g_u := f(x_0,x_u) - f \otimes f, \ u > 0 : \exists u_1 \geq u_0 > 0 : \forall u \in [u_0, +\infty[, \ |g_u|_\infty \leq \pi(u) \) for a bounded function \( \pi, \) decreasing and integrable over \( [u_1, +\infty[ ;
\]
(v) \( \exists \gamma_0 \in ]0, +\infty[ : f(x_0,x_u)(y,z) \leq M(x,y,z,u)u^{-\gamma_0}, \ \forall u \in \mathcal{V} \times ]0, u_0[ \)
with \( M \) continuous at \( (x,x) \) and \( \mathcal{V} \) denoting an open neighbourhood of \( x. \)

Let us give some precisions concerning Assumption 2.1. In condition A2.1-(ii) we have a quite weak assumption of stationarity, whereas conditions A2.1-(iii), A2.1-(iv) should be linked with properties of asymptotic independence for \( \{X_t, t \in \mathbb{R}\}. \) Note also that A2.1-(iii) should imply A2.1-(iv) in some situations (for example when \( \{X_t\} \) is a real stationary Gaussian process with an appropriate correlation function \( \rho(u) \) bounded away from 1 outside some neighbourhood of 0). We refer to Veretennikov (1999), Comte and Merlevède (2002b) for a detailed discussion about condition A2.1-(iv). Note also that this condition, in relation with A2.1-(v), can be weakened for values \( \gamma_0 \geq 1 \) (see Blanke and Bosq, 2000). Moreover to simplify the presentation of our results we do not consider here the case of arithmetic strong mixing processes but note that our results may be extended to such processes with slight modifications. Condition A2.1-(v) appears as less usual but it is quite typical of the continuous time framework since it should be linked with local regularity of sample paths. More precisely if \( Y_u := \left( \frac{X_u^{(1)} - X_{u_1}^{(1)}}{u_1^{d}}, \ldots, \frac{X_u^{(d)} - X_{u_d}^{(d)}}{u_d} \right) \) (where \( X_t^{(i)} \) denotes the \( i \)-th component of \( (X_t) = (X_t^{(1)}, \ldots, X_t^{(d)}) \) and the \( \gamma_i, \ 0 < \gamma_i \leq 1, \ i = 1, \ldots, d \) are Hölder type coefficients), condition A2.1-(v) is equivalent to
\[
f(x_0,x_u)(y, \frac{z - y}{u^\gamma}) \leq M(y,z), \quad (y,z,u) \in \mathcal{V}^2 \times ]0, u_0[ \quad (2.1)
\]
where \( \frac{z - y}{u^\gamma} := \left( \frac{z_1 - y_1}{u_1^{d}}, \ldots, \frac{z_d - y_d}{u_d} \right) \) and provided that \( \gamma_0 = \sum_{i=1}^{d} \gamma_i. \) Roughly speaking \( \gamma_0 \) is linked with the ‘total’ sample path regularity of \( \{X_t, t \in \mathbb{R}\}. \) We have postponed at the end of this section concrete classes of processes fulfilling Assumption 2.1.

For study of bias, some additional assumptions are required on \( f \) and \( K. \) First we introduce condition \( C^{(r)}_K \) (for \( K \in \mathcal{K}_E \) and positive integer \( r) : \)
\[
C^{(r)}_K \left\{ \begin{array}{l}
f_E u_1^{\alpha_1} \ldots u_d^{\alpha_d} K(u_1, \ldots, u_d) du_1 du_2 \ldots du_d = 0,

f_E |u_1|^{\beta_1} \ldots |u_d|^{\beta_d} ||u|| K(u_1, \ldots, u_d) du_1 \ldots du_d < +\infty,

\forall \alpha_i : \alpha_1 + \ldots + \alpha_d = k, k = 1, \ldots, r - 1, \forall \beta_i : \beta_1 + \ldots + \beta_d = r
\end{array} \right.
\]

\[4\]
and we also use the following condition:

\[
\mathcal{C}_f^{(r)} \left\{ \begin{array}{l}
\text{\( f \) is \( r \) - times differentiable with either \( f^{(r)} \) a bounded continuous function}
\text{or for } j_1 + \cdots + j_d = r \text{ and } (\ell, \lambda) \in [0, +\infty] \times [0, 1]:
\quad \left| \frac{\partial^{(r)} f}{\partial x_1^{j_1} \cdots \partial x_d^{j_d}}(y) - \frac{\partial^{(r)} f}{\partial y_1^{j_1} \cdots \partial y_d^{j_d}}(z) \right| \leq \ell \| y - z \|^\lambda.
\end{array} \right.
\]

**Assumption 2.2 (A2.2)**

(i) \( f \) fulfills condition \( \mathcal{C}_f^{(r_0)} \) for \( (y,z) \in \mathcal{V}^2 \);
(ii) \( K \) fulfills condition \( \mathcal{C}_K^{(r_0)} \).

### 2.2 Main results

At this stage, we suppose that observations are delivered at sampling times \( t_{1,n}, \ldots, t_{n,n} \) with sequences \((t_{j,n})_n\) such that \( t_{j+1,n} - t_{j,n} = \delta_n, j = 1, \ldots, n \). We give the stochastic deviation of kernel estimator both for fixed design (\( \delta_n \equiv \delta, \delta > 0 \)) and more accurate ones (\( \delta_n \to 0, n\delta_n \to \infty \)).

**Theorem 2.1**

(a) Set \( \delta_n \equiv \delta \) then under conditions A2.1(i)-(iv) and for \( h_n \) such that \( \frac{n h_n^d}{(\ln n)^3} \to \infty \), one obtains

\[
\limsup_{n \to +\infty} \sqrt{\frac{n h_n^d}{\ln n}} \left| \hat{f}_n(x) - E \hat{f}_n(x) \right| \leq 2^{3/2} f^{1/2}(x) \| K \|_2 \text{ a.s.}
\]

(b) if in addition condition A2.1(v) holds and \( \delta_n \to 0 \) such that \( \frac{n h_n^d}{(\ln n)^3} \to \infty \), the result is preserved as soon as \( \delta_n/\delta_n^*(\gamma_0) \to \infty \) where

\[
\delta_n^*(\gamma_0) = \begin{cases} 
\frac{h_n^d}{\ln(1/h_n)} & \text{if } \gamma_0 < 1, \\
\frac{h_n^d}{\ln(1/h_n)} & \text{if } \gamma_0 = 1, \\
\frac{h_n^d}{\ln(1/h_n)} & \text{if } \gamma_0 > 1.
\end{cases}
\]

We infer that results of Theorem 2.1(a) \( (\delta_n \equiv \delta > 0) \) are sharp : the rate and the constant are close to the i.i.d. case (see Stute, 1984). Part (b) shows that for high sampling rate \( \delta_n \to 0 \) one obtains the same result with \( \delta_n^*(\gamma_0) \) a non-decreasing function of \( \gamma_0 \). Note that for the limit case \( (\delta_n \sim \delta_n^*(\gamma_0)) \) the rate \( \sqrt{\frac{\ln n}{n h_n}} \) is preserved but with a more complicated asymptotic constant:

\[
2c^{-d/2} \left( 2 f(x) \| K \|_2^2 + d_{\gamma_0}(x) \right)^{1/2},
\]

where the \( d_{\gamma_0}(x) \) may be find along the proof of Theorem 2.1. Recall that in the case \( \gamma_0 < 1 \), Theorem 2.1(b) is already
derived in Bosq (1997), see also works by Leblanc (1995), Bosq and Cheze-Payaud (1999) for respective frameworks of wavelet and regression estimators. Finally for general $\gamma_0$, we refer to Blanke and Pumo (2003) where same sampling scheme is obtained for the mean-square error criterion.

Next lemma gives the asymptotic behaviour of the bias term.

**Lemma 2.1** Under Assumption 2.2, \[ \lim_{n \to +\infty} h_n^{-r_0} \left( E \hat{f}_n(x) - f(x) \right) = b_{r_0}(x) \]
with $b_{r_0}(x)$ defined by
\[
\sum_{j_1 + \ldots + j_d = r_0} \frac{1}{j_1! \ldots j_d!} \frac{\partial^{(r_0)} f}{\partial x_1^{j_1} \ldots \partial x_d^{j_d}}(x) \int_E u_1^{j_1} \ldots u_d^{j_d} K(u_1, \ldots, u_d) \, du_1 \ldots du_d.
\]

Collecting results of Theorem 2.1 and Lemma 2.1, we obtain the following corollary.

**Corollary 2.1** (a) Let $\delta_n \equiv \delta$ and $h_n = c \left( \frac{\ln n}{n} \right)^{1/(\gamma_0 + d)} (c > 0)$, then under conditions A2.1(i)-(iv) and Assumption 2.2, one obtains
\[
\limsup_{n \to \infty} \left( \frac{n}{\ln n} \right)^{r_0/\gamma_0 + d} \left| \hat{f}_n(x) - f(x) \right| \leq C_1 \quad a.s.
\]
with
\[
C_1 = 2^{3/2} c^{-d/2} f^{1/2}(x) \| K \|_2 + c^{r_0} |b_{r_0}(x)|
\]
for $b_{r_0}(x)$ given by (2.3).

(b) Now if $\delta_n \to 0$, $r_0 > \frac{d}{\max(1, \gamma_0)}$ and moreover condition A2.1(v) holds, the last result is preserved for all $\delta_n$ satisfying to $\delta_n/\delta_n^*(\gamma_0) \to \infty$ with same $\delta_n^*(\gamma_0)$ as in relations (2.2). Finally in the case $r_0 \leq \frac{d}{\max(1, \gamma_0)}$, the choice of $\delta_n$ such that $\delta_n = \left( \frac{\ln n}{n} \right)^{3/2} a_n$, $a_n \to \infty$, gives the same result.

Corollary 2.1 shows that one may expect same rates as in the i.i.d. case for small $\delta_n$ and an adequate choice of $h_n$. For practical choices of $h_n$ in the dependent case, we refer to further literature: Hart and Vieu (1990), Hall, Lahiri and Truong (1995), Kim (1997) and Sköld (2001b).

The rate obtained in Corollary 2.1(b) is sharp in the following senses: if $T_n$ represents the minimal time of experiment, $T_n = n\delta_n^*(\gamma_0)$, one may express the rate $(\ln n/n)^{r_0/[2r_0+d]}$ in terms of $T_n$. Thereby, for $\gamma_0 < 1$, $\gamma_0 = 1$ and $\gamma_0 > 1$ the corresponding rates are $\left( \frac{\ln T_n}{T_n} \right)^{1/2}$, $\frac{\ln T_n}{T_n^{2/3}}$ and $\left( \frac{\ln T_n}{T_n} \right)^{2r_0/[2r_0+d(\gamma_0-1)]}$. These are the almost sure rates obtained when the whole sample path is observed over
[0, T_n] (see Blanke, 2003). Moreover, up to a factor ‘ln n’ due to the pointwise framework, we recognize the optimal mean-square error rates obtained in Blanke and Bosq (2000) for density estimation in continuous time. For ‘irregular case’ γ_0 < 1, our results are similar to those given by Bosq (1997, 1998) with the same condition r_0 > d, coming from the constraint \[ \frac{n \delta^2 h^d}{(\ln n)^y} \rightarrow \infty \] in Theorem 2.1. Such restriction reduces with γ_0 until disappearing in the case δ_n ≡ δ (γ_0 → ∞).

A natural consequence of our results is that the total time of observation may be seen as a non-decreasing function of γ_0. In other words, more γ_0 is small more observations can be selected close to each other: this emphasizes the fact that irregular sample paths bring much more information to the statistician than more regular ones (when the correlation between two successive variables is stronger). In the borderline case δ_n ≃ δ^* (γ_0), Corollary 2.1 remains true but with a modified asymptotic constant.

### 2.3 Examples of processes satisfying Assumption 2.1

**Example 1** Case γ_0 = 1/2. Typically, a wide class of stationary real diffusion processes fulfill such condition, see Leblanc (1997) and Veretennikov (1999). Let us recall that a real continuous time diffusion process is defined as the solution of a stochastic differential equation: \[ dX_t = m(X_t) dt + \sigma(X_t) dW_t, \quad t \geq 0 \] (with standard notations). Now consider the class of Markov diffusions with transition density \( p_u(y, z) \) (in this case, \( f(x_0, x_u) (y, z) = f(y) p_u(y, z) \)). Leblanc (1997) gives conditions on m and σ to get strictly stationary and geometrically β-mixing solutions satisfying to the CL’s condition (namely with γ_0 = 1/2). One may also find in Veretennikov (1999) the following condition: if σ is not degenerate, σ and m bounded and of class \( C^\alpha \)-Hölder continuous then \( p_u(y, z) \leq C u^{-1/2}, \quad 0 < u \leq 1 \) (this bound holds true for the Ornstein-Uhlenbeck (O.U.) process too). So A2.1-(v) is fulfilled with γ_0 = 1/2. In the last cited paper, technical conditions upon m and σ are made to ensure that Assumption 2.1 is fully satisfied. Note also that Kutoyants (1997, 1999), Bosq and Davydov (1999) propose different estimators with parametric rate \( 1/T \) and not requiring directly CL’s condition.

**Example 2** Case 0 < γ_0 < 1. Using bivariate Levy processes, Barndorff-Nielsen and Pérez-Abreu (1999) give examples and properties of stationary and H-self-similar processes, \( 0 < H < 1 \), with one-dimensional marginal law of type G. Typically in the Gaussian case, one may show that Assumption 2.1 is fulfilled with γ_0 = H.

**Example 3** Case 0 < γ_0 ≤ 1. One may also exhibit a class of processes with γ_0 ≤ 1. Consider for example the real continuous-time fractional ARMA
processes studied in Viano, Deniau and Oppenheim (1994). These processes are based on the Brownian motion and defined by \( X_t = \int_{-\infty}^{t} f(t-s) \, dW(s), \) \( t \in \mathbb{R} \) where the impulse function \( f \) belongs to \( L^2(\mathbb{R}^+) \) and with Laplace transform \( F \) such that \( F(s) = \prod_{k=1}^{K} (s - a_k)^{d_k} \) for \( a_k, d_k \in \mathbb{C} \) and Re\((s) > a := \max(\text{Re}(a_k)/k \in E^*) \), \( E^* \) denoting the index set of singular points of \( F \). If \( D := \sum_{k=1}^{K} d_k < -1/2 \) and \( a < 0 \), \( \{X_t\} \) is a zero-mean stationary and strongly mixing Gaussian process. Furthermore it may be shown that : \( \gamma_0 = -D - 1/2 \) if \(-3/2 < D < -1/2 \) and \( \gamma_0 = 1 \) if \( D \leq -3/2 \) (in this last case, there exist an equivalent process such that almost all sample paths are of class \( C^1 \)).

**Example 4** Case \( \gamma_0 = 1 \). First this case should be associated with one-dimensional processes satisfying to (2.1) (see e.g. Blanke and Bosq 1997, Sköld and Hössjer, 1999) : typically real and mean-square differentiable Gaussian processes fulfill such condition. Combining the previous cases, one may also construct examples in a multidimensional setting. Let \( \{X_t^{(1)}\} \) and \( \{X_t^{(2)}\} \) be two real and independent processes with \( \gamma_1 = \gamma_2 = 1/2 \) : one obtains condition A2.1-(v) with \( \gamma_0 = \gamma_1 + \gamma_2 = 1 \).

**Example 5** Case \( \gamma_0 > 1 \). One has e.g. \( \gamma_0 = \sum_{i=1}^{d} \gamma_i \) for a \( d \)-multivariate process with independent components \( \{X_t^{(i)}, t \in \mathbb{R}\} i = 1, \ldots, d \) : this case will typically occur in dimensions \( d > 1 \) (Blanke and Bosq, 2000, Sköld, 2001a). Moreover one gets \( \gamma_0 = \frac{d}{2} \) \((d \geq 2)\) for an homogeneous \( d \)-dimensional diffusion process with bounded drift vector and constant diffusion coefficient, see Qian, Russo and Zheng (2003).

3 Adaptive procedure : \( \gamma_0 \) known but \( r_0 \) unknown

Our framework is the following : one has at his disposal some device to be calibrated during a learning period. Here we consider that \( \gamma_0 \) is known. This situation may occur in particular when the condition (2.1) is satisfied that is \( \gamma_0 = \sum_{i=1}^{d} \gamma_i \) with known \( 0 < \gamma_i \leq 1 \) (see examples in the previous section). Our goal is to construct an adaptive density estimator relatively to \( r_0 \) and converging on the smallest possible learning period.

We compute the adaptive selection of \( r_0 \) by testing different candidates in
\( \Delta_n = \{1, 2, \ldots, r_n\}. \) Then

\[
    r_0^* = \max \left\{ r_1 \in \Delta_n : \forall r_2 \in \Delta_n, r_2 \leq r_1, \left( \frac{n}{\ln n} \right)^{2r_2/d} |\tilde{f}_{r_2, \gamma_0}(x) - \tilde{f}_{r_1, \gamma_0}(x)| \leq \tilde{\eta}(r_2, \gamma_0) \right\} \quad (3.1)
\]

where \( \tilde{\eta}(r, \gamma_0) \) is a random quantity to be defined later and \( \tilde{f}_{r, \gamma_0} \) is given by

\[
    \tilde{f}_{r, \gamma_0}(x) := \frac{1}{n h_n^d(r)} \sum_{i=1}^{n} K(r) \left( \frac{x - X_i \delta_h(r, \gamma_0)}{h_n(r)} \right),
\]

with \( h_n(r) = c \left( \frac{\ln n}{r^2} \right)^{1/2} \), \( K(r) \) satisfying to \( C_r \) and \( \delta_h(r, \gamma) \) such that

\[
    \begin{cases}
    \delta & \text{if } \gamma = \gamma_\infty \\
    h_n^d(r) \ln_p(n) & \text{if } \gamma < 1, r > d \\
    h_n^d(r) \ln(1/h_n(r)) \ln_p(n) & \text{if } \gamma = 1, r > d \\
    h_n^d(r) \ln_p(n) & \text{if } \gamma > 1, r > d/\gamma \\
    \left( \frac{\ln n}{nh_n^2(r)} \right)^{1/2} \ln_p(n) & \text{if } r \leq d/\max(\gamma, 1).
    \end{cases}
\] \quad (3.2)

Note that \( \ln_p(n) = \ln \cdots \ln n, p \geq 2, \) is used to work with an easier asymptotic constant and \( \gamma_0 = \gamma_\infty \) represents the case \( \delta_n \equiv \delta \).

Finally, we propose the following adaptive estimator of \( f \) :

\[
    \tilde{f}_{r_0^*, \gamma_0}(x) = \frac{1}{nh_n^d(r_0^*)} \sum_{i=1}^{n} K(r_0^*) \left( \frac{x - X_i \delta_h(r_0^*, \gamma_0)}{h_n(r_0^*)} \right)
\]

with \( h_n(r_0^*) = c \left( \frac{\ln n}{r_0^*} \right)^{1/2} \delta_n \) and \( \delta_h(r_0^*, \gamma) \) given by (3.2). Note that by setting \( \gamma_0 = \gamma_\infty \) one obtains adaptive estimation of \( f \) relatively to \( r_0 \) in the case of the fixed design \( \delta_n \equiv \delta \).

In order to establish the asymptotic convergence of this estimator we need some auxiliary results. Here we introduce and study the quantity \( \tilde{\eta}(r, \gamma_0) \) intervening in the definition of \( r_0^* \) in (3.1):

\[
    \tilde{\eta}(r, \gamma) = a \left( 2^{3/2} c^{-d/2} \sqrt{\tilde{f}_{r, \gamma}(x)} \left\| K(r) \right\|_2 + c^r \left\| \tilde{h}_{r, \gamma}(x) \right\| \right) \quad (3.3)
\]
where $a$ is a positive real and $\tilde{f}_{r,\gamma_0}(x)$, $\tilde{b}_{r,\gamma_0}(x)$ are preliminary estimators for $f(x)$ and $b_r(x)$, see (2.3), respectively defined by

$$
\tilde{f}_{r,\gamma_0}(x) = \frac{1}{n h_n^d(r)} \sum_{i=1}^n K(r) \left( \frac{x - X_{ib}(r,\gamma_0)}{h_n(r)} \right)
$$

$$
\tilde{b}_{r,\gamma_0}(x) = \sum_{j_1 + \cdots + j_d = r} \frac{\tilde{f}_{j_1,\ldots,j_d;\gamma_0}(x)}{j_1! \cdots j_d!} \int u_1^{j_1} \cdots u_d^{j_d} K_r(u_1, \ldots, u_d) \, du_1 \cdots du_d
$$

with the same $h_n(r)$, $\delta_k(r, \gamma_0)$ as previously and

$$
\tilde{f}_{j_1,\ldots,j_d;\gamma_0}(x) = \frac{1}{n h_n^{d+r}(r)} \sum_{i=1}^n \frac{\partial^r \tilde{K}(r)}{\partial x_1^{j_1} \cdots \partial x_d^{j_d}} \left( \frac{x - X_{ib}(r,\gamma_0)}{h_n(r)} \right).
$$

Here $K(r) \in K_E^+$, the set of positive kernels $K$ belonging to $K_E$, and $\tilde{K}(r) \in K_r^*$:

$$
K_r^* = \left\{ \tilde{K} : \tilde{K} = \prod_{i=1}^d K_i, K_i \in K_E, E = [a_i, b_i], K_i \text{ continuous over } E, \right. \\
\left. \left\| K_i(r) \right\| \infty < \infty, \left\| K_i(r) \right\|_1 < \infty, \forall p = 0, \ldots, r - 1 : K_i^{(p)}(a_i) = K_i^{(p)}(b_i) = 0 \right\}
$$

where $K_i^{(p)}$ denotes the $p$-th derivative of $K_i$. Moreover both $K(r), \tilde{K}(r)$ fulfill conditions $C_K(r)$.

**Lemma 3.1** Under Assumption A2.1 and if condition A2.2(ii) is satisfied:

(i) $\left| \mathbb{E} \hat{f}_{(r,\gamma)}(x) - f(x) \right| \xrightarrow{n \to \infty} 0$; $\left| \mathbb{E} \hat{b}_{(r,\gamma)}(x) - b_r(x) \right| \xrightarrow{n \to \infty} 0$, $r_k \leq r_0$

(ii) $\forall \varepsilon > 0, \sum_n \sum_{r \in \Delta, r_k \leq r_0} P \left( \left| \hat{f}_{(r,\gamma)}(x) - \mathbb{E} \hat{f}_{(r,\gamma)}(x) \right| > \varepsilon \right) < \infty$,

(iii) $\forall \varepsilon > 0, \sum_n \sum_{r \in \Delta, r_k \leq r_0} P \left( \left| \hat{b}_{(r,\gamma)}(x) - \mathbb{E} \hat{b}_{(r,\gamma)}(x) \right| > \varepsilon \right) < \infty$.

for all $\gamma \geq \gamma_0$ and as soon as $r_n \to \infty$ with $r_n = \mathcal{O}\left( \frac{\ln n}{\ln \ln n} \right)^2$, $\nu_2 > 1$.

For all $\gamma \geq \gamma_0$, note that Lemma 3.1 implies in particular

$$
\tilde{\eta}(r_0, \gamma) \xrightarrow{n \to \infty} a C_1.
$$

(3.4)

Finally, we may state our main result:

**Theorem 3.1** If conditions of Lemma 3.1 are fulfilled and for $\tilde{\eta}(r_k, \gamma_0)$ defined by (3.3) with $a > 2$, one obtains

$$
\limsup_{n \to \infty} \left( \frac{n}{\ln n} \right)^{\frac{\nu_2}{1 + \nu_2}} \left| \hat{f}_{r_0,\gamma_0}(x) - f(x) \right| \leq (a + 1)C_1 \quad \text{a.s.}
$$
where \( C_1 \) is given by (2.4).

**Remarks**
- Here the construction of the adaptive estimator requires the sequence of observations \( X^n = (X_{i\delta_n(r,\gamma_0)}, \ i = 1, \ldots, n, \ ; r = O(r_n)) \). Since by Lemma 3.1, one has \( r_n = O(\ln n) \) we can deduce that the loss of rate is at most logarithmic with respect to an estimator using the whole data: it is not a surprise in this context of local pointwise adaptive results.
- Note that construction of \( r^*_0 \) implies that the maximal learning period is \( n\delta_n(r^*_0 + 1, \gamma_0) \). Actually we have no result of convergence for \( r^*_0 \) (only the property \( \sum P(r^*_0 < r_0) < \infty \) is established along the proof) but we infer that our adaptive estimator should converge for a time of experiment not too far from \( n\delta_n(r_0, \gamma_0) \). A further step should consist in making some numerical studies to confirm such behaviour.
- As noticed at the beginning of this section, the condition \( \gamma_0 \) known is not so restrictive in specific situations. Moreover Theorem 3.1 remains true if one works only with an upper bound for \( \gamma_0 \), say \( \gamma_0 \), since then \( \delta_n(\gamma_0) \geq \delta^*_n(\gamma_0) \), the minimal sampling rate. Note that for processes satisfying to (2.1) one has in a natural way \( \gamma_0 = d \) ! In the next section, we propose a procedure handling the case where the both parameters \( r_0, \gamma_0 \) are unknown.

## 4 Double adaptive procedure : \( r_0 \) and \( \gamma_0 \) unknown

Our framework is still some device to be calibrated during a learning period. Various sampling rates \( \delta_n \) can be adjusted (including the fixed one \( \delta_n \equiv \delta \)). The goal is to keep the minimal one, say \( \delta^*_n \), satisfying to \( \delta^*_n \geq \delta^*_n(\gamma_0) \) in order to reduce e.g. total time of future experiments.

First denote by \( \Gamma_n \), the grid of values tested for \( \gamma_0 \). Since for all \( \gamma_0 < 1 \), \( \delta^*_n(\gamma_0) \) takes the same value, there is no use to compute \( \gamma_0 \) exactly in this case. Moreover the fixed design \( \delta_n \equiv \delta \) is a borderline case of \( \delta_n(\gamma) = h^{d/\gamma} \) with \( \gamma \rightarrow \infty \). Consequently we build \( \Gamma_n \) with ordered values \( \gamma_{j,n} \) and limit points \( \gamma_0, \gamma_\infty \) : \( \Gamma_n = \{ \gamma_0, \gamma_{1,n}, \gamma_{2,n}, \ldots, \gamma_{N_n,n}, \gamma_\infty \} \) where \( 0 < \gamma_0 < 1, \gamma_{1,n} = 1, \gamma_{j+1,n} - \gamma_{j,n} = \tau_{j,n}, j = 1, \ldots, N_n - 1 \). Here and in all the following, the sequences \( (\tau_{j,n}), (N_n) \) will be supposed to satisfy :

\[
\begin{align*}
\sum_{j=1}^{N_n-1} \tau_{j,n} &= o(\ln n), \\
\sup_{j=1, \ldots, N_n} \tau_{j,n} &\leq \tau \\
\text{and } \tau_{j,n} &\geq \frac{(\ln \ln n)^{\nu_1}}{\ln n}, \quad (\nu_1 > 1), \quad j = 1, \ldots, N_n - 1.
\end{align*}
\]

These last conditions ensure that one can discriminate well between values of \( \delta_n(\gamma) \) for different candidates \( \gamma \) in \( \Gamma_n \). In other words, one has \( \frac{\delta_n(\gamma_{j+1,n})}{\delta_n(\gamma_{j,n})} \rightarrow 0 \)
for \( j = 1, \ldots, N_n - 1 \) and \( \delta_n(\gamma_{N_n,n}) \to 0 \) as \( n \to \infty \).

The procedure is in two steps: first we compute the adaptive selection of \( r_0 \) when the sampling rate \( \delta_n \) is fixed \( (\delta_n \equiv \delta) \) and next we select a candidate for \( \gamma_0 \) in \( \Gamma_n \) with the help of \( r_0^* \) defined in (3.1) (with \( \gamma_0 \) replaced by \( \gamma_\infty \)). We get

\[
r_0^* = \max \left\{ r_1 \in \Delta_n : \forall r_2 \in \Delta_n, r_2 \leq r_1, \right. \\
\left. \left( \frac{n}{\ln n} \right)^{\frac{r_2}{2r_2 + d}} \left| \hat{f}_{r_2, \gamma_\infty}(x) - \hat{f}_{r_1, \gamma_\infty}(x) \right| \leq \hat{\eta}(r_2, \gamma_\infty) \right\} \tag{4.2}
\]

and

\[
\gamma_0^* = \min \left\{ \gamma_1 \in \Gamma_n : \forall \gamma_2 \in \Gamma_n, \gamma_2 \geq \gamma_1, \right. \\
\left. \left( \frac{n}{\ln n} \right)^{\frac{r_0^*}{2r_0^* + d}} \left| \hat{f}_{r_0^*, \gamma_2}(x) - \hat{f}_{r_0^*, \gamma_1}(x) \right| \leq \zeta \right\} \tag{4.3}
\]

where \( \zeta \) denotes any positive constant, \( \hat{\eta}(r, \gamma_\infty) \) is given by (3.3) and \( \hat{f}(r, \gamma) \) by

\[
\hat{f}_{r, \gamma}(x) := \frac{1}{nh_n^d(r)} \sum_{i=1}^n K(r) \left( \frac{x - X_i\delta_h(r, \gamma)}{h_n(r)} \right),
\]

Now we propose the following doubly adaptive estimator of \( f \) :

\[
\hat{f}_{r_0^*, \gamma_0^*}(x) = \frac{1}{nh_n^d(r_0^*)} \sum_{i=1}^n K(r_0^*) \left( \frac{x - X_i\delta_h(r_0^*, \gamma_0^*)}{h_n(r_0^*)} \right)
\]

with \( h_n(r_0^*) = c \left( \frac{\ln n}{n} \right)^{\frac{1}{2r_0^* + d}} \) and \( \delta_h(r_0^*, \gamma_0^*) \) as in (3.2). Finally, we may state our main result:

**Theorem 4.1** If conditions of Lemma 3.1 are fulfilled and for \( \hat{\eta}(r, \gamma_\infty) \) defined by (3.3) with \( a > 2 \), one obtains

\[
\limsup_{n \to \infty} \left( \frac{n}{\ln n} \right)^{\frac{r_0^*}{2r_0^* + d}} \left| \hat{f}_{r_0^*, \gamma_0^*}(x) - f(x) \right| \leq (a + 1)C_1 + \zeta \quad \text{a.s.}
\]

where \( C_1, \zeta \) are respectively given by (2.4), (4.3).

**Remarks**

- Here the construction of the adaptive estimator requires the sequence of observations \( X^n = \left( X_{i\delta_h(r, \gamma_0)}, i = 1, \ldots, n; s = 0, 1, \ldots, N_n + 1, r = 1, \ldots, r_n \right) \).

Since by conditions (4.1) and Lemma 3.1, one has \( N_n = o((\ln n)^2), r_n = \)
any positive Since the two terms should be equivalently handled, it is sufficient to study using the whole $N_n \times r_n \times n$ data. But note that such an estimator could be rather inconsistent since it involves variables with high local correlation.

- It is noteworthy that Theorem 4.1 is true for all sequences $(\tau_{j,n})$, $N_n$ satisfying to (4.1) : e.g. $\tau_{j,n} \equiv \tau$, $N_n \equiv N$ are valid choices. This means that ‘rough’ grids $\Gamma_n$ (but always including $\gamma_\infty$) are allowed : by this way the numerical implementation of $\gamma_0^*$ can be quite fast by testing only privileged $\gamma_i$. Naturally, approximation of $\gamma_0$ will be obviously better if one considers sharper grids!

- The double adaptive procedure is only well adapted to the case where one has a device to be calibrated during a learning period. Thus one obtains with $(r_0^*, \gamma_0^*)$ an indication on regularity of sample paths and density at the same time. We do not have for the moment results relating to the validation of the estimator after the learning period. Indeed our work should be seen as a pioneer one in this field. Further investigation to better understanding the behaviour of our estimators, will consist to study them outside the learning period, both in a theoretical and numerical way.

5 Proofs

5.1 Proof of Theorem 2.1 (sketch)

This proof involves the popular method of blocks and coupling : we outline here the main parts. For $p_n$, $q_n$ such that $2p_nq_n = n$, let us set $V_n(j) = \int_{(j-1)q_n}^{jq_n} Z_t \, dt$, $j = 1, \ldots, 2[p_n]$ and $V_n(2[p_n] + 1) = \int_{2p_nq_n}^{\infty} Z_t \, dt$ with $Z_t = K_h \left( x - X_{[t+1]\delta_n} \right) - E K_h \left( x - X_{[t+1]\delta_n} \right)$, $K_h(\cdot) := \frac{1}{n^2} K \left( \frac{\cdot}{n} \right)$.

One may write $|\hat{f}_n(x) - E \hat{f}_n(x)| = \frac{1}{n} \left| \sum_{j=1}^{[p_n]+1} V_n(2j-1) + \sum_{j=1}^{[p_n]} V_n(2j) \right|$ and

$$P \left( |\hat{f}_n(x) - E \hat{f}_n(x)| > \varepsilon \right) \leq P \left( \left| \sum_{j=1}^{[p_n]+1} V_n(2j-1) \right| > \frac{n\varepsilon}{2} \right) + P \left( \left| \sum_{j=1}^{[p_n]} V_n(2j) \right| > \frac{n\varepsilon}{2} \right).$$

Since the two terms should be equivalently handled, it is sufficient to study one of them. We use Rio’s (2000) coupling result recursively to approximate $V_n(1), \ldots, V_n(2[p_n]+1)$ by independent variables $V_n^*(1), \ldots, V_n^*(2[p_n]+1)$. For any positive $\kappa$ (introduced to reduce the asymptotic constant), we have

$$P \left( \left| \sum_{j=1}^{[p_n]+1} V_n(2j-1) \right| > \frac{n\varepsilon}{2} \right) \leq P \left( \left| \sum_{j=1}^{[p_n]+1} V_n^*(2j-1) \right| > \frac{n\varepsilon}{2(1+\kappa)} \right)$$

$$+ P \left( \left| \sum_{j=1}^{[p_n]} V_n(2j-1) - V_n^*(2j-1) \right| > \frac{n\varepsilon\kappa}{2(1+\kappa)} \right) =: I_1 + I_2.$$
First,
\[
\begin{cases}
|V_n(j)| \leq 2 \frac{q_n}{h_n^2} \|K\|_{\infty} & \text{for } j = 1, \ldots, 2[p_n] \\
|V_n(2[p_n] + 1)| \leq 4 \frac{q_n}{h_n^2} \|K\|_{\infty}
\end{cases}
\]  
(5.2)

Since the $V_n^*(2j - 1)$ are independent, Bernstein’s inequality (McDiarmid, 1998) implies $I_1 \leq 2 \exp \left(- \frac{\sum_{j=1}^{[p_n]} 1 + E(V_n^2(2j - 1)) \|K\|_{\infty}^2}{8(1+\kappa)^2} \right)$. Now a similar calculation to Blanke and Pumo (2003) gives following bounds for $E(V_n^2(2j - 1))$, $j = 1, \ldots, [p_n]$:

- If $\delta_n \equiv \delta$, Cauchy-Schwarz’s inequality and conditions A2.1(i),(ii),(iv) imply

\[
EV_n^2(2j - 1) \leq 2f(x) \|K\|_2^2 \frac{q_n}{h_n^2} \left(1 + o(1)\right)
\]  
(5.3)

- For $\delta_n \rightarrow 0$, using in particular Bochner’s lemma together with condition A2.1(v), one obtains

\[
EV_n^2(2j - 1) \leq \frac{q_n}{h_n^2} \left(2f(x) \|K\|_2^2 + d_{\gamma_0}(x)\right) \left(1 + o(1)\right)
\]  
(5.4)

for all $\delta_n \geq \delta_n^*(\gamma_0)$ with $\delta_n^*(\gamma_0)$ given by (2.2). Moreover $d_{\gamma_0}(x) \equiv 0$ for all $\delta_n$ such that $\delta_n/\delta_n^*(\gamma_0) \rightarrow \infty$ whereas if $\delta_n = \delta_n^*(\gamma_0)$, $d_{\gamma_0}(x)$ is given by

- for $\gamma_0 < 1$,

\[
d_{\gamma_0}(x) = 2 \|K\|_1^2 \left(\frac{M(x,x)u_1^{1-\gamma_0}}{1-\gamma_0} + u_0 \|f\|_2^2 + (u_1 - u_0) \|\pi\|_\infty + \int u_1 \pi(u) du\right),
\]

- for $\gamma_0 = 1$, $d_{\gamma_0}(x) = 2dM(x,x) \|K\|_1^2$,

- for $\gamma_0 > 1$, $d_{\gamma_0}(x) = \frac{2(2\gamma_0-1)M(x,x)\|K\|_1^2}{(\gamma_0-1)}$.

Let us turn to evaluation of term $I_2$ defined in relation (5.1). By Rio’s result:

\[
EV_n^*(2j - 1) - V_n(2j - 1) \leq 4\|V_n(2j - 1)\|_{\infty} \left(\sup |P(AB) - P(A)P(B)|\right)
\]  
(5.5)

where supremum is taken over all sets $A$ and $B$ belonging to $\sigma$-algebras of events generated by respectively $\{V_n(1), \ldots, V_n(2j - 3)\}$ and $V_n(2j - 1)$. Now with Markov’s inequality, relations (5.2),(5.5) and condition A2.1-(iii) we get

\[
I_2 \leq \frac{4^{(\kappa+1)} \|K\|_{\infty} \frac{1}{\epsilon h_n}}{\kappa e^{\frac{q_n}{h_n} \ln(1/\rho)}} \left(1 + o(1)\right).
\]
Now, collecting all these results, one obtains for any $\delta_n \geq \delta^*_n(\gamma_0)$,

$$
P \left( \left| \hat{f}_n(x) - E \hat{f}_n(x) \right| > \varepsilon_n \right) \leq \frac{16(k+1)}{\kappa} \| K \|_\infty e^{-q_n \delta_n \ln(1/\rho)} \left( 1 + o(1) \right)
$$

$$+ 4 \exp \left( - \frac{n h_n^d \varepsilon_n^2}{4(1+\kappa)^2 \left( 2 f(x) \| K \|_2^2 + d \gamma_0(x) \right) (1 + o(1)) + \frac{16}{3} \| K \|_\infty q_n \delta_n \ln(1/\rho)} \right).
$$

The choices $\varepsilon_n = \eta \left( \frac{\ln n}{nh_n} \right)^{1/2}$, $q_n = M \delta_n^{-1} \ln n$ imply for any $\delta_n \geq \delta^*_n(\gamma_0)$

$$
P \left( \sqrt{nh_n} \left| \hat{f}_n(x) - E \hat{f}_n(x) \right| > \eta \right) \leq \frac{8(k+1)}{\eta \kappa} \| K \|_\infty \frac{n^{\frac{1}{2}} h_n^{-\frac{3}{2}} \varepsilon_n^2}{(\ln n)^{\frac{5}{2}}} e^{M \ln \rho} \left( 1 + o(1) \right)
$$

$$+ 4 \exp \left( - \frac{\eta^2 \ln n}{4(1+\kappa)^2 \left( 2 f(x) \| K \|_2^2 + d \gamma_0(x) \right) (1 + o(1)) + \frac{16(1+\kappa) \eta \| K \|_\infty \ln(1/\rho)}{3n^{1/2} h_n^d \delta_n (\ln n)^{-3/2}} \right).
$$

Finally Borel-Cantelli’s lemma leads to the desired result.

5.2 Proof of Lemma 2.1 and Corollary 2.1

Straightforward from Taylor’s Formula, Theorem 2.1 and Lemma 2.1.

5.3 Proof of Lemma 3.1

(i) The bias is independent of $\gamma_0$ and moreover $f$ is $r_0$-times differentiable, so one easily obtains that for all $r_k \leq r_0$ and $\gamma > 0$,

$$h_n^{-r_k} (r_k) \left( E \tilde{f}_{r_k, \gamma} (x) - f(x) \right)
$$

$$\xrightarrow{n \to \infty} \sum_{j_1 + \cdots + j_d = r_k} \frac{1}{j_1! \cdots j_d!} \frac{\partial^{(r_k)} f(x)}{\partial x_1^{j_1} \cdots \partial x_d^{j_d}} \int K_{(r_k)} (z) z_1^{j_1} \cdots z_d^{j_d} \, dz.
$$

Moreover for any combination of $j_1, \ldots, j_d$ with $j_1 + \cdots + j_d = r_k$ one has

$$E \tilde{f}_{j_1, \ldots, j_d, \gamma}^{(r_k)} (x) = \frac{1}{h_n^{r_k} (r_k)} \int \frac{\partial^{r_k} K_{(r_k)}}{\partial x_1^{j_1} \cdots \partial x_d^{j_d}} (z) f(x - h_n(r_k)z) \, dz.
$$
Using properties of $\mathcal{K}(r_k)$, we proceed to $r_k$ successive integrations by parts to get the following result:

$$E \frac{\partial^{r_k} f}{\partial x_1^{r_k} \cdots x_d^{r_k}}(x) = \int \mathcal{K}(r_k)(z) A_{r_k, h_n(r_k)}(x, z) \, dz$$

where we have set $A_{r_k, h_n(r_k)}(x, z) = \frac{\partial^{r_k} f}{\partial x_1^{r_k} \cdots x_d^{r_k}}(x - h_n(r_k)z) - \frac{\partial^{r_k} f}{\partial x_1^{r_k} \cdots x_d^{r_k}}(x)$.

First, if one has $r_k = r_0$ the condition A2.2(i) yields

$$E f_{j_1, \ldots, j_d; \gamma}(x) - \frac{\partial^{r_k} f}{\partial x_1^{r_k} \cdots x_d^{r_k}}(x) \to 0.$$

On the other hand, for $r_k < r_0$ one may apply Taylor’s formula up to order $r_0 - r_k$ to get $h_n^{-1}(r_k) \left| E f_{j_1, \ldots, j_d; \gamma}(x) - \frac{\partial^{r_k} f}{\partial x_1^{r_k} \cdots x_d^{r_k}}(x) \right| = O(1)$ using conditions $C_f(r_0)$ and $C_K(r_k)$.

(ii) The proof is similar to the proof of Theorem 2.1. We only indicate here the main changes. Notations are the same excepting $h_n$ replaced by $h_n(r_k)$, $\delta_n$ by $\delta_n(r_k, \gamma)$, $\gamma \geq \gamma_0$, so $Z_t$ is defined by $Z_t = \frac{1}{h^d_n(r_k) K(r_k)} \left( x - \frac{X_t(1+\delta_n(r_k, \gamma))}{h_n(r_k)} \right) - E \frac{1}{h^d_n(r_k) K(r_k)} \left( x - \frac{X_t(1+\delta_n(r_k, \gamma))}{h_n(r_k)} \right)$. Considering (5.3)-(5.6), one gets for all $\gamma \geq \gamma_0$

$$\sum_{r_k \in \Delta_n, r_k \leq r_0} P \left( \left| \tilde{f}_{r_k, \gamma}(x) - E \tilde{f}_{r_k, \gamma}(x) \right| > \varepsilon \right) \leq \text{Card} \left( \Delta_n \right) \cdot 4 \exp \left( - \frac{M \ln n}{\delta_n(r_k, \gamma)} \varepsilon \right)$$

with the choice $q_n = M \frac{\ln n}{\delta_n(r_k, \gamma)}$ and for positive constants $c_i$. For any $\varepsilon > 0$, $r_k \leq r_0$ and $\gamma \geq \gamma_0$, relation (3.2) implies $\frac{(\ln n)^2}{nh^d_n(r_k) \delta_n(r_k, \gamma)} \to 0$, so the series is convergent (with $M$ large enough).

(iii) Similar as above since $\frac{(\ln n)^2}{nh^d_n(r_k) \delta_n(r_k, \gamma)} \to 0$ for all $r_k \leq r_0, \gamma \geq \gamma_0$.

5.4 Proof of Theorem 3.1

We make use of the decomposition

$$\left( \frac{n}{\ln n} \right)^{r_0 \over 2r_0 + d} \left| \tilde{f}(r_0, \gamma_0)(x) - f(x) \right| \leq \left( \frac{n}{\ln n} \right)^{r_0 \over 2r_0 + d} \left| \tilde{f}(r_0, \gamma_0)(x) - \hat{f}(r_0, \gamma_0)(x) \right| + \left( \frac{n}{\ln n} \right)^{r_0 \over 2r_0 + d} \left| \hat{f}(r_0, \gamma_0)(x) - f(x) \right|$$

$$=: A + B$$

16
• Study of term A. Let us decompose $A$ into $A_1 + A_2$ with

$$A_1 := \left(\frac{n}{\ln n}\right)^{r_0/(2r+d)} \left| \tilde{f}(r_0, r_0) - \tilde{f}(r_0, r_0) \right| I_{\{r_0 < r_0\}}$$

and

$$A_2 := \left(\frac{n}{\ln n}\right)^{r_0/(2r+d)} \left| \tilde{f}(r_0, r_0) - \tilde{f}(r_0, r_0) \right| I_{\{r_0 \geq r_0\}}.$$  

We first prove that $\forall \varepsilon > 0, \sum_n P(A_1 > \varepsilon) < \infty$.

We have $P(A_1 > \varepsilon) \leq P(r_0 < r_0) \leq A_{10} + A_{11} + A_{12} + A_{13} + A_{14}$ with

$$A_{10} = \sum_{r_k < r_0} P(\tilde{b}(r_k, r_0) \leq D_1(r_k))$$

$$A_{11} = \sum_{r_k < r_0} P\left(\left(\frac{n}{\ln n}\right)^{r_k/(2r+d)} |E \tilde{f}(r_k, r_0) - f(x) - \frac{D_1(r_k)}{2(1 + \kappa)}\right) > \frac{D_1(r_k)}{2(1 + \kappa)}\right)$$

$$A_{12} = \sum_{r_k < r_0} P\left(\left(\frac{n}{\ln n}\right)^{r_k/(2r+d)} |E \tilde{f}(r_0, r_0) - f(x) - \frac{D_1(r_k)}{2(1 + \kappa)}\right) > \frac{D_1(r_k)}{2(1 + \kappa)}\right)$$

$$A_{13} = \sum_{r_k < r_0} P\left(\left(\frac{n}{\ln n}\right)^{r_k/(2r+d)} |\tilde{f}(r_k, r_0) - E \tilde{f}(r_k, r_0) - \frac{D_1(r_k)}{2(1 + \kappa)}\right) > \frac{D_1(r_k)}{2(1 + \kappa)}\right)$$

$$A_{14} = \sum_{r_k < r_0} P\left(\left(\frac{n}{\ln n}\right)^{r_k/(2r+d)} |\tilde{f}(r_0, r_0) - E \tilde{f}(r_0, r_0) - \frac{D_1(r_k)}{2(1 + \kappa)}\right) > \frac{D_1(r_k)}{2(1 + \kappa)}\right)$$

for a positive constant $D_1(r_k)$ defined by $D_1(r_k) = \xi_1 + 2(1 + \kappa)\left(c^{r_k} |b_{r_k}(x)| + 2^{3/2}c^{-d/2} f^{1/2}(x) \left\| K(r_k) \right\|_2\right)$ with a positive $\xi_1$. Now using classical inequalities $|\sqrt{u_1} - \sqrt{u_2}| \leq \sqrt{|u_1 - u_2|}$ and $|u_1| - |u_2| \leq |u_1 - u_2| \leq |u_1| + |u_2|$, one may bound $A_{10}$ by

$$\sum_{r_k < r_0} P \left(\left| E\bar{b}(r_k, r_0) - \bar{b}(r_k, r_0) \right| \geq \frac{E_1}{2a\max(c^{r_k}, c)} \right)$$

$$+ P \left(\left| E\tilde{f}(r_k, r_0) - \tilde{f}(r_k, r_0) \right| \geq \frac{E_1^2}{32a^2c^{-d}\left\| K(r_k) \right\|_2^2} \right)$$

where $E_1 = -D_1(r_k) + a\left(c^{r_k} \left| E\tilde{b}(r_k, r_0) \right| + 2^{3/2}c^{-d/2}\left\| K(r_k) \right\|_2 \sqrt{E\tilde{f}(r_k, r_0)}\right)$. Now since $r_k < r_0$, we get $E\tilde{b}(r_k, r_0) \rightarrow b_1(x)$, $E\tilde{f}(r_k, r_0) \rightarrow f(x)$ by Lemma 3.1 and moreover $a > 2$ implies $E_1 > 0$. So for all $n \geq N$, with a fixed non random $N$, and for some positive $\tilde{\varepsilon}$,

$$\sum_n A_{10} \leq 2N + \sum_{n \geq N} \sum_{r_k < r_0} P\left(\left| E\bar{b}(r_k, r_0) - \bar{b}(r_k, r_0) \right| > \tilde{\varepsilon}\right)$$

$$+ \sum_{n \geq N} \sum_{r_k < r_0} P\left(\left| E\tilde{f}(r_k, r_0) - \tilde{f}(r_k, r_0) \right| > \tilde{\varepsilon}\right) < \infty.$$
and \( r_k < r_0 \), one has \( \left( \frac{n}{\ln n} \right)^{\frac{r_0}{2r_k + d}} |E \hat{f}_{(r_k, \gamma_0)}(x) - f(x)| \xrightarrow{n \to \infty} c^{r_k} |b_{r_k}(x)| \). So we may deduce that for each \( r_k \), the term \( A_{11} \) will be identically null for sufficiently large \( n \) (and so \( \sum_n A_{11} < \infty \)) since by definition \( D_1(r_k) > 2(1 + \kappa)c^{r_k} |b_{r_k}(x)| \).

By Lemma 2.1, we get for \( A_{12} : h_n^{-r_0}(r_0) |E \hat{f}_{(r_0, \gamma_0)}(x) - f(x)| \to c^{r_0} b_{r_0}(x) \) and \( r_k < r_0 \) implies \( \left( \frac{n}{\ln n} \right)^{\frac{r_k}{2r_k + d}} \times h_n^{r_0}(r_0) \to 0 \), so \( \sum_n A_{12} < \infty \).

In a similar way,
\[
A_{13} \leq \text{Card}(\Delta_n) \max_{r_k < r_0} P \left( \left( \frac{n}{\ln n} \right)^{\frac{r_0}{2r_k + d}} |\hat{f}_{(r_k, \gamma_0)}(x) - E \hat{f}_{(r_k, \gamma_0)}(x)| > \frac{D_1(r_k)}{2(1 + \kappa)} \right).
\]

Relations (3.2) and (5.7) imply \( \sum_n A_{13} < +\infty \), since \( D_1(r_k) > 2(1 + \kappa) \left( 2^{3/2} c^{-d/2} f^{1/2}(x) \| K(r_k) \|_2 \right) \).

Finally for term \( A_{14} \), the same conclusion holds : \( \sum_n A_{14} < \infty \) because of
\[
\left( \frac{n}{\ln n} \right)^{\frac{r_0}{2r_0 + d}} \left( \frac{n}{\ln n} \right)^{-r_0} \to 0 \quad \text{for all } r_k < r_0.
\]

Now considering \( A_2 \), we just use the definition of \( r^*_0 \) to get
\[
\{ r^*_0 \geq r_0 \} \implies \left( \frac{n}{\ln n} \right)^{\frac{r_0}{2r_0 + d}} \left| \hat{f}_{(r^*_0, \gamma_0)}(x) - \hat{f}_{(r_0, \gamma_0)}(x) \right| \leq \eta(r_0, \gamma_0) \}
\]
so with the help of (3.4), \( \limsup_{n \to \infty} A_2 \leq a \quad \text{C}_1 \text{ a.s., } \quad \text{C}_1 \text{ given by (2.4).} \)

**Study of term B.**

By Corollary 2.1, we easily obtain that \( \limsup_{n \to \infty} B \leq C_1 \text{ a.s.} \).

Collecting all these results we finally arrive at
\[
\limsup_{n \to \infty} \left( \frac{n}{\ln n} \right)^{\frac{r_0}{2r_0 + d}} |\hat{f}_{(r^*_0, \gamma_0)}(x) - f(x)| \leq (a + 1) C_1 \quad \text{a.s.} \]

5.5 Proof of Theorem 4.1

We make use of the decomposition
\[
\left( \frac{n}{\ln n} \right)^{\frac{r_0}{2r_0 + d}} |\hat{f}_{(r^*_0, \gamma_0)}(x) - f(x)| \leq \left( \frac{n}{\ln n} \right)^{\frac{r_0}{2r_0 + d}} \left( |\hat{f}_{(r^*_0, \gamma_0)}(x) - \hat{f}_{(r_0, \gamma_0)}(x)| \right.
\]
\[
+ |\hat{f}_{(r_0, \gamma_0)}(x) - \hat{f}_{(r_0, \gamma_0)}(x)| + |\hat{f}_{(r_0, \gamma_0)}(x) - f(x)|
\]
\[
=: A + B + C.
\]
• **Study of term A.** Let us decompose $A$ into $A_1 + A_2$ with

$$A_1 := \left( \frac{n}{\ln n} \right)^{r_0} \left| \frac{\ln n}{\ln n} \right|^d \left| f(r_0, \gamma_0)(x) - \hat{f}(r_0, \gamma_0)(x) \right| I \{r_0 < r_0\} \quad \text{and}$$

$$A_2 := \left( \frac{n}{\ln n} \right)^{r_0} \left| \frac{\ln n}{\ln n} \right|^d \left| f(r_0, \gamma_0)(x) - \hat{f}(r_0, \gamma_0)(x) \right| I \{r_0 \geq r_0\}.$$

First, $P(A_1 > \varepsilon) \leq P(r_0^* < r_0) < \infty$ (same proof as Theorem 3.1 with $\gamma_0 = \gamma_\infty$).

Now for $A_2$, since $\gamma_0^* \leq \gamma_\infty$ one has

$$\left( \frac{n}{\ln n} \right)^{r_0} \left| \frac{\ln n}{\ln n} \right|^d \left| f(r_0, \gamma_\infty)(x) - \hat{f}(r_0, \gamma_\infty)(x) \right| \leq \zeta$$

and $r_0^* \geq r_0$ implies in turn

$$\limsup_{n \to \infty} \left( \frac{n}{\ln n} \right)^{r_0} \left| \frac{\ln n}{\ln n} \right|^d \left| f(r_0, \gamma_\infty)(x) - \hat{f}(r_0, \gamma_\infty)(x) \right| I \{r_0 \geq r_0\} \leq \zeta.$$

**Study of term B.**

We split $B$ into $B_1 := \left( \frac{n}{\ln n} \right)^{r_0} \left| \frac{\ln n}{\ln n} \right|^d \left| f(r_0, \gamma_\infty)(x) - \hat{f}(r_0, \gamma_\infty)(x) \right| I \{r_0 < r_0\}$

and $B_2 := \left( \frac{n}{\ln n} \right)^{r_0} \left| \frac{\ln n}{\ln n} \right|^d \left| f(r_0, \gamma_\infty)(x) - \hat{f}(r_0, \gamma_\infty)(x) \right| I \{r_0 \geq r_0\}$.

For $B_1$, we just use the definition of $r_0^*$ to get

$$\{r_0^* \geq r_0\} \implies \left\{ \left( \frac{n}{\ln n} \right)^{r_0} \left| \frac{\ln n}{\ln n} \right|^d \left| f(r_0, \gamma_\infty)(x) - \hat{f}(r_0, \gamma_\infty)(x) \right| \leq \eta(r_0, \gamma_\infty) \right\}$$

so with the help of (3.4), $\limsup_{n \to \infty} B_1 \leq a C_1$ a.s., $C_1$ given by (2.4). Term $B_2$ is negligible since for all positive $\varepsilon$, $P(B_2 > \varepsilon) \leq P(r_0^* < r_0)$ and similarly to $A_1$, $\sum P(B_2 > \varepsilon) < \infty$, so that $\limsup_{n \to \infty} B \leq a C_1$ a.s. 

**Study of term C.**

By Corollary 2.1(a), we easily obtain that $\limsup_{n \to \infty} C \leq C_1$ a.s. 

Collecting all these results we finally arrive at

$$\limsup_{n \to \infty} \left( \frac{n}{\ln n} \right)^{r_0} \left| \frac{\ln n}{\ln n} \right|^d \left| f(r_0, \gamma_0^*)(x) - f(x) \right| \leq (a + 1) C_1 + \zeta \quad \text{a.s.}$$

**References**


Comte, F. and Merlevède, F.: 2002b, Density estimation for a class of continuous time or discretely observed processes. Submitted.


Kutoyants, Y. A.: 1999, Efficiency density estimation for ergodic diffusion process, 

Springer Series In Statistics, Springer.

Leblanc, F.: 1995, Discretized wavelet density estimators for continuous time 

Leblanc, F.: 1997, Density estimation for a class of continuous time processes, 

to inhomogeneous smoothness: an approach based on kernel estimates with 

Lepskii, O. V.: 1990, On a problem of adaptive estimation in Gaussian white noise, 


Qian, Z., Russo, F. and Zheng, W.: 2003, Comparison theorem and estimates for 
transition probability densities of diffusion processes, *Probab. Theory Related 

Mathématiques & Applications, (31), Springer-Verlag.

Sköld, M.: 2001a, The asymptotic variance of the continuous-time kernel estimator 

Sköld, M.: 2001b, A bias correction for cross-validation bandwidth selection when a 

Sköld, M. and Hössjer, O.: 1999, On the asymptotic variance of the continuous-time 


Veretennikov, A. Y.: 1999, On Castellana-Leadbetter’s condition for diffusion 

Viano, M. C., Deniau, C. and Oppenheim, G.: 1994, Continuous-time fractional 

Wu, B.: 1997, Kernel density estimation under weak dependence with sampled data, 