Asymptotics for function derivatives estimators based on stationary and ergodic discrete time processes

Salim Bouzebda[1] · Mohamed Chaouch[2] · Sultana Didi Biha[3]

Abstract
The main purpose of the present work is to investigate kernel-type estimate of a class of function derivatives including parameters such as the density, the conditional cumulative distribution function and the regression function. The uniform strong convergence rate is obtained for the proposed estimates and the central limit theorem is established under mild conditions. Moreover, we study the asymptotic mean integrated square error of kernel derivative estimator which plays a fundamental role in the characterization of the optimal bandwidth. The obtained results in this paper are established under a general setting of discrete time stationary and ergodic processes. A simulation study is performed to assess the performance of the estimate of the derivatives of the density function as well as the regression function under the framework of a discretized stochastic processes. An application to financial asset prices is also considered for illustration.

Keywords Nonparametric estimation · Density estimation · Regression estimation · Kernel estimation · Function derivative · Ergodic discrete time processes · Martingale differences

✉ Salim Bouzebda
    salim.bouzebda@utc.fr

    Mohamed Chaouch
    mchaouch@qu.edu.qa

    Sultana Didi Biha
    s.biha@qu.edu.sa

[1] Université de Technologie de Compiègne, L.M.A.C., Sorbonne Universités, 57 Avenue de Landshut CS 60319, 60203 Compiegne Cedex, France

[2] Program of Statistics, Department of Mathematics, Statistics and Physics, College of Arts and Sciences, Qatar University, P.O. Box 2713, Doha, Qatar

[3] College of Sciences, Qassim University, PO Box 6688, 51452 Buraydah, Saudi Arabia

1 Introduction

Over years ago, Parzen (1962) studied some properties of kernel density estimators introduced by Akaike (1954) and Rosenblatt (1956). Since then, nonparametric estimation of the density and regression functions received an intense investigation by statisticians for several years and a large variety of estimation methods were developed. Kernel-based nonparametric function estimation methods have received the interest of the statistics community and several theoretical results along with interesting applications, including economics, finance, biology and environmental science, were considered in the literature. For an exhaustive discussion on the topic the reader can be referred to the following pioneer papers Tapia and Thompson (1978), Wertz (1978), Devroye and Györfi (1985), Devroye (1987), Nadaraya (1989), Härdle (1990), Wand and Jones (1995), Eggermont and LaRiccia (2001), Devroye and Lugosi (2001) and the references therein.

The estimation of function derivatives is a versatile tool in statistical data analysis. For instance, Genovese et al. (2013) introduced a test statistics for the modes of a density based on the second order density derivative. Noh et al. (2018) showed that the optimal bandwidth of kernel density estimation depends on the second-order density derivative. Moreover, as discussed in Silverman (1986) and Wand and Jones (1995), the optimal choice of the bandwidth for a local constant estimator of the density depends on the second derivative of the density function. Notice also that the estimation of density derivatives is an instrumental tool in statistical data analysis in many applications. For example, the first-order density derivative is the fundamental feature for the mean shift clustering seeks modes of the data density, see Fukunaga and Hostetler (1975), Yizong (1995) and Comaniciu and Meer (2002). A statistical test for modes of the data density is based on the second order density derivative (Genovese et al. 2013). The second-order density derivative appears also in the bias of nearest-neighbor Kullback–Leibler divergence estimation, for details refer to Noh et al. (2018). Härdle et al. (1990) and Chacón and Duong (2013) consider the problem of estimating the the density derivative and obtained the optimal bandwidth selection. Sasaki et al. (2016) proposed a novel method that directly estimates density derivatives without going through density estimation. In short, the estimation of the density derivatives is a subject of great interest and received a lot of attention, we can refer to Meyer (1977), Silverman (1978), Cheng (1982), Karunamuni and Mehra (1990), Jones (1994), Abdous et al. (2002), Horová et al. (2002), Henderson and Parmeter (2012a, 2012b), Wu et al. (2014), Schuster (1969).

More applications in fundamental statistical problems such as regression, Fisher information estimation, parameter estimation, and hypothesis testing are discussed in Singh (1976, 1977, 1979). For instance the conditional bias, the conditional variance and the optimal local bandwidth selection of the local polynomial regression estimator depend on high order derivatives of the regression function [see Fan and Gijbels (1995) for more details]. Yu and Jones (1998) showed that the mean square error of the local linear quantile regression estimator, and consequently the choice of the optimal bandwidth, depend on the second

derivative, with respect to the covariate, of the conditional cumulative distribution function. Most of the time, when it comes to the numerical implementation of the optimal choice of the bandwidth, we plug-in the derivative of the above mentioned quantities (density, regression or CDF) by their empirical versions without necessarily deeply studying the properties of the estimates of those derivatives.

Furthermore, it has been noted that the estimation of the first- or higher-order derivatives of the regression function is also important for practical implementations including, but not limited to, the modeling of human growth data (Ramsay and Silverman 2002), kidney function for a lupus nephritis patient (Ramsay and Silverman 2005), and Raman spectra of bulk materials (Charnigo et al. 2011). Derivative estimation is also needed in nonparametric regression to construct confidence intervals for regression functions (Eubank and Speckman 1993), to select kernel bandwidths (Ruppert et al. 1995), and to compare regression curves (Park and Kang 2008). Härdle and Gasser (1985) considered an homoscedastic regression model and proposed kernel M-estimators to estimate nonparametrically the first derivative of the regression function. They heuristically extend their proposal to higher order derivatives. The derivative of the regression function, that is used in modal regression, which is an alternative approach to the usual regression methods for exploring the relationship between a response variable and a predictor variable, we may refer to Herrmann and Ziegler (2004); Ziegler (2001, 2002, 2003) and to Bouzebda and Didi (2021) for recent references. The estimation of the regression function was considered from theoretical and practical point of view by Nadaraja (1969), Rice and Rosenblatt (1983), Gasser and Müller (1984), Georgiev (1984) and Delecroix and Rosa (1996). However, less attention was devoted to the study of the derivatives of the regression function.

In the present work, we are interested in studying the asymptotic properties of function derivatives nonparametric estimates. We do not assume anything beyond the stationarity and the ergodicity of the underlying process. For more details about ergodicity assumption, one can refer the reader to Bouzebda et al. (2015), Bouzebda et al. (2016), Bouzebda and Didi (2017a, 2017b) and Krebs (2019) among others. Notice that in the statistical literature, it is commonly assumed that the data are either independent or satisfy a certain form of mixing assumption. Mixing condition can be seen as some kind of asymptotic independence assumption which can be unrealistic and excludes several stochastic processes characterized by a strong dependence structure (such as long memory processes) or their mixing coefficient does not vanish asymptotically (for instance an autoregressive model with discrete innovation). Moreover, one of the arguments invoked by Leucht and Neumann (2013) motivating the usage of the ergodicity assumption is the existence of example of classes of processes where the ergodicity property is much easier to prove than the mixing one. Hence, the ergodicity condition seems to be more natural to adopt as far as it provides a more general dependence framework which includes non-mixing stochastic processes such those generated by noisy chaos.

In the following we illustrate the discussion above through an example of processes which are ergodic but do not necessarily satisfy the mixing condition. For this let $\{(T_i, \lambda_i) : i \in \mathbb{Z}\}$ be a strictly stationary process such that $T_i \mid \mathcal{T}_{i-1}$ is a Poisson process

with parameter λ_i, where $\mathcal{T}_i$ be the σ-field generated by $(T_i, \lambda_i, T_{i-1}, \lambda_{i-1}, \ldots)$. Assume that $\lambda_i = f(\lambda_{i-1}, T_{i-1})$, where $f : [0, \infty) \times \mathbb{N} \to (0, \infty)$ is a given function . This process is not mixing in general (see Remark 3 in Neumann 2011). It is known that any sequence $(\varepsilon_i)_{i \in \mathbb{Z}}$ of i.i.d. random variables is ergodic. Consequently, one can observe that $(Y_i)_{i \in \mathbb{Z}}$ with $Y_i = \vartheta((\ldots, \varepsilon_{i-1}, \varepsilon_i), (\varepsilon_{i+1}, \varepsilon_{i+2}, \ldots))$, for some Borel-measurable function $\vartheta(\cdot)$, is also ergodic (see Proposition 2.10, page 54 in Bradley 2007 for more details).

To the best of our knowledge, the results presented here, respond to a problem that has not been studied systematically up to the present, which was the basic motivation of the paper. To prove our results, we base our methodology upon the martingale approximation which allows to provide an unified nonparametric time series analysis setting enabling one to launch systematic dependent data studies, which are quite different of existing procedures in the i.i.d. setting.

The remainder of the paper is organized as follows. General notation and definitions of the kernel derivatives estimators are given in Sect. 2. The assumptions and asymptotic properties of the kernel derivative estimators are given in Sect. 3, which includes the uniform strong convergence rates, the asymptotic normality and the AMISE of the family of nonparametric function derivative estimators. Section 4 is devoted to an application for the regression function derivatives. The performance of the proposed procedures is evaluated through simulations in the context of the regression derivatives in Sect. 5. In Sect. 6, we illustrate the estimation methodology on real data. Some concluding remarks and future developments are given in Sect. 7. To avoid interrupting the flow of the presentation, all mathematical developments are relegated to the Sect. 8.

2 Problem formulation and estimation

We start by giving some notation and definitions that are needed for the forthcoming sections. Let $(\mathbf{X}, \mathbf{Y})$ be a random vector, where $\mathbf{X} = (X_1, \ldots, X_p) \in \mathbb{R}^p$ and $\mathbf{Y} = (Y_1, \ldots, Y_q) \in \mathbb{R}^q$. The joint distribution function [df] of $(\mathbf{X}, \mathbf{Y})$ is defined as $F(\mathbf{x}, \mathbf{y}) := \mathbb{P}(\mathbf{X} \leq \mathbf{x}, \mathbf{Y} \leq \mathbf{y})$, for $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{y} \in \mathbb{R}^q$. In the sequel, for $\mathbf{v}' = (v'_1, \ldots, v'_r) \in \mathbb{R}^r$ and $\mathbf{v}'' = (v'', \ldots, v'') \in \mathbb{R}^r$, we set $\mathbf{v}' \leq \mathbf{v}''$ whenever $v'_j \leq v''_j$ for $j = 1, \ldots, r$. We denote by $\mathbf{I}$ and $\mathbf{J}$ two fixed subsets of $\mathbb{R}^p$ such that $\mathbf{I} = \prod_{j=1}^p [a_j, b_j] \subset \mathbf{J} = \prod_{j=1}^p [c_j, d_j] \subset \mathbb{R}^p$, where $-\infty < c_j < a_j < b_j < d_j < \infty$ for $j = 1, \ldots, p$. We assume that $(\mathbf{X}, \mathbf{Y})$ has a joint density function defined as

$$f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) := \frac{\partial^{p+q}}{\partial x_1 \ldots \partial x_p \partial y_1 \ldots \partial y_q} F(\mathbf{x}, \mathbf{y}) \text{ on } \mathbf{J} \times \mathbb{R}^q,$$

with respect to the Lebesgue measure $\mathrm{d}\mathbf{x} \times \mathrm{d}\mathbf{y}$, and denote

$$f_{\mathbf{X}}(\mathbf{x}) = \int_{\mathbb{R}^q} f(\mathbf{x}, \mathbf{y}) \mathrm{d}\mathbf{y}, \text{ for } \mathbf{x} \in \mathbf{J},$$

the marginal density of $\mathbf{X}$ (which is only assumed to exist on $\mathbf{J}$). For a nonnegative integer vector $\mathbf{s} = (s_1, \dots, s_p) \in (\{0\} \cup \mathbb{N})^p$, define $|\mathbf{s}| := s_1 + \cdots + s_p$ and

$$D^{|\mathbf{s}|} := \frac{\partial^{|\mathbf{s}|}}{\partial x_1^{s_1} \cdots \partial x_d^{s_p}}.$$

The operator $D^{|\mathbf{s}|}$ is assumed to be well defined and interchange with integration in our setting. Let $\psi : \mathbb{R}^q \to \mathbb{R}$ be a measurable function. In this paper, we are primarily interested in the estimation of the following derivatives

$$D^{|\mathbf{s}|}r(\psi;\mathbf{x}) := D^{|\mathbf{s}|}\mathbb{E}(\psi(\mathbf{Y}) \mid \mathbf{X} = \mathbf{x})f_{\mathbf{X}}(\mathbf{x}) = \int_{\mathbb{R}^q} \psi(\mathbf{y})D^{|\mathbf{s}|}f(\mathbf{x}, \mathbf{y})d\mathbf{y}, \text{ and } D^{|\mathbf{s}|}r(1;\mathbf{x}) = D^{|\mathbf{s}|}f_{\mathbf{X}}(\mathbf{x}).$$

An extension to the derivative of the regression function $m(\mathbf{x}, \psi) = \mathbb{E}(\psi(\mathbf{Y}) \mid \mathbf{X} = \mathbf{x})$, whenever it exists, will be considered.

2.1 Kernel-type estimation

Let $\{\mathbf{X}_i, \mathbf{Y}_i\}_{i \geq 1}$ be a $\mathbb{R}^p \times \mathbb{R}^q$-valued strictly stationary ergodic process defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We now introduce a kernel function $\{K(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^p\}$, fulfilling the conditions below.

(K.i) $\int_{\mathbb{R}^p} K(\mathbf{t})d\mathbf{t} = 1$.
(K.ii) For given $\mathbf{s} \in (\{0\} \cup \mathbb{N})^p$, the partial derivative $D^{|\mathbf{s}|}K : \mathbb{R}^p \to \mathbb{R}$ exists and

$$\sup_{\mathbf{t} \in \mathbb{R}^p} |D^{|\mathbf{s}|}K(\mathbf{t})| < \infty.$$

We need the smoothness condition (K.ii) on the kernel function $K(\cdot)$ in order to make the operator $D^{|\mathbf{s}|}$ well defined and interchange with integration. The conditions (K.i) and (K.ii) will be assumed tacitly in the sequel. For each $n \geq 1$, and for each choice of the bandwidth $h_n > 0$, we define the kernel estimators

$$f_{\mathbf{X};n}(\mathbf{x}, h_n) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right), \quad r_n(\psi;\mathbf{x}, h_n) = \frac{1}{nh_n}\sum_{i=1}^{n} \psi(\mathbf{Y}_i)K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right).$$

Notice that h_n is a positive sequence of real numbers such that

$$(i) \lim_{n\to\infty} h_n = 0, \ (ii) \lim_{n\to\infty} nh_n = +\infty, \text{ or } (iii) \lim_{n\to\infty} \frac{nh_n}{\log n} = +\infty.$$

The condition (i) is used to obtain the asymptotic unbiasedness of the kernel (density or regression) type estimators. We need more restrictive assumption on h_n for the consistency, this is given by the condition (ii), one can refer to Parzen (1962). In general, the strong consistency fails to hold when either (i) or (iii) is not satisfied.

Remark 1 For notational convenience, we have chosen the same bandwidth sequence for each margins. This assumption can be dropped easily. If one wants to make use

of the vector bandwidths (see, in particular, Chapter 12 of Devroye and Lugosi 2001). With obvious changes of notation, our results and their proofs remain true when h_n is replaced by a vector bandwidth $\mathbf{h}_n = (h_n^{(1)}, \ldots, h_n^{(p)})$, where $\min h_n^{(i)} > 0$. In this situation we set $h_n = \prod_{i=1}^p h_n^{(i)}$ and, for any vector $\mathbf{v} = (v_1, \ldots, v_p)$, we replace $\mathbf{v}/h_n$ by $(v_1/h_n^{(1)}, \ldots, v_1/h_n^{(p)})$. For ease of presentation we chose to use real-valued bandwidths throughout.

Our aim is to provide estimators of the $|\mathbf{s}|$-th derivatives $D^{|\mathbf{s}|} f_{\mathbf{X}}(\mathbf{x})$ and $D^{|\mathbf{s}|} r(\psi;\mathbf{x})$, respectively, and to establish their asymptotic properties. The natural choices for these estimators are (for a suitable choice of $h_n > 0$) the $|\mathbf{s}|$-th derivatives of $f_{\mathbf{X};n}(\mathbf{x}, h_n)$ and $r_n(\psi;\mathbf{x}, h_n)$, respectively defined as:

$$D^{|\mathbf{s}|} f_{\mathbf{X};n}(\mathbf{x}, h_n) = \frac{1}{n h_n^{1+|\mathbf{s}|/p}} \sum_{i=1}^n D^{|\mathbf{s}|} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right), \tag{1}$$

$$D^{|\mathbf{s}|} r_n(\psi;\mathbf{x}, h_n) = \frac{1}{n h_n^{1+|\mathbf{s}|/p}} \sum_{i=1}^n \psi(\mathbf{Y}_i) D^{|\mathbf{s}|} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right). \tag{2}$$

Notice that the kernel density derivative estimators $D^{|\mathbf{s}|} f_{\mathbf{X};n}(\mathbf{x}, h_n)$ is a particular case of $D^{|\mathbf{s}|} r_n(\psi;\mathbf{x}, h_n)$, that is

$$D^{|\mathbf{s}|} f_{\mathbf{X};n}(\mathbf{x}, h_n) = D^{|\mathbf{s}|} r_n(1;\mathbf{x}, h_n). \tag{3}$$

Remark 2 The general kernel-type estimator of $m(\cdot, \psi) = \mathbb{E}(\psi(Y) \mid \mathbf{X} = \cdot)$ is given, for $\mathbf{x} \in \mathbb{R}^p$, by

$$\widehat{m}_{n;h_n}(\mathbf{x}, \psi) := \frac{\sum_{i=1}^n \psi(\mathbf{Y}_i) K\left(\dfrac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right)}{\sum_{i=1}^n K\left(\dfrac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right)}. \tag{4}$$

By setting, for $q = 1$, $\psi(y) = y$ (or $\psi(y) = y^k$) into (4) we get the classical Nadaraya–Watson (1964, 1964) kernel regression function estimator of $m(\mathbf{x}) := \mathbb{E}(Y \mid \mathbf{X} = \mathbf{x})$ given by

$$\widehat{m}_{n;h_n}(\mathbf{x}) := \frac{\sum_{i=1}^n Y_i K\left(\dfrac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right)}{\sum_{i=1}^n K\left(\dfrac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right)}, \quad \text{or} \quad \widehat{m}_{n;h_n}(\mathbf{x}) := \frac{\sum_{i=1}^n Y_i^k K\left(\dfrac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right)}{\sum_{i=1}^n K\left(\dfrac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right)}.$$

Nadaraya (1964) established similar results to those of Parzen (1962) for $\widehat{m}_{n;h_n}(\mathbf{x})$ as an estimator for $\mathbb{E}(Y \mid \mathbf{X} = \mathbf{x})$.

Remark 3 By setting $\psi_{\mathbf{t}}(\mathbf{y}) = \mathbb{1}[\mathbf{y} \leq \mathbf{t}]$, for $\mathbf{t} \in \mathbb{R}^q$, into (4) we obtain the kernel estimator of the conditional distribution function $F(\mathbf{t}|\mathbf{x}) := \mathbb{P}(\mathbf{Y} \leq \mathbf{t}|\mathbf{X} = \mathbf{x})$, given by

$$\widehat{F}_{n;h_n}(\mathbf{t}|\mathbf{x}) := \frac{\sum_{i=1}^{n} \mathbb{1}(\mathbf{Y}_i \leq \mathbf{t}) K\left(\dfrac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right)}{\sum_{i=1}^{n} K\left(\dfrac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right)}.$$

These examples motivate the introduction of the function $\psi(\cdot)$ in our setting, refer to Deheuvels (2011) for more discussion.

Remark 4 Local polynomial regression has emerged as a dominant method for nonparametric estimation and inference. The local linear variant was proposed by Stone (1977) and Cleveland (1979), see Fan (1992) and Fan and Gijbels (1996) for an extensive treatment of the local polynomial estimator. However, Racine (2016) mentioned that one feature of local polynomial estimators that may not be widely appreciated is that the local polynomial derivative estimator does not, in general, coincide with the analytic derivative of the local polynomial regression estimator infinite-sample settings. This can cause problems, particularly in the context of shape constrained estimation. The problem arises when the object of interest is the regression function itself and constraints are to be imposed on derivatives of the regression function, however the regression estimate and derivative estimate are not internally consistent, i.e., the derivative of the local polynomial regression estimate does not coincide with the local polynomial derivative estimate.

3 Assumptions and asymptotic properties

We will denote by $\mathcal{F}_i$ the σ−field generated by $(\mathbf{X}_1, \ldots, \mathbf{X}_i)$. For any $i = 1, \ldots, n$ define $f_{\mathbf{X}_i}^{\mathcal{F}_{i-1}}(\cdot)$ as the conditional density of $\mathbf{X}_i$ given the σ−field $\mathcal{F}_{i-1}$. Let $\mathcal{G}_n$ be the σ-field generated by $\{(\mathbf{X}_i, \mathbf{Y}_i) : 1 \leq i \leq n\}$, and let $f_{\mathbf{X},\mathbf{Y}}^{\mathcal{G}_{i-1}}(\cdot)$ be the conditional density of $(\mathbf{X}, \mathbf{Y})$ given the σ−field $\mathcal{G}_{i-1}$. Let us define the σ−field $\mathcal{S}_n = \sigma((\mathbf{X}_k, \mathbf{Y}_k); (\mathbf{X}_{n+1}) : 1 \leq k \leq n)$, and the projection operator

$$\mathcal{P}_k \xi = \mathbb{E}(\xi \mid \mathcal{G}_k) - \mathbb{E}(\xi \mid \mathcal{G}_{k-1}).$$

Moreover, if $\zeta(\cdot)$ is a real-valued random function which satisfies $\zeta(u)/u \to 0$ a.s. as $u \to 0$, we write $\zeta(u) = o_{\text{a.s.}}(u)$. In the same way, we say that $\zeta(u)$ is $O_{\text{a.s.}}(u)$ if $\zeta(u)/u$ is a.s. bounded as $u \to 0$.

The following assumptions will be needed throughout the paper.

(K.1)

    (i)   The kernel $K(\cdot)$ is a symmetric compactly supported probability density function,

(ii)   The kernel derivatives $D^{|s|}K(\cdot)$, $s = 0, 1, \ldots$, are assumed to be Lipschitz function with ratio $C_{K,s} < \infty$ and order γ, i.e., $|D^{|s|}K(\mathbf{x}) - D^{|s|}K(\mathbf{x}')| \leq C_{K,s}\|\mathbf{x} - \mathbf{x}'\|^{\gamma}$,   for $(\mathbf{x}, \mathbf{x}') \in \mathbb{R}^{2p}$;

(iii) $\int_{\mathbb{R}^p} \|\mathbf{x}\| D^{|s|}K(\mathbf{x})d\mathbf{x} < \infty$,   for $s = 0, 1, \ldots$

(iv) $\int_{\mathbb{R}^p} \left(D^{|s|}K(\mathbf{v})\right)^2 d\mathbf{v} < \infty$.

(C.1)   (i)   The conditional density $f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}(\cdot)$ exists and belongs to the space $\mathcal{C}^{|s|}(\mathbb{R})$, here $\mathcal{C}^{|s|}(\mathbb{R}^p)$ denotes the space of all continuous real-valued functions that are $|s|$-times continuously differentiable on $\mathbb{R}^p$;

(ii)   The partial derivative $D^{|s|}f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}(\cdot)$ is continuous and has bounded partial derivatives of order $\Bbbk$, that is, there exists a constant $0 < \mathfrak{C}_1 < \infty$ such that

$$\sup_{\mathbf{x} \in \mathbf{J}} \left| \frac{\partial^{\Bbbk} D^{|s|}f^{\mathcal{G}_{i-1}}(\mathbf{x})}{\partial x_1^{k_1} \ldots \partial x_d^{k_p}} \right| \leq \mathfrak{C}_1, \quad k_1, \ldots, k_p \geq 0, \quad 0 < k_1 + \cdots + k_p = \Bbbk;$$

(C.2)   For any $\mathbf{x} \in \mathbb{R}^p$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}(\mathbf{x}) = f(\mathbf{x}), \qquad \text{in the } a.s. \text{ and } L^2 \text{ sense}.$$

(C.3)   (i)   The density $f_{\mathbf{X}}(\cdot)$ is continuous and has bounded partial derivatives of order $\mathbf{r}$, that is, there exists a constant $0 < \mathfrak{C}_2 < \infty$ such that

$$\sup_{\mathbf{x} \in \mathbf{J}} \left| \frac{\partial^r f_{\mathbf{X}}(\mathbf{x})}{\partial x_1^{k_1} \ldots \partial x_d^{k_p}} \right| \leq \mathfrak{C}_2, \quad k_1, \ldots, k_p \geq 0, \quad 0 < k_1 + \cdots + k_p = \mathbf{r};$$

(ii)   The density $f_{\mathbf{X},\mathbf{Y}}(\cdot, \cdot)$ is continuous and has bounded partial derivatives of order ℓ, that is, there exists a constant $0 < \mathfrak{C}_3 < \infty$ such that

$$\sup_{\mathbf{x} \in \mathbf{J}} \left| \frac{\partial^{\ell} f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})}{\partial x_1^{k_1} \ldots \partial x_d^{k_p}} \right| \leq \mathfrak{C}_3, \quad k_1, \ldots, k_p \geq 0, \quad 0 < k_1 + \cdots + k_p = \ell;$$

(C.4)   There exists a positive constant $f_{\star} < \infty$ such that

$$\sup_{\mathbf{x} \in \mathbb{R}^p} D^{|s|}f_{\mathbf{X}_1}^{\mathcal{G}_0}(\mathbf{x}) \leq f_{\star},$$

holds with probability 1.

(C.5) $\sup_{\mathbf{x}} \sum_{i=1}^{\infty} \left\| \mathcal{P}_1 D^{|s|}f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}(\mathbf{x}) \right\|^2 < \infty$.

(R.1)   (i) $\mathbb{E}(|\psi(\mathbf{Y}_i)||\mathcal{S}_{i-1}) = \mathbb{E}(|\psi(\mathbf{Y}_r)| \mid \mathbf{X}_i) = m(\mathbf{X}_i, |\psi|);$

         (ii)   there exist constants $C_\psi > 0$ and $\beta > 0$ such that, for any couple $(\mathbf{x}, \mathbf{x}') \in \mathbb{R}^{2p}$,

$$\left|m(\mathbf{x}, \psi) - m(\mathbf{x}', \psi)\right| \leq C_\psi \|\mathbf{x} - \mathbf{x}'\|^\beta;$$

         (iii)   For any $k \geq 2$, $\mathbb{E}(|\psi^k(\mathbf{Y}_i)||\mathcal{S}_{i-1}) = \mathbb{E}(|\psi^k(\mathbf{Y}_i)||\mathbf{X}_i)$, and the function

$$\Psi_k(\mathbf{x}, \psi) = \mathbb{E}(|\psi^k(\mathbf{Y})||\mathbf{X} = \mathbf{x}),$$

         is continuous in the neighborhood of $\mathbf{x}$.

(H)   (i) $h_n \to 0, nh_n^{1+2\left(\frac{|\mathbf{s}|}{p}\right)} \to \infty;$

             (ii) $h_n \to 0, \frac{nh^{1+s/p}}{\log n} \to \infty.$

3.1 Comments on the conditions

Conditions (K.1) are very common in nonparametric function estimation literature. They set some kind of regularity upon the kernels used in our estimates. In particular, by imposing the condition (K.1)(iii), the kernel function exploits the smoothness of the function $D^{|\mathbf{s}|}r(\psi;\mathbf{x})$. Notice that the transformation of the stationary ergodic process $(\mathbf{X}_i, \mathbf{Y}_i)_{i\geq 1}$ into the process $(\psi^2(\mathbf{Y}_i))_{i\geq 1}$ is a measurable function. Therefore, making use of Proposition 4.3 of Krengel (1985) and then the ergodic theorem, we obtain $\lim_{n\to\infty} n^{-1} \sum_{i=1}^n \psi^2(\mathbf{Y}_i) = \mathbb{E}\left[\psi^2(\mathbf{Y}_1)\right]$ almost surely. Conditions (C.1) and (C.3) impose the needed regularity upon the joint, marginal and the conditional densities to reach the rates of convergence given below. Conditions (C.2) involves the ergodic nature of the data as given, for instance, see Proposition 4.3 and Theorem 4.4 of Krengel (1985) and Delecroix (1987) (Lemma 4 and Corollary 1 together with their proofs). The assumption (C.5) is assumed by Wu (2003) which is satisfied by various processes including linear as well as many nonlinear ones. For more details and examples, see Wu (2003) and Wu et al (2010). We refer also to the recent paper of Wu et al (2010) for more details on conditions (C.4). The conditions (R.1)(i) and (R.1)(iii) is usual in the literature dealing with the study of ergodic processes. The condition (R.1)(ii) is a regularity condition upon the regression function.

Remark 5 Our results remain valid when replacing the condition that the kernel function $K(\cdot)$ has compact support in (K.1)(i) with another condition (K.1)(i)' whose content is as follows:

(K.1)(i)' There exists a sequence of positive real numbers a_n such that $a_n h_n^d$ tends to zero when n tends to infinity, and

$$\sqrt{n} \int_{\{\|\mathbf{v}\|>a_n\}} |K(\mathbf{v})| d\mathbf{v} \to 0.$$

3.2 Almost sure uniform consistency rates

In the following theorems, we will give the uniform convergence with rate of $D^{|\mathbf{s}|}r_n(\psi;\mathbf{x}, h_n)$ defined in (2).

Theorem 1 *Assume that the assumptions H(ii), K(i)–(ii), (K.1)(i)–(iii), (C.1), (C.4), (C.5) and (R.1) are fulfilled. We have, as $n \to \infty$,*

$$\sup_{\mathbf{x}\in\mathbf{J}} \left| D^{|\mathbf{s}|}r_n(\psi;\mathbf{x}, h_n) - \mathbb{E}D^{|\mathbf{s}|}r_n(\psi;\mathbf{x}, h_n) \right| = O_{a.s.}\left(\sqrt{\frac{\log n}{nh^{1+s/p}}} \right).$$

Theorem 2 *Suppose that the assumptions H(ii), K(i)–(ii), (K.1)(i)–(iii), (C.1), (C.3) (ii), (C.4), (C.5) and (R.1) are satisfied. We have, as $n \to \infty$,*

$$\sup_{\mathbf{x}\in\mathbf{J}} \left| D^{|\mathbf{s}|}r_n(\psi;\mathbf{x}, h_n) - D^{|\mathbf{s}|}r(\psi;\mathbf{x}) \right| = O_{a.s.}\left(\sqrt{\frac{\log n}{nh^{1+s/p}}} \right) + O\left(h_n^{\ell/p}\right).$$

3.3 Asymptotic distribution

Let us now state the following theorem, which gives the weak convergence rate of the estimator $D^{|\mathbf{s}|}r_n(\psi;\mathbf{x}, h_n)$ defined in (2). Below, we write $Z \overset{\mathcal{D}}{=} N(\mu, \sigma^2)$ whenever the random variable Z follows a normal law with expectation μ and variance σ^2.

Theorem 3 *Assume that the conditions H(i), K(i)–(ii), (K.1), (C.1), (C.2) (C.3), (C.4), (C.5) and (R.1) hold. We have, as $n \to \infty$*

$$\sqrt{nh_n^{1+2\left(\frac{|\mathbf{s}|}{p}\right)}} \left(D^{|\mathbf{s}|}r_n(\psi;\mathbf{x}, h_n) - \mathbb{E}D^{|\mathbf{s}|}r_n(\psi;\mathbf{x}, h_n) \right) \to N(0, \sigma_\psi^2(\mathbf{x})),$$

where

$$\sigma_\psi^2(\mathbf{x}) = \Psi_2(\mathbf{x})f(\mathbf{x})\left(\int_{\mathbb{R}^p} \left(D^{|\mathbf{s}|}K(\mathbf{v})\right)^2 d\mathbf{v} \right), \text{ and } \Psi_2(\mathbf{x}, \psi) = \mathbb{E}(\psi^2(\mathbf{Y})|\mathbf{X} = \mathbf{x}).$$

Theorem 4 *Assume that the conditions H(i), K(i)–(ii), (K.1), (C.1), (C.2) (C.3), (C.4), (C.5) and (R.1) hold. In addition we assume*

$$n^{1/2}h_n^{(|\mathbf{s}|+\ell)/p+1/2} \to 0 \quad \text{as} \quad n \to \infty.$$

Then, we have, as $n \to \infty$

$$\sqrt{nh_n^{1+2\left(\frac{|s|}{p}\right)}}\left(D^{|s|}r_n(\psi;\mathbf{x},h_n) - D^{|s|}r(\psi;\mathbf{x})\right) \to N(0,\sigma_\psi^2(\mathbf{x})).$$

3.4 Asymptotic mean square error

In the following, we will give asymptotic mean integrated squared error (AMISE) of the estimator $D^{|s|}r_n(\psi;\mathbf{x},h_n)$.

Theorem 5 *Assume that the conditions K(i)–(ii), (K.1), (C.1), (C.2), (C.3) (ii), (C.4), (C.5) and (R.1)(i)–(ii) hold. We have, as $n \to \infty$*

$$\text{AMISE}\left(D^{|s|}r_n(\psi;\mathbf{x},h_n)\right) = \int_{\mathbb{R}^p} \text{Bias}\left\{D^{|s|}r_n(\psi;\mathbf{x},h_n)\right\}^2 d\mathbf{x} + \int_{\mathbb{R}^p} \text{Var}\left(D^{|s|}r_n(\psi;\mathbf{x},h_n)\right) d\mathbf{x}$$

$$= O\left(h_n^{2\ell/p}\right) + O\left(\frac{1}{nh_n^{1+2\frac{|s|}{p}}}\right).$$

Remark 6 Keeping in mind the relation (3), one can easily deduce the following results concerning the density function derivative, that is,

$$\sup_{\mathbf{x}\in\mathbf{J}}\left|D^{|s|}f_{\mathbf{X};n}(\mathbf{x},h_n) - D^{|s|}f_{\mathbf{X}}(\mathbf{x})\right| = O_{a.s.}\left(\sqrt{\frac{\log n}{nh^{1+s/p}}}\right) + O\left(h_n^{\ell/p}\right),$$

$$\sqrt{nh_n^{1+2\left(\frac{|s|}{p}\right)}}\left(D^{|s|}f_{\mathbf{X};n}(\mathbf{x},h_n) - D^{|s|}f_{\mathbf{X}}(\mathbf{x})\right) \to N\left(0,f(\mathbf{x})\left(\int_{\mathbb{R}^p}\left(D^{|s|}K(\mathbf{v})\right)^2 d\mathbf{v}\right)\right),$$

and

$$\text{AMISE}\left(D^{|s|}f_{\mathbf{X};n}(\mathbf{x},h_n)\right) = \int_{\mathbb{R}^p} \text{Bias}\left\{D^{|s|}f_{\mathbf{X};n}(\mathbf{x},h_n)\right\}^2 d\mathbf{x} + \int_{\mathbb{R}^p} \text{Var}\left(D^{|s|}f_{\mathbf{X};n}(\mathbf{x},h_n)\right) d\mathbf{x}$$

$$= O\left(h_n^{2\ell/p}\right) + O\left(\frac{1}{nh_n^{1+2\frac{|s|}{p}}}\right).$$

3.5 Confidence intervals

The asymptotic variance in the central limit theorem depends on the unknown functions, which should be estimated in practice. Let us introduce $\widehat{\Psi}_{2,n}(\mathbf{x},\psi)$ a kernel estimator of $\Psi_2(\mathbf{x},\psi)$ defined by

$$\widehat{\Psi}_{2,n}(\mathbf{x}, \psi) := \frac{\sum_{i=1}^{n} \psi^2(\mathbf{Y}_i) K\left(\dfrac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right)}{\sum_{i=1}^{n} K\left(\dfrac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right)}.$$

This permits to estimate asymptotic variance $\sigma_\psi^2(\mathbf{x})$ by

$$\widehat{\sigma}_\psi^2(\mathbf{x}) = \widehat{\Psi}_{2,n}(\mathbf{x}, \psi) f_{\mathbf{X};n}(\mathbf{x}, h_n) \left(\int_{\mathbb{R}^p} \left(D^{|\mathbf{s}|} K(\mathbf{v}) \right)^2 d\mathbf{v} \right).$$

Furthermore, from Theorem 4, the approximate confidence interval of $D^{|\mathbf{s}|} r(\psi;\mathbf{x})$ can be obtained as

$$D^{|\mathbf{s}|} r(\psi;\mathbf{x}) \in \left[D^{|\mathbf{s}|} r_n(\psi;\mathbf{x}, h_n) \pm c_\alpha \frac{\widehat{\sigma}_\psi(\mathbf{x})}{\sqrt{nh_n^{1+2\left(\frac{|\mathbf{s}|}{p}\right)}}} \right],$$

where c_α, denotes the $(1 - \alpha)$–quantile of the normal distribution.

Remark 7 An alternative approach based on resampling techniques might be used to estimate confidence intervals. In contrast to the asymptotic confidence intervals, the main advantage of such approach is that avoids the estimation of the variance of estimators. Below, we give a brief description of the bootstrap-based confidence intervals approach. Let $\{Z_i\}$ be a sequence of random variables satisfying the following assumption:

B. The $\{Z_i\}$ are independent and identically distributed, with distribution function P_Z, mean zero and variance 1.

We assume that the bootstrap weights Z_i's are independent from the data (X_i, Y_i), $i = 1, \ldots, n$. Define,

$$D^{|\mathbf{s}|} r_n^*(\psi;\mathbf{x}, h_n) = \frac{1}{nh_n^{1+|\mathbf{s}|/p}} \sum_{i=1}^{n} Z_i \psi(\mathbf{Y}_i) D^{|\mathbf{s}|} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right).$$

Let

$$\alpha_n^* = \sqrt{nh_n^{1+2\left(\frac{|\mathbf{s}|}{p}\right)}} \left(D^{|\mathbf{s}|} r_n^*(\psi;\mathbf{x}, h_n) - D^{|\mathbf{s}|} r_n(\psi;\mathbf{x}, h_n) \right).$$

Let $\mathfrak{N}$, be a large integer and $Z_1^k, \dots, Z_n^k$, $k = 1, \dots, \mathfrak{N}$ independent copies of Z. Let $\alpha_n^{*(k)}$ be the bootstrapped copies of α_n^*. In order to approximate c_α, one can use the sampling estimator $\widehat{c}_\alpha$, of c_α, as the smallest $z \geq 0$ such that $\frac{1}{\mathfrak{N}} \sum_{k=1}^{\mathfrak{N}} \mathbb{1}_{\left\{\alpha_n^{*(k)} \leq z\right\}} \geq 1 - \alpha$.

4 Application to the regression derivatives

In this section, we will follow the same notation as in Deheuvels and Mason (2004). We will consider especially the conditional expectation of $\psi(Y)$ given $X = x$, for $p = q = 1$. Recall that

$$m(x, \psi) = \mathbb{E}(\psi(Y) \mid X = x) = \frac{1}{f_X(x)} \int_{\mathbb{R}} \psi(y) f_{X,Y}(x, y) \mathrm{d}y = \frac{r(\psi, x)}{f_X(x)}.$$

The kernel estimator is given by

$$m_n(x, \psi) = \begin{cases} \dfrac{r_n(\psi; x, h_n)}{f_{X;n}(x, h_n)} & \text{if } f_{X;n}(x, h_n) \neq 0, \\[3mm] \dfrac{1}{n} \sum_{i=1}^{n} \psi(Y_i) & \text{if } f_{X;n}(x, h_n) = 0. \end{cases}$$

Recall the following derivatives

$$m_\psi'(x) = \frac{r'(\psi, x)}{f_X(x)} - \frac{r(\psi, x) f_X'(x)}{f_X^2(x)}, \tag{5}$$

and

$$m_\psi''(x) = \frac{r''(\psi, x)}{f_X(x)} - \frac{2r'(\psi, x) f_X'(x)}{f_X^2(x)} + \frac{r(\psi, x)\{2(f_X'(x))^2 - f_X(x) f_X''(x)\}}{f_X^3(x)}.$$

In order to estimate the derivatives of $m_\psi'(x)$ in (5) and $m_\psi''(x)$ in (6) by replacing $f_X(\cdot)$, $f_X'(\cdot)$, $f_X''(\cdot)$, $r(\psi;\cdot)$, $r'(\psi;\cdot)$ and $r''(\psi;\cdot)$ by $f_{X;n}(\cdot, h_n)$, $f_{X;n}'(\cdot, h_n)$, $f_{X;n}''(\cdot, h_n)$, $r_n(\psi;\cdot, h_n)$, $r_n'(\psi;\cdot, h_n)$ and $r_n''(\psi;\cdot, h_n)$. We so define $m_{\psi,n}'(x; h_n)$ and $m_{\psi,n}''(x; h_n)$ when $f_{X;n}(x, h_n) \neq 0$. The definition of $m_{\psi,n}'(x; h_n)$ and $m_{\psi,n}''(x; h_n)$ is completed by setting $m_{\psi,n}'(x; h_n) = m_{\psi,n}''(x; h_n) = 0$ when $f_{X;n}(x, h_n) = 0$.

The following theorem is more or less a straightforward consequence of Theorem 2.

Corollary 1 *Under the assumptions of Theorem* 2, *we have*

$$\sup_{x \in J} \left| m'_{\psi,n}(x;h_n) - m'_{\psi}(x) \right| = o_{\mathbb{P}}(1),$$

$$\sup_{x \in J} \left| m''_{\psi,n}(x;h_n) - m''_{\psi}(x) \right| = o_{\mathbb{P}}(1).$$

Remark 8 We note that, when $|\mathbf{s}| \geq 2$, $m_{\psi,n}^{(|\mathbf{s}|)}(x, h_n) = D^{|\mathbf{s}|}(m_{\psi,n}(x, h_n))$ may be obtained likewise through the usual Leibniz expansion of derivatives of products given by

$$m_{\psi,n}^{(|\mathbf{s}|)}(x, h_n) = \sum_{|\mathbf{j}|=0}^{|\mathbf{s}|} C_{|\mathbf{s}|}^{|\mathbf{j}|} r_n^{(|\mathbf{j}|)}(\psi;x,h_n) \left\{ f_{X;n}^{-1}(x,h_n) \right\}^{(|\mathbf{s}|-|\mathbf{j}|)}, \quad f_{X;n}(x,h_n) \neq 0.$$

5 Simulation study

The first part of this section investigates the estimation of the first derivative of the density function whenever X is, respectively, a unidimensional and bidimensional stochastic process. Then, we focus on the study of the estimation of the first derivative of the regression function when the data is generated according to a specific stochastic regression model. Motivated by the extension of numerical results obtained in Blanke and Pumo (2003) and Chaouch and Laïb (2019), we suppose in the sequel that X is an Ornstein-Uhlenbeck (OU) process (it can be unidimensional or bidimensional as will be discussed below) solution of the following stochastic differential equation (SDE):

$$dX_t = -a X_t \, dt + b \, dW_t, \quad \text{for} \quad a > 0, b > 0, \tag{6}$$

where $(W_t)_{t \geq 0}$ is a standard Wiener process. Thus, for $0 \leq t \leq T$, the solution of the SDE given in (6) can be expressed as $X_t = e^{-at}X_0 + b \int_0^t e^{-a(t-s)}dW_s$, where $X_0 \sim N(0,1)$ independent of W. In the sequel, and following Blanke and Pumo (2003), we consider $b = \sqrt{2}$ and $a = 1$, since then X_t has a density of a $N(0,1)$. As one can observe the OU process $\{X_t; 0 \leq t \leq T\}$ is a continuous time process used to model the dynamic of several random variable. For instance OU processes are widely used in finance to model and predict asset prices. In real life the process X_t cannot be observed at any time between $[0, T]$, it is rather observed at a specific grid, say $0 = \tau_0 < \tau_1 < \cdots < \tau_n = T$, of time representing a discretization of the interval $[0, T]$. Therefore, the simulation of an OU process can be achieved by considering, for instance, the iterative Euler–Maruyama scheme, see for instance Kloeden and Platen (1992), which allows to build an approximate solution $\{\widetilde{X}_t; t = \tau_0, \tau_1, \ldots, \tau_n\}$ of the original process $\{X_t, 0 \leq t \leq T\}$. The discretized version of the above SDE in (6) is given as follows:

$$\widetilde{X}_{\tau_{j+1}} = \widetilde{X}_{\tau_j} - \widetilde{X}_{\tau_j}(\tau_{j+1} - \tau_j) + \sqrt{2}\left(W_{\tau_{j+1}} - W_{\tau_j}\right), \quad j = 0, \ldots, n. \tag{7}$$

In this simulation, study a deterministic equidistant discretization scheme, i.e., $\tau_{j+1} - \tau_j = T/n =: \delta_n$, called the sampling mesh, is considered. Figure 1 displays
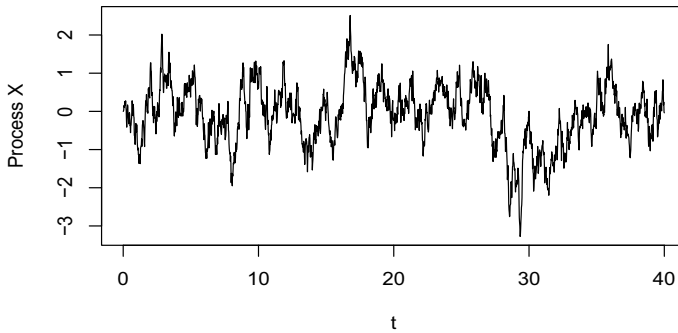
Fig. 1 An example of a univariate OU sample path where $\delta_n = 0.4$ and $n = 105$

an example of OU sample path for $n = 105$ and $\delta_n = 0.4$. Notice that, as discussed in Blanke and Pumo (2003), the sampling mesh δ_n plays an important role in the estimation of the density function on an OU process. Indeed Blanke and Pumo (2003) discussed, theoretically as well as via simulations, the selection of the optimal mesh that minimizes the mean integrated square error (MISE). Chaouch and Laïb (2019) discussed the selection of the sampling mesh for the estimation of the regression function when the response variable is affected by a missing at random phenomena. In this section, we are interested in extending the numerical results obtained by Blanke and Pumo (2003) and Chaouch and Laïb (2019) to the first derivative of the density and the regression function, respectively. More precisely, we will discuss the numerical selection of the optimal mesh, say $\delta_n^{\star}$, which allows to obtain a consistent (in the sense of minimizing the MISE) estimator of the first derivative of the density and the regression function. This section contains two parts: in the first one we study the optimal selection of the sampling mesh of the first derivative of the density function of a one-dimensional OU process then we extend the study to the bidimensional OU processes. The second part of the simulation deals with the first derivative of the univariate regression function.

5.1 Estimation of the first derivative of density function

5.2 Case of the one-dimensional discretized diffusion processes

Now, we consider that X_t is a one-dimensional OU process (see Fig. 1 for an example of a sample path obtained with Euler–Maruyama discretization scheme). As discussed in Blanke and Pumo (2003), the OU process $(X_t)_{t \geq 0}$, as a solution of the SDE (6), has a Gaussian density function $N(0, 1)$. Our purpose in this subsection is to find the optimal (in terms of minimizing the Mean Integrated Square Error, MISE) sampling mesh needed to accurately estimate the first derivative of the density function of X_t. For this, we consider a sequence of sampling mesh δ, a grid of 50 values of x taken between $[-4, 4]$ where the density is locally estimated. Based on $N = 1000$ independent replications, we define the MISE as follows:

$$\text{MISE}(\delta) := \frac{1}{N} \sum_{k=1}^{N} \int \left(f'_{X,n,k}(x, \delta) - f'_X(x) \right)^2 \mathrm{d}x,$$

where $\delta := \delta_n = \tau_{j+1} - \tau_j$, $n = 105$ and $f'_{X,n,k}(x, \delta)$ is the estimator of first derivative of the density function $f'_X(x)$ at the point x in the grid obtained with the kth simulated sample path with a specific mesh δ.

To estimate nonparametrically the first derivative of density function of the OU process X_t, we consider as a kernel the Gaussian density function and the cross-validation technique is used as a tool to select the optimal bandwidth that is:

$$h_{opt} = \arg \min_h \sum_{x \in \mathcal{S}} \left(f'_{X,n,h}(x, \delta) - f'_X(x) \right)^2, \tag{8}$$

where $\mathcal{S}$ is a grid of randomly fixed values of x where the first derivative of the density is estimated. Fig. 2 displays the evolution of the MISE for different values of δ. One can observe that the optimal sampling mesh minimizing the MISE is $\delta^\star = 0.4$. In other words, in practice, one should sample the underlying OU process with a frequency 0.4 to obtain an estimate of the first derivative of the density with a minimum MISE. Moreover, Fig. 2 tells that sampling the OU process with a frequency less than 0.4 will lead to an inaccurate estimate of $f'_X(x)$ because of the high correlation between the observations $\widetilde{X}_{\tau_j+1}$ and $\widetilde{X}_{\tau_j}$, for $j = 0, \dots, n$. Whereas sampling the underlying process X_t with a frequency higher than 0.4 will not improve the quality of the estimate since one can see that the MISE becomes stable at some level. One notices that the discussion made about the interpretation of the MISE plot remains valid for the similar graphs in this paper.

Figure 3a displays the shape the density function of the OU process X and Fig. 3b shows the true first derivative of the density function as well as its nonparametric estimate based on the optimal sampling mesh $\delta^\star = 0.4$. It is worth noting that this first simulation study generalizes some of the results obtained by Blanke and Pumo
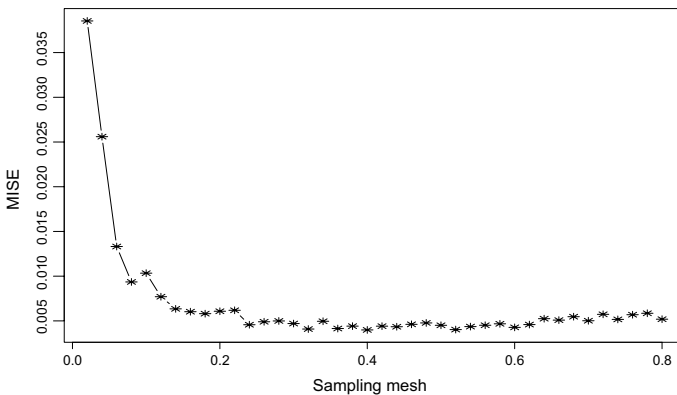

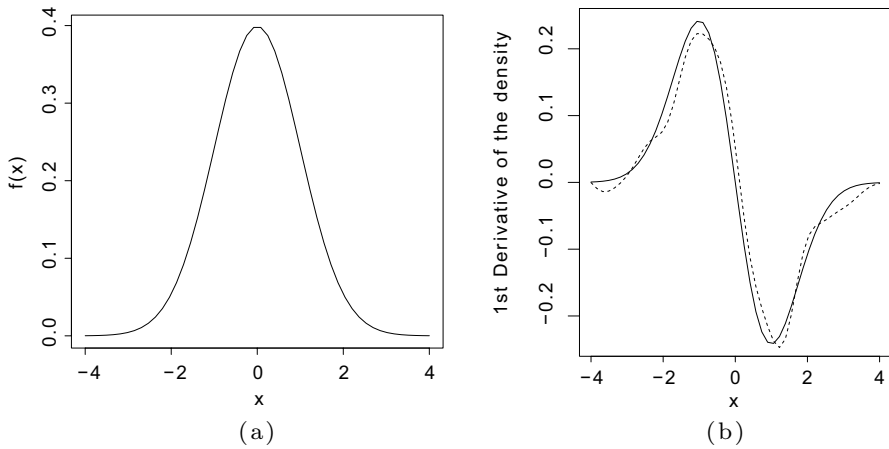**Fig. 2** MISE(δ) for a one-dimensional OU process where $n = 105$

Fig. 3 a the Gaussian density of an OU process. **b** Dark bold line displays the first derivative of the density function and dotted line its estimation

(2003) to the case of the first derivative of the univariate density function. The following simulation aims to extend the results to the bivariate case.

5.3 Case of the two-dimensional discretized diffusion processes

In this simulation, we are interested in studying the estimation of the first derivative of the density function of a bidimensional OU process $\mathbf{X}_t := (X_{1,t}, X_{2,t})$, where X_1 and X_2 are generated independently. Following the same description made above $X_{1,t}$ and $X_{2,t}$ are solutions of the SDE (6) and numerically simulated according to the discretization scheme given in (7). An example of simulated sample path of the vector of OU processes $\mathbf{X}$ is displayed in Fig. 4. Because of the independence between $X_{1,t}$ and $X_{2,t}$ the density function of $\mathbf{X}$ will be the product of the marginals which are both Gaussian. Therefore
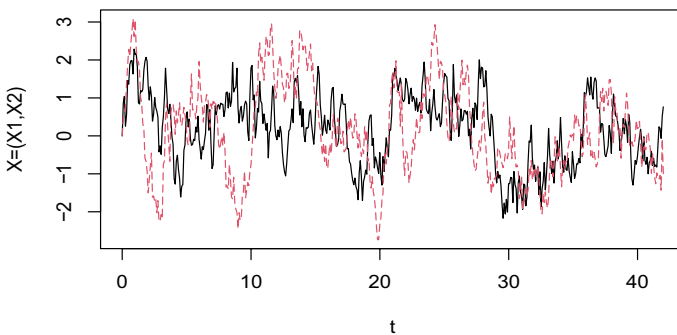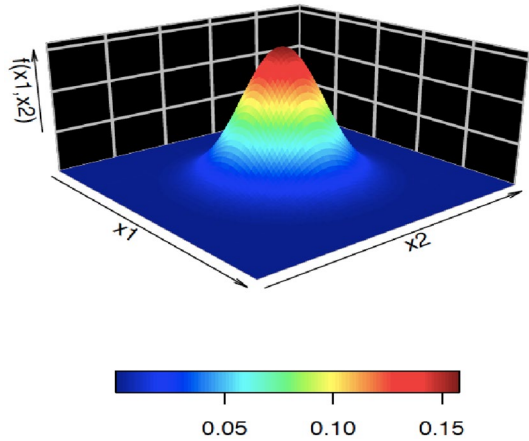

Fig. 4 An example of a bidimensional OU path

Fig. 5 The joint density function of a bidimensional OU process

$$f_{\mathbf{X}}(x_1, x_2) = \frac{1}{2\pi} \exp(-(x_1^2 + x_2^2)/2), \text{ for } (x_1, x_2) \in \mathbb{R}^2.$$

Figure 5 displays the joint distribution of the pair (X_1, X_2). The true first derivative of the joint density function is given as follows:

$$f''_{X_1 X_2}(x_1, x_2) := \frac{\partial^2 f_{\mathbf{X}}}{\partial x_1 \partial x_2}(x_1, x_2) = \frac{x_1 x_2}{2\pi} \exp(-(x_1^2 + x_2^2)/2), \text{ for } (x_1, x_2) \in \mathbb{R}^2.$$

Figure 6a displays the shape of $\partial^2 f_{\mathbf{X}}/\partial x_1 \partial x_2$. The kernel considered in the formula of the nonparametric estimator (2) is a product kernel where

$$K\left(\frac{\mathbf{x} - \mathbf{X}}{h_n^{1/2}}\right) = K\left(\frac{x_1 - X_1}{h_n^{1/2}}\right) \times K\left(\frac{x_2 - X_2}{h_n^{1/2}}\right).$$
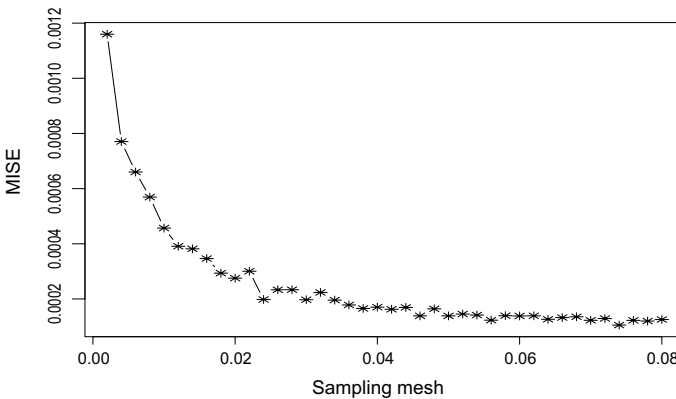

Fig. 6 MISE(δ) for bidimensional OU processes with $n = 105$

For simplicity and without lack of generalization, the same kernel (Gaussian in this case) and the same bandwidth are considered for X_1 and X_2. The bandwidth is selected according to the cross-validation criterion given in (8) and adapted to the two dimensional case, that is

$$h_{opt} = \arg\min_{h} \sum_{x_1 \in \mathcal{S}} \sum_{x_2 \in \mathcal{S}} \left(f''_{X_1 X_2, n, h}(x_1, x_2, \delta) - f''_{X_1 X_2}(x_1, x_2) \right)^2. \tag{9}$$

Moreover, the selection of the optimal sampling mesh is based on the following definition of the MISE:

$$\text{MISE}(\delta) := \frac{1}{N} \sum_{k=1}^{N} \int \int \left(f''_{X_1 X_2, n, k}(x_1, x_2, \delta) - f''_{X_1 X_2}(x_1, x_2) \right)^2 dx_1 dx_2.$$

Figure 6 displays the evolution of the MISE as a function of δ and one can observe that the optimal sampling mesh is $\delta^\star = 0.074$. Compared to the estimation of the first derivative of the density function for a one-dimensional OU, high frequency
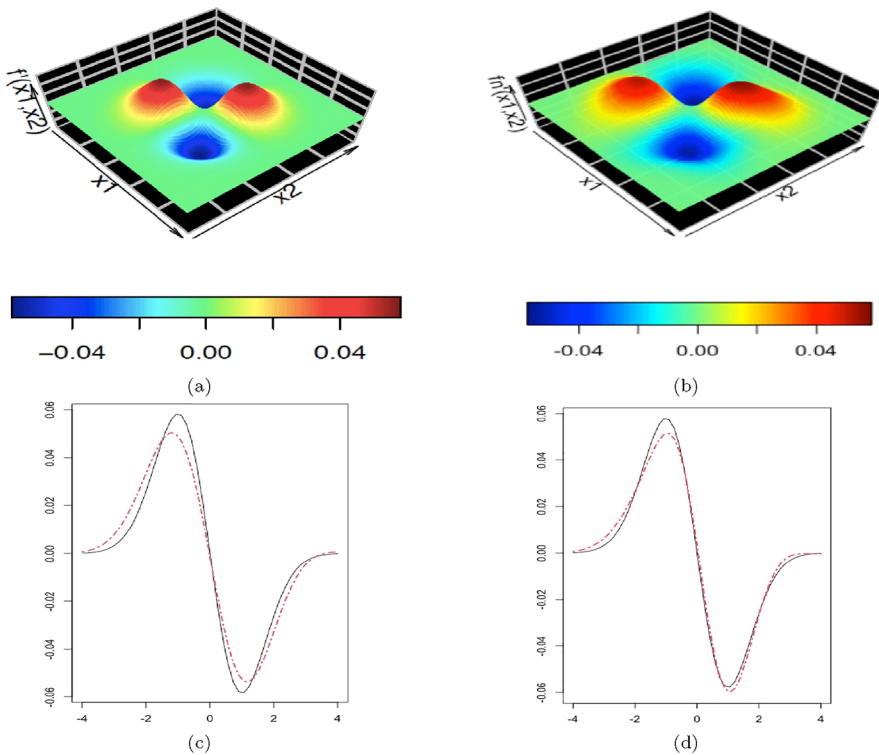

Fig. 7 **a** The true first derivative of the joint density function $f'(x_1, x_2)$. **b** The estimate of the first derivative of the joint density function $f'_n(x_1, x_2)$. **c** the solid line for $x_1 \to f'(x_1, x_2)$ versus the estimator $x_1 \to f'_n(x_1, x_2)$ in red dotted line for a fixed x_2. **d** the solid line for $x_2 \to f'(x_1, x_2)$ versus the estimator $x_2 \to f'_n(x_1, x_2)$ in red dotted line for a fixed x_1 (color figure online)

sampling of the bidimensional OU process is required to perform a consistent estimate of the first derivative of the joint density function $f''_{X_1 X_2}(x_1, x_2)$. Figure 7b displays the nonparametric estimate of $f''_{X_1 X_2}(x_1, x_2)$ based on the obtained optimal mesh $\delta^\star = 0.074$. Moreover, Fig. 7c (resp. (d)) shows the estimation of the first derivative of the marginal of X_1 (resp. X_2).

5.4 Estimation of the first derivative of the regression function

In this subsection, we are interested in the estimation of the first derivative of the regression function. For this, let us consider X_t an OU process solution of equation (6) and numerically generated as per equation (7). We also suppose that $\psi(Y_t) = Y_t$, where the responses Y_t are generated following regression model: $Y_{\tau_j} = m(X_{\tau_j}) + \epsilon_{\tau_j}$, $j = 0, 1, \ldots, n$, where ϵ's are generated from a standard normal distribution and $m(x) := \dfrac{1}{1 + x^2}$. As discussed in (5), the true first derivative of the regression function is define as:

$$m'(x) = \frac{r'(x)}{f(x)} - \frac{r(x)f'(x)}{f^2(x)}. \tag{10}$$

A natural estimator of $m'(x)$, say $m'_n(x)$, can be defined by plugging-in the above formula (10) $r'_n(x), f'_n(x), f_n(x)$ and $r_n(x)$ nonparametric estimators of $r'(x), f'(x), f(x)$ and $r(x)$, respectively. The calculation of $r'_n(x), f'_n(x), f_n(x)$ and $r_n(x)$ can be obtained as described in Sect. 2. One can easily notice that the case of estimating the first derivative if the regression function is more complicated than the estimation of the first derivative of univariate or bivariate density function. Indeed, in the last case we have to select only one bandwidth (or two in the bivariate case), whereas four different bandwidths should be selected for the estimation of the first derivative of the regression function. This makes the estimation task harder. In this simulation a separate cross-validation technique is used to select the optimal bandwidth for $r'_n(x), f'_n(x), f_n(x)$ and $r_n(x)$.

Remark 9 Another approach of selecting a *global bandwidth* for $m'_n(x)$ could be obtained by considering the following cross-validation criterion:

$$h_{opt} = \arg \min_h \sum_{x \in \mathcal{S}} \left(m'_{n,h}(x) - m'(x) \right)^2. \tag{11}$$

In contrast to the estimation of the regression function, where the cross-validation criterion is expressed as a function of the observed values of the response variable Y_i and the estimator of the regression function $m_{n,h}(X_i)$, the true value of the gradient is typically not observed. this makes the problem of bandwidth selection more difficult. Rice (1986) suggested the use of a differencing operator and a criterion which was shown to be a nearly unbiased estimator of the Mean Integrated Square Error (MISE) between the estimated derivative and the oracle. Müller et al. (1987) used Rice's noise-corrupted suggestion to select the bandwidth based on the
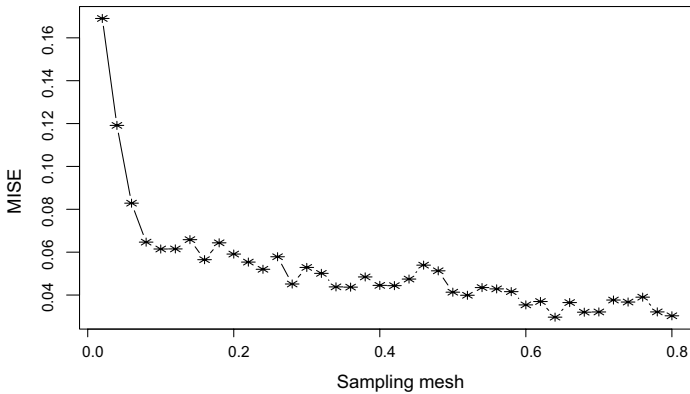
Fig. 8 MISE(δ) for the first derivative of the regression function with $n = 105$

natural extension of the least squares cross-validation. More recently, Henderson et al. (2015) generalized the previous approaches to the multivariate setting where local polynomial estimator was used.

Figure 8 shows that the optimal (in the sense of minimizing the mean squared error) sampling mesh for the estimator of the first derivative of the regression function is $\delta^\star = 0.64$ and the corresponding MISE(δ) for the first derivative regression is defined as follows:

$$\text{MISE}(\delta) := \frac{1}{N} \sum_{k=1}^{N} \int \left(m'_{n,k}(x,\delta) - m'(x) \right)^2 dx.$$

Figure 9a displays the shape of the regression function, whereas Fig. 9b shows the true first derivative as well as its estimate based on the obtained optimal sampling mesh.
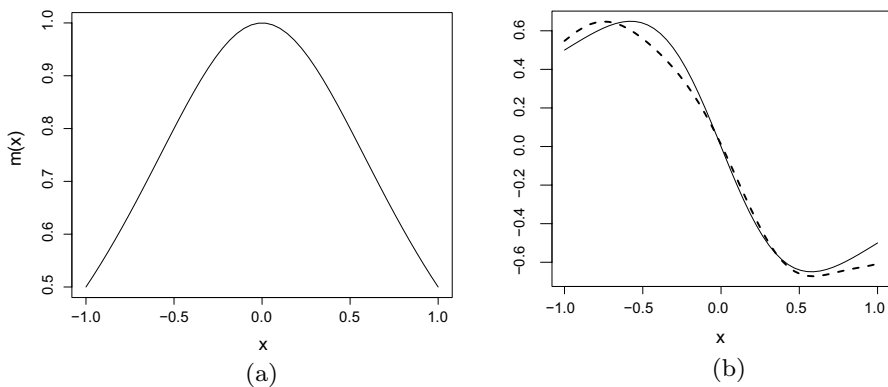

Fig. 9 **a** The true regression function $m(x)$. **b** Dark bold line displays the first derivative of the regression function and dotted line its estimation

6 Application to real data

In this section, we are interested in illustrating the estimation methodology on real data. For this one considers two asset prices which are the oil price (WTI) and the gold price. Figure 10 displays the daily time series of those asset prices from 02/01/1986 to 28/02/2018. One can observe a high correlation between the price of oil and the price of gold which is translated by a correlation coefficient equal to 0.8. It is well known that in most of the financial market analysis one can be interested in the log-return of the asset price rather that the price itself. For this reason we consider:

$$X_{1,t} = \ln(oil_t) - \ln(oil_{t-1}) \qquad \text{and} \qquad X_{2,t} = \ln(gold_t) - \ln(gold_{t-1}), \quad \text{for} \quad t = 2, \dots, n,$$

where n is the number of days from 02/01/1986 to 28/02/2018. Observe that the log-return processes of oil and gold are stationary.

We are interested in estimating the first derivative of the density function of oil and gold separately. Then, one considers the estimation of their joint density function. In this application to real data section, we consider a Gaussian kernel and select the bandwidth according to the cross-validation criterion, as tuning parameters. Figure 11a shows the kernel-type estimation of the density function of $X_{1,t}$ and Fig. 11b displays the estimation of its first derivative. Similarly, Fig. 12a, b correspond to the nonparametric estimate of the density function of $X_{2,t}$ and its first derivative respectively. Finally, one considers the pair of log-return of oil and gold $(X_{1,t}, X_{2,t})$ and we are interested in the nonparametric estimation of the first derivative of the joint density function. Figure 13 displays the shape of the first derivative, with respect to x_1 and x_2, of the joint pdf of log-return of oil and gold prices. One can observe that the derivative of the joint density is positive high values of oil and gold log-returns or whenever the log-return of oil is around zero and the log-return of gold is negative. In the contrary, the first derivative of the joint pdf is negative for negative log-returns of oil and gold or oil log-return is null and gold log-return is positive. Otherwise, the first derivative of the joint pdf is around zero.
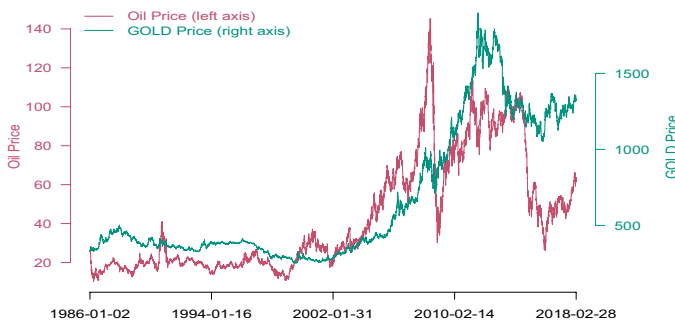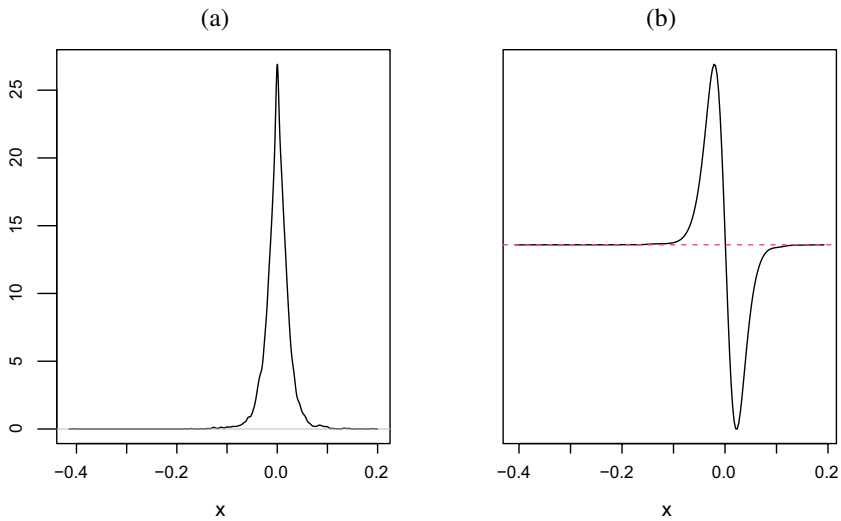

Fig. 10 Oil and Gold prices

Fig. 11 a The estimate of density function of the log-return of oil price. **b** The estimate of the first derivative of its density function. The red dotted line corresponds to the y-coordinate equal zero (color figure online)
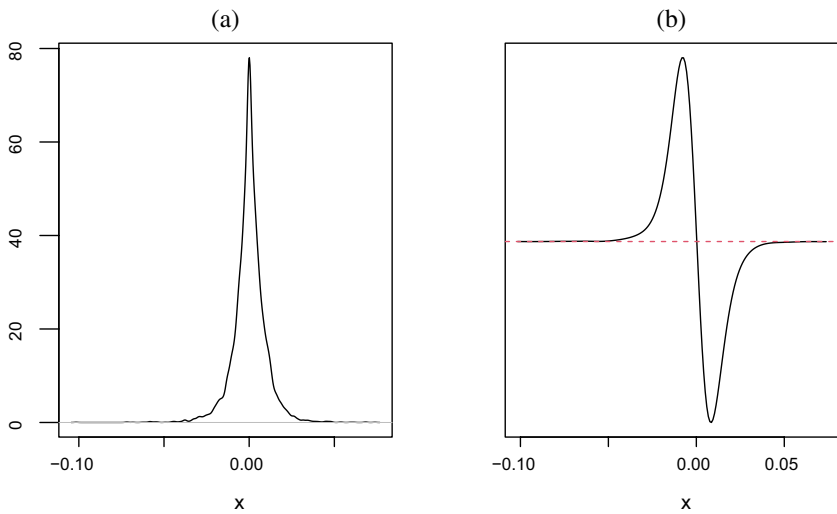

Fig. 12 a The estimate of density function of the log-return of gold price. **b** The estimate of the first derivative of its density function. The red dotted line corresponds to the *y*-coordinate equal zero (color figure online)

7 Concluding remarks

In the present paper, we have considered kernel type derivative estimators. We have extended and completed the existing work by relaxing the dependence assumption by assuming only the ergodicity of the process. We have obtained the almost sure

Fig. 13 The kernel-type estimate of the first derivative of the joint density function $f''_{x_1 x_2}(x_1, x_2)$.
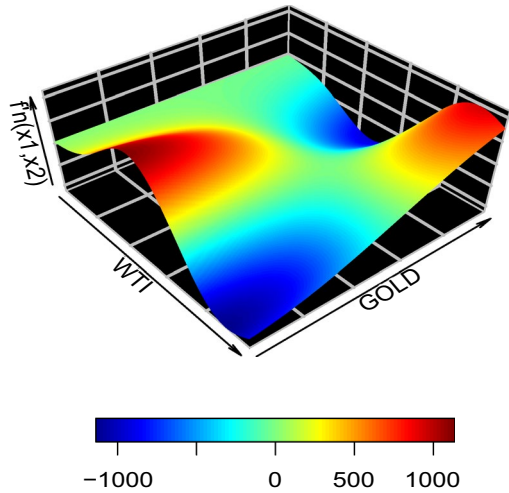
convergence rate that is close the i.i.d. framework. We have established the limiting distribution of the proposed estimators. An application concerning the regression derivatives is discussed theoretically as well as numerically. It would be interesting to extend our work to the case of censored data, which requires non trivial mathematics, this would go well beyond the scope of the present paper. Another direction of research is to enrich our results by considering the uniformity in terms of the bandwidth, that is an important question arising in some practical applications.

8 Mathematical developments

This section is devoted to the proofs of our results. The previously presented notation continues to be used in the following. The following technical lemma will be instrumental in the proof of our theorems.

Lemma 1 *Let $(Z_n)_{n \geq 1}$ be a sequence of real martingale differences with respect to the sequence of σ-fields $(\mathcal{F}_n = \sigma(Z_1, \ldots, Z_n))_{n \geq 1}$, where $\sigma(Z_1, \ldots, Z_n)$ is the σ-field generated by the random variables $Z_1, \ldots, Z_n$. Set $S_n = \sum_{i=1}^n Z_i$. For any $v \geq 2$ and any $n \geq 1$, assume that there exist some nonnegative constants C and d_n such that $\mathbb{E}(|Z_n|^v | \mathcal{F}_{n-1}) \leq C^{v-2} v! d_n^2$, almost surely. Then, for any $\epsilon > 0$, we have*

$$\mathbb{P}(|S_n| > \epsilon) \leq 2 \exp\left\{-\frac{\epsilon^2}{2(D_n + C\epsilon)}\right\},$$

where $D_n = \sum_{i=1}^n d_i^2$.

The proof follows as a particular case of Theorem 8.2.2 due to de la Peña and Giné (1999).

Lemma 2 *For any* $\mathbf{x} \in \mathbb{R}^p$, *we let*

$$H_n(\mathbf{x}) = \sum_{i=1}^n D^{|\mathbf{s}|} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}(\mathbf{x}) - n D^{|\mathbf{s}|} f(\mathbf{x}).$$

Under assumptions (*C*.4) *and* (*C*.5), *we have*

$$\sup_{\mathbf{x} \in \mathbb{R}^p} \|H_n(\mathbf{x})\|^2 = O(n).$$

Proof Following Wu (2003) and Wu et al (2010), and making use of the Cauchy-Schwarz inequality, one obtains

$$\left\|H_n(\mathbf{x})\right\|^2 \leq n\left(\sum_{i=1}^n \left\|\mathcal{P}_1 D^{|\mathbf{s}|} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}(\mathbf{x})\right\|^2\right).$$

Making use of the assumption (C.5), one infer that

$$\sup_{\mathbf{x} \in \mathbb{R}^p} \|H_n(\mathbf{x})\|^2 \leq n\left(\sup_{\mathbf{x} \in \mathbb{R}^p} \sum_{i=1}^\infty \left\|\mathcal{P}_k D^{|\mathbf{s}|} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}(\mathbf{x})\right\|\right) = O(n).$$

Hence the proof is complete.   □

Proposition 1 *Under the assumptions* (*K*.1)(*i*)–(*ii*), (*C*.1), (*R*.1), (*R*.2), *we have*

$$\sup_{\mathbf{x} \in \mathbf{J}} \left|D^{|\mathbf{s}|} r_n(\psi;\mathbf{x}, h_n) - \mathbb{E} D^{|\mathbf{s}|} r_n(\psi;\mathbf{x}, h_n)\right| = O\left(\sqrt{\frac{\log n}{n h^{1+s/p}}}\right). \tag{12}$$

8.1 Proof of Proposition 1.

Let us introduce the following notation

$$\widetilde{D}^{|\mathbf{s}|} r_n(\psi;\mathbf{x}, h_n) = \frac{1}{n h_n^{1+|\mathbf{s}|/p}} \sum_{i=1}^n \mathbb{E}\left[\psi(\mathbf{Y}_i) D^{|\mathbf{s}|} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right) \mid \mathcal{G}_{i-1}\right].$$

We next consider the following decomposition

$$\sup_{\mathbf{x} \in \mathbf{J}} \left|D^{|\mathbf{s}|} r_n(\psi;\mathbf{x}, h_n) - \mathbb{E}\left(D^{|\mathbf{s}|} r_n(\psi;\mathbf{x}, h_n)\right)\right|$$

$$\leq \sup_{\mathbf{x} \in \mathbf{J}} \left|D^{|\mathbf{s}|} r_n(\psi;\mathbf{x}, h_n) - \widetilde{D}^{|\mathbf{s}|} r_n(\psi;\mathbf{x}, h_n)\right|$$

$$+ \sup_{\mathbf{x} \in \mathbf{J}} \left|\widetilde{D}^{|\mathbf{s}|} r_n(\psi;\mathbf{x}, h_n) - \mathbb{E} D^{|\mathbf{s}|} r_n(\psi;\mathbf{x}, h_n)\right| = D_{n,1} + D_{n,2}.$$

Let $\{\mathbf{x}_k, k = 1, \ldots, l\} \subset \mathbf{J}$. Consider a the partition $\{\mathcal{S}_k\}_{1 \le k \le \ell}$ of the compact set $\mathbf{J}$ by a finite number l of spheres $\mathcal{S}_k$ centered upon by $\mathbf{x}_k$, with radius, for a positive constant a, $\mathbf{r} = a\left(\frac{h_n^{1/p}}{n}\right)^{1/\gamma}$. We have then $\mathbf{J} \subset \bigcup_{k=1}^{l} \mathcal{S}_k$. We readily infer that

$$
\begin{aligned}
D_{n,1} &= \sup_{\mathbf{x} \in \mathbf{J}} \left| D^{|\mathbf{s}|} r_n(\psi; \mathbf{x}, h_n) - \widetilde{D}^{|\mathbf{s}|} r_n(\psi; \mathbf{x}, h_n) \right| \\
&\le \max_{1 \le k \le l} \sup_{\mathbf{x} \in \mathcal{S}_k} \left| D^{|\mathbf{s}|} r_n(\psi; \mathbf{x}, h_n) - D^{|\mathbf{s}|} r_n(\psi; \mathbf{x}_k, h) \right| \\
&\quad + \max_{1 \le k \le l} \left| D^{|\mathbf{s}|} r_n(\psi; \mathbf{x}_k, h) - \widetilde{D}^{|\mathbf{s}|} r_n(\psi; \mathbf{x}_k, h) \right| \\
&\quad + \max_{1 \le k \le l} \sup_{\mathbf{x} \in \mathcal{S}_k} \left| \widetilde{D}^{|\mathbf{s}|} r_n(\psi; \mathbf{x}_k, h) - \widetilde{D}^{|\mathbf{s}|} r_n(\psi; \mathbf{x}, h) \right| \\
&= D_{n,1,1} + D_{n,1,2} + D_{n,1,3}.
\end{aligned}
\tag{13}
$$

Consider the first term of (13). Making use of the Cauchy-Schwarz inequality we readily obtain

$$
\begin{aligned}
&\left| D^{|\mathbf{s}|} r_n(\psi; \mathbf{x}, h_n) - D^{|\mathbf{s}|} r_n(\psi; \mathbf{x}_k, h) \right| \\
&= \left| \frac{1}{n h_n^{1+|\mathbf{s}|/p}} \sum_{i=1}^{n} \psi(\mathbf{Y}_i) \left(D^{|\mathbf{s}|} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right) - D^{|\mathbf{s}|} K\left(\frac{\mathbf{x}_k - \mathbf{X}_i}{h_n^{1/p}}\right) \right) \right| \\
&= \frac{1}{\sqrt{n} h_n^{1+|\mathbf{s}|/p}} \left(\frac{1}{n} \sum_{i=1}^{n} \psi^2(\mathbf{Y}_i) \right)^{1/2} \left(\sum_{i=1}^{n} \left(D^{|\mathbf{s}|} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}}\right) - D^{|\mathbf{s}|} K\left(\frac{\mathbf{x}_k - \mathbf{X}_i}{h_n^{1/p}}\right) \right)^2 \mathrm{d}t \right)^{1/2}.
\end{aligned}
$$

Keeping in mind the condition (K.1)(ii), we obtain that, almost surely,

$$
\begin{aligned}
&\left| D^{|\mathbf{s}|} r_n(\psi; \mathbf{x}, h_n) - D^{|\mathbf{s}|} r_n(\psi; \mathbf{x}_k, h) \right| \\
&\le \frac{1}{\sqrt{n} h_n^{1+|\mathbf{s}|/p}} \left(\mathbb{E}\left[\psi^2(\mathbf{Y}_1)\right] \right)^{1/2} \left(C_{K,s}^2 \sum_{i=1}^{n} \frac{\|\mathbf{x} - \mathbf{x}_k\|^{2\gamma}}{h_n^{2/p}} \right)^{1/2} \\
&\le \frac{C_{K,s}}{\sqrt{n} h_n^{1+|\mathbf{s}|/p}} \left(\mathbb{E}\left[\psi^2(\mathbf{Y}_1)\right] \right)^{1/2} \left(\frac{n \left(a \left(h_n^{1/p}/n \right)^{1/\gamma} \right)^{2\gamma}}{h_n^{2/p}} \right)^{1/2} \\
&\le \frac{a^\gamma C_{K,s}}{n h_n^{1+|\mathbf{s}|/p}} \left(\mathbb{E}\left[\psi^2(\mathbf{Y}_1)\right] \right)^{1/2}.
\end{aligned}
$$

Therefore, by considering the following choice $\epsilon_n = \left(\frac{\log n}{n h_n^{1+|\mathbf{s}|/p}} \right)^{1/2}$, we have

$$
\epsilon_n^{-1} D_{n,1,1} = O_{a.s}\left(\frac{1}{n h_n^{1+|\mathbf{s}|/p} \log n} \right)^{1/2}.
\tag{14}
$$

In view of the condition (K.1)(ii), we infer that, almost surely,

$$
\left| \widetilde{D}^{|s|} r_n(\psi; \mathbf{x}, h_n) - \widetilde{D}^{|s|} r_n(\psi; \mathbf{x}_k, h) \right|
$$

$$
\leq \frac{1}{h_n^{1+|s|/p}} \left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\psi(\mathbf{Y}_i) \mid \mathcal{G}_{i-1} \right] \right) \left(C_{K,s} \frac{\|\mathbf{x} - \mathbf{x}_k\|^{\gamma}}{h_n^{1/p}} \right)
$$

$$
\leq \frac{1}{h_n^{1+|s|/p}} \left(\mathbb{E}\left[\psi(\mathbf{Y}_1) \right] \right) \left(C_{K,s} \frac{\left(a\left(h_n^{1/p}/n \right)^{1/\gamma} \right)^{\gamma}}{h_n^{1/p}} \right) \leq \frac{a^{\gamma} C_{K,s}}{n h_n^{1+|s|/p}} \left(\mathbb{E}\left[\psi(\mathbf{Y}_1) \right] \right).
$$

We have then

$$
\epsilon_n^{-1} D_{n,1,3} = O_{a.s} \left(\frac{1}{n h_n^{1+|s|/p} \log n} \right)^{1/2}. \tag{15}
$$

We now deal with the term $D_{n,1,2}$ of the decomposition give, in equation (13). We first observe that we have

$$
D_{n,1,2} = \frac{1}{n h_n^{1+|s|/p}} \sum_{i=1}^{n} \left(\psi(\mathbf{Y}_i) D^{|s|} K\left(\frac{\mathbf{x}_k - \mathbf{X}_i}{h_n^{1/p}} \right) - \mathbb{E}\left[\psi(\mathbf{Y}_i) D^{|s|} K\left(\frac{\mathbf{x}_k - \mathbf{X}_i}{h_n^{1/p}} \right) \mid \mathcal{G}_{i-1} \right] \right)
$$

$$
= \frac{1}{n h_n^{1+|s|/p}} \sum_{i=1}^{n} R_i(\mathbf{x}_k),
$$

where

$$
R_i(\mathbf{x}_k) = \psi(\mathbf{Y}_i) D^{|s|} K\left(\frac{\mathbf{x}_k - \mathbf{X}_i}{h_n^{1/p}} \right) - \mathbb{E}\left[\psi(\mathbf{Y}_i) D^{|s|} K\left(\frac{\mathbf{x}_k - \mathbf{X}_i}{h_n^{1/p}} \right) \mid \mathcal{G}_{i-1} \right].
$$

We observe that the sequence $\{ R_i(\mathbf{x}_k) \}_{0 \leq i \leq n}$ is a sequence of martingale differences. For $v \geq 2$, we have

$$
R_i^{v}(\mathbf{x}_k) = \sum_{j=0}^{v} C_v^{j} \left(\psi(\mathbf{Y}_i) D^{|s|} K\left(\frac{\mathbf{x}_k - \mathbf{X}_i}{h_n^{1/p}} \right) \right)^{j} (-1)^{v-j} \left(\mathbb{E}\left[\psi(\mathbf{Y}_i) D^{|s|} K\left(\frac{\mathbf{x}_k - \mathbf{X}_i}{h_n^{1/p}} \right) \mid \mathcal{G}_{i-1} \right] \right)^{v-j},
$$

thus, we have

$$
\left| \mathbb{E}\left[R_i^{v}(\mathbf{x}_k) \mid \mathcal{G}_{i-1} \right] \right|
$$

$$
= \sum_{j=0}^{v} C_v^{j} \mathbb{E}\left[\left| \psi(\mathbf{Y}_i) D^{|s|} K\left(\frac{\mathbf{x}_k - \mathbf{X}_i}{h_n^{1/p}} \right) \right|^{j} \mid \mathcal{G}_{i-1} \right] \left(\mathbb{E}\left[\left| \psi(\mathbf{Y}_i) D^{|s|} K\left(\frac{\mathbf{x}_k - \mathbf{X}_i}{h_n^{1/p}} \right) \right| \mid \mathcal{G}_{i-1} \right] \right)^{v-j}.
$$

Making use of Jensen's inequality, we have

$$\mathbb{E}\left[\left|\psi(\mathbf{Y}_i)D^{|\mathbf{s}|}K\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_n^{1/p}}\right)\right|^j \mid \mathcal{G}_{i-1}\right]\left(\mathbb{E}\left[\left|\psi(\mathbf{Y}_i)D^{|\mathbf{s}|}K\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_n^{1/p}}\right)\right| \mid \mathcal{G}_{i-1}\right]\right)^{\nu-j}$$

$$\leq \mathbb{E}\left[\left|\psi(\mathbf{Y}_i)D^{|\mathbf{s}|}K\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_n^{1/p}}\right)\right|^j \mid \mathcal{G}_{i-1}\right]\mathbb{E}\left[\left|\psi(\mathbf{Y}_i)D^{|\mathbf{s}|}K\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_n^{1/p}}\right)\right|^{\nu-j} \mid \mathcal{G}_{i-1}\right].$$

Observe that for any $m \geq 1$, under assumption (R.1)(iii), we have

$$\mathbb{E}\left[\left|\psi(\mathbf{Y}_i)D^{|\mathbf{s}|}K\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_n^{1/p}}\right)\right|^m \mid \mathcal{G}_{i-1}\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[|\psi(\mathbf{Y}_i)|^m\left(D^{|\mathbf{s}|}K\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_n^{1/p}}\right)\right)^m \mid \mathcal{S}_i\right] \mid \mathcal{G}_{i-1}\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[|\psi(\mathbf{Y}_i)|^m \mid \mathcal{S}_i\right]\left(D^{|\mathbf{s}|}K\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_n^{1/p}}\right)\right)^m \mid \mathcal{G}_{i-1}\right]$$

$$= \mathbb{E}\left[\Psi_m(\mathbf{X}_i)\left(D^{|\mathbf{s}|}K\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_n^{1/p}}\right)\right)^m \mid \mathcal{G}_{i-1}\right]$$

$$\leq \mathbb{E}\left[|\Psi_m(\mathbf{X}_i)-\Psi_m(\mathbf{x})|\left(D^{|\mathbf{s}|}K\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_n^{1/p}}\right)\right)^m \mid \mathcal{G}_{i-1}\right]$$

$$+ \mathbb{E}\left[\Psi_m(\mathbf{x})\left(D^{|\mathbf{s}|}K\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_n^{1/p}}\right)\right)^m \mid \mathcal{G}_{i-1}\right]$$

$$\leq \mathbb{E}\left[\left(D^{|\mathbf{s}|}K\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_n^{1/p}}\right)\right)^m \mid \mathcal{G}_{i-1}\right]\left(\Psi_m(\mathbf{x})+\sup_{\mathbf{u}\in B(\mathbf{x},h_n^{1/p})}|\Psi_m(\mathbf{u})-\Psi_m(\mathbf{x})|\right)$$

$$= C_{0,\psi}\mathbb{E}\left[\left(D^{|\mathbf{s}|}K\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_n^{1/p}}\right)\right)^m \mid \mathcal{G}_{i-1}\right],$$

where $C_{0,\psi}$ is a positive constant. By a simple change of variable, we obtain that

$$\mathbb{E}\left[\left(D^{|\mathbf{s}|}K\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_n^{1/p}}\right)\right)^m \mid \mathcal{G}_{i-1}\right] = \int_{\mathbb{R}^p}\left(D^{|\mathbf{s}|}K\right)^m\left(\frac{\mathbf{x}_k-\mathbf{u}}{h_n^{1/p}}\right)f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}(\mathbf{u})\mathrm{d}\mathbf{u}$$

$$= h_n\int_{\mathbb{R}^p}\left(D^{|\mathbf{s}|}K\right)^m(\mathbf{v})f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}(\mathbf{x}_k-h_n^{1/p}\mathbf{v})\mathrm{d}\mathbf{v}.$$

In the light of the assumption (K.1)(i), we know that the kernel $K(\cdot)$ is a compactly supported, this implies that $D^{|\mathbf{s}|}K(\mathbf{x}) \leq \Gamma_K$. Making use of the assumption (C.1) in combination with an integration by parts repeated $|\mathbf{s}|$ times and a Taylor's expansion of order 1, implies that we have

$$\mathbb{E}\left[\left(D^{|\mathbf{s}|}K\left(\frac{\mathbf{x}_k - \mathbf{X}_i}{h_n^{1/p}}\right)\right)^m \mid \mathcal{G}_{i-1}\right]$$

$$\leq h_n \Gamma_K^{m-1} \int_{\mathbb{R}^p} D^{|\mathbf{s}|}K(\mathbf{v}) f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}(\mathbf{x}_k - h_n^{1/p}\mathbf{v}) d\mathbf{v}$$

$$= \Gamma_K^{m-1} h_n^{1+|\mathbf{s}|/p} \int_{\mathbb{R}^p} K(\mathbf{v}) D^{|\mathbf{s}|} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}(\mathbf{x}_k - h_n^{1/p}\mathbf{v}) d\mathbf{v}$$

$$= \Gamma_K^{m-1} h_n^{1+|\mathbf{s}|/p} \int_{\mathbb{R}^p} K(\mathbf{v})\left(D^{|\mathbf{s}|} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}(\mathbf{x}_k) + O\left(h_n^{\Bbbk/p}\right)\right) d\mathbf{v}$$

$$= \Gamma_K^{m-1} h_n^{1+|\mathbf{s}|/p}\left(D^{|\mathbf{s}|} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}(\mathbf{x}_k) + O\left(h_n^{\Bbbk/p}\right)\right).$$

Since $h_n^{2(1+|\mathbf{s}|/p)} \leq h_n^{1+|\mathbf{s}|/p}$, we readily obtain that

$$\begin{aligned}
\left|\mathbb{E}\left[R_I^\nu(\mathbf{x}_k) \mid \mathcal{G}_{i-1}\right]\right| &= C_{0,\psi}^2 \Gamma_K^{\nu-2} h_n^{2(1+|\mathbf{s}|/p)}\left(\left(D^{|\mathbf{s}|} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}\right)^2(\mathbf{x}_k) + O\left(h_n^{\Bbbk/p}\right)\right) \sum_{j=0}^\nu C_\nu^j \\
&= 2^\nu C_{0,\psi}^2 \Gamma_K^{\nu-2} h_n^{1+|\mathbf{s}|/p}\left(\left(D^{|\mathbf{s}|} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}\right)^2(\mathbf{x}_k) + O\left(h_n^{\Bbbk/p}\right)\right) \\
&= \nu! \Gamma_K^{\nu-2} h_n^{1+|\mathbf{s}|/p}\left[\left(D^{|\mathbf{s}|} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}\right)^2(\mathbf{x}_k) + O\left(h_n^{\Bbbk/p}\right)\right] \\
&= \nu! C^{\nu-2} h_n^{1+|\mathbf{s}|/p}\left[\left(D^{|\mathbf{s}|} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}\right)^2(\mathbf{x}_k) + O\left(h_n^{\Bbbk/p}\right)\right],
\end{aligned}$$

where $C = \Gamma_K$. By choosing that

$$d_i^2 = h_n^{1+|\mathbf{s}|/p}\left[\left(D^{|\mathbf{s}|} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}\right)^2(\mathbf{x}_k) + O\left(h_n^{\Bbbk/p}\right)\right],$$

we have

$$D_n = \sum_{i=1}^n d_i^2 = n h_n^{1+|\mathbf{s}|/p}\left[\frac{1}{n}\sum_{i=1}^n \left(D^{|\mathbf{s}|} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}\right)^2(\mathbf{x}_k) + O\left(h_n^{\Bbbk/p}\right)\right],$$

where $\left(\left(D^{|\mathbf{s}|} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}\right)^2(\mathbf{x}_k)\right)_{1\leq i\leq n}$ is a sequence of stationary and ergodic sequence. Hence we obtain that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \left(D^{|\mathbf{s}|} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}\right)^2(\mathbf{x}) = \left(D^{|\mathbf{s}|}f\right)^2(\mathbf{x}).$$

The use of the assumption (C.3) implies that

$$D_n = n h_n^{1+|\mathbf{s}|/p}\left[\left(D^{|\mathbf{s}|}f\right)^2(\mathbf{x}_k) + O\left(h_n^{\Bbbk/p}\right)\right] = O\left(n h_n^{1+|\mathbf{s}|/p}\right).$$

Now, taking $\epsilon_n = \epsilon_0 \left(\dfrac{\log n}{n h_n^{1+|\mathbf{s}|/p}} \right)^{1/2}$ and an application of Lemma 1 gives that

$$\mathbb{P}\{|D_{n,1,2}| > \epsilon_n\} = \mathbb{P}\left\{ \left| \max_{1 \le k \le l} \sum_{i=1}^{n} R_i(\mathbf{x}_k) \right| > \epsilon_n (n h_n^{1+|\mathbf{s}|/p}) \right\}$$

$$\le \sum_{i=1}^{\ell} \exp\{ - \frac{\epsilon_n^2 \left(n h_n^{1+|\mathbf{s}|/p} \right)^2}{2 \left(D_n + C \left(n h_n^{1+|\mathbf{s}|/p} \right) \epsilon_0 \left(\dfrac{\log n}{n h_n^{1+|\mathbf{s}|/p}} \right)^{1/2} \right)} \}$$

$$\le 2\ell \exp\{ - \frac{\epsilon_0^2 \left(n h_n^{1+|\mathbf{s}|/p} \right) \log n}{O\left(n h_n^{1+|\mathbf{s}|/p} \right) \left(1 + C \epsilon_0 \left(\dfrac{\log n}{n h_n^{1+|\mathbf{s}|/p}} \right)^{1/2} \right)} \}$$

$$= 2\ell \exp\left\{ \log n^{-\epsilon_0^2 C_1} \right\} = 2\ell n^{-\epsilon_0^2 C_1},$$

where C_1 is a positive constant. Hence, for ϵ large enough, we obtain

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{ \max_{1 \le k \le l} \left| \sum_{i=1}^{n} R_i(\mathbf{x}_k) \right| > \epsilon_n (n h_n^{1+|\mathbf{s}|/p}) \right\} < \infty.$$

The proof is completed by a routine application of Borel–Cantelli lemma. This, in turn, implies that

$$\frac{1}{n h_n^{1+|\mathbf{s}|/p}} \max_{1 \le k \le l} \left| \sum_{i=1}^{n} R_i(\mathbf{x}_k) \right| = O\left(\left(\frac{\log n}{n h_n^{1+|\mathbf{s}|/p}} \right)^{1/2} \right), \quad \text{a.s}.$$

Hence, we have

$$D_{n,1,2} = O_{a.s} \left(\frac{\log n}{n h_n^{1+|\mathbf{s}|/p}} \right)^{1/2}. \tag{16}$$

By combining the statements (14), (15) and (16), we obtain

$$D_{n,1} = O_{a.s} \left(\frac{\log n}{n h_n^{1+|\mathbf{s}|/p}} \right)^{1/2}. \tag{17}$$

Consider now, the second term of the decomposition given in equation (13), under assumption (R.1)(i), we infer that

$$D_{n,2} = \sup_{\mathbf{x} \in \mathbf{J}} \left| \widetilde{D}^{|\mathbf{s}|} r_n(\psi; \mathbf{x}, h_n) - \mathbb{E} D^{|\mathbf{s}|} r_n(\psi; \mathbf{x}, h_n) \right|$$

$$= \frac{1}{n h_n^{1+|\mathbf{s}|/p}} \sup_{\mathbf{x} \in \mathbf{J}} \left| \sum_{i=1}^{n} \left(\mathbb{E} \left[\mathbb{E} \left[\psi(\mathbf{Y}_i) D^{|\mathbf{s}|} K \left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}} \right) \mid \mathcal{S}_{i-1} \right] \mid \mathcal{G}_{i-1} \right] \right. \right.$$

$$\left. \left. - \mathbb{E} \left[\mathbb{E} \left[\psi(\mathbf{Y}_i) D^{|\mathbf{s}|} K \left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}} \right) \mid \mathcal{S}_{i-1} \right] \right] \right) \right|$$

$$= \frac{1}{n h_n^{1+|\mathbf{s}|/p}} \sup_{\mathbf{x} \in \mathbf{J}} \left| \sum_{i=1}^{n} \left(\mathbb{E} \left[\mathbb{E} \left[\psi(\mathbf{Y}_i) \mid \mathcal{S}_{i-1} \right] D^{|\mathbf{s}|} K \left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}} \right) \mid \mathcal{G}_{i-1} \right] \right. \right.$$

$$\left. \left. - \mathbb{E} \left[\mathbb{E} \left[\psi(\mathbf{Y}_i) \mid \mathcal{S}_{i-1} \right] D^{|\mathbf{s}|} K \left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}} \right) \right] \right) \right|$$

$$= \frac{1}{n h_n^{1+|\mathbf{s}|/p}} \sup_{\mathbf{x} \in \mathbf{J}} \left| \sum_{i=1}^{n} \left(\mathbb{E} \left[m(\mathbf{X}_i) D^{|\mathbf{s}|} K \left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}} \right) \mid \mathcal{G}_{i-1} \right] \right. \right.$$

$$\left. \left. - \mathbb{E} \left[m(\mathbf{X}_i) D^{|\mathbf{s}|} K \left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n^{1/p}} \right) \right] \right) \right|.$$

A simple change of variables and making use of the assumption (R.1)(ii), we obtain

$$D_{n,2} = \frac{1}{n h_n^{1+|\mathbf{s}|/p}} \sup_{\mathbf{x} \in \mathbf{J}} \left| \sum_{i=1}^{n} \int_{\mathbb{R}^p} m(\mathbf{u}, \psi) D^{|\mathbf{s}|} K \left(\frac{\mathbf{x} - \mathbf{u}}{h_n^{1/p}} \right) \left(f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}(\mathbf{u}) - f(\mathbf{u}) \right) d\mathbf{u} \right|$$

$$\leq \frac{1}{n h_n} \sup_{\mathbf{x} \in \mathbf{J}} \left| \sup_{\|\mathbf{u} - \mathbf{x}\| \leq \lambda h_n} |m(\mathbf{u}, \psi) - m(\mathbf{x}, \psi)| + m(\mathbf{x}, \psi) \right|$$

$$\sup_{\mathbf{x} \in \mathbf{J}} \left| \sum_{i=1}^{n} \int_{\mathbb{R}^p} K \left(\frac{\mathbf{x} - \mathbf{u}}{h_n^{1/p}} \right) \left(D^{|\mathbf{s}|} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}(\mathbf{u}) - D^{|\mathbf{s}|} f(\mathbf{u}) \right) d\mathbf{u} \right|$$

$$= \frac{1}{n h_n} \left(\sup_{\mathbf{x} \in \mathbf{J}} |m(\mathbf{x}, \psi)| + O(h_n) \right)$$

$$\times \sup_{\mathbf{x} \in \mathbf{J}} \left| \sum_{i=1}^{n} \int_{\mathbb{R}^p} K \left(\frac{\mathbf{x} - \mathbf{u}}{h_n^{1/p}} \right) \left(D^{|\mathbf{s}|} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}(\mathbf{u}) - D^{|\mathbf{s}|} f(\mathbf{u}) \right) d\mathbf{u} \right|$$

$$= \frac{1}{n h_n} \mathbf{C_m} \sup_{\mathbf{x} \in \mathbf{J}} \left| \int_{\mathbb{R}^p} K \left(\frac{\mathbf{x} - \mathbf{u}}{h_n^{1/p}} \right) \left(\sum_{i=1}^{n} \left(D^{|\mathbf{s}|} f_{\mathbf{X}_i}^{\mathcal{G}_{i-1}}(\mathbf{u}) - D^{|\mathbf{s}|} f(\mathbf{u}) \right) \right) d\mathbf{u} \right|$$

$$\leq \frac{1}{n h_n} C_{\mathbf{m}} \sup_{\mathbf{x} \in \mathbf{J}} \left| \int_{\mathbb{R}^p} K \left(\frac{\mathbf{x} - \mathbf{u}}{h_n^{1/p}} \right) \| H_n(\mathbf{x}) \| d\mathbf{u} \right|$$

$$\leq O(n^{-1/2}) C_{\mathbf{m}} \sup_{\mathbf{x} \in \mathbf{J}} \left| \frac{1}{h_n} \int_{\mathbb{R}^p} K \left(\frac{\mathbf{x} - \mathbf{u}}{h_n^{1/p}} \right) d\mathbf{u} \right|,$$

where $C_{\mathbf{m}} = \left(\sup\limits_{\mathbf{x} \in \mathbf{J}} |m(\mathbf{x}, \psi)| + O(h_n) \right)$. Hence we have

$$D_{n,2} = O(n^{-1/2}). \tag{18}$$

Combining the statements (17) and (18), we obtain the desired result given in (12).

<div align="right">□</div>

References

Abdous, B., Germain, S., Ghazzali, N. (2002). A unified treatment of direct and indirect estimation of a probability density and its derivatives. *Statistics & Probability Letters, 56*(3), 239–250.

Akaike, H. (1954). An approximation to the density function. *Annals of the Institute of Statistical Mathematics, 6*, 127–132.

Blanke, D., Pumo, B. (2003). Optimal sampling for density estimation in continuous time. *Journal of Time Series Analysis, 24*, 1–23.

Bouzebda, S., Didi, S. (2017a). Multivariate wavelet density and regression estimators for stationary and ergodic discrete time processes: Asymptotic results. *Communications in Statistics-Theory and Methods, 46*(3), 1367–1406.

Bouzebda, S., Didi, S. (2017b). Additive regression model for stationary and ergodic continuous time processes. *Communications in Statistics-Theory and Methods, 46*(5), 2454–2493.

Bouzebda, S., Didi, S. (2021). Some asymptotic properties of kernel regression estimators of the mode for stationary and ergodic continuous time processes. *Revista Matemática Complutense, 34*(3), 811–852.

Bouzebda, S., Didi, S., El Hajj, L. (2015). Multivariate wavelet density and regression estimators for stationary and ergodic continuous time processes: Asymptotic results. *Mathematical Methods of Statistics, 24*(3), 163–199.

Bouzebda, S., Chaouch, M., Laïb, N. (2016). Limiting law results for a class of conditional mode estimates for functional stationary ergodic data. *Mathematical Methods of Statistics, 25*(3), 168–195.

Bradley, R. C. (2007). *Introduction to strong mixing conditions* (Vol. 1). Heber City, UT: Kendrick Press.

Chacón, J. E., Duong, T. (2013). Data-driven density derivative estimation, with applications to nonparametric clustering and bump hunting. *Electronic Journal of Statistics, 7*, 499–532.

Chaouch, M., Laïb, N. (2019). Optimal asymptotic MSE of kernel regression estimate for continuous time processes with missing at random response. *Statistics & Probability Letters, 154*(2), 161–178.

Charnigo, R., Hall, B., Srinivasan, C. (2011). A generalized C_p criterion for derivative estimation. *Technometrics, 53*(3), 238–253.

Cheng, K. F. (1982). On estimation of a density and its derivatives. *The Annals of Mathematical Statistics, 34*(3), 479–489.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74*(368), 829–836.

Comaniciu, D., Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(5), 603–619.

Deheuvels, P. (2011). One bootstrap suffices to generate sharp uniform bounds in functional estimation. *Kybernetika (Prague), 47*(6), 855–865.

Deheuvels, P., Mason, D. M. (2004). General asymptotic confidence bands based on kernel-type function estimators. *Statistical Inference for Stochastic Processes, 7*(3), 225–277.

de la Peña, V. H., Giné, E. (1999). *Decoupling. From dependence to independence, Randomly stopped processes. U-statistics and processes. Martingales and beyond. Probability and its Applications* (*New York*). New York: Springer.

Delecroix, M. (1987). Sur l'estimation et la prévision non-paramétrique des processus ergodiques. Doctorat d'État. Université des sciences de Lille, Flandre-Artois.

Delecroix, M., Rosa, A. C. (1996). Nonparametric estimation of a regression function and its derivatives under an ergodic hypothesis. *Journal of Nonparametric Statistics, 6*(4), 367–382.

Devroye, L. (1987). *A course in density estimation, vol. 14 of Progress in Probability and Statistics*. Boston, MA: Birkhäuser Boston, Inc.

Devroye, L., Györfi, L. (1985). *Nonparametric density estimation*. The L_1view. Wiley Series in Probability and Mathematical Statistics: *Tracts on Probability and Statistics*. New York: Wiley, Inc.

Devroye, L., Lugosi, G. (2001). *Combinatorial methods in density estimation*. Springer Series in Statistics. New York: Springer.

Eggermont, P. P. B., LaRiccia, V. N. (2001). *Maximum penalized likelihood estimation. Density estimation. Vol. I. Springer Series in Statistics*. New York: Springer.

Eubank, R. L., Speckman, P. L. (1993). Confidence bands in nonparametric regression. *Journal of the American Statistical Association, 88*(424), 1287–1301.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association, 87*(420), 998–1004.

Fan, J., Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society*: Series B (*Methodological*), *57*(2), 371–394.

Fan, J., Gijbels, I. (1996). *Local polynomial modelling and its applications, vol. 66 of Monographs on Statistics and Applied Probability*. London: Chapman & Hall.

Fukunaga, K., Hostetler, L. D. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, IT-21, 32–40.

Gasser, T., Müller, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics, 11*(3), 171–185.

Genovese, C. R., Perone-Pacifico, M., Verdinelli, I., Wasserman, L. A. (2013). Nonparametric inference for density modes. *CoRR*, abs/1312.7567.

Georgiev, A. A. (1984). Speed of convergence in nonparametric kernel estimation of a regression function and its derivatives. *Annals of the Institute of Statistical Mathematics, 36*(3), 455–462.

Härdle, W. (1990). *Applied nonparametric regression, vol. 19 of Econometric Society Monographs*. Cambridge: Cambridge University Press.

Härdle, W., Gasser, T. (1985). On robust kernel estimation of derivatives of regression functions. *Scandinavian Journal of Statistics, 12*(3), 233–240.

Härdle, W., Marron, J. S., Wand, M. P. (1990). Bandwidth choice for density derivatives. *Journal of the Royal Statistical Society*: Series B (*Methodological*), *52*(1), 223–232.

Henderson, D. J., Parmeter, C. F. (2012a). Canonical higher-order kernels for density derivative estimation. *Statistics & Probability Letters, 82*(7), 1383–1387.

Henderson, D. J., Parmeter, C. F. (2012b). Normal reference bandwidths for the general order, multivariate kernel density derivative estimator. *Statistics & Probability Letters, 82*(12), 2198–2205.

Henderson, D. J., Li, Q., Parmeter, C. F., Yao, S. (2015). Gradient-based smoothing parameter selection for nonparametric regression estimation. *Journal of Econometrics, 184*, 233–241.

Herrmann, E., Ziegler, K. (2004). Rates on consistency for nonparametric estimation of the mode in absence of smoothness assumptions. *Statistics & Probability Letters, 68*(4), 359–368.

Horová, I., Vieu, P., Zelinka, J. (2002). Optimal choice of nonparametric estimates of a density and of its derivatives. *Statistics & Risk Modeling, 20*(4), 355–378.

Jones, M. C. (1994). On kernel density derivative estimation. *Communications in Statistics-Theory and Methods, 23*(8), 2133–2139.

Karunamuni, R. J., Mehra, K. L. (1990). Improvements on strong uniform consistency of some known kernel estimates of a density and its derivatives. *Statistics & Probability Letters, 9*(2), 133–140.

Kloeden, P., Platen, E. (1992). *Numerical solution of stochastic differential equations. Applications of Mathematics*, 23, Berlin: Springer.

Krebs, J. T. N. (2019). The bootstrap in kernel regression for stationary ergodic data when both response and predictor are functions. *Journal of Multivariate Analysis, 173*, 620–639.

Krengel, U. (1985). *Ergodic theorems, vol. 6 of de Gruyter Studies in Mathematics*. Berlin: Walter de Gruyter & Co. With a supplement by Antoine Brunel.

Leucht, A., Neumann, M. H. (2013). Degenerate *U*- and *V*-statistics under ergodicity: Asymptotics, bootstrap and applications in statistics. *Annals of the Institute of Statistical Mathematics, 65*(2), 349–386.

Meyer, T. G. (1977). Bounds for estimation of density functions and their derivatives. *The Annals of Statistics, 5*(1), 136–142.

Müller, G. H., Stadmüller, U., Schmitt, T. (1987). Bandwidth choice and confidence intervals for derivatives of noisy data. *Biometrika, 74*(4), 743–749.

Nadaraya, E. (1964). On estimating regression. *Theory of Probability & Its Applications, 9*, 157–159.

Nadaraja, E. A. (1969). Nonparametric estimates of the derivatives of a probability density and a regression function. *Sakharth. SSR Mecn. Akad. Moambe, 55*, 29–32.

Nadaraya, E. A. (1989). *Nonparametric estimation of probability densities and regression curves, volume 20 of Mathematics and its Applications* (*Soviet Series*). Dordrecht: Kluwer Academic Publishers Group. Translated from the Russian by Samuel Kotz.

Neumann, M. H. (2011). Absolute regularity and ergodicity of Poisson count processes. *Bernoulli, 17*(4), 1268–1284.

Noh, Y., Sugiyama, M., Liu, S., d. Plessis, M. C., Park, F. C., Lee, D. D. (2018). Bias reduction and metric learning for nearest-neighbor estimation of kullback-leibler divergence. *Neural Computation, 30*(7), 1930–1960.

Park, C., Kang, K.-H. (2008). SiZer analysis for the comparison of regression curves. *Computational Statistics & Data Analysis, 52*(8), 3954–3970.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics, 33*, 1065–1076.

Racine, J. (2016). Local polynomial derivative estimation: Analytic or Taylor? *Advances in Econometrics, 36*, 617–633.

Ramsay, J. O, Silverman, B. W. (2002). *Applied functional data analysis. Methods and case studies*. Springer Series in Statistics. New York: Springer.

Ramsay, J. O., Silverman, B. W. (2005). *Functional data analysis. Springer Series in Statistics, second edition*. New York: Springer.

Rice, J. S. (1986). Bandwidth choice for differentiation. *Journal of Multivariate Analysis, 19*, 251–264.

Rice, J. S., Rosenblatt, M. (1983). Smoothing splines: Regression, derivatives and deconvolution. *The Annals of Statistics, 11*(1), 141–156.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics, 27*, 832–837.

Ruppert, D., Sheather, S. J., Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association, 90*(432), 1257–1270.

Sasaki, H., Noh, Y.-K., Niu, G., Sugiyama, M. (2016). Direct density derivative estimation. *Neural Computation, 28*(6), 1101–1140.

Schuster, E. F. (1969). Estimation of a probability density function and its derivatives. *The Annals of Mathematical Statistics, 40*, 1187–1195.

Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics, 6*(1), 177–184.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. London: Chapman & Hall.

Singh, R. S. (1976). Nonparametric estimation of mixed partial derivatives of a multivariate density. *Journal of Multivariate Analysis, 6*(1), 111–122.

Singh, R. S. (1977). Applications of estimators of a density and its derivatives to certain statistical problems. *Journal of the Royal Statistical Society*: Series B, *39*(3), 357–363.

Singh, R. S. (1979). Mean squared errors of estimates of a density and its derivatives. *Biometrika, 66*(1), 177–180.

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics*, *5*(4), 595–645. With discussion and a reply by the author.

Tapia, R. A., Thompson, J. R. (1978). *Nonparametric probability density estimation, vol. 1 of Johns Hopkins Series in the Mathematical Sciences*. Baltimore, MD: Johns Hopkins University Press.

Yizong, C. (1995). Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 17*(8), 790–799.

Yu, K., Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association, 93*(441), 228–237.

Wand, M. P., Jones, M. C. (1995). *Kernel smoothing, vol. 60 of Monographs on Statistics and Applied Probability*. London: Chapman and Hall, Ltd.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Series A, 26*, 359–372.

Wertz, W. (1978). Statistical density estimation: A survey, vol. 13 of Angewandte Statistik und Ökonometrie [Applied Statistics and Econometrics]. Göttingen: Vandenhoeck & Ruprecht. With German and French summaries.

Wu, T.-J., Hsu, C.-Y., Chen, H.-Y., Yu, H.-C. (2014). Root *n* estimates of vectors of integrated density partial derivative functionals. *Annals of the Institute of Statistical Mathematics, 66*(5), 865–895.

Wu, W. B. (2003). *Nonparametric estimation for stationary processes*. Technical Report 536, University of Chicago.

Wu, W. B., Huang, Y., Huang, Y. (2010). Kernel estimation for time series: An asymptotic theory. *Stochastic Processes and their Applications, 120*, 2412–2431.

Ziegler, K. (2001). On bootstrapping the mode in the nonparametric regression model with random design. *Metrika, 53*(2), 141–170.

Ziegler, K. (2002). On nonparametric kernel estimation of the mode of the regression function in the random design model. *Journal of Nonparametric Statistics, 14*(6), 749–774.

Ziegler, K. (2003). On the asymptotic normality of kernel regression estimators of the mode in the nonparametric random design model. *Journal of Statistical Planning and Inference, 115*(1), 123–144.