Theory of Stochastic Processes
Vol. 26 (42), no. 1, 2022, pp. 27–59

SAHAR SLAMA, YOUSRI SLAOUI, AND HAMDI FATHALLAH

THE STOCHASTIC APPROXIMATION METHOD FOR SEMI-RECURSIVE MULTIVARIATE KERNEL-TYPE REGRESSION ESTIMATION

In this research paper, we elaborate an extension of the semi-recursive kernel-type regression function estimator. We investigate the asymptotic properties of this estimator and compare them with non-recursive Nadaraya Watson regression estimator. From this perspective, we first calculate the bias and the variance of the proposed estimator which strongly depend on the choice of three parameters, namely the stepsizes (β_n) and (γ_n) as well as the bandwidth (h_n) chosen using one of the best methods of bandwidth selection, the bootstrap approach compared to the plug-in method. An appropriate choice of those parameters yields that, under some conditions, the MSE (Mean Squared Error) of the proposed estimator can be smaller than that of Nadaraya Watson's estimator. We corroborate our theoretical results through simulations studies and by considering two real dataset applications, the French Hospital Data of COVID-19 epidemic as well as the Plasmodium Falciparum Parasite Load (PL).

1. Introduction

Let $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ and $(X_1, Y_1), ..., (X_n, Y_n)$ be independent random vectors identically distributed as (X, Y) with a joint density function $g(x, y)$ and let f denote the probability density of X. For a chosen measurable function φ and $x \in \mathbb{R}^d$ the regression function, whenever it exists, is defined by

$$r_\varphi(x) = \mathbb{E}[\varphi(Y)|X = x] = \frac{1}{f(x)} \int_{\mathbb{R}} \varphi(y) g(x, y) dy.$$

Regression analysis stands for the study of how a response variable depends on one or more predictors. In fact, it's a reliable method of identifying which variables have impact on a topic of interest. The process of performing a regression allows us to confidently determine which factors matter most, which factors can be ignored, and how these factors influence each other. Regression problems can be usefully solved using nonparametric regression methods, which correspond to a category of regression analysis where the predictor does not take a predetermined form but is constructed according to information derived from the data. From this perspective, we have multiple methods of nonparametric estimation, such as Gaussian process regression (Kriging), kernel regression and regression trees.

In this work, our central focus is upon kernel-type regression function estimation which is a non-parametric technique in statistics to estimate the conditional expectation of a random variable. For the recursive approach, Kiefer and Wolfwitz (1952) [8] set forward the stochastic approximation algorithm for recursive regression function estimation. Their work was extended with Nadaraya (1964) [14] and Watson(1964) [22].

2010 *Mathematics Subject Classification.* Primary 62H12. 62G08. 62G09. 62L20. 65D10.

Key words and phrases. Multivariate estimation; Nonparametric regression; Resampling methods; Stochastic approximation algorithm; Smoothing, curve fitting.

An estimator of the r_φ regression was first developed by Roussas (1990) [16] and improved by Einmahl and Mason (2000) [5] to determine exact rates of uniform strong consistency of kernel-type function estimators.

Later, Deheuvels and Mason (2004) [3] established uniform and non-uniform asymptotic simultaneous confidence bands for functionals of the distribution based on kernel-type estimators.

The classical recursive regression estimator was addressed in Mokaddem *et al.* (2009b) [12] for univariate framework and a multivariate extension of this estimator was carried out by Mokaddem and Pelletier (2016) [13]. Subsequently, Slaoui (2016) [20] established the semi-recursive case and introduced a new estimator which is the fraction of a recursive regression by a recursive density function. As far as this research is concerned, our basic objective is to extend this estimator for kernel-type estimation with large choice of parameters and properties in a multivariate case. Note that, recently, Bouzebda and Slaoui (2020) [2] explored general kernel type estimators for spatial data defined by the stochastic approximation algorithm.

Let us start with the presentation of our stochastic approximation method. To build up a stochastic algorithm, which estimates recursively the regression function

$$a_\varphi : x \longmapsto r_\varphi(x)f(x) = \int_{\mathbb{R}} \varphi(y)g(x,y)dy$$

at a given vector x, we follow the approach of Révész (1977) [15] and Tsybakov (1990) [21]. Therefore, the stochastic approximation algorithm can be expressed as follows :

$$a_{\varphi_n}(x) = (1-\beta_n)a_{\varphi_{n-1}}(x) + \beta_n\,\varphi(Y_n)\,h_n^{-d}\mathbf{K}\left(\frac{x-X_n}{h_n}\right),$$

where the bandwidth (h_n) and the stepsize (β_n) are positive sequences of real numbers decreasing towards zero and $\mathbf{K}$ is a multivariate kernel.
Here, we consider that $a_0(x) = 0$, then by a recurrence, we get

$$a_{\varphi_n}(x) = Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k\,\varphi(Y_k)h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right).$$

Moreover, we use the recursive multivariate probability density estimator of the density function f defined in Mokkadem *et al.* (2009a) [11] and provided by

$$f_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k\,h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right),$$

where (γ_n) is a positive sequence of real numbers decreasing towards zero.

Through this paper, we consider the general multivariate kernel-type estimator for the regression function $r : x \longmapsto \mathbb{E}[\varphi(Y)|X=x]$ at the vector x

$$r_{\varphi_n}(x) = \begin{cases} \dfrac{a_{\varphi_n}(x)}{f_n(x)} & \text{if} \quad f_n(x) \neq 0 \\ 0 & \text{if} \quad f_n(x) = 0 \end{cases}. \tag{1}$$

Our first aim is to examine the asymptotic properties of our proposed semi-recursive estimator of a multivariate regression function. Then, we prove its performance.
We shall also compare our estimator to the generalized non-recursive kernel regression estimator of Nadaraya-Watson [14] and [22] $\widetilde{r}_{\varphi_n}$ indicated by

$$\widetilde{r}_{\varphi_n}(x) = \begin{cases} \dfrac{\widetilde{a}_{\varphi_n}(x)}{\widetilde{f}_n(x)} & \text{if} \quad \widetilde{f}_n(x) \neq 0 \\ 0 & \text{if} \quad \widetilde{f}_n(x) = 0 \end{cases}, \tag{2}$$

with

$$\widetilde{a}_{\varphi_n}(x) = \frac{1}{nh_n^d}\sum_{k=1}^{n}\varphi(Y_k)\mathbf{K}\left(\frac{x-X_k}{h_n}\right) \text{ and } \widetilde{f}_n(x) = \frac{1}{nh_n^d}\sum_{k=1}^{n}\mathbf{K}\left(\frac{x-X_k}{h_n}\right).$$

Particular cases:

(1) For $\varphi(y) := \mathbb{I}(y) = y$, we have the classical regression function

$$r_{\mathbb{I}}(x) = \mathbb{E}[Y|X=x].$$

A recursive estimator of $r_{\mathbb{I}}$ was reported in Slaoui (2015) [19].

(2) For $\varphi(y) := \mathbb{I}(y) = y^m$, $m \in \mathbb{N}$, we have the conditional moments

$$r_{\mathbb{I}}(x) = \mathbb{E}[Y^m|X=x].$$

(3) For $\varphi(y) := \chi_t(y) = \mathbb{1}_{\{y\leqslant t\}}$, $t \in \mathbb{R}$, we have the conditional cumulative distribution function

$$r_{\chi_t}(y) = \pi(t|x) = \mathbb{P}[Y \leqslant t|X=x].$$

A recursive estimator of r_{χ_t} was identified in Slama *et al.* (2020) [17].

2. Assumptions and Notations

For our theoretical main results, we need the following technical assumptions.
Assumptions:

(A_1) $\mathbf{K} : \mathbb{R}^d \longrightarrow \mathbb{R}_+$ is a continuous bounded function satisfying:

$$\int_{\mathbb{R}^d}\mathbf{K}(u)du = 1\ ,\ \forall j \in \{1,\ldots,d\}, \int_{\mathbb{R}^d} u_j\mathbf{K}(u)du = 0 \ \text{ and } \int_{\mathbb{R}^d} u_j^2\mathbf{K}(u)du < \infty.$$

(A_2) (i) $(\beta_n)_{n\geq1} \in \mathcal{GS}(-\beta)$, with $\beta \in \left(\frac{1}{2},1\right]$.
(ii) $(\gamma_n)_{n\geq1} \in \mathcal{GS}(-\alpha)$, with $\alpha \in \left(\frac{1}{2},1\right]$.
(iii) $(h_n)_{n\geq1} \in \mathcal{GS}(-a)$, with $a \in (0,1]$.
(iv) $\lim_{n\to+\infty}(n\beta_n) \in \left(\min\{2a, \frac{\beta-ad}{2}\}, \infty\right]$.
(v) $\lim_{n\to+\infty}(n\gamma_n) \in \left(\min\{2a, \frac{\alpha-ad}{2}\}, \infty\right]$.

(A_3) (i) f is bounded, twice differentiable and $\forall i,j \in \{1,\ldots,d\}, f_{ij}^{(2)} := \dfrac{\partial^2 f}{\partial x_i \partial x_j}$ is bounded.
(ii) a_φ is bounded, twice differentiable and $\forall i,j \in \{1,\ldots,d\}, a_{\varphi_{ij}}^{(2)} := \dfrac{\partial^2 a_\varphi}{\partial x_i \partial x_j}$ is bounded.
(iii) The function $s \longmapsto \int_{\mathbb{R}}\varphi(t)^g(s,t)dt$ is bounded continuous at $s = x$.
(iv) For all $p > 0$, $s \longmapsto \int_{\mathbb{R}}|\varphi(t)|^{2+p}g(s,t)dt$ is a bounded function.

Throughout this paper, the following notations are used :

$$\xi_\beta = \lim_{n\to+\infty}(n\beta_n)^{-1},\ Q_n = \prod_{j=1}^{n}(1-\beta_j),\ \xi_{\beta,\alpha} = \lim_{n\to+\infty}(\beta_n\gamma_n^{-1}),\ \mu_i(\mathbf{K}) = \int_{\mathbb{R}^d} z_i^2\mathbf{K}(z)dz,$$

$$\xi_\alpha = \lim_{n\to+\infty}(n\gamma_n)^{-1},\ \Pi_n = \prod_{j=1}^{n}(1-\gamma_j),\ \xi_{\alpha,\beta} = \lim_{n\to+\infty}(\gamma_n\beta_n^{-1}),\ R(\mathbf{K}) = \int_{\mathbb{R}^d}\mathbf{K}^2(z)\,dz,$$

$$I_1 = \int_{\mathbb{R}^d}\left(\sum_{j=1}^{d}\mu_j(\mathbf{K})a_{\varphi_{jj}}^{(2)}(x)\right)^2 f(x)dx, \quad I_4 = \int_{\mathbb{R}^d}\mathbb{E}[\varphi(Y)^2|X=x]f^2(x)dx,$$

$$I_2 = \int_{\mathbb{R}^d} \left(\sum_{j=1}^{d} \mu_j(\mathbf{K}) a^{(2)}_{\varphi_{jj}}(x) \right) \left(\sum_{j=1}^{d} \mu_j(\mathbf{K}) f^{(2)}_{jj}(x) \right) r_\varphi(x) f(x) dx,$$

$$I_3 = \int_{\mathbb{R}^d} \left(\sum_{j=1}^{d} \mu_j(\mathbf{K}) f^{(2)}_{jj}(x) \right)^2 r^2_\varphi(x) f(x) dx, \quad I_5 = \int_{\mathbb{R}^d} r^2_\varphi(x) f^2(x) dx.$$

First of all, let us set the following definition of the class of regularly varying sequences introduced by Galambos and Seneta (1973) [7].

Definition 2.1. Let $(v_n)_{n\geq 1}$ be a nonrandom positive sequence and $\gamma \in \mathbb{R}$. We state that

$$(v_n)_{n\geq 1} \in \mathcal{GS}(\gamma) \text{ if } \lim_{n\to+\infty} n \left[1 - \frac{v_{n-1}}{v_n} \right] = \gamma.$$

3. Main results of r_{φ_n}

In order to investigate the asymptotic properties of our estimator r_{φ_n}, we need to first introduce the following proposition which provide the bias and the variance of a_{φ_n}.

Proposition 3.1. *(Bias and variance of a_{φ_n})*
Under the assumptions $(A_1) - (A_3)$, and assuming that, for all $i, j \in \{1, \dots, d\}$, $a^{(2)}_{\varphi_{ij}}$ is continuous at x, we obtain

(1) *If $a \in \left(0, \frac{\beta}{d+4}\right]$, then*

$$\mathbb{E}[a_{\varphi_n}(x)] - a_\varphi(x) = \frac{h_n^2}{2(1-2a\xi_\beta)} \sum_{j=1}^{d} \mu_j(\mathbf{K}) a^{(2)}_{\varphi_{jj}}(x) + o\left(h_n^2\right). \tag{3}$$

If $a \in \left(\frac{\beta}{d+4}, 1\right)$, then

$$\mathbb{E}[a_{\varphi_n}(x)] - a_\varphi(x) = o\left(\sqrt{\beta_n h_n^{-d}}\right). \tag{4}$$

(2) *If $a \in \left(0, \frac{\beta}{d+4}\right)$, then*

$$Var[a_{\varphi_n}(x)] = o\left(h_n^4\right). \tag{5}$$

If $a \in \left[\frac{\beta}{d+4}, 1\right)$, then

$$Var[a_{\varphi_n}(x)] = \frac{\beta_n}{h_n^d} \frac{\mathbb{E}[\varphi(Y)^2|X=x]}{2-(\beta-ad)\xi_\beta} f(x) R(\mathbf{K}) + o\left(\beta_n h_n^{-d}\right). \tag{6}$$

Our main result rests on the following theorem, which yields the bias and the variance of r_{φ_n}.

Theorem 3.1. *(Bias and variance of r_{φ_n})*
Under the assumptions $(A_1) - (A_3)$, and assuming that, for all $i, j \in \{1, \dots, d\}$, $a^{(2)}_{\varphi_{ij}}$ and $f^{(2)}_{ij}$ are continuous at x, we obtain

(1) *If $a \in \left(0, \frac{\min(\beta,\alpha)}{d+4}\right]$, then*

$$\mathbb{E}[r_{\varphi_n}(x)] - r_\varphi(x) = \frac{h_n^2}{f(x)} \left(\frac{\sum_{j=1}^{d} \mu_j(\mathbf{K}) a^{(2)}_{\varphi_{jj}}(x)}{2(1-2a\xi_\beta)} - \frac{r_\varphi(x) \sum_{j=1}^{d} \mu_j(\mathbf{K}) f^{(2)}_{jj}(x)}{2(1-2a\xi_\alpha)} \right) + o\left(h_n^2\right). \tag{7}$$

If $a \in \left(\frac{\min(\beta,\alpha)}{d+4}, 1\right)$, *then*

$$\mathbb{E}[r_{\varphi_n}(x)] - r_\varphi(x) = o\left(\sqrt{\beta_n h_n^{-d}}\right)\mathbf{1}_{\{\beta\leq\alpha\}} + o\left(\sqrt{\gamma_n h_n^{-d}}\right)\mathbf{1}_{\{\alpha<\beta\}}. \tag{8}$$

(2) *If* $a \in \left(0, \frac{\min(\beta,\alpha)}{d+4}\right)$, *then*

$$Var[r_{\varphi_n}(x)] = o\left(h_n^4\right). \tag{9}$$

If $a \in \left[\frac{\min(\beta,\alpha)}{d+4}, 1\right)$, *then*

$$\begin{aligned} Var[r_{\varphi_n}(x)] = {} & \frac{\beta_n}{h_n^d}\frac{R(\mathbf{K})}{f(x)}\left[\frac{\mathbb{E}[\varphi(Y)^2|X=x]}{2-(\beta-ad)\xi_\beta}\right. \\ & \left. -r_\varphi^2(x)\left(\frac{2}{1-(\beta-ad-\xi_\beta^{-1})\xi_\alpha} - \frac{\xi_{\alpha,\beta}}{2-(\alpha-ad)\xi_\alpha}\right) + o\left(\frac{\beta_n}{h_n^d}\right)\right]\mathbf{1}_{\{\beta\leqslant\alpha\}} \\ & + \frac{\gamma_n}{h_n^d}\frac{R(\mathbf{K})}{f(x)}\left[\frac{\mathbb{E}[\varphi(Y)^2|X=x]\xi_{\beta,\alpha}}{2-(\beta-ad)\xi_\beta}\right. \\ & \left. -r_\varphi^2(x)\left(\frac{2}{1-(\alpha-ad-\xi_\alpha^{-1})\xi_\beta} - \frac{1}{2-(\alpha-ad)\xi_\alpha}\right) + o\left(\frac{\gamma_n}{h_n^d}\right)\right]\mathbf{1}_{\{\alpha<\beta\}}. \end{aligned} \tag{10}$$

Therefore, the bias and the variance of the estimator r_{φ_n} defined by the stochastic approximation algorithm (1) depend heavily on the choice of the stepsizes (β_n) and (γ_n).

Remark 3.1. Notice that, for the case where $(\gamma_n) = (\beta_n)$ and then $\alpha = \beta$, the expression (10) will be written as follows

$$Var[r_{\varphi_n}(x)] = \frac{\beta_n}{h_n^d}\frac{R(\mathbf{K})}{f(x)}\frac{Var[\varphi(Y)|X=x]}{2-(\beta-ad)\xi_\beta} + o\left(\beta_n h_n^{-d}\right).$$

The asymptotic normality of the generalized semi-recursive estimator r_{φ_n} is indicated by the following theorem. Note that $\xrightarrow[n\to+\infty]{\mathcal{D}}$ denotes convergence in distribution, $\mathcal{N}$ corresponds to the normal distribution and $\xrightarrow[n\to+\infty]{\mathcal{P}}$ stands for convergence in probability.

3.1. Weak pointwise convergence rate of r_{φ_n}.

Theorem 3.2. *Under the assumptions* $(A_1)-(A_3)$, *we obtain:*

(1) ***For the case*** $\beta \leq \alpha$***:***

(a) *If there exists* $c \geq 0$ *such that* $\beta_n^{-1}h_n^{d+4} \xrightarrow[n\to+\infty]{} c$, *then*

$$\sqrt{\beta_n^{-1}h_n^d}\left(r_{\varphi_n}(x) - r_\varphi(x)\right) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\Big(\sqrt{c}\mathbf{M}_\beta(x)\ ,\ \mathbf{\Sigma}_\beta(x)\Big), \tag{11}$$

with

$$\mathbf{M}_\beta(x) = \frac{1}{2f(x)}\left(\frac{\sum_{j=1}^d \mu_j(\mathbf{K})a_{\varphi_{jj}}^{(2)}(x)}{(1-2a\xi_\beta)} - \frac{r_\varphi(x)\sum_{j=1}^d \mu_j(\mathbf{K})f_{jj}^{(2)}(x)}{(1-2a\xi_\alpha)}\right) \tag{12}$$

and

$$\mathbf{\Sigma}_\beta(x) = \frac{R(\mathbf{K})}{f(x)}\left[\frac{\mathbb{E}[\varphi(Y)^2|X=x]}{2-(\beta-ad)\xi_\beta}\right. \tag{13}$$

$$\left. - r_{\varphi}^2(x)\left(\frac{2}{1-(\beta-ad-\xi_{\beta}^{-1})\xi_{\alpha}} - \frac{\xi_{\alpha,\beta}}{2-(\alpha-ad)\xi_{\alpha}}\right)\right].$$

(b) *If* $\beta_n^{-1}h_n^{d+4} \underset{n\to+\infty}{\longrightarrow} \infty$, *then*

$$\frac{1}{h_n^2}\left(r_{\varphi_n}(x) - r_{\varphi}(x)\right) \xrightarrow[n\to+\infty]{\mathcal{P}} \mathbf{M}_{\beta}(x).$$

(2) ***For the case*** $\beta > \alpha$***:***

(a) *If there exists* $c \geq 0$ *such that* $\gamma_n^{-1}h_n^{d+4} \underset{n\to+\infty}{\longrightarrow} c$, *then*

$$\sqrt{\gamma_n^{-1}h_n^d}\left(r_{\varphi_n}(x) - r_{\varphi}(x)\right) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\Big(\sqrt{c}\mathbf{M}_{\gamma}(x)\ ,\ \mathbf{\Sigma}_{\gamma}(x)\Big), \tag{14}$$

with

$$\mathbf{M}_{\gamma}(x) = \frac{1}{2f(x)}\left(\frac{\sum_{j=1}^{d}\mu_j(\mathbf{K})a_{\varphi_{jj}}^{(2)}(x)}{(1-2a\xi_{\beta})} - \frac{r_{\varphi}(x)\sum_{j=1}^{d}\mu_j(\mathbf{K})f_{jj}^{(2)}(x)}{(1-2a\xi_{\alpha})}\right) \tag{15}$$

and

$$\mathbf{\Sigma}_{\gamma}(x) = \frac{R(\mathbf{K})}{f(x)}\left[\frac{\mathbb{E}[\varphi(Y)^2|X=x]\xi_{\beta,\alpha}}{2-(\alpha-ad)\xi_{\alpha}}\right. \tag{16}$$
$$\left. - r_{\varphi}^2(x)\left(\frac{2}{1-(\alpha-ad-\xi_{\alpha}^{-1})\xi_{\beta}} - \frac{1}{2-(\alpha-ad)\xi_{\alpha}}\right)\right].$$

(b) *If* $\gamma_n^{-1}h_n^{d+4} \underset{n\to+\infty}{\longrightarrow} \infty$, *then*

$$\frac{1}{h_n^2}\left(r_{\varphi_n}(x) - r_{\varphi}(x)\right) \xrightarrow[n\to+\infty]{\mathcal{P}} \mathbf{M}_{\gamma}(x).$$

The following theorem demonstrates the strong pointwise convergence rate of our estimator r_{φ_n}.

3.2. Strong pointwise convergence rate of r_{φ_n}.

Theorem 3.3. *Under the assumptions* $(A_1)-(A_3)$, *we get:*

(1) ***For the case*** $\beta \leq \alpha$***:***

(a) *If there exists* $b \geq 0$ *such that* $\frac{\beta_n^{-1}h_n^{d+4}}{\ln\left(\sum_{i=1}^{n}\beta_i\right)} \underset{n\to+\infty}{\longrightarrow} b$, *then with probability one, the sequence*

$$\left(\sqrt{\frac{\beta_n^{-1}h_n^d}{2\ln\left(\sum_{i=1}^{n}\beta_i\right)}}\left(r_{\varphi_n}(x) - r_{\varphi}(x)\right)\right)$$

is relatively compact and its limit set is the interval

$$\left[\sqrt{\frac{b}{2}}\mathbf{M}_{\beta}(x) - \sqrt{\mathbf{\Sigma}_{\beta}(x)}\, ,\, \sqrt{\frac{b}{2}}\mathbf{M}_{\beta}(x) + \sqrt{\mathbf{\Sigma}_{\beta}(x)}\right].$$

(b) *If* $\frac{\beta_n^{-1}h_n^{d+4}}{\ln\left(\sum_{i=1}^n \beta_i\right)} \underset{n\to+\infty}{\longrightarrow} \infty$*, then, with probability one,*

$$\lim_{n\to+\infty} \frac{1}{h_n^2}\left(r_{\varphi_n}(x) - r_\varphi(x)\right) = \mathbf{M}_\beta(x).$$

(2) ***For the case*** $\beta > \alpha$***:***

(a) *If there exists* $b \geq 0$ *such that* $\frac{\gamma_n^{-1}h_n^{d+4}}{\ln\left(\sum_{i=1}^n \gamma_i\right)} \underset{n\to+\infty}{\longrightarrow} b$*, then with probability one, the sequence*

$$\left(\sqrt{\frac{\gamma_n^{-1}h_n^d}{2\ln\left(\sum_{i=1}^n \gamma_i\right)}}\left(r_{\varphi_n}(x) - r_\varphi(x)\right)\right)$$

is relatively compact and its limit set is the interval

$$\left[\sqrt{\frac{b}{2}}\mathbf{M}_\gamma(x) - \sqrt{\mathbf{\Sigma}_\gamma(x)}\,,\,\sqrt{\frac{b}{2}}\mathbf{M}_\gamma(x) + \sqrt{\mathbf{\Sigma}_\gamma(x)}\right].$$

(b) *If* $\frac{\gamma_n^{-1}h_n^{d+4}}{\ln\left(\sum_{i=1}^n \gamma_i\right)} \underset{n\to+\infty}{\longrightarrow} \infty$*, then, with probability one,*

$$\lim_{n\to+\infty} \frac{1}{h_n^2}\left(r_{\varphi_n}(x) - r_\varphi(x)\right) = \mathbf{M}_\gamma(x).$$

In what follows, we clarify the choices of the stepsizes (β_n) as well as (γ_n) and the bandwidth (h_n) based on the $MWISE$ of the recursive estimator minimization, and then enact a comparison with Nadaraya Watson's estimator.

4. Optimal choice of the stepsizes

In order to measure the optimal choice of the couple of stepsizes (β_n, γ_n), we need to minimize the Mean Weighted Integrated Squared Error ($MWISE$) of the semi-recursive estimator r_{φ_n}.

The $MWISE$ of the estimator r_{φ_n} is determined by the following expression

$$MWISE[r_{\varphi_n}] = \int_{\mathbb{R}^d}\left(\mathbb{E}[r_{\varphi_n}(x)] - r_\varphi(x)\right)^2 f^3(x)dx + \int_{\mathbb{R}^d} Var[r_{\varphi_n}(x)]f^3(x)dx.$$

Proposition 4.1. *We first note,*

$$C_1 = \frac{I_1}{(1-2a\xi_\beta)^2} - \frac{2I_2}{(1-2a\xi_\beta)(1-2a\xi_\alpha)} + \frac{I_3}{(1-2a\xi_\alpha)^2}, \quad C_2 = \frac{I_4}{2-(\beta-ad)\xi_\beta},$$

$$C_3 = \frac{I_4\xi_{\beta,\alpha}}{2-(\beta-ad)\xi_\beta}, \qquad C_4 = I_5\left(\frac{2}{1-(\beta-ad-\xi_\beta^{-1})\xi_\alpha} - \frac{\xi_{\alpha,\beta}}{2-(\alpha-ad)\xi_\alpha}\right),$$

$$C_5 = I_5\left(\frac{2}{1-(\alpha-ad-\xi_\alpha^{-1})\xi_\beta} - \frac{1}{2-(\alpha-ad)\xi_\alpha}\right).$$

(1) ***For the case*** $\beta \leq \alpha$***:***

$$MWISE[r_{\varphi_n}] = \begin{cases} \frac{1}{4}C_1h_n^4 + o(h_n^4) & \text{if} \quad a \in \left(0, \frac{\beta}{d+4}\right) \\ (C_2 - C_4)R(\mathbf{K})\beta_n h_n^{-d} + \frac{1}{4}C_1h_n^4 + o(h_n^4) & \text{if} \quad a = \frac{\beta}{d+4} \\ (C_2 - C_4)R(\mathbf{K})\beta_n h_n^{-d} + o\left(\beta_n h_n^{-d}\right) & \text{if} \quad a \in \left(\frac{\beta}{d+4}, 1\right) \end{cases}.$$

(2) ***For the case*** $\beta > \alpha$***:***

$$MWISE[r_{\varphi_n}] = \begin{cases} \frac{1}{4}C_1 h_n^4 + o(h_n^4) & \text{if} \quad a \in \left(0, \frac{\alpha}{d+4}\right) \\ (C_3 - C_5)R(\mathbf{K})\gamma_n h_n^{-d} + \frac{1}{4}C_1 h_n^4 + o(h_n^4) & \text{if} \quad a = \frac{\alpha}{d+4} \\ (C_3 - C_5)R(\mathbf{K})\gamma_n h_n^{-d} + o\left(\gamma_n h_n^{-d}\right) & \text{if} \quad a \in \left(\frac{\alpha}{d+4}, 1\right) \end{cases}.$$

The following corollary ensures that the bandwidth which minimizes the $MWISE$ of r_{φ_n} depends on the choice of the stepsizes (β_n) and (γ_n) and then the corresponding $MWISE$ depends in turn on (β_n) and (γ_n).

Corollary 4.1. *Let assumptions $(A_1) - (A_3)$ hold. To minimize the MWISE of r_{φ_n}, the bandwidth (h_n) needs to be equal to the following expressions.*

(1) ***For the case*** $\beta \leq \alpha$***:***

$$h_n = d^{\frac{1}{d+4}} \left(\frac{C_2 - C_4}{C_1}\right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} \beta_n^{\frac{1}{d+4}}.$$

Hence, the corresponding $MWISE$ is specified by

$$MWISE[r_{\varphi_n}] = \frac{(d+4)}{4d^{\frac{d}{d+4}}} C_1^{\frac{d}{d+4}} (C_2 - C_4)^{\frac{4}{d+4}} R(\mathbf{K})^{\frac{4}{d+4}} \beta_n^{\frac{4}{d+4}} + o\left(\beta_n^{\frac{4}{d+4}}\right).$$

(2) ***For the case*** $\alpha < \beta$:

$$h_n = d^{\frac{1}{d+4}} \left(\frac{C_3 - C_5}{C_1}\right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} \gamma_n^{\frac{1}{d+4}}.$$

Thus, the corresponding $MWISE$ is expressed by

$$MWISE[r_{\varphi_n}] = \frac{(d+4)}{4d^{\frac{d}{d+4}}} C_1^{\frac{d}{d+4}} (C_3 - C_5)^{\frac{4}{d+4}} R(\mathbf{K})^{\frac{4}{d+4}} \gamma_n^{\frac{4}{d+4}} + o\left(\gamma_n^{\frac{4}{d+4}}\right).$$

The following corollary is provided in the special cases, where (β_n) is chosen as $(\beta_n) = (\beta_0 n^{-1})$ in order to minimize the $MWISE[a_{\varphi_n}]$ and (γ_n) is chosen as $(\gamma_n) = (\gamma_0 n^{-1})$ in order to minimize the $MWISE[f_n]$.

Proposition 4.2. *Let assumptions $(A_1) - (A_3)$ hold. It is worth noting,*

$$\Theta_1 = \frac{\beta_0 I_4}{(d+4)\beta_0 - 2} - \frac{2\gamma_0 I_5}{(d+4)(\frac{\gamma_0+\beta_0}{2}) - 2} + \frac{\beta_0^{-1}\gamma_0^2 I_5}{(d+4)\gamma_0 - 2},$$
$$\Theta_2 = \frac{\gamma_0 I_5}{(d+4)\gamma_0 - 2} - \frac{2\beta_0 I_5}{(d+4)(\frac{\gamma_0+\beta_0}{2}) - 2},$$
$$\Theta_3 = \frac{\beta_0^2 I_1}{((d+4)\beta_0 - 2)^2} - \frac{2\beta_0\gamma_0 I_2}{((d+4)\beta_0 - 2)((d+4)\gamma_0 - 2)} + \frac{\gamma_0^2 I_3}{((d+4)\gamma_0 - 2)^2}.$$

To minimize the $MWISE$ of r_{φ_n}, we need to choose the stepsize (γ_n) in $\mathcal{GS}(-1)$ such that $\lim_{n\to\infty}(n\gamma_n) = \gamma_0$ and the stepsize (β_n) in $\mathcal{GS}(-1)$ such that $\lim_{n\to\infty}(n\beta_n) = \beta_0$. As a matter of fact,

(1) ***For the case*** $\beta \leq \alpha$***:*** *The bandwidth (h_n) needs to be equal to*

$$\left(\beta_0^{\frac{1}{d+4}} \left(\frac{d}{2(d+4)}\right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{\frac{-1}{d+4}} \left(\frac{\Theta_1}{\Theta_3}\right)^{\frac{1}{d+4}}\right).$$

Consequently, the corresponding $MWISE$ is determined by

$$MWISE[r_{\varphi_n}] = \frac{(d+4)^{\frac{3d+8}{d+4}}}{4^{\frac{d+6}{d+4}} d^{\frac{d}{d+4}}} \beta_0^{\frac{4}{d+4}} \Theta_1^{\frac{4}{d+4}} \Theta_3^{\frac{d}{d+4}} R(\mathbf{K})^{\frac{4}{d+4}} n^{\frac{-4}{d+4}} + o\left(n^{\frac{-4}{d+4}}\right).$$

(2) ***For the case*** $\alpha < \beta$ *: The bandwidth (h_n) needs to be equal to*

$$\left(\gamma_0^{\frac{1}{d+4}} \left(\frac{d}{2(d+4)}\right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{\frac{-1}{d+4}} \left(\frac{\Theta_2}{\Theta_3}\right)^{\frac{1}{d+4}}\right).$$

Therefore, the corresponding $MWISE$ is specified by

$$MWISE[r_{\varphi_n}] = \frac{(d+4)^{\frac{3d+8}{d+4}}}{4^{\frac{d+6}{d+4}} d^{\frac{d}{d+4}}} \gamma_0^{\frac{4}{d+4}} \Theta_2^{\frac{4}{d+4}} \Theta_3^{\frac{d}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{\frac{-4}{d+4}} + o\left(n^{\frac{-4}{d+4}}\right).$$

Additionally, the minimum of $MWISE[r_{\varphi_n}]$ is achieved at $(\beta_0, \gamma_0) = (1, 1)$. From this perspective, the optimal bandwidth (h_n) must be equal to

$$\left(\left(\frac{d(d+2)}{2(d+4)}\right)^{\frac{1}{d+4}} \left(\frac{I_4 - I_5}{I_1 - 2I_2 + I_3}\right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{\frac{-1}{d+4}}\right). \tag{17}$$

Thus, the corresponding $MWISE$ is indicated by

$$\begin{aligned} & MWISE[r_{\varphi_n}] \\ &= \frac{(d+4)^{\frac{3d+8}{d+4}}}{4^{\frac{d+6}{d+4}} d^{\frac{d}{d+4}} (d+2)^{\frac{d+6}{d+4}}} (I_1 - 2I_2 + I_3)^{\frac{d}{d+4}} (I_4 - I_5)^{\frac{4}{d+4}} R(\mathbf{K})^{\frac{4}{d+4}} n^{\frac{-4}{d+4}} + o\left(n^{\frac{-4}{d+4}}\right). \end{aligned}$$

Remark 4.1. Note that for the particular case where the stepsize (β_n) is in $\mathcal{GS}(-1)$ such that $\lim_{n\to\infty} (n\beta_n) = 1$, (γ_n) is in $\mathcal{GS}(-1)$ such that $\lim_{n\to\infty} (n\gamma_n) = 1$ and the bandwidth (h_n) is chosen such that $\lim_{n\to\infty} nh_n^{d+4} = 0$ (which corresponds to undersmoothing), the asymptotic normality of the proposed estimator is represented as follows

$$\sqrt{nh_n^d}\left(r_{\varphi_n}(x) - r_\varphi(x)\right) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0\,,\ \frac{1}{a+d} R(\mathbf{K}) \frac{Var[\varphi(Y)|X=x]}{f(x)}\right).$$

5. Main results of $\widetilde{r}_{\varphi_n}$

The main properties of the generalized non-recursive regression function estimator $\widetilde{r}_{\varphi_n}$ are displayed in the following proposition.

Proposition 5.1. *Let assumptions (A_1) and (A_3) hold. Then, the asymptotic properties of Nadaraya-Watson's estimator are denoted as follows.*

- *The bias of $\widetilde{r}_{\varphi_n}$:*

$$\mathbb{E}[\widetilde{r}_{\varphi_n}(x)] - r_\varphi(x) = \frac{1}{2f(x)} h_n^2 \left(\sum_{j=1}^{d} \mu_j(\mathbf{K}) a_{\varphi jj}^{(2)}(x) - r_\varphi(x) \sum_{j=1}^{d} \mu_j(\mathbf{K}) f_{jj}^{(2)}(x)\right) + o\left(h_n^2\right).$$

- *The variance of $\widetilde{r}_{\varphi_n}$:*

$$Var[\widetilde{r}_{\varphi_n}(x)] = \frac{1}{nh_n^d} \frac{1}{f(x)} Var[\varphi(Y)|X=x] R(\mathbf{K}) + o\left(\frac{1}{nh_n^d}\right).$$

- *The $MWISE$ of $\widetilde{r}_{\varphi_n}$:*

$$MWISE[\widetilde{r}_{\varphi_n}] = \frac{1}{4}\left(I_1 - 2I_2 + I_3\right) h_n^4 + \frac{1}{nh_n^d}\left(I_4 - I_5\right) R(\mathbf{K}) + o\left(h_n^4 + \frac{1}{nh_n^d}\right).$$

To minimize the $MWISE$ of $\widetilde{r}_{\varphi_n}$, the bandwidth (h_n) must be equal to

$$\left(d^{\frac{1}{d+4}}\left(\frac{I_4-I_5}{I_1-2I_2+I_3}\right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{-\frac{1}{d+4}}\right). \tag{18}$$

Therefore, the corresponding $MWISE$ is expressed by

$$MWISE[\widetilde{r}_{\varphi_n}] = \frac{(d+4)}{4d^{\frac{d}{d+4}}}\left(I_4-I_5\right)^{\frac{4}{d+4}}\left(I_1-2I_2+I_3\right)^{\frac{d}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{-\frac{4}{d+4}} + o\left(n^{-\frac{4}{d+4}}\right).$$

- *The asymptotic normality of $\widetilde{r}_{\varphi_n}$:*
  Suppose that $nh_n^{d+4} \underset{n\to+\infty}{\longrightarrow} 0$. Thus,

$$\sqrt{nh_n^d}\left(\widetilde{r}_{\varphi_n}(x)-r_\varphi(x)\right) \underset{n\to+\infty}{\overset{\mathcal{D}}{\longrightarrow}} \mathcal{N}\left(0\ ,\ R(\mathbf{K})\frac{Var[\varphi(Y)|X=x]}{f(x)}\right).$$

- *The weak pointwise convergence rate of $\widetilde{r}_{\varphi_n}$:*
  If $nh_n^{d+4} \underset{n\to+\infty}{\longrightarrow} \infty$, then

$$\frac{1}{h_n^2}\left(\widetilde{r}_{\varphi_n}(x)-r_\varphi(x)\right) \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} 0.$$

6. Bandwidth selection

Kernel smoothing in non-parametric statistics requires the choice of a bandwidth parameter. There are numerous methods for bandwidth selection, namely the cross-validation method, the bootstrap procedure and the second generation plug-in approach.

First of all, for the sake of simplicity, the kernel $\mathbf{K}$ we shall use is considered as a product of univariate kernels K satisfying $\int_{\mathbb{R}} K(x)dx = 1$. Hence, we have

$$r_{\varphi n}(x) = \frac{Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k \varphi(Y_k)\, h_k^{-d} \prod_{i=1}^{d} K\left(\frac{x_i - X_{k_i}}{h_k}\right)}{\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k\, h_k^{-d} \prod_{i=1}^{d} K\left(\frac{x_i - X_{k_i}}{h_k}\right)}$$

and

$$\widetilde{r}_{\varphi_n}(x) = \frac{\left(nh_n^d\right)^{-1} \sum_{k=1}^{n} \varphi(Y_k) \prod_{i=1}^{d} K\left(\frac{x_i - X_{k_i}}{h_n}\right)}{\left(nh_n^d\right)^{-1} \sum_{k=1}^{n} \prod_{i=1}^{d} K\left(\frac{x_i - X_{k_i}}{h_n}\right)}.$$

Let us start by introducing our bandwidth selection methods.

6.1. **Plug-in method.** In statistics, Altman and Leger (1995) [1] set forward an efficient method of bandwidth selection, a plug-in estimate which minimizes an estimate of the mean weighted integrated squared error, using the density function as a weight function. Since the $MWISE$ depends on the unknown quantities I_j, $j = 1,\dots,5$, we attempt to construct an asymptotic unbiased estimator of those quantities.
In order to estimate the optimal bandwidth (17), we have to estimate I_j, $j = 1,\dots,5$, by using the approach of Altman and Leger (1995) [1], called "Plug-in estimate". For this reason, we introduce $(b_n)_{n\geq 1} \in \mathcal{GS}(-\delta)$, $\delta \in (0,1)$.
In practice, we take $b_n = n^{-\delta}\min\left\{\widehat{s}, \frac{Q_3-Q_1}{1.349}\right\}$, with $\widehat{s}$ being the sample standard deviation and Q_1, Q_3 being the first and third quartiles.
At this stage of analysis, our choice of the parameter δ is based on the work of Slaoui (2014) [18] and Slaoui (2016) [20]. Therefore, we recall that K_b is a kernel and b_n is the

associated bandwidth, such that $\delta = 2/5$, and $K_{b'}^{(2)}$ is the second derivative of a kernel $K_{b'}$ with the associated bandwidth $b_n^{'}$ such that $\delta = 3/14$.

6.1.1. *Semi-Recursive estimator r_{φ_n}:* To estimate the optimal bandwidth (17), we need to estimate I_j, $j = 1, \ldots, 5$.

Estimation of I_1, I_2 and I_3: Here, the plug-in estimate gives

$$\widehat{I'}_1 = \frac{Q_n^2}{n} \sum_{\substack{i,j,k=1 \\ i\neq j\neq k}}^{n} Q_j^{-1} Q_k^{-1} \beta_j \beta_k {b'}_j^{-(d+2)} {b'}_k^{-(d+2)} \times \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{it} - X_{jt}}{b_j^{'}} \right) \prod_{\substack{l=1 \\ l\neq t}}^{d} K_b \left(\frac{X_{il} - X_{jl}}{b_j} \right) \right] \times \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{it} - X_{kt}}{b_k^{'}} \right) \prod_{\substack{l=1 \\ l\neq t}}^{d} K_b \left(\frac{X_{il} - X_{kl}}{b_k} \right) \right] \varphi(Y_j)\varphi(Y_k),$$

$$\widehat{I'}_2 = \frac{Q_n \Pi_n}{n} \sum_{\substack{i,j,k=1 \\ i\neq j\neq k}}^{n} Q_j^{-1} \Pi_k^{-1} \beta_j \gamma_k {b'}_j^{-(d+2)} {b'}_k^{-(d+2)} \times \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{it} - X_{jt}}{b_j^{'}} \right) \prod_{\substack{l=1 \\ l\neq t}}^{d} K_b \left(\frac{X_{il} - X_{jl}}{b_j} \right) \right] \times \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{it} - X_{kt}}{b_k^{'}} \right) \prod_{\substack{l=1 \\ l\neq t}}^{d} K_b \left(\frac{X_{il} - X_{kl}}{b_k} \right) \right] \varphi(Y_j)\varphi(Y_i),$$

$$\widehat{I'}_3 = \frac{\Pi_n^2}{n} \sum_{\substack{i,j,k,m=1 \\ i\neq j\neq k\neq m}}^{n} \Pi_j^{-1} \Pi_k^{-1} \gamma_j \gamma_k {b'}_j^{-(d+2)} {b'}_k^{-(d+2)} \times \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{it} - X_{jt}}{b_j^{'}} \right) \prod_{\substack{l=1 \\ l\neq t}}^{d} K_b \left(\frac{X_{il} - X_{jl}}{b_j} \right) \right] \times \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{it} - X_{kt}}{b_k^{'}} \right) \prod_{\substack{l=1 \\ l\neq t}}^{d} K_b \left(\frac{X_{il} - X_{kl}}{b_k} \right) \right] \varphi(Y_i)\varphi(Y_m),$$

Therefore, we obtain

$$\widehat{I}_i = \mu^2(\mathbf{K}) \widehat{I'}_i, \quad i = 1 \ldots 3.$$

Estimation of I_4 and I_5:

$$\widehat{I}_4 = \frac{\Pi_n}{n} \sum_{\substack{i,k=1 \\ i\neq k}}^{n} \Pi_k^{-1} \gamma_k b_k^{-d} \prod_{l=1}^{d} K_b \left(\frac{X_{il} - X_{kl}}{b_k} \right) \varphi(Y_i)^2$$

and

$$\widehat{I}_5 = \frac{Q_n}{n} \sum_{\substack{i,k=1 \\ i \neq k}}^{n} Q_k^{-1} \beta_k b_k^{-d} \prod_{l=1}^{d} K_b \left(\frac{X_{il} - X_{kl}}{b_k} \right) \varphi(Y_i) \varphi(Y_k),$$

As a result, the plug-in estimator of (17) is denoted in terms of :

$$h_n = \left(\left(\frac{d(d+2)}{2(d+4)} \right)^{\frac{1}{d+4}} \left(\frac{\widehat{I}_4 - \widehat{I}_5}{\widehat{I}_1 - 2\widehat{I}_2 + \widehat{I}_3} \right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{\frac{-1}{d+4}} \right), \tag{19}$$

Finally, an estimator of $MWISE[r_{\varphi_n}]$ is expressed as

$$\begin{aligned} &\widehat{MWISE}[r_{\varphi_n}] \\ &= \frac{(d+4)^{\frac{3d+8}{d+4}}}{4^{\frac{d+6}{d+4}} d^{\frac{d}{d+4}} (d+2)^{\frac{d+6}{d+4}}} \left(\widehat{I}_1 - 2\widehat{I}_2 + \widehat{I}_3 \right)^{\frac{d}{d+4}} \left(\widehat{I}_4 - \widehat{I}_5 \right)^{\frac{4}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{\frac{-4}{d+4}} + o\left(n^{\frac{-4}{d+4}} \right). \end{aligned}$$

6.1.2. *Non-Recursive estimator* $\widetilde{r}_{\varphi_n}$*:* To estimate the optimal bandwidth (18), we need to estimate I_j, $j = 1, \ldots, 5$.

Estimation of I_1, I_2 and I_3: For the non-recursive case, the plug-in estimate yields

$$\begin{aligned} \widetilde{I'}_1 &= \frac{1}{n^3 {b'}_n^{2(d+2)}} \sum_{\substack{i,j,k=1 \\ i \neq j \neq k}}^{n} \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{it} - X_{jt}}{b'_n} \right) \prod_{\substack{l=1 \\ l \neq t}}^{d} K_b \left(\frac{X_{il} - X_{jl}}{b_n} \right) \right] \\ &\quad \times \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{it} - X_{kt}}{b'_n} \right) \prod_{\substack{l=1 \\ l \neq t}}^{d} K_b \left(\frac{X_{il} - X_{kl}}{b_n} \right) \right] \varphi(Y_j) \varphi(Y_k), \\ \widetilde{I'}_2 &= \frac{1}{n^3 {b'}_n^{2(d+2)}} \sum_{\substack{i,j,k=1 \\ i \neq j \neq k}}^{n} \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{it} - X_{jt}}{b'_n} \right) \prod_{\substack{l=1 \\ l \neq t}}^{d} K_b \left(\frac{X_{il} - X_{jl}}{b_n} \right) \right] \\ &\quad \times \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{it} - X_{kt}}{b'_n} \right) \prod_{\substack{l=1 \\ l \neq t}}^{d} K_b \left(\frac{X_{il} - X_{kl}}{b_n} \right) \right] \varphi(Y_j) \varphi(Y_i), \\ \widetilde{I'}_3 &= \frac{1}{n^4 {b'}_n^{2(d+2)}} \sum_{\substack{i,j,k,m=1 \\ i \neq j \neq k \neq m}}^{n} \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{it} - X_{jt}}{b'_n} \right) \prod_{\substack{l=1 \\ l \neq t}}^{d} K_b \left(\frac{X_{il} - X_{jl}}{b_n} \right) \right] \\ &\quad \times \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{it} - X_{kt}}{b'_n} \right) \prod_{\substack{l=1 \\ l \neq t}}^{d} K_b \left(\frac{X_{il} - X_{kl}}{b_n} \right) \right] \varphi(Y_i) \varphi(Y_m), \end{aligned}$$

Therefore, we obtain

$$\widetilde{I}_i = \mu^2(\mathbf{K}) \widetilde{I'}_i, \quad i = 1 \ldots 3.$$

Estimation of I_4 and I_5:

$$\widetilde{I}_4 = \frac{1}{n^2 b_n^d} \sum_{\substack{i,k=1 \\ i \neq k}}^{n} \prod_{l=1}^{d} K_b \left(\frac{X_{il} - X_{kl}}{b_n} \right) \varphi(Y_i)^2$$

and

$$\widetilde{I}_5 = \frac{1}{n^2 b_n^d} \sum_{\substack{i,k=1 \\ i \neq k}}^{n} \prod_{l=1}^{d} K_b \left(\frac{X_{il} - X_{kl}}{b_n} \right) \varphi(Y_i) \varphi(Y_k),$$

As a consequence, the plug-in estimator of (18) is indicated by

$$h_n = \left(\left(\frac{\widetilde{I}_4 - \widetilde{I}_5}{\widetilde{I}_1 - 2\widetilde{I}_2 + \widetilde{I}_3} \right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{-\frac{1}{d+4}} \right), \tag{20}$$

Finally, a non-recursive estimator of $MWISE[r_{\varphi n}]$ is determined by

$$\widetilde{MWISE}[\widetilde{r}_{\varphi_n}] = \frac{5}{4} \left(\widetilde{I}_4 - \widetilde{I}_5 \right)^{\frac{4}{d+4}} \left(\widetilde{I}_1 - 2\widetilde{I}_2 + \widetilde{I}_3 \right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{-\frac{4}{d+4}} + o\left(n^{-\frac{4}{d+4}} \right).$$

6.2. **Wild Bootstrap approach.** The basic idea of the wild bootstrap introduced in Hardle and Marron (1991) [6] lies in resampling from the estimated residuals

$$\varepsilon_i = \varphi(Y_i) - r_n(X_i)$$

instead of resampling from the pairs $(Y_i, X_i)_{i=1}^n$ and then investing the obtained data to construct an estimator whose distribution will approximate the distribution of the original estimator. Notice that each bootstrapped residual ε_i is drawn from a two-point distribution, such that

$$\mathbb{E}(\varepsilon_i^*) = 0, \qquad \mathbb{E}(\varepsilon_i^{*2}) = \hat{\varepsilon}_i^2 \qquad \text{and} \quad \mathbb{E}(\varepsilon_i^{*3}) = \hat{\varepsilon}_i^3.$$

Such distribution is expressed by

$$G_i^* = \left(\frac{5+\sqrt{5}}{10} \right) \delta_{\hat{\varepsilon}_i \frac{(1-\sqrt{5})}{2}} + \left(\frac{5-\sqrt{5}}{10} \right) \delta_{\hat{\varepsilon}_i \frac{(1+\sqrt{5})}{2}}.$$

Our adapted procedure for bandwidth selection to estimate the operator r_φ recursively relies on three steps:

(1) Giving the bootstrapped residuals ε_i^* drawn from the distribution G_i^*.
(2) Resampling new observations $\varphi(Y_i^*) = r_n(X_i, g) + \varepsilon_i^*$ such that g should be oversmoothed (g needs to be larger than h).
(3) Computing the kernel regression estimator $r_n^*(X_i, h)$, based on the bootstrapped data $(X_i, Y_i^*)_{i=1}^n$.

The bootstrapped bandwidth h^* is then indicated by:

$$h^* = \underset{h \in H}{argmin} \left(\frac{1}{N_B} \sum_{i=1}^{N_B} (r_n^*(X_i, h) - r_n(X_i, g))^2 \right), \tag{21}$$

where H is a fixed set of bandwidths and N_B is the number of replications.
In order to ameliorate the performance of the bootstrap procedure over the plug-in method, we set $H =]h_n - \epsilon, h_n + \epsilon[$, where h_n is the plug-in bandwidth and ϵ is quite close to zero.

7. Confidence intervals

Now, let ϕ denote the distribution function of the standard normal distribution, and let $t_{\lambda/2}$ be such that $\phi\left(t_{\frac{\lambda}{2}}\right) = 1 - \frac{\lambda}{2}$ with $\lambda \in (0,1)$. We set

$$I_{r_n} = \left[r_n(x) - t_{\frac{\lambda}{2}} \Lambda, r_n(x) + t_{\frac{\lambda}{2}} \Lambda \right],$$

with

$$\Lambda = \sqrt{C_f(r_n) \left[C_\sigma(r_n) \sigma_n^2(x) - C_r(r_n) r_n^2(x) \right]}, \qquad \sigma_n^2(x) = \frac{1}{n} \sum_{i=1}^n (\varphi(Y_i) - r_n(X_i))^2$$

and

r_n	case	$C_f(r_n)$	$C_\sigma(r_n)$	$C_r(r_n)$
$r_{\varphi n}$	$\alpha > \beta$	$\frac{\beta_n R(\mathbf{K})}{h_n^d f_n(x)}$	$\frac{1}{2-(\beta-ad)\xi_\beta}$	$\frac{2}{1-(\beta-ad-\xi_\beta^{-1})\xi_\alpha} - \frac{\xi_{\alpha,\beta}}{2-(\alpha-ad)\xi_\alpha} - \frac{1}{2-(\beta-ad)\xi_\beta}$
$r_{\varphi n}$	$\alpha < \beta$	$\frac{\gamma_n R(\mathbf{K})}{h_n^d f_n(x)}$	$\frac{\xi_{\beta,\alpha}}{2-(\beta-ad)\xi_\beta}$	$\frac{2}{1-(\alpha-ad-\xi_\alpha^{-1})\xi_\beta} - \frac{1}{2-(\alpha-ad)\xi_\alpha} - \frac{\xi_{\beta,\alpha}}{2-(\beta-ad)\xi_\beta}$
$r_{\varphi n}$	$\gamma_n = \beta_n = \frac{1}{n}$	$\frac{R(\mathbf{K})}{nh_n^d f_n(x)}$	$\frac{1}{1+ad}$	0
$\widetilde{r}_{\varphi n}$	$\frac{1}{n}$	$\frac{R(\mathbf{K})}{nh_n^d \widetilde{f}_n(x)}$	1	0

In fact, since we have (32) and considering that

$$\mathbf{\Sigma}_{\beta,n}(x) = \frac{R(\mathbf{K})}{f_n(x)}\left[\frac{\sigma_n^2(x)}{2-(\beta-ad)\xi_\beta} - r_n^2(x)\left(\frac{2}{1-(\beta-ad-\xi_\beta^{-1})\xi_\alpha} - \frac{\xi_{\alpha,\beta}}{2-(\alpha-ad)\xi_\alpha} - \frac{1}{2-(\beta-ad)\xi_\beta}\right)\right]$$

is an estimator of (13), the Confidence Intervals for means with unknown standard deviation approach ensure

$$\mathbb{P}\left[-t_{\frac{\lambda}{2}} < \sqrt{\beta_n^{-1}h_n^d}\left(\frac{r_n(x)-\mathbb{E}[r_n(x)]}{\sqrt{\mathbf{\Sigma}_{\beta,n}(x)}}\right) < t_{\frac{\lambda}{2}}\right] = 1-\lambda.$$

Therefore, a confidence interval for the coverage error is given by

$$I_{r_n} = \left[r_n(x) - t_{\frac{\lambda}{2}}\sqrt{\frac{\mathbf{\Sigma}_{\beta,n}(x)}{\beta_n^{-1}h_n^d}}\ ,\ r_n(x) + t_{\frac{\lambda}{2}}\sqrt{\frac{\mathbf{\Sigma}_{\beta,n}(x)}{\beta_n^{-1}h_n^d}}\right].$$

8. Numerical applications

The main target of this section is to perform a simulation study comparing the performance of our semi-recursive estimator (1) to that of Nadaraya-Watson (2) from confidence interval point of view. Throughout this section, we consider the regression model defined as

$$\varphi(Y) = r_\varphi(X) + \varepsilon,$$

where X follows the multivariate normal distribution $\mathcal{N}(0_d, \sigma I_d)$ and ε follows the normal distribution $\mathcal{N}(0, \sigma_\varepsilon)$, with σ and σ_ε are two positive constants smaller than 2.

8.1. **Simulation studies.** We shall start by specifying our kernel function $\mathbf{K}$. The Gaussian kernel is considered. This choice is not carried out at random but according to several criteria. The Gaussian kernel has as an expression

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right), \text{ for all } x \in \mathbb{R} \text{ with } \phi(K) = \frac{1}{\sqrt{\pi}} \text{ and } \mu(K) = 1.$$

When applying our estimator r_{φ_n}, we must choose three quantities:

- The Gaussian kernel $\mathbf{K}$.
- The stepsizes $(\beta_n, \gamma_n) = \left(\beta_0 n^{-1}, \gamma_0 n^{-1}\right)$, where $\beta_0 = 1$ and $\gamma_0 = 1$.
- The bandwidth (h_n) which is chosen to be equal to (19) for plug-in recursive estimator (resp. (21) for bootstrapped recursive one).

For this special case, we set

$$\widehat{I}_n = \left[r_{\varphi_n}(x) - 1.96\sqrt{\frac{R(\mathbf{K})\sum_{i=1}^{n}\left(\varphi(Y_i) - r_{\varphi_n}(X_i)\right)^2}{(1+ad)n^2 h_n^d f_n(x)}}, \right.$$
$$\left. r_{\varphi_n}(x) + 1.96\sqrt{\frac{R(\mathbf{K})\sum_{i=1}^{n}\left(\varphi(Y_i) - r_{\varphi_n}(X_i)\right)^2}{(1+ad)n^2 h_n^d f_n(x)}} \right].$$

When applying our estimator $\widetilde{r}_{\varphi_n}$, we have to opt for two quantities:

- The Gaussian kernel $\mathbf{K}$.
- The bandwidth (h_n) which is chosen to be equal to (20) for plug-in non-recursive estimator (resp. (21) for bootstrapped non-recursive one).

For this special case, we set

$$\widetilde{I}_n = \left[\widetilde{r}_{\varphi_n}(x) - 1.96\sqrt{\frac{R(\mathbf{K})\sum_{i=1}^{n}\left(\varphi(Y_i) - \widetilde{r}_{\varphi_n}(X_i)\right)^2}{n^2 h_n^d \widetilde{f}_n(x)}}, \right.$$
$$\left. \widetilde{r}_{\varphi_n}(x) + 1.96\sqrt{\frac{R(\mathbf{K})\sum_{i=1}^{n}\left(\varphi(Y_i) - \widetilde{r}_{\varphi_n}(X_i)\right)^2}{n^2 h_n^d \widetilde{f}_n(x)}} \right].$$

In what follows, we denote by r_i^* the reference regression, by r_i the test regression and by L_i the average length of the test confidence interval, then we compute the following measures:

- Mean squared error: $MSE = \frac{1}{n}\sum_{i=1}^{n}(r_i - r_i^*)^2$.
- The linear correlation: $Cor = Cov(r_i, r_i^*)\sigma(r_i)^{-1}\sigma(r_i^*)^{-1}$.
- Mean amplitude of the confidence interval: $MAIC = \frac{1}{Np}\sum_{i=1}^{Np} L_i$.

Aiming to compare the proposed semi-recursive estimator to the non-recursive Nadaraya-Watson one, we consider four sample sizes: n = 50, 100, 200 and 500, a fixed number of simulations : N=500 and two models:

- Model 1: X follows the normal distribution $\mathcal{N}(0,5)$ and $r_\varphi(x) = \frac{1}{1+\exp(-x)}$.

- Model 2: X follows the standard bivariate normal distribution $\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1&0\\0&1\end{pmatrix}\right)$ and $r_\varphi(x_1, x_2) = \exp(-x_1^2) + \sin(x_2)$.

				Plug-in		Bootstrap	
Model	σ_ε		n	Nadaraya's estimator	Recursive estimator	Nadaraya's estimator	Recursive estimator
Model 1	0.01	MSE	50	0.00316	0.00282	0.00307	**0.00273**
			100	0.00176	0.00165	0.00171	**0.00159**
			200	0.00111	0.00103	0.00108	**0.00099**
			500	0.00090	0.00085	0.00088	**0.00083**
		Cor	50	0.99499	0.99573	0.99512	**0.99585**
			100	0.99690	0.99723	0.99699	**0.99731**
			200	0.99797	0.99819	0.99802	**0.99825**
			500	0.99828	0.99841	0.99832	**0.99845**
		$MAIC$	50	0.05291	0.05051	0.04785	**0.04556**
			100	0.03647	0.03533	0.02995	**0.02911**
			200	0.02599	0.02479	0.02181	**0.02103**
			500	0.02089	0.02039	0.01663	**0.01614**
Model 2	0.01	MSE	50	0.04793	0.04395	0.04683	**0.04317**
			100	0.04366	0.03475	0.04086	**0.03142**
			200	0.01696	0.00593	0.01698	**0.00592**
			500	0.01596	0.00438	0.01534	**0.00436**
		Cor	50	0.98491	0.98493	0.98295	**0.98600**
			100	0.98738	0.98756	0.98746	**0.98772**
			200	0.98387	0.99434	0.98379	**0.99423**
			500	0.99639	0.99739	0.99640	**0.99739**
		$MAIC$	50	0.53178	0.50550	0.52328	**0.47993**
			100	0.43686	0.41610	0.42394	**0.37903**
			200	0.45451	0.36258	0.42906	**0.35669**
			500	0.36321	0.32708	0.34213	**0.30261**

TABLE 1. Quantitative comparison between Nadaraya-Watson estimator and the proposed estimator with stepsizes $(\beta_n, \gamma_n) = (n^{-1}, n^{-1})$ through a plug-in method and a bootstrap one for both models.

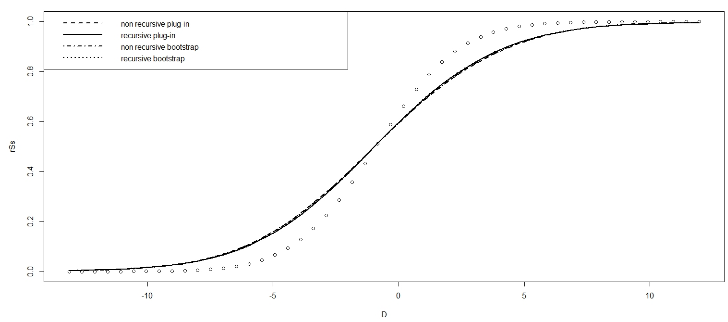

FIGURE 1. Qualitative comparison between the Nadaraya-Watson estimator and the recursive estimator for Model 1 with n=50 and $\sigma_\varepsilon = 0.01$.

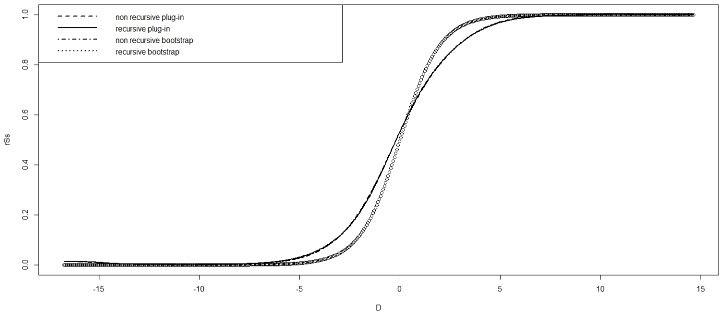

FIGURE 2. Qualitative comparison between the Nadaraya-Watson estimator and the recursive estimator for Model 1 with n=500 and $\sigma_\varepsilon = 0.01$.

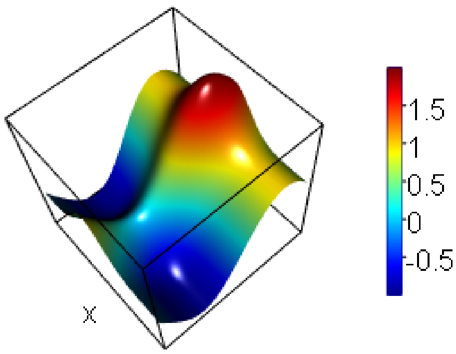

FIGURE 3. The reference regression function for Model 2 for one simple simulation with $n = 500$.

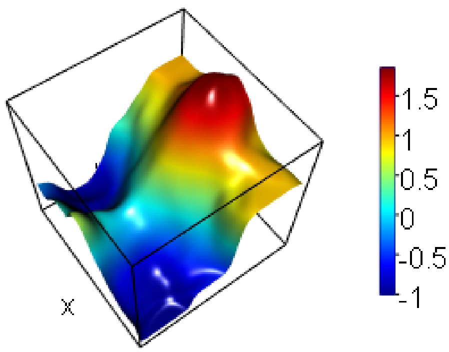

FIGURE 4. The recursive regresion estimator for Model 2 for one simple simulation with $n = 500$.

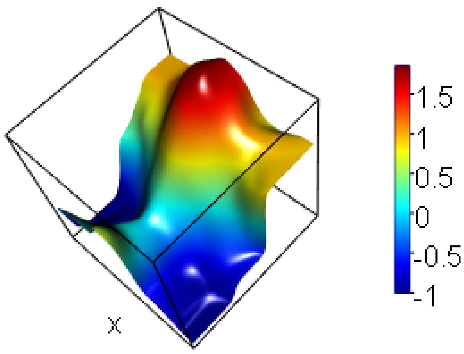

FIGURE 5. The non-recursive regression estimator for Model 2 for one simple simulation with $n = 500$.

Numerical interpretation:
Owing to the specific choice of the bandwidth interval, $H =]h_n - \epsilon, h_n + \epsilon[$, with an appropriate choice of the plug-in bandwidth h_n, the proposed estimator (1) often provides better results compared to the non-recursive Nadaraya Watson's one in terms of estimation error. Thereafter, the use of our recursive estimator enables us to get closer to the true regression function rather than non-recursive one. Meanwhile, even if the modified bootstrap approach outperforms the plug-in method, it's not quite accurate to assert that one method is better than the other. We recommend the reader to consult Delaigle and Gijbels (2004) [4] for a detailed comparison of practical bandwidth selection procedures. They are indistinguishable and it has been widely proven that they behave similarly.

8.2. **Real Datasets.**

8.2.1. *Application 1: French Hospital Data of COVID19.*

The French Hospital data of the COVID-19 epidemic are found in https://www.data.gouv.fr/fr/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-covid-19/. The **Santé publique France**'s mission is devoted to improve and protect the health of population. During the health crisis related to the COVID-19 epidemic, **Santé publique France** has been in charge of monitoring and understanding the dynamics of the epidemic, anticipating the different scenarios and implementing actions so as to prevent and limit the spread of this virus on the national territory.

Description of the dataset

This dataset provides information on the hospital situation regarding the COVID-19 epidemic. We have chosen the first proposed file:

Hospital data related to the COVID-19 epidemic by department (dep) and sex (sex) of the patient: number of hospitalized patients (hosp), number of persons currently in intensive care or resuscitation (rea), number of persons currently in follow-up and rehabilitation care (SSR) or long-term care units (USLD), number of persons currently in conventional hospitalization (HospConv), number of persons currently hospitalized in another type of service (autres) or cumulative number of persons having returned home (rad), cumulative number of persons who died (dc).

The data are daily updated. For the current application, we have selected the data of

28/07/2021, with a total of 150894 observations. For simplicity reasons, we opted for focusing just on the department of 'Paris' database.
As a matter of fact, our application rests upon a dataframe of 1494 observations and 6 variables. The following two models are considered :

- Model 1: $X =$ rea, $Y =$ hosp and $\varphi : y \longmapsto y$.
- Model 2: $X_1 =$ rea, $X_2 =$ dc, $Y =$ hosp and $\varphi : y \longmapsto y$.

	Plug-in		Bootstrap	
Model	Nadaraya's estimator	Recursive estimator	Nadaraya's estimator	Recursive estimator
Model 1	3.64817	3.64737	3.64819	**3.64736**
Model 2	2.85212	2.51463	2.85266	**2.51417**

Table 2. Quantitative comparison between Nadaraya-Watson estimator and the proposed one with stepsizes $(\beta_n, \gamma_n) = (n^{-1}, n^{-1})$ through plug-in method and the bootstrap one.

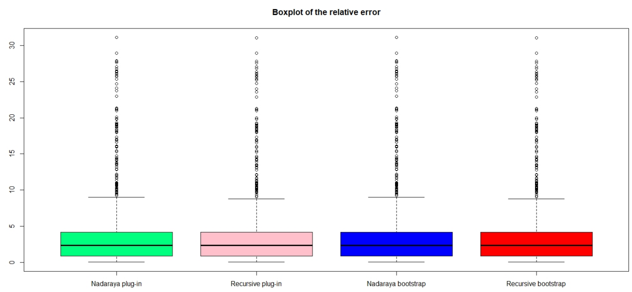

Figure 6. Box-plot of the relative error estimation of the four considered estimators for the bivariate COVID-19 application Model 1.

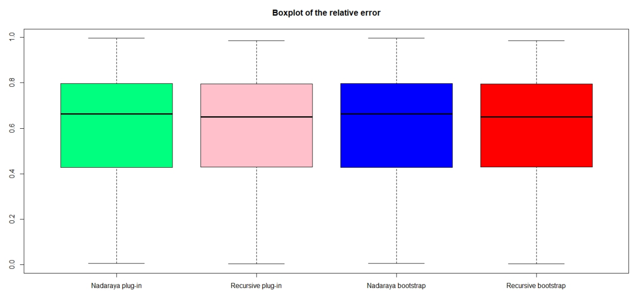

Figure 7. Box-plot of the relative error estimation of the four considered estimators for the bivariate COVID-19 application Model 1.

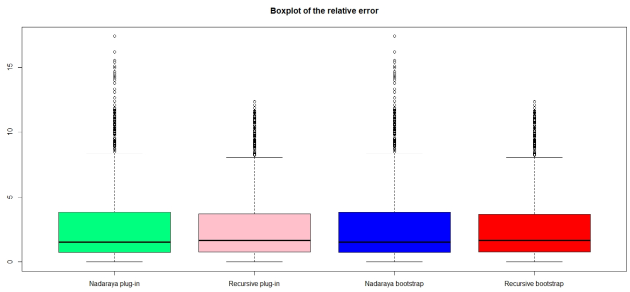

FIGURE 8. Box-plot of the relative error estimation of the four considered estimators for the bivariate COVID-19 application Model 2.

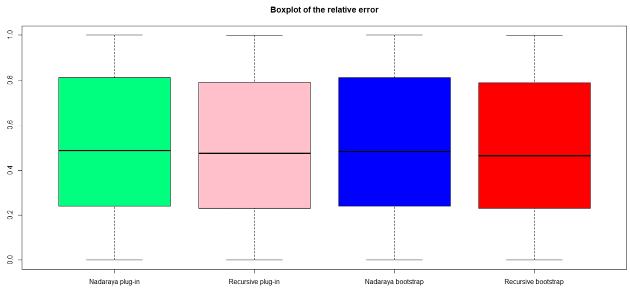

FIGURE 9. Box-plot of the relative error estimation of the four considered estimators for the bivariate COVID-19 application Model 2.

8.2.2. *Application 2: Plasmodium falciparum Parasite Load.*
As far as our application is concerned, we considered a dataset of 176 families belonging to Senegal, living in two villages of Niakhar (Diohine and Toucar), with 505 children aged between 2 and 19 years old. The total number of observations was 6986. We measured Plasmodium falciparum Parasite Load (PL) from thick blood smears obtained by finger-prick during two different seasons and regularly over a three-year observation period (2001–2003). The number of measurements per child ranged from 1 to 15. For more details about the data, we refer the reader to consult Milet *et al.* (2010) [9]. This application relies upon the following variables:

- PL : Parasite Load, as our response variable Y.
- malariae : The presence of co-infection with P. malariae, a factor with two levels (infected: 1 or not infected: 0).
- sex : A factor with two levels (a boy: 0 or a girl: 1).
- age : Age of the child in years between 2 and 19.
- season : A factor with two levels (July-October and October-March).

Therefore, for our selection we have a dataframe of 500 observations and 3 variables. The following two models are considered :

- Model 3: $X_1 = \text{sex}$, $X_2 = \text{age}$, $Y = \text{PL}$ and $\varphi : y \longmapsto \log(y+1)$.
- Model 4: $X_1 = \text{age}$, $X_2 = \text{malariae}$, $X_3 = \text{season}$, $Y = \text{PL}$ and $\varphi : y \longmapsto \log(y+1)$.

	Plug-in		Bootstrap	
Model	Nadaraya's estimator	Recursive estimator	Nadaraya's estimator	Recursive estimator
Model 3	0.80578	**0.80153**	0.80587	0.80194
Model 4	0.80199	0.79629	0.80253	**0.79547**

TABLE 3. Quantitative comparison between Nadaraya-Watson estimator and the proposed one with stepsizes $(\beta_n, \gamma_n) = (n^{-1}, n^{-1})$ through plug-in method and the bootstrap one.

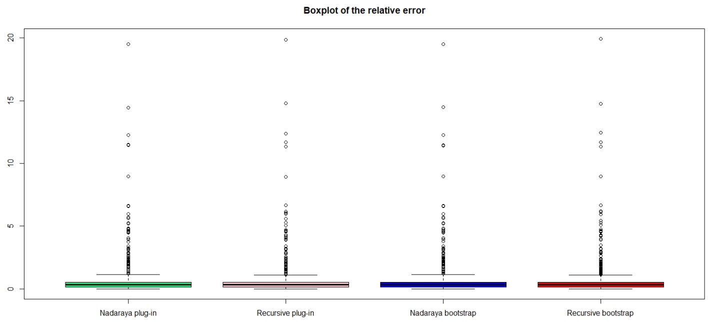

FIGURE 10. Box-plot of the relative error estimation of the four considered estimators for the multivariate PL application Model 3.

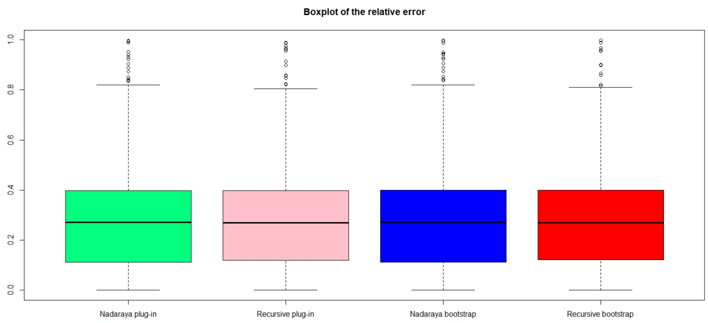

FIGURE 11. Box-plot of the relative error estimation of the four considered estimators for the multivariate PL application Model 3.

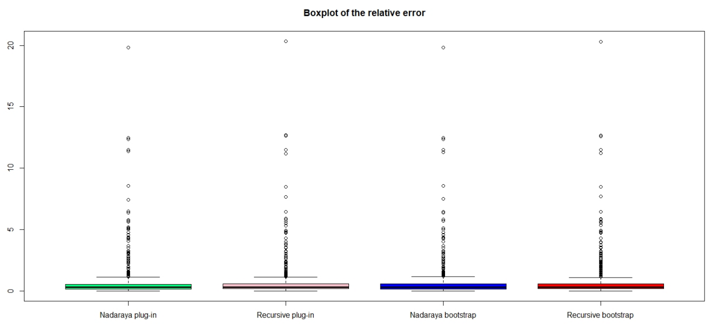

FIGURE 12. Box-plot of the relative error estimation of the four considered estimators for the multivariate PL application Model 4.

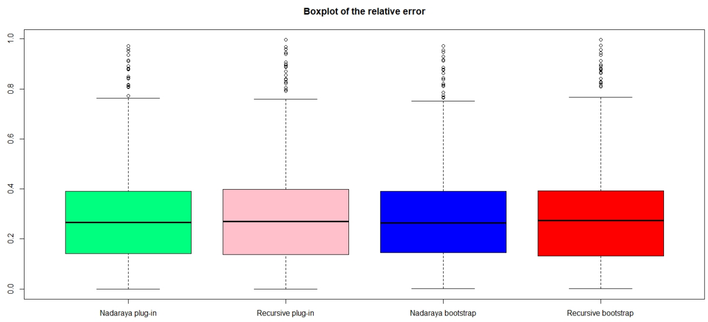

FIGURE 13. Box-plot of the relative error estimation of the four considered estimators for the multivariate PL application Model 4.

9. Conclusion

This paper reports an extension of the semi-recursive regression function estimator. Initially, we tackled the asymptotic properties of the proposed estimator in order to demonstrate that our estimator asymptotically follows a normal distribution. The proposed estimator was compared to the non-recursive multivariate Nadaraya Watson regression estimator. Basically, we revealed that using a specific bandwidth selection, the plug-in approach as well as the bootstrap procedure, and particular stepsizes couple $(\gamma_n, \beta_n) = \left(n^{-1}, n^{-1}\right)$; the proposed estimator (1) often provides better results compared to the non-recursive Nadaraya Watson's one in terms of estimation error. The simulation studies and real datasets illustrate our findings. In conclusion, the use of our recursive estimator, with an appropriate choice of the bandwidth, enables us to get closer to the true regression function rather than non-recursive one.

10. proofs

Throughout this section, we will need the following notations:

$$\mathcal{Z}_n(x) = h_n^{-d}\varphi(Y_n)\mathbf{K}\left(\frac{x - X_n}{h_n}\right) \text{ and } \mathcal{W}_n(x) = h_n^{-d}\mathbf{K}\left(\frac{x - X_n}{h_n}\right), \text{ for all } x \in \mathbb{R}^d.$$

First of all, we introduce a lemma that will be widely used for the study of our estimator r_{φ_n}. This lemma's proof was recorded in Mokkadem *et al.* (2009a) [11].

Lemma 10.1. *Let $(v_n)_{n\geq 1} \in \mathcal{GS}(v^*), (\gamma_n)_{n\geq 1} \in \mathcal{GS}(-\alpha)$ and let $m > 0$ such that $m - v^*\xi > 0$. Hence,*

$$\lim_{n\to+\infty} v_n \Pi_n^m \sum_{k=1}^{n} \Pi_k^{-m} \frac{\gamma_k}{v_k} = \frac{1}{m - v^*\xi}.$$

Furthermore, for any positive sequence $(\alpha_n)_{n\geq 1}$ such that $\lim_{n\to+\infty} \alpha_n = 0$ and all $C \in \mathbb{R}$,

$$\lim_{n\to+\infty} v_n \Pi_n^m \left[\sum_{k=1}^{n} \Pi_k^{-m} \frac{\gamma_k}{v_k}\alpha_k + C\right] = 0.$$

In the next paragraph, we shall depict the asymptotic properties of the multivariate density estimator f_n developed in Mokkadem *et al.* (2009a) [11].

Proposition 10.1. *(Bias and variance of f_n) Under the assumptions $(A_1) - (A_3)$, and assuming that, for all $i, j \in \{1, \ldots, d\}$, $f_{ij}^{(2)}$ is continuous at x; we obtain*

(1) *If* $a \in \left(0, \frac{\alpha}{d+4}\right]$, *then*

$$\mathbb{E}[f_n(x)] - f(x) = \frac{h_n^2}{2(1-2a\xi_\alpha)} \sum_{j=1}^{d} \mu_j(\mathbf{K}) f_{jj}^{(2)}(x) + o\left(h_n^2\right). \tag{22}$$

If $a \in \left(\frac{\alpha}{d+4}, 1\right)$, *then*

$$\mathbb{E}[f_n(x)] - f(x) = o\left(\sqrt{\gamma_n h_n^{-d}}\right). \tag{23}$$

(2) *If* $a \in \left(0, \frac{\alpha}{d+4}\right)$, *then*

$$Var[f_n(x)] = o\left(h_n^4\right). \tag{24}$$

If $a \in \left[\frac{\alpha}{d+4}, 1\right)$, *then*

$$Var[f_n(x)] = \frac{\gamma_n}{h_n^d} \frac{1}{2-(\alpha-ad)\xi_\alpha} f(x) R(\mathbf{K}) + o\left(\gamma_n h_n^{-d}\right). \tag{25}$$

Proof. (Proposition 3.1)
This proof is mainly based on the same concept as Slama *et al.* (2020) [17], by assuming $a := a_\varphi$.
To this extent, we just briefly outline the proof. We have

$$a_{\varphi n}(x) - a_\varphi(x) = Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k (\mathcal{Z}_k(x) - a_\varphi(x)) + Q_n[a_0(x) - a_\varphi(x)].$$

Hence,

$$\mathbb{E}[a_{\varphi n}(x)] - a_\varphi(x) = Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k (\mathbb{E}[\mathcal{Z}_k(x)] - a_\varphi(x)) + Q_n[a_0(x) - a_\varphi(x)].$$

Bias of $a_{\varphi n}$: Resting upon the assumptions (A_1) and (A_3) and by applying Taylor's development formula for a_φ, we deduce that

$$\begin{aligned}
\mathbb{E}[\mathcal{Z}_k(x)] - a_\varphi(x) &= \int_{\mathbb{R}^d} h_k^{-d} \mathbf{K}\left(\frac{x-y}{h_k}\right) \mathbb{E}[\varphi(Y)|X=y] f(y) dy - \int_{\mathbb{R}^d} \mathbf{K}(y)\, a_\varphi(x) dy \\
&= \int_{\mathbb{R}^d} \mathbf{K}(z) \left[a_\varphi(x - zh_k) - a_\varphi(x)\right] dz \\
&= \int_{\mathbb{R}^d} \mathbf{K}(z) \left[\sum_{i=1}^{d} \frac{\partial a}{\partial x_i}(x) z_i h_k + \int_0^1 (1-t) \sum_{i,j=1}^{d} \frac{\partial^2 a}{\partial x_i \partial x_j}(x + tzh_k) z_i z_j h_k^2 dt\right] dz \\
&= \frac{h_k^2}{2} \sum_{j=1}^{d} \mu_j(\mathbf{K}) a_{\varphi jj}^{(2)}(x) + h_k^2 \eta_k(x),
\end{aligned}$$

where $\eta_k(x) = \sum_{i,j=1}^{d} \int_{\mathbb{R}^d} \int_0^1 (1-t) \left[a_{\varphi ij}^{(2)}(x + tzh_k) - a_{\varphi ij}^{(2)}(x)\right] z_i z_j \mathbf{K}(z) dt dz.$
We thus get

$$\begin{aligned}
\mathbb{E}[a_{\varphi n}(x)] - a_\varphi(x) &= \frac{1}{2} \sum_{j=1}^{d} \mu_j(\mathbf{K}) a_{\varphi jj}^{(2)}(x) Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k h_k^2 + Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k h_k^2 \eta_k(x) \\
&\quad + Q_n[a_0(x) - a_\varphi(x)].
\end{aligned}$$

For the case $a > \beta/(d+4)$, we have $\lim_{+\infty}(n\beta_n) > \frac{\alpha-a}{2}$, which ensures that $h_n^2 = o\left(\sqrt{\beta_n h_n^{-d}}\right)$. Therefore, the application of lemma 10.1 entails

$$\begin{aligned}\mathbb{E}[a_{\varphi_n}(x)] - a_\varphi(x) &= \frac{1}{2}\sum_{j=1}^{d}\mu_j(\mathbf{K})a_{\varphi_{jj}}^{(2)}(x)Q_n\sum_{k=1}^{n}Q_k^{-1}\beta_k o\left(\sqrt{\beta_k h_k^{-d}}\right)\\ &\quad + Q_n\sum_{k=1}^{n}Q_k^{-1}\beta_k o\left(\sqrt{\beta_k h_k^{-d}}\right) + O(Q_n)\\ &= o\left(\sqrt{\beta_n h_n^{-d}}\right).\end{aligned}$$

For the case $a \leqslant \beta/(d+4)$, we have $\lim_{+\infty}(n\beta_n) > 2a$ and then $1 - 2a\xi_\beta > 0$. Hence, the application of lemma 10.1 provides

$$\begin{aligned}&\mathbb{E}[a_{\varphi_n}(x)] - a_\varphi(x)\\ &= \frac{1}{2}\sum_{j=1}^{d}\mu_j(\mathbf{K})a_{\varphi_{jj}}^{(2)}(x)Q_n\sum_{k=1}^{n}Q_k^{-1}\beta_k h_k^2 + Q_n\sum_{k=1}^{n}Q_k^{-1}\beta_k o\left(h_k^2\right) + O(Q_n)\\ &\qquad = \frac{h_n^2}{2(1-2a\xi_\beta)}\sum_{j=1}^{d}\mu_j(\mathbf{K})a_{\varphi_{jj}}^{(2)}(x) + o\left(h_n^2\right) + o(1) + O(Q_n).\end{aligned}$$

Variance of a_{φ_n} : For the variance, we infer that

$$\begin{aligned}Var[a_{\varphi_n}(x)] &= Var\left[Q_n\sum_{k=1}^{n}Q_k^{-1}\beta_k\mathcal{Z}_k(x)\right]\\ &= Q_n^2\sum_{k=1}^{n}Q_k^{-2}\beta_k^2\left(\mathbb{E}[\mathcal{Z}_k^2(x)] - \mathbb{E}[\mathcal{Z}_k(x)]^2\right).\end{aligned}$$

We have

$$\begin{aligned}\mathbb{E}[\mathcal{Z}_k^2(x)] &= \int_{\mathbb{R}^d} h_k^{-2d}\mathbb{E}[\varphi(Y)^2|X=y]\mathbf{K}^2\left(\frac{x-y}{h_k}\right)f(y)dy\\ &= \int_{\mathbb{R}^d} h_k^{-d}\mathbf{K}^2(z)\,\mathbb{E}[\varphi(Y)^2|X=x-zh_k]f(x-zh_k)dz\\ &= h_k^{-d}\left[\mathbb{E}[\varphi(Y)^2|X=x]f(x)\int_{\mathbb{R}^d}\mathbf{K}^2(z)\,dz + \nu_k(x)\right],\end{aligned}$$

with

$$\nu_k(x) = \int_{\mathbb{R}^d}\mathbf{K}^2(z)\left[\mathbb{E}[\varphi(Y)^2|X=x-zh_k]f(x-zh_k) - \mathbb{E}[\varphi(Y)^2|X=x]f(x)\right]dz.$$

Thus,

$$\begin{aligned}Var[a_{\varphi_n}(x)] &= Q_n^2\sum_{k=1}^{n}Q_k^{-2}\beta_k^2 h_k^{-d}\\ &\qquad\times\left[\mathbb{E}[\varphi(Y)^2|X=x]f(x)\int_{\mathbb{R}^d}\mathbf{K}^2(z)\,dz + \nu_k(x) - h_k^d\eta_k(x)\right],\end{aligned}$$

where $\eta_k(x) = \left(\int_{\mathbb{R}^d}\mathbf{K}(z)\,a_\varphi(x-zh_k)dz\right)^2$.

For the case $a \geqslant \beta/(d+4)$, we have $\lim\limits_{+\infty}(n\beta_n) > \frac{\alpha-ad}{2}$ and therefore $1-2a\xi_\beta > 0$. Since we have $\lim\limits_{k\to+\infty} h_k\eta_k(x) = 0$ and Taylor's expansions for

$$x \longmapsto \mathbb{E}[\varphi(Y)^2|X=x]f(x) = \int_{\mathbb{R}} \varphi(y)^2 g(x,y)dy$$

ensures that $\nu_k(x) = o(1)$, then the application of lemma 10.1 yields

$$\begin{aligned}
Var[a_{\varphi_n}(x)] &= Q_n^2 \sum_{k=1}^{n} Q_k^{-2}\beta_k^2 h_k^{-d}\left[\mathbb{E}[\varphi(Y)^2|X=x]f(x)R(\mathbf{K}) + \nu_k(x) - h_k^d\eta_k(x)\right] \\
&= Q_n^2 \sum_{k=1}^{n} Q_k^{-2}\beta_k^2 h_k^{-d}\left[\mathbb{E}[\varphi(Y)^2|X=x]f(x)R(\mathbf{K}) + o\left(1\right)\right] \\
&= \frac{\mathbb{E}[\varphi(Y)^2|X=x]}{2-(\alpha-ad)\xi_\beta}\frac{\beta_n}{h_n}[f(x)R(\mathbf{K}) + o\left(1\right)].
\end{aligned}$$

Thus, the result is indicated in terms of

$$Var[a_{\varphi_n}(x)] = \frac{\mathbb{E}[\varphi(Y)^2|X=x]}{2-(\alpha-a)\xi_\beta}\frac{\beta_n}{h_n}f(x)R(\mathbf{K}) + o\left(\frac{\beta_n}{h_n}\right).$$

For the case $a < \beta/(d+4)$, we have $\lim\limits_{+\infty}(n\beta_n) > 2a$, which ensures that $\beta_n h_n^{-d} = o\left(h_n^4\right)$. By applying lemma 10.1, we obtain

$$\begin{aligned}
Var[a_{\varphi_n}(x)] &= Q_n^2 \sum_{k=1}^{n} Q_k^{-2}\beta_k^2 h_k^{-d}\left[\mathbb{E}[\varphi(Y)^2|X=x]f(x)R(\mathbf{K}) + o\left(1\right)\right] \\
&= Q_n^2 \sum_{k=1}^{n} Q_k^{-2}\beta_k o\left(h_k^4\right) \\
&= o\left(h_n^4\right).
\end{aligned}$$

$\square$

Proof. (Theorem 3.1)
This proof is based on the following observation

$$r_{\varphi_n}(x) - r_\varphi(x) = D_n(x)\frac{f(x)}{f_n(x)},\ f_n \neq 0 \tag{26}$$

with

$$D_n(x) = \frac{1}{f(x)}\left(a_{\varphi_n}(x) - a_\varphi(x)\right) - \frac{r_\varphi(x)}{f(x)}\left(f_n(x) - f(x)\right).$$

The only remaining point concerns the asymptotic behaviour of $r_{\varphi_n}(x) - r_\varphi(x)$, which can be deduced from that of $D_n(x)$. Hence, we can state

$$\mathbb{E}[D_n(x)] = \frac{1}{f(x)}\left(\mathbb{E}[a_{\varphi_n}(x)] - a_\varphi(x)\right) - \frac{r_\varphi(x)}{f(x)}\left(\mathbb{E}[f_n(x)] - f(x)\right).$$

Combining the bias of $a_{\varphi_n}(x)$ ((3) and (4)) as well as that of $f_n(x)$ ((22) and (23)) yields the desired results (7) and (8).
For the variance, we get

$$Var[D_n(x)] = \frac{1}{(f(x))^2}Var[a_{\varphi_n}(x)] - \frac{(r_\varphi(x))^2}{(f(x))^2}Var[f_n(x)] - 2\frac{r_\varphi(x)}{(f(x))^2}Cov(a_{\varphi_n}(x), f_n(x)).$$

(1) **For the case $\beta \leq \alpha$:**
Since X_k's are independent, then for all $i \neq k, Cov(\mathcal{Z}_k(x), \mathcal{W}_i(x)) = 0$ and by applying lemma 10.1, classical computations entail

$$Cov(a_{\varphi_n}(x), f_n(x)) = r_\varphi(x) f(x) R(\mathbf{K}) \beta_n h_n^{-d} \left(\frac{1}{1-(\beta - ad - \xi_\beta^{-1})\xi_\alpha} + o\,(1) \right). \tag{27}$$

In fact, we have

$$\begin{aligned}
&Cov(a_{\varphi_n}(x), f_n(x)) \\
&= Cov\left(Q_n \sum_{k=1}^n Q_k^{-1} \beta_k \,\varphi(Y_k)\, h_k^{-d} \mathbf{K}\left(\frac{x-X_k}{h_k}\right), \Pi_n \sum_{i=1}^n \Pi_i^{-1} \gamma_i h_i^{-d} \mathbf{K}\left(\frac{x-X_i}{h_i}\right) \right) \\
&= Q_n \sum_{k=1}^n Q_k^{-1} \beta_k \Pi_n \sum_{i=1}^n \Pi_i^{-1} \gamma_i \, Cov\left(\varphi(Y_k)\, h_k^{-d} \mathbf{K}\left(\frac{x-X_k}{h_k}\right), h_i^{-d} \mathbf{K}\left(\frac{x-X_i}{h_i}\right) \right) \\
&= Q_n \Pi_n \sum_{k=1}^n \Pi_k^{-1} Q_k^{-1} \gamma_k \beta_k Cov\left(\varphi(Y_k)\, h_k^{-d} \mathbf{K}\left(\frac{x-X_k}{h_k}\right), h_k^{-d} \mathbf{K}\left(\frac{x-X_k}{h_k}\right) \right) \\
&= Q_n \Pi_n \sum_{k=1}^n \Pi_k^{-1} Q_k^{-1} \gamma_k \beta_k \left(\mathbb{E}\left[\varphi(Y_k)\, h_k^{-2d} \mathbf{K}^2\left(\frac{x-X_k}{h_k}\right) \right] \right. \\
&\qquad \left. - \mathbb{E}\left[\varphi(Y_k)\, h_k^{-d} \mathbf{K}\left(\frac{x-X_k}{h_k}\right) \right] \mathbb{E}\left[h_k^{-d} \mathbf{K}\left(\frac{x-X_k}{h_k}\right) \right] \right) \\
&= Q_n \Pi_n \sum_{k=1}^n \Pi_k^{-1} Q_k^{-1} \gamma_k \beta_k \Big(\mathbf{E}[\varphi(Y)|X=x] f(x) R(\mathbf{K}) h_k^{-d} + o\left(h_k^{-d}\right) \\
&\qquad - \mathbf{E}[\varphi(Y)|X=x] f^2(x) + o(1) \Big) \\
&= Q_n \Pi_n \sum_{k=1}^n \Pi_k^{-1} Q_k^{-1} \gamma_k \beta_k h_k^{-d} \Big(r_\varphi(x) f(x) R(\mathbf{K}) + o\,(1) \Big) \\
&= \frac{\beta_n h_n^{-d}}{1-(\beta - ad - \xi_\beta^{-1})\xi_\alpha} r_\varphi(x) f(x) R(\mathbf{K}) + o\left(\beta_n h_n^{-d}\right).
\end{aligned}$$

Consequently, (9) and (10) follow from the combination of the variance of $a_{\varphi_n}(x)$ ((5) and (6)), as well as from that of $f_n(x)$ ((24) and (25)) and the covariance expression (27).
It is noteworthy that, for the case $a \geq \beta/(d+4)$, we deduce

$$\begin{aligned}
&Var[r_{\varphi_n}(x)] \\
&= \frac{1}{f(x)} \frac{\beta_n}{h_n^d} \frac{\mathbb{E}[\varphi(Y)^2|X=x]}{2-(\beta-ad)\xi_\beta} R(\mathbf{K}) + o\left(\beta_n h_n^{-d}\right) + \frac{r_\varphi(x)^2}{f(x)} \frac{\gamma_n}{h_n^d} \frac{1}{2-(\alpha-ad)\xi_\alpha} R(\mathbf{K}) \\
&\quad + o\left(\gamma_n h_n^{-d}\right) - 2\frac{r_\varphi(x)}{f(x)^2} \frac{\beta_n h_n^{-d}}{1-(\beta-ad-\xi_\beta^{-1})\xi_\alpha} r_\varphi(x) f(x) R(\mathbf{K}) + o\left(\beta_n h_n^{-d}\right) \\
&= \frac{1}{f(x)} \frac{\beta_n}{h_n^d} \frac{\mathbb{E}[\varphi(Y)^2|X=x]}{2-(\beta-ad)\xi_\beta} R(\mathbf{K}) + o\left(\beta_n h_n^{-d}\right) + \frac{r_\varphi(x)^2}{f(x)} \frac{\beta_n}{h_n^d} \frac{\xi_{\alpha,\beta}}{2-(\alpha-ad)\xi_\alpha} R(\mathbf{K}) \\
&\qquad - 2\frac{r_\varphi(x)^2}{f(x)} \frac{\beta_n h_n^{-d}}{1-(\beta-ad-\xi_\beta^{-1})\xi_\alpha} R(\mathbf{K}) + o\left(\beta_n h_n^{-d}\right) \\
&= \frac{\beta_n}{h_n^d} \frac{R(\mathbf{K})}{f(x)} \left[\frac{\mathbb{E}[\varphi(Y)^2|X=x]}{2-(\beta-ad)\xi_\beta} - r_\varphi(x)^2 \left(\frac{2}{1-(\beta-ad-\xi_\beta^{-1})\xi_\alpha} - \frac{\xi_{\alpha,\beta}}{2-(\alpha-ad)\xi_\alpha} \right) \right]
\end{aligned}$$

$$+ o\left(\beta_n h_n^{-d}\right).$$

(2) **For the case $\alpha < \beta$:**
Similarly to the first case, and taking the stepsize (γ_n) as a reference, we infer the result.

□

Proof. (Theorem 3.2)
We have

$$\begin{aligned} D_n(x) - \mathbb{E}[D_n(x)] &= \frac{1}{f(x)}[a_{\varphi_n}(x) - \mathbb{E}[a_{\varphi_n}(x)]] - \frac{r_\varphi(x)}{f(x)}[f_n(x) - \mathbb{E}[f_n(x)]] \\ &= \frac{1}{f(x)} Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k\left(T_k(x) - \mathbb{E}[T_k(x)]\right), \end{aligned}$$

with

$$T_k(x) = \mathcal{Z}_k(x) - r_\varphi(x) Q_n^{-1}\Pi_n\Pi_k^{-1}Q_k\beta_k^{-1}\gamma_k\mathcal{W}_k(x).$$

We note

$$S_k(x) = Q_k^{-1}\beta_k\left(T_k(x) - \mathbb{E}[T_k(x)]\right). \tag{28}$$

Hence, we can write

$$D_n(x) - \mathbb{E}[D_n(x)] = \frac{1}{f(x)} Q_n \sum_{k=1}^{n} S_k(x). \tag{29}$$

Now, we are trying to apply Lyapunov's theorem for $S_k(x)$. For this reason, we assume

$$\begin{aligned} \upsilon_n^2 &= \sum_{k=1}^{n} Var[S_k(x)] \\ &= \sum_{k=1}^{n} Q_k^{-2}\beta_k^2 Var[T_k(x)] \\ &= \sum_{k=1}^{n} Q_k^{-2}\beta_k^2 Var\left[\mathcal{Z}_k(x)\right] + r_\varphi(x)^2 Q_n^{-2}\Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 Var\left[\mathcal{W}_k(x)\right] \\ &\quad - 2r_\varphi(x) Q_n^{-1}\Pi_n \sum_{k=1}^{n} Q_k^{-1}\beta_k\Pi_k^{-1}\gamma_k Cov\left(\mathcal{Z}_k(x), \mathcal{W}_k(x)\right). \end{aligned}$$

Here, we consider the case $\beta \leq \alpha$. Since we have

$$\begin{aligned} Var\left[\mathcal{Z}_k(x)\right] &= h_k^{-d}\Big(\mathbb{E}[\varphi(Y)^2|X = x] f(x) R(\mathbf{K}) + o(1)\Big), \\ Var\left[\mathcal{W}_k(x)\right] &= h_k^{-d}\Big(f(x) R(\mathbf{K}) + o(1)\Big), \\ Cov\left(\mathcal{Z}_k(x), \mathcal{W}_k(x)\right) &= h_k^{-d}\Big(r_\varphi(x) f(x) R(\mathbf{K}) + o(1)\Big), \end{aligned}$$

then, the application of lemma 10.1 ensures that

$$\begin{aligned} \upsilon_n^2 &= \sum_{k=1}^{n} Q_k^{-2}\beta_k^2 h_k^{-d}\Big(\mathbb{E}[\varphi(Y)^2|X = x] f(x) R(\mathbf{K}) + o(1)\Big) \\ &+ r_\varphi(x)^2 Q_n^{-2}\Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 h_k^{-d}\Big(f(x) R(\mathbf{K}) + o(1)\Big) \\ &- 2r_\varphi(x) Q_n^{-1}\Pi_n \sum_{k=1}^{n} Q_k^{-1}\beta_k\Pi_k^{-1}\gamma_k h_k^{-d}\Big(r_\varphi(x) f(x) R(\mathbf{K}) + o(1)\Big) \end{aligned}$$

$$= \frac{\beta_n}{h_n^d} \frac{f(x)^2}{Q_n^2} [\mathbf{\Sigma}_\beta(x) + o(1)].$$

On the other side, we have

$$\forall p > 0, \quad \mathbb{E}[|T_k(x)|^{2+p}] = O\left(\frac{1}{h_k^{d(1+p)}}\right).$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[|S_k(x)|^{2+p}\right] &= Q_k^{-2-p}\beta_k^{2+p}\mathbb{E}\left[|T_k(x) - \mathbb{E}[T_k(x)]|^{2+p}\right] \\ &\leq 2Q_k^{-2-p}\beta_k^{2+p}\mathbb{E}\left[|T_k(x)|^{2+p}\right].\end{aligned}$$

Hence,

$$\mathbb{E}\left[|S_k(x)|^{2+p}\right] = O\left(Q_k^{-2-p}\beta_k^{2+p}\frac{1}{h_k^{d(1+p)}}\right). \tag{30}$$

As a consequence,

$$\sum_{k=1}^n \mathbb{E}[|S_k(x)|^{2+p}] = O\left(\sum_{k=1}^n Q_k^{-2-p}\beta_k^{2+p}\frac{1}{h_k^{d(1+p)}}\right).$$

In what follows, let us assume that there is $p > 0$, such that

$$\lim_{n\to+\infty}(n\beta_n) > \frac{1+p}{2+p}(\beta - ad).$$

The application of lemma 10.1 yields

$$\sum_{k=1}^n \mathbb{E}[|S_k(x)|^{2+p}] = O\left(\frac{\beta_n^{1+p}}{Q_n^{2+p}h_k^{d(1+p)}}\right).$$

Hence,

$$\frac{1}{\upsilon_n^{2+p}}\sum_{k=1}^n \mathbb{E}[|S_k(x)|^{2+p}] = O\left(\frac{\beta_n^{1+p}}{\upsilon_n^{2+p}Q_n^{2+p}h_n^{d(1+p)}}\right).$$

Thus, we deduce

$$\frac{1}{\upsilon_n^{2+p}}\sum_{k=1}^n \mathbb{E}[|S_k(x)|^{2+p}] = O\left(\left(\frac{\beta_n}{h_n^d}\right)^{p/2}\right) = o(1).$$

In addition, since we have

$$\lim_{n\to+\infty}\frac{1}{\upsilon_n^{2+p}}\sum_{k=1}^n \mathbb{E}\left[|S_k(x) - \mathbb{E}[S_k(x)]|^{2+p}\right] = \lim_{n\to+\infty}\frac{1}{\upsilon_n^{2+p}}\sum_{k=1}^n \mathbb{E}[|S_k(x)|^{2+p}] = 0,$$

therefore, by applying the Lyapunov theorem, we get

$$\frac{1}{\sqrt{\upsilon_n^2}}\sum_{k=1}^n(S_k(x) - \mathbb{E}[S_k(x)]) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}(0,1),$$

which implies

$$\frac{1}{\upsilon_n}\sum_{k=1}^n S_k(x) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}(0,1).$$

Moreover, (26) and (29) ensure that

$$f(x)Q_n^{-1}\upsilon_n^{-1}\left(r_{\varphi_n}(x) - \mathbb{E}[r_{\varphi_n}(x)]\right) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}(0,1). \tag{31}$$

Given that $v_n^2 = \dfrac{\beta_n}{h_n^d}\dfrac{f(x)^2}{Q_n^2}[\mathbf{\Sigma}_\beta(x) + o\,(1)]$, where $\mathbf{\Sigma}_\beta(x)$ is defined in (12) and by replacing v_n with its value in (31), we conclude that

$$\sqrt{\beta_n^{-1}h_n^d}\left(r_{\varphi_n}(x) - \mathbb{E}[r_{\varphi_n}(x)]\right) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0\,,\mathbf{\Sigma}_\beta(x)\right). \tag{32}$$

The convergence in (11) then follows from the application of Lyapounov's Theorem and the combination between (7), (8) and (32).

For the convergence in probability, by applying the Bienaymé–Chebyshev inequality, we get

$$\mathbb{P}\left[\left|\frac{r_{\varphi_n}(x) - r_\varphi(x)}{h_n^2} - \mathbb{E}\left[\frac{r_{\varphi_n}(x) - r_\varphi(x)}{h_n^2}\right]\right| \geq \epsilon\right] \leq \frac{Var[r_{\varphi_n}(x)]}{h_n^4\epsilon^2}.$$

Since we have $\beta_n^{-1}h_n^{d+4} \underset{n\to+\infty}{\longrightarrow} \infty$, then we deduce that

$$\frac{1}{h_n^2}\left(r_{\varphi_n}(x) - r_\varphi(x)\right) \xrightarrow[n\to+\infty]{\mathcal{P}} \mathbf{M}_\beta(x),$$

with $\mathbf{M}_\beta(x)$ is provided in (12). □

Proof. (Theorem 3.3)

For this proof, we state

$$\zeta_n(x) := \sum_{k=1}^{n} S_k(x) = f(x)Q_n^{-1}\left(D_n(x) - \mathbb{E}[D_n(x)]\right),$$

where S_k is given in (28).

Here, we consider the case $\beta \leq \alpha$ and we suppose that $a \geq \frac{\beta}{d+4}$. Set $\beta_0 = h_0 = 1$ and $H_n^2 = Q_n^2\beta_n^{-1}h_n^d$, then we get

$$\begin{aligned}
\ln(H_n^{-2}) &= \ln\left(Q_n^{-2}\right) + \ln(\beta_n h_n^{-d}) \\
&= -2\ln\left(Q_n\right) + \ln\left(\prod_{k=1}^{n}\frac{\beta_{k-1}^{-1}h_{k-1}^{-d}}{\beta_k^{-1}h_k^{-d}}\right) \\
&= -2\sum_{k=1}^{n}\ln\left(1-\beta_k\right) + \sum_{k=1}^{n}\ln\left(1 - \frac{\beta - ad}{k} + o\left(\frac{1}{k}\right)\right) \\
&= -2\sum_{k=1}^{n}\left(-\beta_k + o(\beta_k)\right) + \sum_{k=1}^{n}\left(-(\beta - ad)\beta_k\xi + o(\beta_k)\right) \\
&= \sum_{k=1}^{n}\left(2\beta_k - (\beta - ad)\,\beta_k\xi + o(\beta_k)\right).
\end{aligned}$$

Hence, using the notation $s_n = \sum\limits_{k=1}^{n}\beta_k$, we can write

$$\ln(H_n^{-2}) = \left(2 - (\beta - ad)\xi\right)s_n + o\left(s_n\right). \tag{33}$$

Since $2 - (\beta - ad)\xi > 0$ and s_∞ diverges, then we deduce that $\lim\limits_{n\to+\infty} H_n^{-2} = \infty$ and $\lim\limits_{n\to+\infty}\frac{H_{n-1}^{-2}}{H_n^{-2}} = 1$. Moreover, we have $\sum\limits_{k=1}^{n} Var[S_k(x)] = \dfrac{\beta_n}{h_n^d}\dfrac{f(x)^2}{Q_n^2}[\mathbf{\Sigma}_\beta(x) + o\,(1)]$, where $\mathbf{\Sigma}_\beta(x)$ is defined in (13).

From this perspective, it's obvious that

$$\lim_{n\to+\infty} H_n^2\sum_{k=1}^{n} Var[S_k(x)] = f(x)^2\mathbf{\Sigma}_\beta(x).$$

Considering the particular case of $p=1$ in (30), we have $\mathbb{E}\left[|S_k(x)|^3\right] = O\left(Q_k^{-3}\beta_k^3 h_k^{-2d}\right)$ and then we deduce that

$$\begin{aligned}
\frac{1}{n\sqrt{n}}\sum_{k=1}^{n}\mathbb{E}\left[|H_nS_k(x)|^3\right] &= O\left(\frac{H_n^3}{n\sqrt{n}}Q_k^{-3}\beta_k^3h_k^{-2d}\right)\\
&= O\left(\frac{H_n^3}{n\sqrt{n}}Q_k^{-3}\beta_k\ o\left([\beta_k h_k^{-d}]^{\frac{3}{2}}\right)\right)\\
&= O\left(\frac{H_n^3}{n\sqrt{n}}Q_n^{-3}\ o\left([\beta_n h_n^{-d}]^{\frac{3}{2}}\right)\right)\\
&= o\left(\frac{H_n^3}{n\sqrt{n}}Q_n^{-3}[\beta_n h_n^{-d}]^{\frac{3}{2}}\right)\\
&= o\left(\frac{1}{n\sqrt{n}}\right)\\
&= o\left([\ln(H_n^{-2})]^{-1}\right).
\end{aligned}$$

The application of the LIL Theorem1 in Mokaddem and Pelletier (2007) [10] then ensures that, with probability one, the sequence

$$\left(\frac{H_n\zeta_n(x)}{\sqrt{2\ln\ln(H_n^{-2})}}\right) = \left(\frac{\sqrt{\beta_n^{-1}h_n^d}f(x)\left(D_n(x)-\mathbb{E}[D_n(x)]\right)}{\sqrt{2\ln\ln(H_n^{-2})}}\right)$$

is relatively compact and its limit set is the interval

$$\left[-f(x)\sqrt{\mathbf{\Sigma}_\beta(x)},\, f(x)\sqrt{\mathbf{\Sigma}_\beta(x)}\right].$$

On account of (33), we have $\lim_{n\to+\infty}\frac{\ln\ln(H_n^{-2})}{\ln(sn)} = 1$, and referring to (26) and (29), we deduce that

$$\left(\frac{\sqrt{\beta_n^{-1}h_n^d}\left(r_{\varphi_n}(x)-\mathbb{E}[r_{\varphi_n}(x)]\right)}{\sqrt{2\ln s_n}}\right)$$

is relatively compact and its limit set is the interval

$$\left[-\sqrt{\mathbf{\Sigma}_\beta(x)},\, \sqrt{\mathbf{\Sigma}_\beta(x)}\right].$$

The combination between (7) and (8) then entails

$$\left(\sqrt{\frac{\beta_n^{-1}h_n^d}{2\ln(s_n)}}\left(r_{\varphi_n}(x)-r_\varphi(x)\right)\right)$$

is relatively compact and its limit set is the interval

$$\left[\sqrt{\frac{b}{2}}\mathbf{M}_\beta(x)-\sqrt{\mathbf{\Sigma}_\beta(x)}\,,\,\sqrt{\frac{b}{2}}\mathbf{M}_\beta(x)+\sqrt{\mathbf{\Sigma}_\beta(x)}\right],$$

where $\mathbf{M}_\beta(x)$ is defined in (12) and $\mathbf{\Sigma}_\beta(x)$ is provided in (13).
Now we suppose that $a<\frac{\beta}{d+4}$. Set $H_n^{-2} = Q_n^{-2}h_n^4(\ln\ln(Q_n^{-2}h_n^4)^{-1})$, then we get

$$\begin{aligned}
\ln(Q_n^{-2}h_n^4) &= \ln\left(Q_n^{-2}\right)+\ln(h_n^4)\\
&= -2\ln\left(Q_n\right)+\ln\left(\prod_{k=1}^{n}\frac{h_{k-1}^{-4}}{h_k^{-4}}\right)
\end{aligned}$$

$$= -2\sum_{k=1}^{n} \ln(1-\beta_k) + \sum_{k=1}^{n} \ln\left(1 - \frac{4a}{k} + o\left(\frac{1}{k}\right)\right)$$
$$= -2\sum_{k=1}^{n} \left(-\beta_k + o(\beta_k)\right) + \sum_{k=1}^{n} \left(-4a\beta_k\xi + o(\beta_k)\right).$$

Hence, using the notation $s_n = \sum_{k=1}^{n} \beta_k$, we can write

$$\ln(H_n^{-2}) = (2-4a\xi)\, s_n + o\left(s_n\right). \tag{34}$$

Since $2-4a\xi > 0$ and s_∞ diverges, then we deduce that $\lim_{n\to+\infty} H_n^{-2} = +\infty$ and $\lim_{n\to+\infty} \frac{H_{n-1}^{-2}}{H_n^{-2}} = 1$. Moreover, we have

$$H_n^2 \sum_{k=1}^{n} Var[S_k(x)] = O\left(Q_n^2 h_n^{-4} \ln\ln(Q_n^{-2}h_n^4) \frac{\beta_n}{h_n^d Q_n^2}\right)$$
$$= o(1).$$

Considering the particular case of $p=1$ in (30), we have $\mathbb{E}\left[|S_k(x)|^3\right] = O\left(Q_k^{-3}\beta_k^3 h_k^{-2d}\right)$ and then we deduce that

$$\frac{1}{n\sqrt{n}} \sum_{k=1}^{n} \mathbb{E}\left[|H_n S_k(x)|^3\right] = O\left(\frac{H_n^3}{n\sqrt{n}} \sum_{k=1}^{n} Q_k^{-3}\beta_k^3 h_k^{-2d}\right)$$
$$= O\left(\frac{H_n^3}{n\sqrt{n}} \sum_{k=1}^{n} Q_k^{-3}\beta_k \, o\left(h_k^6\right)\right)$$
$$= o\left(\frac{H_n^3}{n\sqrt{n}} Q_n^{-3} h_n^6\right)$$
$$= o\left([\ln(H_n^{-2})]^{-1}\right).$$

The application of the LIL Theorem then ensures that, with probability one, the sequence

$$\left(\frac{H_n\zeta_n(x)}{\sqrt{2\ln\ln(H_n^{-2})}}\right) = \left(\frac{h_n^{-2}\sqrt{\ln\ln(Q_n^{-2}h_n^4)}f(x)\left(D_n(x) - \mathbb{E}[D_n(x)]\right)}{\sqrt{2\ln\ln(H_n^{-2})}}\right)$$

is relatively compact and its limit set is 0.

On account of (33), we have $\lim_{n\to+\infty} \frac{\ln\ln(H_n^{-2})}{\ln\ln(Q_n^{-2}h_n^4)} = 1$, and referring to (26) and (29), we deduce that

$$\lim_{n\to+\infty} h_n^{-2}\left(r_{\varphi_n}(x) - \mathbb{E}[r_{\varphi_n}(x)]\right) = 0.$$

The combination between (7) and (8) then entails

$$\lim_{n\to+\infty} h_n^{-2}\left(r_{\varphi_n}(x) - r_\varphi(x)\right) = \mathbf{M}_\beta(x),$$

where $\mathbf{M}_\beta(x)$ is defined in (12). □

References

1. N. Altman, and C. Leger, *Bandwidth Selection for Kernel Distribution Function Estimation*, *J. Statist. Plann. Inference.* **46** (1995), 195–214. https://doi.org/10.1016/0378-3758(94)00102-2
2. S. Bouzebda, and Y. Slaoui, *Nonparametric Recursive Method for Kernel-Type Function Estimators for Censored Data*, *J. Stoch. Anal.* **3** (2020), Art. 4, 19. https://doi.org/10.31390/josa.1.3.04

3. P. Deheuvels, and D. M. Mason, *General asymptotic confidence bands based on kernel-type function estimators*, *Stat. Inference Stoch. Process.* **7** (2004), 225–277. https://doi.org/10.1023/B:SISP.0000049092.55534.af
4. A. Delaigle, and I. Gijbels, *Practical bandwidth selection in deconvolution kernel density estimation*, *Comput. Statist. Data Anal.* **45** (2004), 249–267. https://doi.org/10.1016/S0167-9473(02)00329-8
5. U. Einmahl, and D. M. Mason, *An empirical process approach to the uniform consistency of kernel-type function estimators*, *J. Theoret. Probab.* **13** (2000), 1–37. https://doi.org/10.1023/A:1007769924157
6. W. Hardle, and J. S. Marron, *Bootstrap Simultaneous Error Bars for Nonparametric Regression*, *The Ann. Statist.* **19** (1991), 778–796. https://doi.org/10.1214/aos/1176348120
7. J. Galambos, and E. Seneta, *Regularly Varying Sequences*, *Proc. Amer. Math. Soc.* **41** (1973), 110–116. https://doi.org/10.1090/S0002-9939-1973-0323963-5
8. J. Kiefer, and J. Wolfowitz, *Stochastic Estimation of the Maximum of a Regression Function*, *Ann. Math. Statist.* **23** (1952), 462–466. https://doi.org/10.1214/aoms/1177729392
9. J. Milet, AND G. Nuel, AND L. Watier, AND D. Courtin, AND Y. Slaoui, AND P. Senghor, AND F. Migot-Nabias, AND O. Gaye, AND A. Garcia, *Genome Wide Linkage Study, Using a 250K SNP Map, of Plasmodium falciparum Infection and Mild Malaria Attack in a Senegalese Population*, *PLOS ONE* **5** (2010), 1–11. https://doi.org/10.1371/journal.pone.0011616
10. A. Mokkadem, and M. Pelletier, *Compact law of the iterated logarithm for matrix-normalized sums of random vectors*, *Teor. Veroyatn. Primen.* **52** (2007), 752–767. https://doi.org/10.4213/tvp1532
11. A. Mokkadem, M. Pelletier, and Y. Slaoui, *The stochastic approximation method for the estimation of a multivariate probability density*, *J. Statist. Plann. Inference.* **139** (2009a), 2459–2478. https://doi.org/10.1016/j.jspi.2008.11.012
12. A. Mokkadem, M. Pelletier, and Y. Slaoui, *Revisiting Révész's Stochastic Approximation Method for the Estimation of a Regression Function*, *ALEA Lat. Amer.J. Probab.Math. Statist.* **6** (2009b), 63–114.
13. A. Mokkadem, and M. Pelletier, *The Multivariate Révész's Online Estimator of a Regression Function and Its Averaging*, *M. Math. Meth. Stat.* **25** (2016), 151–167. https://doi.org/10.3103/S1066530716030017
14. E. A. Nadaraya, *On Estimating Regression*, *Theory Probab. Appl.* **9** (1964), no. 1, 141–142. https://doi.org/10.1137/1109020
15. P. Révész, *How to Apply the Method of Stochastic Approximation in the Non parametric Estimation of a Regression Function*, *Math. Operationsforsch. Statist. Ser. Statistics.* **8** (1977), 119–126. https://doi.org/10.1080/02331887708801361
16. G. Roussas, *Nonparametric regression estimation under mixing conditions*, *Stochastic Process. Appl.* **36** (1990), 107–116. https://doi.org/10.1016/0304-4149(90)90045-T
17. S. Slama, Y. Slaoui, and H. Fathallah, *Statistical inference for multivariate conditional cumulative distribution function estimation by stochastic approximation method*, hal-03090711.
18. Y. Slaoui, *Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method*, *J. Probab. Stat.* **2014** (2014), 1–11. https://doi.org/10.1155/2014/739640
19. Y. Slaoui, *Plug-in bandwidth selector for recursive kernel regression estimators defined by stochastic approximation method*, *Stat. Neerl.***69** (2015), 483–509. https://doi.org/10.1111/stan.12069
20. Y. Slaoui, *Optimal bandwidth selection for semi-recursive kernel regression estimators*, *Stat. Interface* **9** (2016), 375–388. https://doi.org/10.4310/SII.2016.v9.n3.a11
21. A. B. Tsybakov, *Recurrent Estimation of the Mode of a Multidimensional Distribution*, *Probl. Inf. Transm.* **8** (1990), 119–126.
22. G.S. Watson, *Smooth regression analysis, Sankhya Ser. A.* **26** (1964), 359–372.

Université de Sousse, LAMMDA: Laboratoire de Mathématiques Modélisation Déterministe et Aléatoire, Hammam Sousse, Tunisie
E-mail address: sahar.slama@essths.u-sousse.tn

Université de Poitiers, LMA Laboratoire des Mathématiques et Applications, Futuroscope Chasseneuil, France
E-mail address: Yousri.Slaoui@math.univ-poitiers.fr

Université de Sousse, LAMMDA: Laboratoire de Mathématiques Modélisation Déterministe et Aléatoire, Hammam Sousse, Tunisie
E-mail address: hamdi.fathallah@essths.u-sousse.tn