Theory of Stochastic Processes
Vol. 26 (42), no. 1, 2022, pp. 60–95

Y. SLAOUI AND A. JMAEI

RECURSIVE AND NON-RECURSIVE REGRESSION ESTIMATORS USING BERNSTEIN POLYNOMIALS

If a regression function has a bounded support, the kernel estimates often exceed the boundaries and are therefore biased on and near these limits. In this paper, we focus on mitigating this boundary problem. We apply Bernstein polynomials and the Robbins-Monro algorithm to construct a non-recursive and recursive regression estimator. We study the asymptotic properties of these estimators, and we compare them with those of the Nadaraya-Watson estimator and the generalized Révész estimator introduced by Mokkadem et al. [21]. In addition, through some simulation studies, we show that our non-recursive estimator has the lowest integrated root mean square error (ISE) in most of the considered cases. Finally, using a set of real data, we demonstrate how our non-recursive and recursive regression estimators can lead to very satisfactory estimates, especially near the boundaries.

1. Introduction

The goal in any data analysis is to extract from raw information the accurate estimation. One of the most important and common questions concerning if there is a statistical relationship between a response variable (Y) and an explanatory variable (X_i). An option to answer this question is to employ regression analysis in order to model this relationship.

Let $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ be independent, identically distributed pairs of random variables with joint density function $g(x, y)$, and let f denote the probability density of X. There were many ways to estimate the regression function $r : x \mapsto \mathbb{E}[Y|X = x]$. The most known are the kernel regression estimators. On the non-recursive approach, we refer, among many others, to the estimator proposed by Nadaraya [23] and Watson [43], the alternative kernel estimators given by Priestley and Chao [24] and Gasser and Müller [8]. On the other hand, the recursive estimation was widely discussed, we refer to the approach of Révész [26, 27] and Tsybakov [41] which was studied by Mokkadem et al. [21], Slaoui [32, 33, 34], also we find the semi-recursive approach introduced by Slaoui [35]. The advantage of the recursive estimator is the update of the estimation, which is computationally cost-effective when new data appear in the sample, this advantage given an additional motivation of the present work. Each of the last estimators has its own particular strengths and weaknesses. However, the common problem is the edge effect. In fact, when the regression function has bounded support, kernel estimates often overspill the boundaries and are consequently biased at and near these edges. To overcome this problem, many works are devoted to reducing the effects, we can list Gasser and Müller [8], Gasser et al. [9], Granovsky and Müller [11] and Müller [22] discuss boundary kernel methods. Djojosugito and Speckman [5] investigated boundary bias reduction based on a finite-dimensional projection in a Hilbert space. In this work, we propose a non-recursive and recursive approach of regression estimation using Bernstein polynomials.

2010 *Mathematics Subject Classification.* 62G08, 62L20.

Key words and phrases. Nonparametric regression; Stochastic approximation algorithm; Bernstein polynomial.

The estimation using Bernstein polynomial for density and distribution functions have been widely discussed in several frameworks. See, for instance, the original work of Vitale [42] and extensions given by Tenbusch [39], Ghosal [10], Kakizawa [14, 15], Igarashi and Kakizawa [12], Rao [25], Leblanc [16, 17, 18], Babu et al. [1], Babu and Chaubey [2], Jmaei et al. [13] and more recently Slaoui and Jmaei [36].

The layout of the present paper is as follows. In Section 2, we list our assumptions and notations. In Section 3, we introduce our non-recursive estimator and we compute its bias, variance, mean squared error (MSE), the mean integrated squared error ($MISE$) and we establish a weak convergence rate. In Section 4 we introduce our recursive estimator and we state the main theoretical results. Section 5 is devoted to some numerical studies : first, a simulation study is presented in Subsection 5.1 and, then, an application to a real dataset is described in Subsection 5.2. Finally, we discuss our conclusion in Section 6. To avoid interrupting the flow of this paper, all mathematical developments are relegated to the Appendix.

2. Assumptions and Notations

Let us first define the class of positive sequences that will be used in the statement of our assumptions.

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

$$\lim_{n\to+\infty} n\left[1-\frac{v_{n-1}}{v_n}\right]=\gamma.$$

This condition was introduced by Galambos and Seneta [7] to define regularly varying sequences (see also Bojanic and Seneta [3]). Typical sequences in $\mathcal{GS}(\gamma)$ are, for $b \in \mathbb{R}$, $n^{\gamma}(\log n)^b$, $n^{\gamma}(\log\log n)^b$, and so on.

To obtain the behavior of our estimators, we make to the following assumptions :

- $(A1)$ $(\gamma_n) \in \mathcal{GS}(-\alpha)$, $\alpha \in \left(\frac{3}{4}, 1\right]$.
- $(A2)$ $(m_n) \in \mathcal{GS}(a)$, $a \in \left(\frac{1-\alpha}{4}, \frac{2}{3}\alpha\right)$.
- $(A3)$ (i) $g(s,t)$ is twice continuously differentiable with respect to s.

  (ii) For $q \in \{0,1,2\}$, $s \mapsto \int_{\mathbb{R}} t^q g(s,t)\,dt$ is a bounded function continuous at $s = x$. For $q \in [2,3]$, $s \mapsto \int_{\mathbb{R}} |t|^q g(s,t)\,dt$ is a bounded function.

  (iii) For $q \in \{0,1\}$, $\int_{\mathbb{R}} |t|^q \left|\frac{\partial g}{\partial x}(x,t)\right| dt < \infty$, and $s \mapsto \int_{\mathbb{R}} t^q \frac{\partial^2 g}{\partial s^2}(s,t)\,dt$ is a bounded function continuous at $s = x$.

Assumption $(A1)$ on the stepsize was used in the recursive framework for the estimation of the density function (see Mokkadem et al. [20] and [29, 30]), for the estimation of the distribution function (see Slaoui [31]) and for the estimation of the regression function (see Mokkadem et al. [21] and [34, 35]). This assumption ensures that $\sum_{n\geq 1}\gamma_n = \infty$ and $\sum_{n\geq 1}\gamma_n^2 < \infty$, which are two classical assumptions for obtaining the convergence of Robbins-Monro's algorithm (see [6]).

Assumption $(A2)$ on (m_n) was introduced similarly to the assumption on the bandwidth used for the recursive kernel regression estimator (see Mokkadem et al. [20, 21]), to ensure the application of the technical lemma given in the appendix A.

Assumption $(A3)$ on the density of the couple (X, Y) was used in the nonrecursive framework for the estimation of the regression function (see Nadaraya [23] and Watson [43]) and in the recursive framework (see Mokkadem et al. [21] and Slaoui [32, 33, 34]).

Remark 2.1. The intuition behind the use of such order (m_n) belonging to $\mathcal{GS}(a)$ is that the ratio m_{n-1}/m_n is equal to $1 - a/n + o(1/n)$, then using such order and using the assumption on the stepsize, which ensures that γ_{n-1}/γ_n is equal to $1 + \alpha/n + o(1/n)$. The application of the technical lemma given in the appendix A ensures that the bias and the variance will depend only on m_n and γ_n and not on $m_1, \dots, m_n$ and $\gamma_1, \dots, \gamma_n$, then the $MISE$ will depend also only on m_n and γ_n, which will be helpful to deduce an optimal order and an optimal stepsize.

Throughout this paper we will use the following notations :

$$\begin{aligned}
\Delta_1(x) &= \frac{1}{2}\left[(1-2x)f'(x)+x(1-x)f^{(2)}(x)\right], \quad \psi(x)=(4\pi x(1-x))^{-1/2},\\
\xi &= \lim_{n\to\infty}(n\gamma_n)^{-1},\\
N(x) &= r(x)\,f(x), \quad \Delta_2(x)=\frac{1}{2}\Big\{(1-2x)N(x)+x(1-x)N'(x)\Big\},\\
\Delta(x) &= \frac{1}{2}\left\{x(1-x)r^{(2)}(x)+\left[(1-2x)+2x(1-x)\frac{f'(x)}{f(x)}\right]r'(x)\right\},\\
C_1 &= \int_0^1 \Delta^2(x)dx, \quad C_2=\int_0^1 \frac{Var[Y|X=x]}{f(x)}\psi(x)dx,\\
K_1 &= \int_0^1 \left\{\frac{\Delta(x)f(x)}{f(x)-a\xi}\right\}^2 dx, \quad K_2=\int_0^1 \frac{2f(x)\psi(x)\mathrm{V}ar\,[Y|X=x]}{4f(x)-(2\alpha-a)\xi}dx.
\end{aligned}$$

Moreover, we denote by o_x the pointwise bound in x (i.e., the error is not uniform in $x\in[0,1]$).

3. Estimators based on the Bernstein polynomials

Let $(X,Y),(X_1,Y_1),\dots,(X_n,Y_n)$ be independent, identically distributed pairs of random variables with joint density function $g(x,y)$, and let f denote the probability density of X which is supported on $[0,1]$. We follow the approach of Vitale [42] and Leblanc [16, 17], used for distribution and density estimation, to define a Bernstein estimator of the regression $r: x\mapsto \mathbb{E}[Y|X=x]$ at a given point $x\in[0,1]$ such that $f(x)\neq 0$

$$\widehat{r}_n(x)=\frac{\sum_{i=1}^n Y_i\sum_{k=0}^{m_n-1}\mathbb{1}_{\left\{\frac{k}{m_n}<X_i\le\frac{k+1}{m_n}\right\}}b_k(m_n-1,x)}{\sum_{i=1}^n\sum_{k=0}^{m_n-1}\mathbb{1}_{\left\{\frac{k}{m_n}<X_i\le\frac{k+1}{m_n}\right\}}b_k(m_n-1,x)}, \tag{1}$$

where $b_k(m,x)=\binom{m}{k}x^k(1-x)^{m-k}$ is the Bernstein polynomial of order m. This estimator can be viewed as a generalization of the estimator proposed in Tenbusch [40], in which the order m_n is chosen to be equal to n.

The following proposition gives the bias, the variance and the MSE of $\widehat{r}_n(x)$, for $x\in[0,1]$ such that $f(x)>0$.

Proposition 3.1. *Let Assumptions (A2) and (A3) hold. For $x\in[0,1]$, such that $f(x)>0$, we have*

$$\mathbb{E}[\widehat{r}_n(x)]-r(x)=\Delta(x)m_n^{-1}+o\left(m_n^{-1}\right). \tag{2}$$

$$Var[\widehat{r}_n(x)]=\begin{cases}\frac{m_n^{1/2}}{n}\frac{Var[Y|X=x]}{f(x)}\psi(x)+o_x\left(\frac{m_n^{1/2}}{n}\right) & \textit{if } x\in(0,1),\\ \frac{m_n}{n}\frac{Var[Y|X=x]}{f(x)}+o_x\left(\frac{m_n}{n}\right) & \textit{if } x=0,1.\end{cases} \tag{3}$$

$MSE[\widehat{r}_n(x)]$

$$= \begin{cases} \Delta^2(x)m_n^{-2} + \frac{m_n^{1/2}}{n}\frac{Var[Y|X=x]}{f(x)}\psi(x) + o\left(m_n^{-2}\right) + o_x\left(\frac{m_n^{1/2}}{n}\right) & \textit{if } x \in (0,1), \\ \Delta^2(x)m_n^{-2} + \frac{m_n}{n}\frac{Var[Y|X=x]}{f(x)} + o\left(m_n^{-2}\right) + o_x\left(\frac{m_n}{n}\right) & \textit{if } x = 0,1. \end{cases}$$

To minimize the MSE of $\widehat{r}_n$, for $x \in [0,1]$ such that $f(x) > 0$, the order m must equal to

$$m_{opt} = \begin{cases} \left[\frac{4\Delta^2(x)f(x)}{Var[Y|X=x]\psi(x)}\right]^{2/5} n^{2/5} & \text{if } x \in (0,1), \\ \left[\frac{2\Delta^2(x)f(x)}{Var[Y|X=x]}\right]^{1/3} n^{1/3} & \text{if } x = 0,1, \end{cases}$$

then

$$MSE\left[\widehat{r}_{n,m_{opt}}(x)\right] = \begin{cases} \frac{5(\Delta(x))^{2/5}(Var[Y|X=x]\psi(x))^{4/5}}{(4f(x))^{4/5}} n^{-4/5} + o\left(n^{-4/5}\right) & \text{if } x \in (0,1), \\ \frac{3(\Delta(x)Var[Y|X=x])^{2/3}}{(2f(x))^{2/3}} n^{-2/3} + o\left(n^{-2/3}\right) & \text{if } x = 0,1. \end{cases}$$

Remark 3.1. Clearly, our first proposed estimator converge to the true regression function. The rate of convergence is bigger near of the edge ($x \in \{0,1\}$) than inside the interval, however, as it was shown, the non-parametric kernel estimation near of the edge fails (see for instance Jmaei et al. [13], Slaoui and Jmaei [36]).

The following proposition gives the $MISE$ of $\widehat{r}_n$

Proposition 3.2. *Let Assumptions* (*A*2) *and* (*A*3) *hold, we have*

$$MISE(\widehat{r}_n) = C_2\frac{m_n^{1/2}}{n} + C_1 m_n^{-2} + o\left(\frac{m_n^{1/2}}{n}\right) + o\left(m_n^{-2}\right). \tag{4}$$

Hence, the asymptotically optimal choice of m is

$$m_{opt} = \left[\frac{4C_1}{C_2}\right]^{2/5} n^{2/5},$$

for which we get

$$MISE\left(\widehat{r}_{n,m_{opt}}\right) = \frac{5C_1^{1/5}C_2^{4/5}}{4^{4/5}} n^{-4/5} + o\left(n^{-4/5}\right).$$

Let us now state the following theorem which gives the weak convergence rate of the estimator $\widehat{r}_n(x)$ defined in (1), for $x \in [0,1]$ such that $f(x) > 0$.

Theorem 3.1. *(Weak pointwise convergence rate). Let Assumptions* (*A*2) *and* (*A*3) *hold.*

When $x \in (0,1)$, and m_n is chosen such that $nm_n^{-5/2} \to c$ for some constant $c \geq 0$, we have

$$n^{1/2}m_n^{-1/4}\left(\widehat{r}_n(x) - r(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\sqrt{c}\Delta(x), \frac{Var\left[Y|X=x\right]\psi(x)}{f(x)}\right).$$

When $x \in \{0,1\}$, and m_n is chosen such that $nm_n^{-3} \to c$ for some constant $c \geq 0$, we have

$$\sqrt{\frac{n}{m_n}}\left(\widehat{r}_n(x) - r(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\sqrt{c}\Delta(x), \frac{Var\left[Y|X=x\right]}{f(x)}\right).$$

When $x \in (0,1)$ and m_n is chosen such that $nm_n^{-5/2} \to \infty$ or when $x \in \{0,1\}$ and m_n is chosen such that $nm_n^{-3} \to \infty$ we have

$$m_n\left(\widehat{r}_n(x) - r(x)\right) \xrightarrow{\mathbb{P}} \Delta(x),$$

where $\overset{\mathcal{D}}{\to}$ denotes the convergence in distribution, $\mathcal{N}$ the Gaussian-distribution and $\overset{\mathbb{P}}{\to}$ the convergence in probability. The next corollary is an immediate consequence of the previous Theorem on which we give the the weak convergence rate of the estimator $\widehat{r}_n(x)$, for $x \in [0,1]$ such that $f(x) > 0$ in the case when m_n is chosen such that $nm_n^{-5/2} \to 0$ for $x \in (0,1)$ and $nm_n^{-3} \to 0$ for $x \in \{0,1\}$.

Corollary 3.1. *Let Assumptions (A2) and (A3) hold.*

When $x \in (0,1)$, and m_n is chosen such that $nm_n^{-5/2} \to 0$, then

$$n^{1/2}m_n^{-1/4}\left(\widehat{r}_n(x) - r(x)\right) \overset{\mathcal{D}}{\to} \mathcal{N}\left(0, \frac{\mathrm{Var}\left[Y|X=x\right]\psi(x)}{f(x)}\right).$$

When $x \in \{0,1\}$, and m_n is chosen such that $nm_n^{-3} \to 0$, then

$$\sqrt{\frac{n}{m_n}}\left(\widehat{r}_n(x) - r(x)\right) \overset{\mathcal{D}}{\to} \mathcal{N}\left(0, \frac{\mathrm{Var}\left[Y|X=x\right]}{f(x)}\right).$$

4. Recursive estimator

In order to construct a stochastic algorithm for the estimation of the regression function $r : x \mapsto \mathbb{E}\left[Y|X=x\right]$ at a point x such as $f(x) \neq 0$, Révész [26] defines an algorithm, which approximates the zero of the function $h : y \mapsto f(x)r(x) - f(x)y$. Using the procedure proposed in Robbins and Monro [28], the considered algorithm is defined by setting $r_0(x) \in \mathbb{R}$ and for $n \geq 1$

$$r_n(x) = r_{n-1}(x) + \gamma_n \mathcal{W}_n(x),$$

where γ_n is the stepsize and $\mathcal{W}_n$ is an observation of the function h at the point $r_{n-1}(x)$. We define $\mathcal{W}_n$, using Bernstein polynomials

$$\begin{aligned}\mathcal{W}_n(x) = m_n Y_n \sum_{k=0}^{m_n-1} \mathbb{1}_{\left\{\frac{k}{m_n} < X_n \leq \frac{k+1}{m_n}\right\}} b_k(m_n-1,x)\\ - m_n \sum_{k=0}^{m_n-1} \mathbb{1}_{\left\{\frac{k}{m_n} < X_n \leq \frac{k+1}{m_n}\right\}} b_k(m_n-1,x) r_{n-1}(x),\end{aligned}$$

then, the estimator r_n can be rewritten as

$$\begin{aligned}r_n(x) &= \left(1 - \gamma_n m_n \sum_{k=0}^{m_n-1} \mathbb{1}_{\left\{\frac{k}{m_n} < X_n \leq \frac{k+1}{m_n}\right\}} b_k(m_n-1,x)\right) r_{n-1}(x)\\ &\quad + \gamma_n m_n Y_n \sum_{k=0}^{m_n-1} \mathbb{1}_{\left\{\frac{k}{m_n} < X_n \leq \frac{k+1}{m_n}\right\}} b_k(m_n-1,x),\\ &= \left(1 - \gamma_n f(x)\right) r_{n-1}(x) + \gamma_n \left(f(x) - m_n \sum_{k=0}^{m_n-1} \mathbb{1}_{\left\{\frac{k}{m_n} < X_n \leq \frac{k+1}{m_n}\right\}} b_k(m_n-1,x)\right) r_{n-1}(x)\\ &\quad + \gamma_n m_n Y_n \sum_{k=0}^{m_n-1} \mathbb{1}_{\left\{\frac{k}{m_n} < X_n \leq \frac{k+1}{m_n}\right\}} b_k(m_n-1,x).\end{aligned}$$

We set

$$\begin{aligned}Z_n(x) &= m_n \sum_{k=0}^{m_n-1} \mathbb{1}_{\left\{\frac{k}{m_n} < X_n \leq \frac{k+1}{m_n}\right\}} b_k(m_n-1,x),\\ W_n(x) &= m_n Y_n \sum_{k=0}^{m_n-1} \mathbb{1}_{\left\{\frac{k}{m_n} < X_n \leq \frac{k+1}{m_n}\right\}} b_k(m_n-1,x).\end{aligned}$$

Then, the proposed algorithm can be rewritten as follows:

(5) $$r_n(x) = (1-\gamma_n f(x))\, r_{n-1}(x) + \gamma_n \left(f(x) - Z_n(x)\right) r_{n-1}(x) + \gamma_n W_n(x).$$

In order, to establish the asymptotic behaviour of r_n, we introduce the auxiliary stochastic approximation algorithm defined by setting $\rho_n(x) = r(x)$ for all $n \leq n_0 - 2$, $\rho_{n_0-1}(x) = r_{n_0-1}(x)$, and, for $n \geq n_0$,

(6) $$\rho_n(x) = (1-\gamma_n f(x))\, \rho_{n-1}(x) + \gamma_n \left(f(x) - Z_n(x)\right) r(x) + \gamma_n W_n(x).$$

We first give the behaviour of ρ_n. Then, we show how the behaviour of r_n can be deduced from that of ρ_n.

4.1. **Within the interval** $[0,1]$**.** To obtain the bias, the variance and the MSE of $r_n(x)$, for $x \in (0,1)$ such that $f(x) > 0$, we set

: $(A4)$ $\lim_{n\to\infty}(n\gamma_n) \in \left(\min\left(\frac{a}{f(x)}, \frac{2\alpha - a}{4f(x)}\right), \infty\right]$.

Proposition 4.1. *Let Assumptions* $(A1)-(A4)$ *hold. For* $x \in (0,1)$*, such that* $f(x) > 0$*, we have*

$$\begin{aligned}\mathbb{E}[r_n(x)] - r(x) &= \frac{f(x)\Delta(x)}{f(x) - a\xi}\mathbb{1}_{\left\{a\in\left(\frac{1-\alpha}{4}, \frac{2}{5}\alpha\right]\right\}} m_n^{-1} + \mathbb{1}_{\left\{a\in\left(\frac{2}{5}\alpha, \frac{2}{3}\right)\right\}} o\left(\sqrt{\gamma_n m_n^{1/2}}\right)\\ &\quad + o\left(m_n^{-1} + \sqrt{\gamma_n m_n^{1/2}}\right),\end{aligned} \tag{7}$$

$$\begin{aligned}Var[r_n(x)] &= \frac{2f(x)\psi(x)\mathrm{Var}\left[Y|X=x\right]}{4f(x) - (2\alpha - a)\xi}\mathbb{1}_{\left\{a\in\left[\frac{2}{5}\alpha, \frac{2}{3}\right)\right\}}\gamma_n m_n^{1/2} + \mathbb{1}_{\left\{a\in\left(\frac{1-\alpha}{4}, \frac{2}{5}\alpha\right)\right\}} o\left(m_n^{-2}\right)\\ &\quad + o\left(\gamma_n m_n^{1/2} + m_n^{-2}\right),\end{aligned} \tag{8}$$

and

$$\begin{aligned}MSE\left[r_n(x)\right] &= \frac{f^2(x)\Delta^2(x)}{(f(x) - a\xi)^2}\mathbb{1}_{\left\{a\in\left(\frac{1-\alpha}{4}, \frac{2}{5}\alpha\right)\right\}} m_n^{-2}\\ &\quad + \frac{2f(x)\mathrm{Var}\left[Y|X=x\right]\psi(x)}{4f(x) - (2\alpha - a)\xi}\mathbb{1}_{\left\{a\in\left(\frac{2}{5}\alpha, \frac{2}{3}\right)\right\}}\gamma_n m_n^{1/2} + o\left(m_n^{-2} + \gamma_n m_n^{1/2}\right).\end{aligned}$$

Remark 4.1. When $\lim_{n\to\infty}(n\gamma_n) > \max\left(\frac{a}{f(x)}, \frac{2\alpha-a}{4f(x)}\right)$, the equations (7) and (8) hold simultaneously.

To minimize the MSE of $r_n(x)$, for $x \in (0,1)$ such that $f(x) > 0$, the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$ and (m_n) must must be in $\mathcal{GS}\left(2/5\right)$ such that

$$\left(4^{3/5}\left(f(x) - \frac{2}{5}\xi\right)^{-2/5}\left[\frac{f(x)\Delta^2(x)}{\mathrm{Var}\left[Y|X=x\right]\psi(x)}\right]^{2/5}\gamma_n^{-2/5}\right),$$

then

$$MSE\left[r_n(x)\right] = \frac{5(f(x))^{6/5}(\Delta(x))^{2/5}(Var\left[Y|X=x\right]\psi(x))^{4/5}}{4^{6/5}\left(f(x) - \frac{2}{5}\xi\right)^{6/5}}\gamma_n^{4/5} + o\left(\gamma_n^{4/5}\right).$$

Let us now state the following theorem, which gives the weak convergence rate of the estimator $r_n(x)$ defined in (5), for $x \in (0,1)$ such that $f(x) > 0$.

Theorem 4.1. *(Weak pointwise convergence rate). Let Assumptions* $(A1)$-$(A4)$ *hold, we have*

(1) *If* $\gamma_n^{-1} m_n^{-5/2} \to c$ *for some constant* $c \geq 0$*, then*

$$\gamma_n^{-1/2} m_n^{-1/4}\left(r_n(x) - r(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\sqrt{c}\frac{f(x)\Delta(x)}{f(x) - a\xi}, \frac{2f(x)\mathrm{Var}\left[Y|X=x\right]\psi(x)}{4f(x) - (2\alpha - a)\xi}\right).$$

(2) *If* $\gamma_n^{-1}m_n^{-5/2} \to \infty$, *then*

$$m_n\left(r_n\left(x\right)-r\left(x\right)\right)\stackrel{\mathbb{P}}{\to}\frac{f(x)\Delta(x)}{f(x)-a\xi},$$

where $\stackrel{\mathcal{D}}{\to}$ *denotes the convergence in distribution,* $\mathcal{N}$ *the Gaussian-distribution and* $\stackrel{\mathbb{P}}{\to}$ *the convergence in probability.*

The next corollary is an immediate consequence of the previous Theorem on which we give the the weak convergence rate of the estimator $r_n(x)$, for $x \in (0,1)$ in the case when m_n is chosen such that $nm_n^{-5/2} \to 0$ for $x \in (0,1)$.

Corollary 4.1. *Let Assumptions* (*A*1)-(*A*4) *hold, when* $x \in (0,1)$, *and* m_n *is chosen such that* $nm_n^{-5/2} \to 0$, *then*

$$\gamma_n^{-1/2}m_n^{-1/4}\left(r_n\left(x\right)-r\left(x\right)\right)\stackrel{\mathcal{D}}{\to}\mathcal{N}\left(0,\frac{2f(x)\mathrm{Var}\left[Y|X=x\right]\psi(x)}{4f(x)-(2\alpha-a)\xi}\right).$$

4.2. **The edges of the interval** $[0,1]$**.** For the case $x \in \{0,1\}$, such that $f(x) > 0$, we need to consider the following additional Assumption

: $(A'4)$ $\quad \lim_{n\to\infty}(n\gamma_n) \in \left(\min\left(\frac{a}{f(x)},\frac{\alpha-a}{2f(x)}\right),\infty\right]$.

The following proposition gives the bias, the variance and the MSE of $r_n(x)$, for $x \in \{0,1\}$.

Proposition 4.2. *Let Assumptions* $(A1)-(A'4)$ *hold. For* $x \in \{0,1\}$, *such that* $f(x) > 0$, *we have*

$$\begin{aligned}\mathbb{E}[r_n(x)]-r(x) &= \frac{f(x)\Delta(x)}{f(x)-a\xi}\mathbb{1}_{\left\{a\in\left(\frac{1-\alpha}{4},\frac{\alpha}{3}\right]\right\}}m_n^{-1}+\mathbb{1}_{\left\{a\in\left(\frac{\alpha}{3},\frac{2}{3}\alpha\right)\right\}}o\left(\sqrt{\gamma_n m_n}\right)\\ &\quad +o\left(m_n^{-1}\right),\end{aligned}\tag{9}$$

$$\begin{aligned}Var[r_n(x)] &= \frac{f(x)\mathrm{Var}\left[Y|X=x\right]}{2f(x)-(\alpha-a)\xi}\mathbb{1}_{\left\{a\in\left[\frac{\alpha}{3},\frac{2}{3}\alpha\right)\right\}}\gamma_n m_n+\mathbb{1}_{\left\{a\in\left(\frac{1-\alpha}{4},\frac{\alpha}{3}\right]\right\}}o\left(m_n^{-2}\right)\\ &\quad +o\left(\gamma_n m_n\right),\end{aligned}\tag{10}$$

and

$$\begin{aligned}MSE\left[r_n(x)\right] &= \frac{f^2(x)\Delta^2(x)}{(f(x)-a\xi)^2}\mathbb{1}_{\left\{a\in\left(\frac{1-\alpha}{4},\frac{\alpha}{3}\right]\right\}}m_n^{-2}\\ &\quad +\frac{f(x)\mathrm{Var}\left[Y|X=x\right]}{2f(x)-(\alpha-a)\xi}\mathbb{1}_{\left\{a\in\left(\frac{1-\alpha}{4},\frac{\alpha}{3}\right]\right\}}\gamma_n m_n+o\left(m_n^{-2}+\gamma_n m_n\right).\end{aligned}$$

Remark 4.2. (1) When $\lim_{n\to\infty}(n\gamma_n) > \max\left(\frac{a}{f(x)},\frac{\alpha-a}{2f(x)}\right)$, the equations (9) and (10) hold simultaneously.

(2) To minimize the MSE of r_n, for $x \in \{0,1\}$ such that $f(x) > 0$, the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$ and (m_n) must be in $\mathcal{GS}\,(1/3)$ such that

$$\left(2^{2/3}\left(f(x)-\frac{1}{3}\xi\right)^{-1/3}\left[\frac{f(x)\Delta^2(x)}{\mathrm{Var}\left[Y|X=x\right]}\right]^{1/3}\gamma_n^{-1/3}\right),$$

then

$$MSE\left[r_n(x)\right]=\frac{3(f(x))^{4/3}(\Delta(x)\mathrm{Var}\left[Y|X=x\right])^{2/3}}{2^{4/3}\left(f(x)-\frac{1}{3}\xi\right)^{4/3}}\gamma_n^{2/3}+o\left(\gamma_n^{2/3}\right).$$

Let us now state the following theorem, which gives the weak convergence rate of the estimator $r_n(x)$ defined in (5), for $x \in \{0,1\}$ such that $f(x) > 0$.

Theorem 4.2. *(Weak pointwise convergence rate). Let Assumption* $(A1)$*-*$(A'4)$ *hold, we have*

(1) *If* $\gamma_n^{-1}m_n^{-3} \to c$ *for some constant* $c \geq 0$*, then*

$$\gamma_n^{-1/2}m_n^{-1/2}\left(r_n\left(x\right)-r\left(x\right)\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(\sqrt{c}\frac{f(x)\Delta(x)}{f(x)-a\xi}, \frac{f(x)\mathrm{Var}\left[Y|X=x\right]\psi(x)}{2f(x)-(\alpha-a)\xi}\right),$$

(2) *If* $\gamma_n^{-1}m_n^{-3} \to \infty$*, then*

$$m_n\left(r_n\left(x\right)-r\left(x\right)\right) \stackrel{\mathbb{P}}{\to} \frac{f(x)\Delta(x)}{f(x)-a\xi},$$

where $\stackrel{\mathcal{D}}{\to}$ *denotes the convergence in distribution,* $\mathcal{N}$ *the Gaussian-distribution and* $\stackrel{\mathbb{P}}{\to}$ *the convergence in probability.*

The next corollary is an immediate consequence of the previous Theorem on which we give the the weak convergence rate of the estimator $r_n(x)$, for $x \in \{0,1\}$ in the case when m_n is chosen such that $nm_n^{-5/2} \to 0$ for $x \in \{0,1\}$.

Corollary 4.2. *Let Assumption* $(A1)$*-*$(A'4)$ *hold, when* $x \in \{0,1\}$*, and* m_n *is chosen such that* $nm_n^{-3} \to 0$*, then*

$$\gamma_n^{-1/2}m_n^{-1/2}\left(r_n\left(x\right)-r\left(x\right)\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(0, \frac{f(x)\mathrm{Var}\left[Y|X=x\right]\psi(x)}{2f(x)-(\alpha-a)\xi}\right).$$

4.3. **The** $MISE$ **of** r_n**.** To obtain the $MISE$ of r_n, we add the following assumption

: $(A''4)$ Set $\varphi = \inf_{x\in[0,1]} f(x) > 0$, we demand that

$$\lim_{n\to\infty}\left(n\gamma_n\right) \in \left(\min\left(\frac{a}{\varphi}, \frac{2\alpha-a}{4\varphi}\right), \infty\right].$$

Proposition 4.3. *Let Assumptions* $(A1)-(A''4)$ *hold, we have*

$$MISE\left(r_n\right) = K_1\mathbb{1}_{\left\{a\in\left(\frac{1-\alpha}{4},\frac{2}{5}\alpha\right]\right\}}m_n^{-2} + K_2\mathbb{1}_{\left\{a\in\left[\frac{2}{5}\alpha,\frac{2}{3}\alpha\right\}\right.}\gamma_n m_n^{1/2} + o\left(m_n^{-2}+\gamma_n m_n^{1/2}\right)$$

The following result is a consequence of the previous proposition which gives the optimal order (m_n) of the estimator r_n introduced in (5) and the corresponding $MISE$.

Corollary 4.3. *Let Assumptions* $(A1)-(A''4)$ *hold. To minimize the* $MISE$ *of* r_n*, the stepsize* (γ_n) *must be chosen in* $\mathcal{GS}(-1)$ *and* (m_n) *must be in* $\mathcal{GS}(2/5)$ *such that*

$$\left(\frac{4K_1}{K_2}\right)^{2/5}\gamma_n^{-2/5},$$

and then

$$MISE\left(r_n\right) = \frac{5K_1^{1/5}K_2^{4/5}}{4^{4/5}}\gamma_n^{4/5} + o\left(\gamma_n^{4/5}\right).$$

Remark 4.3. We can claim that our two proposed estimators converge to the true regression function. It is true that the rate of convergence is bigger near of the edge ($x \in \{0,1\}$) than inside the interval, however, as it was shown previously (see for instance Jmaei et al. [13], Slaoui and Jmaei [36]) the non-parametric kernel estimation near of the edge fails.

5. Applications

We recall the regression function's kernel estimator proposed by Nadaraya [23] and Watson [43], for $x \in \mathbb{R}$ such that $f(x) \neq 0$

$$\widehat{r}_n^{NW}(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)}, \tag{11}$$

where $K : \mathbb{R} \to \mathbb{R}$ is a nonnegative, continuous, bounded function satisfying $\int_{\mathbb{R}} K(z)dz = 1$, $\int_{\mathbb{R}} zK(z)dz = 0$ and $\int_{\mathbb{R}} z^2K(z)dz < \infty$ known as kernel and $h = (h_n)$ is a bandwidth (that is, a sequence of positive real numbers that goes to zero). We also recall the recursive estimator of a regression function which is a generalized version of Révész's estimator (see Révész [26, 27]) and was studied by Mokkadem et al. [21]

$$r_n^{GR}(x) = \left(1 - \gamma_n h_n^{-1} K\left(\frac{x - X_n}{h_n}\right)\right) r_{n-1}^{GR}(x) + \gamma_n h_n^{-1} Y_n K\left(\frac{x - X_n}{h_n}\right). \tag{12}$$

The major limitation of these two estimators occurs at the edges of the support. In fact, these estimators are inconsistent at the boundary. This effectively restricts their application to values of x in the interior of the support of the estimated regression function.

The purpose of this section is to provide a comparative study between Nadaraya-Watson's estimator $\widehat{r}_n^{NW}$ defined in (11), the generalized Révész's estimator r_n^{GR} defined in (12), our non-recursive estimator $\widehat{r}_n$ defined in (1) and our recursive estimator r_n introduced in (5).

5.1. **Simulations.** We consider the regression model

$$Y = r(X) + \varepsilon,$$

where $\varepsilon \sim \mathcal{N}(0, 1)$.

When using the estimators $\widehat{r}_n^{NW}$ and r_n^{GR}, we choose the kernel

$$K(x) = (2\pi)^{-1/2} \exp\left(-x^2/2\right)$$

and the bandwidth equal to $(h_n) = n^{-1/5}(\ln(n + 1))^{-1}$. When using our proposed Bernstein estimators $\widehat{r}_n$ and r_n, we choose the order equal to $m_n = \lfloor n^{2/5}(\ln(n + 1))\rfloor$ and we choose two stepsize $(\gamma_n) = (n^{-0.9})$ and $(\gamma_n) = (n^{-1})$.

We consider three sample sizes $n = 50$, $n = 100$ and $n = 500$, three regression functions

- (a) $r(x) = \cos(x)$,
- (b) $r(x) = 0.3 \exp\left(-x^2/2\right) + 0.7 \exp\left(-(x - 1)^2/2\right)$,
- (c) $r(x) = 1 + 0.6x$,

and three densities of X, the beta density $\mathcal{B}(3, 5)$, the beta mixture density $0.5\mathcal{B}(2, 1) + 0.5\mathcal{B}(1, 4)$ and the truncated standard normal density $\mathcal{N}_{[0,1]}(0, 1)$. We consider six estimators: our non-recursive Bernstein estimator $\widehat{r}_n$ defined in (1), Nadaraya-Watson's estimator $\widehat{r}_n^{NW}$ proposed in (11), two proposed recursive Bernstein estimators $r_{n,1}$ and $r_{n,2}$ introduced in (5) with stepsize $(\gamma_n) = (n^{-1})$ and $(\gamma_n) = (n^{-0.9})$ respectively and finally two Generalized Révész's estimators $r_{n,1}^{GR}$ and $r_{n,2}^{GR}$ defined in (12) using the same stepsizes as the one previously used in r_n. For each model and sample of size n, we approximate the average integrated squared error (ISE) of the estimator using $N = 500$ trials of sample size n

$$\overline{ISE} = \frac{1}{N} \sum_{k=1}^{N} ISE\left[\bar{r}_k\right],$$

Regression function	Density of X	n	$\widehat{r}_n$	$\widehat{r}_n^{NW}$	$r_{n,1}$	$r_{n,2}$	$r_{n,1}^{GR}$	$r_{n,2}^{GR}$
(a)	$\mathcal{B}(3,5)$	50	0.055	0.055	0.104	0.115	**0.038**	0.047
		200	0.032	0.034	0.079	0.081	**0.022**	0.028
		500	0.026	0.028	0.066	0.069	**0.016**	0.019
	$0.5\mathcal{B}(2,1)+0.5\mathcal{B}(1,4)$	50	0.068	**0.067**	0.125	0.134	0.108	0.106
		200	**0.038**	0.042	0.106	0.104	0.090	0.085
		500	**0.030**	0.033	0.099	0.101	0.084	0.075
	$\mathcal{N}_{[0,1]}(0,1)$	50	0.064	0.063	0.064	0.081	**0.041**	0.048
		200	0.038	0.041	0.025	0.033	**0.017**	0.024
		500	0.029	0.031	0.012	0.018	**0.009**	0.014
(b)	$\mathcal{B}(3,5)$	50	0.036	**0.036**	0.092	0.095	0.050	0.052
		200	**0.014**	0.016	0.064	0.063	0.032	0.032
		500	**0.009**	0.011	0.057	0.057	0.027	0.028
	$0.5\mathcal{B}(2,1)+0.5\mathcal{B}(1,4)$	50	**0.030**	0.030	0.085	0.092	0.042	0.048
		200	**0.012**	0.013	0.061	0.063	0.029	0.029
		500	**0.007**	0.008	0.057	0.057	0.024	0.024
	$\mathcal{N}_{[0,1]}(0,1)$	50	0.051	0.050	0.060	0.078	**0.038**	0.048
		200	0.020	0.023	0.023	0.033	**0.017**	0.023
		500	0.013	0.015	0.012	0.018	**0.010**	0.014
(c)	$\mathcal{B}(3,5)$	50	0.067	**0.067**	0.113	0.107	0.120	0.105
		200	**0.041**	0.043	0.078	0.079	0.086	0.076
		500	**0.036**	0.037	0.074	0.070	0.079	0.064
	$0.5\mathcal{B}(2,1)+0.5\mathcal{B}(1,4)$	50	0.055	**0.054**	0.098	0.105	0.094	0.086
		200	**0.039**	0.039	0.079	0.072	0.068	0.070
		500	**0.033**	0.034	0.066	0.063	0.065	0.052
	$\mathcal{N}_{[0,1]}(0,1)$	50	0.082	0.082	0.065	0.085	**0.055**	0.059
		200	0.051	0.053	0.024	0.034	**0.021**	0.025
		500	0.039	0.042	0.012	0.017	**0.010**	0.014

TABLE 1. The average integrated squared error (ISE) of our non-recursive estimator $\widehat{r}_n$, Nadaraya-Watson's estimator $\widehat{r}_n^{NW}$, $r_{n,1}$ and $r_{n,2}$ correspond to our recursive estimator with the choice $(\gamma_n) = (n^{-1})$ and $(\gamma_n) = (n^{-0.9})$ respectively, and $r_{n,1}^{GR}$ and $r_{n,2}^{GR}$ correspond to the generalized Révész's estimator with the choice $(\gamma_n) = (n^{-1})$ and $(\gamma_n) = (n^{-0.9})$ respectively.

where $\bar{r}_k$ is the estimator computed from the k^{th} sample, and

$$ISE\left[\bar{r}_k\right] = \int_0^1 \left\{\bar{r}(x) - r(x)\right\}^2 dx.$$

In Table 1 we give qualitative comparison between our non-recursive estimator $\widehat{r}_n$ defined in (1) and our recursive estimator r_n given in (5) with $(\gamma_n) = (n^{-1})$. We then conclude that:

- In all the considered models, the average ISE of our non-recursive regression estimator $\widehat{r}_n$ defined in (1) is the smallest, except the cases with $X \sim \mathcal{B}(3,5)$ where the the average ISE of Nadaraya-Watson's estimator $\widehat{r}_n^{GR}$ given in (11) is the smallest when the size is $n = 50$ and the cases with $X \sim \mathcal{N}_{[0,1]}(0,1)$ where the average ISE of the generalized Révész's estimator estimator is the smallest.
- In all the models with $X \sim \mathcal{N}_{[0,1]}(0,1)$, the average ISE of our recursive regression estimator r_n defined in (5) with the choice $(\gamma_n) = (n^{-1})$ is smaller than that of our non-recursive regression estimator $\widehat{r}_n$ introduced in (1).
- In all the models, the average ISE of our recursive regression estimator $r_{n,1}$ with the choice $(\gamma_n) = (n^{-1})$ is smaller than that of our recursive regression estimator $r_{n,2}$ with the choice $(\gamma_n) = (n^{-0.9})$.
- The average ISE decreases as the sample size increases.

Figure 1, shows regression estimates plotted for 500 simulated samples from the model $r(x) = \cos(x)$ with $X \sim \mathcal{N}_{[0,1]}(0,1)$ of sizes $n = 100$ (left panel) and $n = 500$ (right panel). From Figure 1, we conclude that:

- Both our estimators are close to the true regression function.

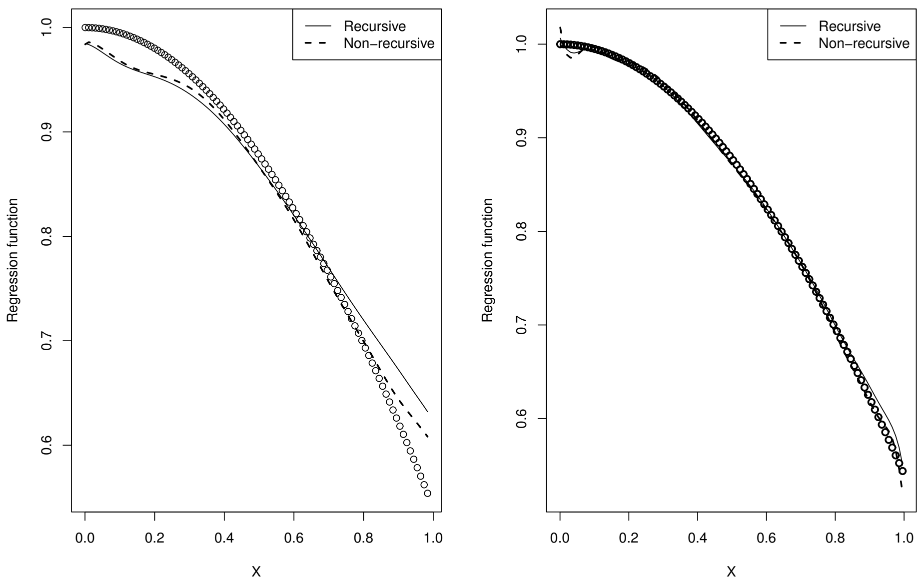

FIGURE 1. Qualitative comparison between the two proposed regression estimators r_n given in (5) with stepsize $(\gamma_n) = (n^{-1})$ (solid line) and $\widehat{r}_n$ given in (1) (dashed line), the true regression function (circle line) for 500 samples respectively of size 100 (left panel) and of size 500 (right panel) of the model $r(x) = \cos(x)$ with $X \sim \mathcal{N}_{[0,1]}(0,1)$.

- Our recursive regression estimator r_n defined in (5) using the stepsize $(\gamma_n) = (n^{-1})$ is closer to the true regression function than that of the proposed non-recursive estimator $\widehat{r}_n$ given in (1) especially with the size $n = 500$.
- When the sample size increases, we get closer estimation of the true regression function.

5.2. **Real dataset.** In order to illustrate our two proposed estimators by using a set of real data. We consider the `CO2` dataset which is available in the `R` package `Stat2Data` and contained 237 observations on two variables; Day and CO2. Scientists at a research station in Brotjacklriegel, Germany recorded CO2 levels, in parts per million, in the atmosphere for each day from the start of April through November in 2001. First, to estimate an unknown regression function, it is critical to have a reliable data-dependent rule for order selection. One popular and practical approach is cross-validation. First, we compute the leave-one-out residuals:

$$\forall i \in \{1,\ldots,n\}, \quad e_{-i} = Y_i - \bar{r}_{-i}(X_i),$$

where $\bar{r}_{-i}$ is the regression estimate without the data point (X_i, Y_i). Then, the smoothing parameter is chosen by minimizing

$$CV(m_n) = \frac{1}{n}\sum_{i=1}^{n} e_{-i}^2.$$

We then apply our proposed estimators $\widehat{r}_n$ defined in (1) and r_n given in (5) on this model. For convenience, we assume that the minimum of days is 90 and the maximum is 335 (the Day data are such that $\min_i(x_i) = 91$ and $\max_i(x_i) = 334$). Finally, we used the Cross-validation method to obtain $m_n = 220$ for our non-recursive estimator $\widehat{r}_n$ and

$m_n = n$ for our recursive estimator. We observe from Figure 2 that the two proposed estimators give better estimation compared to Nadaraya-Watson's (11) especially near the boundaries, which corroborate remarks 3.1 and 4.3f.

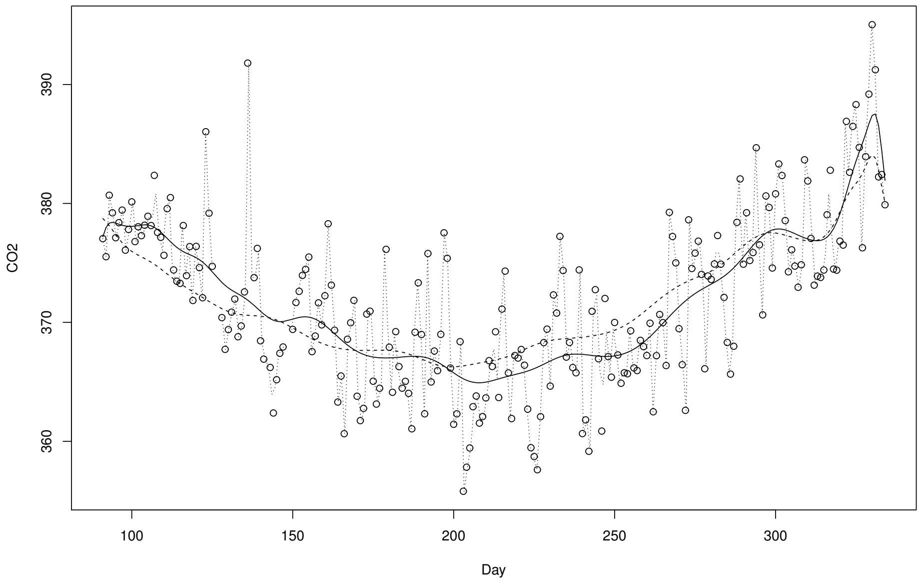

FIGURE 2. The daily carbon dioxide measurements data using Nadaraya-Watson's estimator (11) (dotted line) and our proposed Bernstein estimators $\widehat{r}_n$ defined in (1) (dashed line) and r_n given in (5) with stepsize $(\gamma_n) = (n^{-1})$ (solid line).

6. Conclusion

In this paper, we propose a non-recursive and recursive estimator of regression function based on Bernstein polynomials and stochastic algorithm derived from the Robbins-Monro's scheme. We first study their theoretical behavior. Then, we conduct a simulation study and analyse a real data application on CO2 data. For all the models, the average ISE of our non-recursive regression estimator $\widehat{r}_n$ defined in (1) is the smallest, except the cases with $X \sim \mathcal{B}(3,5)$ where the average ISE of Nadaraya-Watson's estimator $\widehat{r}_n^{GR}$ given in (11) is smaller in the case when the sample size is $n = 50$ and the cases with $X \sim \mathcal{N}_{[0,1]}(0,1)$ where the average ISE of the generalized Révész's estimator give better results in terms of average ISE. In addition, a major advantage of our recursive estimator is that its update, when new sample points are available, requires less computational cost than Nadaraya-Watson estimator. Finally, the two estimators have nice features and satisfactory improvement in comparison to Kernel estimators especially near the boundaries.

In conclusion, the estimation using Bernstein polynomials allowed us to overcome the edge problem and obtain quite similar results as Nadaraya-Watson's estimator. Moreover, we plan to make extensions of our method in the future and to consider the functional data (see, Slaoui [37, 38]) to built a semi-recursive Bernstein estimator for regression function. We plan also to extend our estimators by considering the recursive nonparametric estimation for Bayesian networks (see for instance the recent paper in this subject Boukabour and Masmoudi [4]).

Appendix A. Outlines of the proofs

In this section, we present proofs for the results given in the paper. First, we recall a series of results, which are proven in Leblanc [16], linked to different sums of Bernstein polynomial, defined by

$$S_{m_n}(x) = \sum_{k=0}^{m_n} b_k^2(m_n, x),$$

These results are given in the following lemma.

Lemma A.1. *We have*

- : *(i)* $0 \leq S_{m_n}(x) \leq 1$, $\forall x \in [0,1]$,
- : *(ii)* $S_{m_n}(x) = m_n^{-1/2}\left[\psi(x) + o_x(1)\right]$, $\forall x \in (0,1)$
- : *(iii)* $S_{m_n}(0) = S_{m_n}(1) = 1$.

  Let g *be any continuous function on* $[0,1]$. *Then*
- : *(iv)* $m_n^{1/2}\int_0^1 g(x)S_{m_n}(x)dx = \int_0^1 g(x)\psi(x)dx + o(1)$,

We start by proving the characteristics of our non-recursive estimator $\widehat{r}_n$ defined by (1). To do so, we note

$$N_n(x) = \frac{m_n}{n}\sum_{i=1}^n Y_i \sum_{k=0}^{m_n-1} \mathbb{1}_{\left\{\frac{k}{m_n} < X_i \leq \frac{k+1}{m_n}\right\}} b_k(m_n - 1, x).$$

Then, we may rewrite $\widehat{r}_n$ as

$$\widehat{r}_n(x) = \frac{N_n(x)}{f_n(x)},$$

where f_n is the Vitale's estimator of the density f defined, for all $x \in [0,1]$, by

$$\begin{aligned}
f_n(x) &= \frac{m_n}{n}\sum_{i=1}^n \sum_{k=0}^{m_n-1} \mathbb{1}_{\left\{\frac{k}{m_n} < X_i \leq \frac{k+1}{m_n}\right\}} b_k(m_n - 1, x) \\
&= m\sum_{k=0}^{m_n-1}\left\{F_n\left(\frac{k+1}{m_n}\right) - F_n\left(\frac{k}{m_n}\right)\right\} b_k(m_n - 1, x),
\end{aligned}$$

with F_n is the empirical distribution function of the variable X.

A.1. **Prooof of Proposition 3.1.** We start by giving the bias and the variance of $N_n(x)$

$$\begin{aligned}
\mathbb{E}\left[N_n(x)\right] &= m_n\mathbb{E}\left[Y\sum_{k=0}^{m_n-1} \mathbb{1}_{\left\{\frac{k}{m_n} < X \leq \frac{k+1}{m_n}\right\}} b_k(m_n - 1, x)\right], \\
&= m_n\sum_{k=0}^{m_n-1}\int_{\frac{k}{m_n}}^{\frac{k+1}{m_n}}\left(\int_{\mathbb{R}} yg(z,y)dy\right)dz b_k(m_n - 1, x), \\
&= m_n\sum_{k=0}^{m_n-1}\left(\int_{\frac{k}{m_n}}^{\frac{k+1}{m_n}} r(z)f(z)dz\right) b_k(m_n - 1, x).
\end{aligned}$$

Using Taylor expansion, we have

$$\begin{aligned}
r(z)f(z) &= \left[r(x) + (z-x)r'(x) + \frac{(z-x)^2}{2}r^{(2)}(x) + o\left((z-x)^2\right)\right] \\
&\quad \times\left[f(x) + (z-x)f'(x) + \frac{(z-x)^2}{2}f^{(2)}(x) + o\left((z-x)^2\right)\right], \\
&= r(x)f(x) + (z-x)\left(r'(x)f(x) + r(x)f'(x)\right)
\end{aligned}$$

$$+\frac{(z-x)^2}{2}\left(r^{(2)}(x)f(x)+f^{(2)}(x)r(x)+2r'(x)f'(x)\right)+o\left((z-x)^2\right).$$

Since $N(x)=r(x)f(x)$, then we obtain

$$\begin{aligned}
\mathbb{E}\left[N_n(x)\right] &= r(x)f(x)m_n\sum_{k=0}^{m_n-1}\left(\frac{k+1}{m_n}-\frac{k}{m_n}\right)b_k(m_n-1,x)\\
&+\left(r'(x)f(x)+r(x)f'(x)\right)\frac{m_n}{2}\sum_{k=0}^{m_n-1}\left\{\left(\frac{k+1}{m_n}-x\right)^2-\left(\frac{k}{m_n}-x\right)^2\right\}\\
&\times b_k(m_n-1,x)\\
&+\left(r^{(2)}(x)f(x)+f^{(2)}(x)r(x)+2r'(x)f'(x)\right)\frac{m_n}{6}\sum_{k=0}^{m_n-1}\left\{\left(\frac{k+1}{m_n}-x\right)^3-\left(\frac{k}{m_n}-x\right)^3\right\}\\
&\times b_k(m_n-1,x)\\
&+o\left(m\sum_{k=0}^{m_n-1}\left\{\left(\frac{k+1}{m_n}-x\right)^3-\left(\frac{k}{m_n}-x\right)^3\right\}b_k(m_n-1,x)\right)\\
&= N(x)+\left(r'(x)f(x)+r(x)f'(x)\right)\frac{m_n}{2}\sum_{k=0}^{m_n-1}m_n^{-2}(2k+1-2m_nx)b_k(m_n-1,x)\\
&+\left(r^{(2)}(x)f(x)+f^{(2)}(x)r(x)+2r'(x)f'(x)\right)\frac{m_n}{6}\sum_{k=0}^{m_n-1}m_n^{-3}\Big\{(k+1-m_nx)^2+(k-m_nx)^2\\
&+(k+1-m_nx)(k-m_nx)\Big\}b_k(m_n-1,x)[1+o(1)]\\
&= N(x)+\left(r'(x)f(x)+r(x)f'(x)\right)\frac{m_n^{-1}}{2}\left\{2T_{1,m_n-1}(x)+(1-2x)T_{0,m_n-1}(x)\right\}\\
&+\left(r^{(2)}(x)f(x)+f^{(2)}(x)r(x)+2r'(x)f'(x)\right)\frac{m_n^{-2}}{6}\sum_{k=0}^{m_n-1}\Big\{3(k-m_nx)^2\\
&+3(k-mx)+1\Big\}b_k(m_n-1,x)[1+o(1)]\\
&= N(x)+\left(r'(x)f(x)+r(x)f'(x)\right)\frac{m_n^{-1}}{2}\left\{2T_{1,m_n-1}(x)+(1-2x)T_{0,m_n-1}(x)\right\}\\
&+\left(r^{(2)}(x)f(x)+f^{(2)}(x)r(x)+2r'(x)f'(x)\right)\frac{m_n^{-2}}{6}\Big\{3T_{2,m_n-1}(x)\\
&+3(2x+1)T_{1,m_n-1}(x)+(x^2+3x+1)T_{0,m_n-1}(x)\Big\}[1+o(1)],
\end{aligned}$$

where

$$T_{j,m_n}(x)=\sum_{k=0}^{m_n-1}(k-m_nx)^jb_k(m_n,x),\quad \forall j\in\mathbb{N}.$$

Note that it is easy to obtain

$$T_{0,m_n}(x)=1,\quad T_{1,m_n}(x)=0,\quad T_{2,m_n}(x)=m_nx(1-x),$$

then, we have

$$\mathbb{E}\left[N_n(x)\right] = N(x)+\Delta_2(x)m_n^{-1}+o\left(m_n^{-1}\right). \tag{13}$$

Moreover, we have

$$\mathbb{V}ar\left[N_n(x)\right]=\mathbb{E}\left[N_n^2(x)\right]-\mathbb{E}^2\left[N_n(x)\right],$$

where

$$\begin{aligned} N_n^2(x) &= \frac{m_n^2}{n^2}\sum_{i=1}^n Y_i^2\left(\sum_{k=0}^{m_n-1}\mathbb{1}_{\left\{\frac{k}{m_n}<X_i\leq\frac{k+1}{m_n}\right\}}b_k(m_n-1,x)\right)^2 \\ &\quad+\frac{m_n^2}{n^2}\sum_{\substack{i,j=1\\ i\neq j}}^n Y_iY_j\left(\sum_{k=0}^{m_n-1}\mathbb{1}_{\left\{\frac{k}{m_n}<X_i\leq\frac{k+1}{m_n}\right\}}b_k(m_n-1,x)\right) \\ &\quad\times\left(\sum_{k=0}^{m_n-1}\mathbb{1}_{\left\{\frac{k}{m_n}<X_j\leq\frac{k+1}{m_n}\right\}}b_k(m_n-1,x)\right). \end{aligned}$$

Then, we get

$$\begin{aligned} \mathbb{E}\left[N_n^2(x)\right] &= \frac{m_n^2}{n}\mathbb{E}\left[Y^2\left(\sum_{k=0}^{m_n-1}\mathbb{1}_{\left\{\frac{k}{m_n}<X\leq\frac{k+1}{m_n}\right\}}b_k(m_n-1,x)\right)^2\right] \\ &\quad+\frac{m_n^2n(n-1)}{n^2}\mathbb{E}^2\left[Y\sum_{k=0}^{m_n-1}\mathbb{1}_{\left\{\frac{k}{m_n}<X\leq\frac{k+1}{m_n}\right\}}b_k(m_n-1,x)\right], \\ &= \frac{m_n^2}{n}\mathbb{E}\left[Y^2\left(\sum_{k=0}^{m_n-1}\mathbb{1}_{\left\{\frac{k}{m_n}<X\leq\frac{k+1}{m_n}\right\}}b_k(m_n-1,x)\right)^2\right]+\left(1-\frac{1}{n}\right)\mathbb{E}^2\left[N_n(x)\right], \end{aligned}$$

and

$$\begin{aligned} Var\left[N_n(x)\right] &= \frac{m_n^2}{n}\mathbb{E}\left[Y^2\left(\sum_{k=0}^{m_n-1}\mathbb{1}_{\left\{\frac{k}{m_n}<X\leq\frac{k+1}{m_n}\right\}}b_k(m_n-1,x)\right)^2\right]-\frac{1}{n}\mathbb{E}^2\left[N_n(x)\right], \\ &= \frac{m_n^2}{n}\mathbb{E}\left[Y^2\sum_{k=0}^{m_n-1}\mathbb{1}_{\left\{\frac{k}{m_n}<X\leq\frac{k+1}{m_n}\right\}}b_k^2(m_n-1,x)\right]-\frac{1}{n}\mathbb{E}^2\left[N_n(x)\right], \\ &= \frac{m_n^2}{n}\sum_{k=0}^{m_n-1}\int_{\frac{k}{m_n}}^{\frac{k+1}{m_n}}\left(\int_{\mathbb{R}}y^2g(z,y)dy\right)dz\,b_k^2(m_n-1,x)-\frac{1}{n}\mathbb{E}^2\left[N_n(x)\right], \\ &= \frac{m_n^2}{n}\sum_{k=0}^{m_n-1}\left(\int_{\frac{k}{m_n}}^{\frac{k+1}{m_n}}\mathbb{E}[Y^2|X=z]f(z)dz\right)b_k^2(m_n-1,x)-\frac{1}{n}\mathbb{E}^2\left[N_n(x)\right], \\ &= \frac{m_n}{n}\mathbb{E}[Y^2|X=x]f(x)S_{m_n}(x)-\frac{1}{n}\mathbb{E}^2\left[N_n(x)\right]. \end{aligned}$$

The application of Lemma A.1 (ii) and (iii), ensures that

$$(14)\quad Var\left[N_n(x)\right]=\begin{cases}\frac{m_n^{1/2}}{n}\mathbb{E}[Y^2|X=x]f(x)\psi(x)+o_x\left(\frac{m_n^{1/2}}{n}\right) & \text{for } x\in(0,1),\\ \frac{m_n}{n}\mathbb{E}[Y^2|X=x]f(x)+o_x\left(\frac{m_n}{n}\right) & \text{for } x=0,1.\end{cases}$$

Furthermore, we have

$$\begin{aligned} Cov\left(f_n(x),N_n(x)\right) &= \mathbb{E}\left[f_n(x)N_n(x)\right]-\mathbb{E}\left[f_n(x)\right]\mathbb{E}\left[N_n(x)\right], \\ &= \frac{m_n^2}{n}\mathbb{E}\left[Y\left(\sum_{k=0}^{m_n-1}\mathbb{1}_{\left\{\frac{k}{m_n}<X\leq\frac{k+1}{m_n}\right\}}b_k(m_n-1,x)\right)^2\right] \\ &\quad+\frac{n(n-1)m_n^2}{n^2}\mathbb{E}^2\left[Y\sum_{k=0}^{m_n-1}\mathbb{1}_{\left\{\frac{k}{m_n}<X\leq\frac{k+1}{m_n}\right\}}b_k(m_n-1,x)\right]-\mathbb{E}\left[f_n(x)\right]\mathbb{E}\left[N_n(x)\right], \end{aligned}$$

$$= \frac{m_n^2}{n}\mathbb{E}\left[Y\left(\sum_{k=0}^{m_n-1}\mathbb{1}_{\left\{\frac{k}{m_n}<X\leq\frac{k+1}{m_n}\right\}}b_k(m_n-1,x)\right)^2\right] - \frac{1}{n}\mathbb{E}\left[f_n(x)\right]\mathbb{E}\left[N_n(x)\right],$$

$$= \frac{m_n^2}{n}\sum_{k=0}^{m_n-1}\int_{\frac{k}{m_n}}^{\frac{k+1}{m_n}}\left(\int_{\mathbb{R}} yg(z,y)dy\right)dz b_k^2(m_n-1,x) - \frac{1}{n}\mathbb{E}\left[f_n(x)\right]\mathbb{E}\left[N_n(x)\right],$$

$$= \frac{m_n}{n}r(x)f(x)S_{m_n}(x) - \frac{1}{n}\mathbb{E}\left[f_n(x)\right]\mathbb{E}\left[N_n(x)\right].$$

The application of Lemma A.1 (ii) and (iii), ensures that

$$\text{(15)} \quad Cov\left(f_n(x), N_n(x)\right) = \begin{cases} \frac{m_n^{1/2}}{n}r(x)f(x)\psi(x) + o_x\left(\frac{m_n^{1/2}}{n}\right) & \text{for } x\in(0,1), \\ \frac{m_n}{n}r(x)f(x) + o_x\left(\frac{m_n}{n}\right) & \text{for } x=0,1. \end{cases}$$

To compute the bias of $\widehat{r}_n(x)$, we let $h(x,y) = \frac{y}{x}$ and we apply Taylor's expansion, we get

$$h(x_n,y_n) = h(x,y) + (x_n-x,y_n-y)\nabla h^T(x,y) + \frac{1}{2}(x_n-x,y_n-y)\mathbb{H}(x,y)(x_n-x,y_n-y)^T + o\left(||(x_n-x,y_n-y)||^2\right),$$

where ∇h is the gradient of h and $\mathbb{H}$ is its hessian matrix.

$$\nabla h(x,y) = \left(-\frac{y}{x^2},\frac{1}{x}\right)$$
$$\mathbb{H} = \begin{pmatrix} \frac{2y}{x^3} & -\frac{1}{x^2} \\ -\frac{1}{x^2} & 0 \end{pmatrix}.$$

Then, we have

$$\frac{y_n}{x_n} = \frac{y}{x} - \frac{y}{x^2}(x_n-x) + \frac{1}{x}(y_n-y) + \frac{y}{x^3}(x_n-x)^2 - \frac{1}{x^2}(x_n-x)(y_n-y) + o\left((x_n-x)^2+(x_n-x)(y_n-y)\right).$$

We set $(x_n,y_n) = (f_n(x),N_n(x))$ and $(x,y) = (f(x),N(x))$, we infer that

$$\begin{aligned} \widehat{r}_n(x) &= r(x) - \frac{r(x)}{f(x)}\left(f_n(x)-f(x)\right) + \frac{1}{f(x)}\left(N_n(x)-N(x)\right) \\ &+\frac{r(x)}{\{f(x)\}^2}\left(f_n(x)-f(x)\right)^2 - \frac{1}{\{f(x)\}^2}\left(f_n(x)-f(x)\right)\left(N_n(x)-N(x)\right) \\ &+o\Big(\left(f_n(x)-f(x)\right)^2 + \left(f_n(x)-f(x)\right)\left(N_n(x)-N(x)\right)\Big), \end{aligned}$$

then

$$\begin{aligned} \mathbb{E}\left[\widehat{r}_n(x)\right] &= r(x) - \frac{r(x)}{f(x)}\left(\mathbb{E}\left[f_n(x)\right]-f(x)\right) + \frac{1}{f(x)}\left(\mathbb{E}\left[N_n(x)\right]-N(x)\right) \\ &+ \frac{r(x)}{\{f(x)\}^2}\left(\mathbb{E}\left[f_n(x)\right]-f(x)\right)^2 - \frac{1}{\{f(x)\}^2}\mathbb{E}\left[\left(f_n(x)-f(x)\right)\left(N_n(x)-N(x)\right)\right] \\ &+ o\Big(\mathbb{E}\left[\left(f_n(x)-f(x)\right)^2\right] + \mathbb{E}\left[\left(f_n(x)-f(x)\right)\left(N_n(x)-N(x)\right)\right]\Big). \end{aligned}$$

Let us recall, that for the Vitale's estimator f_n, we have

$$\text{(16)} \qquad \mathbb{E}\left[f_n(x)\right] = f(x) + \frac{\Delta_1(x)}{m_n} + o\left(m_n^{-1}\right), \quad \forall x\in[0,1],$$

and

$$Var\left[f_n(x)\right]=\begin{cases}\frac{m_n^{1/2}}{n}f(x)\psi(x)+o_x\left(\frac{m_n^{1/2}}{n}\right) & \text{for } x\in(0,1),\\ \frac{m_n}{n}f(x)+o_x\left(\frac{m_n}{n}\right) & \text{for } x=0,1.\end{cases}\tag{17}$$

The combination of (16) and (13), ensures that

$$\begin{aligned}\mathbb{E}\left[\widehat{r}_n(x)\right] &= r(x)+\left(\frac{1}{f(x)}\Delta_2(x)-\frac{r(x)}{f(x)}\Delta_1(x)\right)m_n^{-1}+o\left(m_n^{-1}\right),\\ &= r(x)+\Delta(x)m_n^{-1}+o\left(m_n^{-1}\right),\quad \forall x\in[0,1]\end{aligned}$$

and we obtain (2) of Proposition 3.1.

Now, in order to compute the variance of $\widehat{r}_n(x)$, we use the fact that

$$Var\left[h(x_n,y_n)\right]=\nabla h(x,y)Var(x_n,y_n)\nabla h^T(x,y)[1+o(1)],$$

which ensures that

$$Var\left[\widehat{r}_n(x)\right]=\nabla h(x,y)\Sigma\left[f_n(x),N_n(x)\right]\nabla h^T(x,y)[1+o(1)].$$

The combination of (14), (15) and (17), ensures that

$$\begin{aligned}&\Sigma\left[f_n(x),N_n(x)\right]\\ &=\begin{cases}\frac{m_n^{1/2}}{n}\begin{pmatrix}f(x) & f(x)r(x)\\ f(x)r(x) & f(x)\mathbb{E}(Y^2|X=x)\end{pmatrix}\psi(x)+o_x\left(\frac{m_n^{1/2}}{n}\right) & \text{for } x\in(0,1),\\ \frac{m_n}{n}\begin{pmatrix}f(x) & f(x)r(x)\\ f(x)r(x) & f(x)\mathbb{E}(Y^2|X=x)\end{pmatrix}+o\left(\frac{m_n}{n}\right) & \text{for } x=0,1.\end{cases}\end{aligned}$$

We infer that, for $x\in(0,1)$, we have

$$\begin{aligned}Var\left[\widehat{r}_n(x)\right] &= \left(-\frac{N(x)}{\{f(x)\}^2},\frac{1}{f(x)}\right)\times\begin{pmatrix}f(x) & f(x)r(x)\\ f(x)r(x) & f(x)\mathbb{E}(Y^2|X=x)\end{pmatrix}\\ &\quad\times\left(-\frac{N(x)}{\{f(x)\}^2},\frac{1}{f(x)}\right)^T\times\frac{m^{1/2}}{n}\psi(x)+o_x\left(\frac{m_n^{1/2}}{n}\right),\\ &=\left(-\frac{N(x)}{f(x)}+r(x),-\frac{N(x)r(x)}{f(x)}+\mathbb{E}(Y^2|X=x)\right)\times\left(-\frac{N(x)}{\{f(x)\}^2},\frac{1}{f(x)}\right)^T\\ &\quad\times\frac{m_n^{1/2}}{n}\psi(x)+o_x\left(\frac{m_n^{1/2}}{n}\right),\\ &=\frac{1}{f(x)}\left(0,-\mathbb{E}^2(Y|X=x)+\mathbb{E}(Y^2|X=x)\right)\times(-1,1)^T\frac{m_n^{1/2}}{n}\psi(x)+o_x\left(\frac{m_n^{1/2}}{n}\right),\\ &=\frac{1}{f(x)}\left[\mathbb{E}(Y^2|X=x)-\mathbb{E}^2(Y|X=x)\right]\times\frac{m_n^{1/2}}{n}\psi(x)+o_x\left(\frac{m_n^{1/2}}{n}\right),\\ &=\frac{m_n^{1/2}}{n}\frac{Var\left[Y|X=x\right]}{f(x)}\psi(x)+o_x\left(\frac{m_n^{1/2}}{n}\right),\end{aligned}$$

and, For $x\in\{0,1\}$, we have

$$Var\left[\widehat{r}_n(x)\right]=\frac{m_n}{n}\frac{Var\left[Y|X=x\right]}{f(x)}+o_x\left(\frac{m_n}{n}\right).$$

which gives (3) of Proposition 3.1.

A.2. **Proof of Proposition 3.2.** First, we have

$$\begin{aligned} MISE(\widehat{r}_n) &= \int_0^1 \left(Var\left[\widehat{r}_n(x)\right] + Bias^2\left[\widehat{r}_n(x)\right]\right) dx \\ &= \int_0^1 Var\left[\widehat{r}_n(x)\right] dx + C_1 m_n^{-2} + o\left(m_n^{-2}\right). \end{aligned}$$

Moreover, we have

$$\begin{aligned} Var\left[\widehat{r}_n(x)\right] &\simeq \left(-\frac{N(x)}{\{f(x)\}^2}, \frac{1}{f(x)}\right) \times \begin{pmatrix} Var\left[f_n(x)\right] & Cov\left[f_n(x), N_n(x)\right] \\ Cov\left[f_n(x), N_n(x)\right] & Var\left[N_n(x)\right] \end{pmatrix} \\ &\quad \times \left(-\frac{N(x)}{\{f(x)\}^2}, \frac{1}{f(x)}\right)^T, \\ &= \frac{r^2(x)}{\{f(x)\}^2} Var\left[f_n(x)\right] - 2\frac{r(x)}{\{f(x)\}^2} Cov\left[f_n(x), N_n(x)\right] \\ &\quad + \frac{1}{\{f(x)\}^2} Var\left[N_n(x)\right]\left[1 + o(1)\right], \end{aligned}$$

then

$$\begin{aligned} \int_0^1 Var\left[\widehat{r}_n(x)\right] dx &= \int_0^1 r^2(x) dx \frac{Var\left[f_n(x)\right]}{\{f(x)\}^2} dx - 2\int_0^1 r(x) \frac{Cov\left[f_n(x), N_n(x)\right]}{\{f(x)\}^2} dx \\ &\quad + \int_0^1 \frac{Var\left[N_n(x)\right] dx}{\{f(x)\}^2}\left[1 + o(1)\right]. \end{aligned} \tag{18}$$

Since, we have for $x \in [0,1]$,

$$\begin{aligned} Var\left[f_n(x)\right] &= \frac{1}{n}\left[A_{m_n}(x) - f_{m_n}^2(x)\right], \\ f_{m_n}^2(x) &= \mathbb{E}^2\left[f_n(x)\right] = f^2(x) + O\left(m_n^{-1}\right), \\ A_{m_n}(x) &= m_n^2 \sum_{k=0}^{m_n-1}\left[F\left(\frac{k+1}{m_n}\right) - F\left(\frac{k}{m_n}\right)\right] b_k^2(m_n - 1, x), \\ &= m_n\left[f(x) S_{m_n-1}(x) + O\left(H_{m_n-1}(x)\right) + O\left(m_n^{-1}\right)\right], \\ H_{m_n}(x) &= \sum_{k=0}^{m_n-1}\left|\frac{k}{m_n} - x\right| b_k^2(m_n - 1, x). \end{aligned}$$

The application of Cauchy-Schwarz inequality together with the fact that $0 \le b_k(m_n, x) \le 1$ and

$$\sum_{k=0}^{m_n}\left(\frac{k}{m_n} - x\right)^2 b_k(m_n, x) = \frac{x(1-x)}{m_n} \le \frac{1}{4m_n},$$

gives for all $m_n \ge 1$ and $x \in [0,1]$

$$H_{m_n}(x) \le \left[\sum_{k=0}^{m}\left(\frac{k}{m_n} - x\right)^2 b_k(m_n, x)\right]^{1/2}\left[\sum_{k=0}^{m_n} b_k^3(m_n, x)\right]^{1/2} \le \left[\frac{S_{m_n}(x)}{4m_n}\right]^{1/2}.$$

Moreover, the application of Jensen's inequality together with Lemma A.1 (iv), ensures that, for any continuous function g

$$\int_0^1 g(x) H_{m_n}(x) dx \le \int_0^1 g(x)\left[\frac{S_{m_n}(x)}{4m_n}\right]^{1/2} dx,$$

$$\leq \left[\int_0^1 g(x)dx\right]^{1/2}\left[\frac{1}{4m_n^{3/2}}\int_0^1 g(x)\psi(x)dx + o\left(m_n^{-3/2}\right)\right]^{1/2}$$
$$= O\left(m_n^{-3/4}\right).$$

Then, we infer that

$$\begin{aligned}
\int_0^1 r^2(x)\frac{Var\left[f_n(x)\right]}{\{f(x)\}^2}dx &= \frac{1}{n}\int_0^1 r^2(x)\frac{A_{m_n}(x)-f_{m_n}^2(x)}{\{f(x)\}^2}dx,\\
&= \frac{1}{n}\left[\int_0^1 r^2(x)\frac{A_{m_n}(x)}{\{f(x)\}^2}dx - \int_0^1 r^2(x)dx\right] + O\left(\frac{1}{m_n n}\right),\\
&= \frac{m_n}{n}\left[\int_0^1 \frac{r^2(x)}{\{f(x)\}^2}\left(S_{m_n-1}(x) + O\left(H_{m_n-1}(x)\right) + O\left(m_n^{-1}\right)\right)dx\right]\\
&\quad - \frac{1}{n}\int_0^1 r^2(x)dx + O\left(\frac{1}{m_n n}\right),\\
&= \frac{m_n}{n}\left[\int_0^1 \frac{r^2(x)}{f(x)}S_{m_n-1}(x)dx + O\left(m_n^{-3/4}\right)\right] - \frac{1}{n}\int_0^1 r^2(x)dx + O\left(\frac{1}{m_n n}\right).
\end{aligned}$$

Moreover, the application of Lemma A.1 (iv), gives

$$\begin{aligned}
\int_0^1 r^2(x)\frac{Var\left[f_n(x)\right]}{\{f(x)\}^2}dx &= \frac{m_n^{1/2}}{n}\int_0^1 \frac{r^2(x)}{f(x)}\psi(x)dx - \frac{1}{n}\int_0^1 r^2(x)dx\\
&\quad + o\left(\frac{m_n^{1/2}}{n}\right) + O\left(\frac{1}{m_n n}\right).
\end{aligned} \tag{19}$$

Further, we have

$$\begin{aligned}
&Cov\left[f_n(x), N_n(x)\right]\\
&= \frac{1}{n}\left\{m_n^2\sum_{k=0}^{m_n-1}\left(\int_{\frac{k}{m}}^{\frac{k+1}{m_n}} r(z)f(x)dz\right)b_k^2(m_n-1,x) - \mathbb{E}\left[f_n(x)\right]\mathbb{E}\left[N_n(x)\right]\right\},\\
&= \frac{1}{n}\left\{m_n^2\sum_{k=0}^{m_n-1}\left(\int_{\frac{k}{m_n}}^{\frac{k+1}{m_n}} \left[r(x)f(x)+O(z-x)\right]dz\right)b_k^2(m_n-1,x) - f(x)N(x)\right\}\\
&\quad + O\left(\frac{1}{m_n n}\right),\\
&= \frac{m_n}{n}\left[r(x)f(x)S_{m_n-1}(x) + O\left(H_{m_n-1}(x)\right) + O\left(m_n^{-1}\right)\right] - \frac{1}{n}f(x)N(x)\\
&\quad + O\left(\frac{1}{m_n n}\right).
\end{aligned}$$

Then, using the same argument for $H_{m_n-1}(x)$ as previously, we obtain

$$\begin{aligned}
&\int_0^1 r(x)\frac{Cov\left[f_n(x), N_n(x)\right]}{\{f(x)\}^2}dx\\
&= \frac{m_n}{n}\left[\int_0^1 \frac{r^2(x)}{f(x)}S_{m_n-1}(x)dx + O\left(m_n^{-3/4}\right)\right] - \frac{1}{n}\int_0^1 r^2(x)dx\\
&\quad + O\left(\frac{1}{m_n n}\right)
\end{aligned} \tag{20}$$

$$= \frac{m_n^{1/2}}{n}\int_0^1 \frac{r^2(x)}{f(x)}\psi(x)dx - \frac{1}{n}\int_0^1 r^2(x)dx + o\left(\frac{m_n^{1/2}}{n}\right) + O\left(\frac{1}{m_n n}\right).$$

Moreover, we have

$$\begin{aligned}
Var\left[N_n(x)\right] &= \frac{m_n^2}{n}\sum_{k=0}^{m_n-1}\left(\int_{\frac{k}{m_n}}^{\frac{k+1}{m_n}} \mathbb{E}[Y^2|X=z]f(z)dz\right) b_k^2(m_n-1,x) - \frac{1}{n}\mathbb{E}^2\left[N_n(x)\right],\\
&= \frac{m_n^2}{n}\sum_{k=0}^{m_n-1}\left(\int_{\frac{k}{m_n}}^{\frac{k+1}{m_n}} \left[\mathbb{E}[Y^2|X=x]f(x) + O\left(z-x\right)\right]dz\right) b_k^2(m_n-1,x) - \frac{1}{n}N^2(x)\\
&\quad + O\left(\frac{1}{m_n n}\right),\\
&= \frac{m_n}{n}\left[\mathbb{E}[Y^2|X=x]f(x)S_{m_n-1}(x) + O\left(H_{m_n-1}(x)\right) + O\left(m_n^{-1}\right)\right] - \frac{1}{n}N^2(x)\\
&\quad + O\left(\frac{1}{m_n n}\right),
\end{aligned}$$

then,

$$\begin{aligned}
(21)\quad &\int_0^1 \frac{Var\left[N_n(x)\right]}{\left\{f(x)\right\}^2}dx\\
&= \frac{m_n}{n}\left[\int_0^1 \frac{\mathbb{E}[Y^2|X=x]}{f(x)}S_{m_n-1}(x)dx + O\left(m_n^{-3/4}\right)\right] - \frac{1}{n}\int_0^1 r^2(x)dx\\
&\quad + O\left(\frac{1}{m_n n}\right),\\
&= \frac{m_n^{1/2}}{n}\int_0^1 \frac{\mathbb{E}[Y^2|X=x]}{f(x)}\psi(x)dx - \frac{1}{n}\int_0^1 r^2(x)dx + o\left(\frac{m_n^{1/2}}{n}\right)\\
&\quad + O\left(\frac{1}{m_n n}\right).
\end{aligned}$$

Finally, substituting (19), (20) and (21) into (18), gives

$$\begin{aligned}
\int_0^1 Var\left[\widehat{r}_n(x)\right]dx &= \left(\int_0^1 \frac{\mathbb{E}[Y^2|X=x]}{f(x)}\psi(x)dx - \int_0^1 \frac{\mathbb{E}^2[Y|X=x]}{f(x)}\psi(x)dx\right)\frac{m_n^{1/2}}{n}\\
&\quad + o\left(\frac{m_n^{1/2}}{n}\right),\\
&= \int_0^1 \frac{\mathbb{E}[Y^2|X=x] - \mathbb{E}^2[Y|X=x]}{f(x)}\psi(x)dx\frac{m_n^{1/2}}{n} + o\left(\frac{m_n^{1/2}}{n}\right),\\
&= \int_0^1 \frac{Var[Y|X=x]}{f(x)}\psi(x)dx\frac{m_n^{1/2}}{n}\\
&\quad + o\left(\frac{m_n^{1/2}}{n}\right).
\end{aligned}$$

Then, we obtain the result in (4) of Proposition 3.2.

A.3. **Proof of Theorem 3.1.** To prove the convergence, for $x \in (0,1)$, we use the fact that

$$n^{1/2}m_n^{-1/4}\left(\widehat{r}_n(x) - \mathbb{E}\left[\widehat{r}_n(x)\right]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{Var\left[Y|X=x\right]}{f(x)}\psi(x)\right), \tag{22}$$

which will be proved later. We have

$$\begin{aligned} &n^{1/2}m_n^{-1/4}\left(\widehat{r}_n(x) - r(x)\right) \\ &\quad = n^{1/2}m_n^{-1/4}\left(\widehat{r}_n(x) - \mathbb{E}\left[\widehat{r}_n(x)\right] + n^{1/2}m_n^{-1/4}\left(\mathbb{E}\left[\widehat{r}_n(x)\right)\right] - r(x)\right), \\ &\quad\quad = n^{1/2}m_n^{-1/4}\left(\widehat{r}_n(x) - \mathbb{E}\left[\widehat{r}_n(x)\right]\right) + n^{1/2}m_n^{-5/4}\Delta(x)[1+o(1)], \end{aligned}$$

then, when $nm_n^{-5/2} \to c$ for some constant $c \geq 0$, Part 1 of Theorem 3.1 follows immediately.

Now, when $nm_n^{-5/2} \to \infty$, we have

$$\begin{aligned} m_n\left(\widehat{r}_n(x) - r(x)\right) &= m_n\left(\widehat{r}_n(x) - \mathbb{E}\left[\widehat{r}_n(x)\right]\right) + m_n\left(\mathbb{E}\left[\widehat{r}_n(x)\right] - r(x)\right), \\ &= \left(n^{-1/2}m_n^{5/4}\right)n^{1/2}m_n^{-1/4}\left(\widehat{r}_n(x) - \mathbb{E}\left[\widehat{r}_n(x)\right]\right) + \Delta(x)[1+o(1)]. \end{aligned}$$

Since we have $n^{1/2}m_n^{5/4} \to 0$, Part 2 of Theorem 3.1 follows from (22). Now let us prove (22). First, we let

$$w_i = \frac{\sum_{k=0}^{m_n-1} \mathbb{1}_{\left\{\frac{k}{m_n} < X_i \leq \frac{k+1}{m_n}\right\}} b_k(m_n-1,x)}{\sum_{i=1}^{n}\sum_{k=0}^{m_n-1} \mathbb{1}_{\left\{\frac{k}{m_n} < X_n \leq \frac{k+1}{m_n}\right\}} b_k(m_n-1,x)}.$$

Clearly we have

$$\widehat{r}_n(x) = \sum_{i=1}^{n} w_i Y_i,$$

and

$$\widehat{r}_n(x) - \mathbb{E}\left[\widehat{r}_n(x)\right] = \sum_{i=1}^{n}\left(w_iY_i - \mathbb{E}\left[w_iY_i\right]\right).$$

Noting that $0 \leq w_i \leq 1$, for all $p > 0$, we have $\mathbb{E}\left[|w_iY_i|^{2+p}\right] = O(1)$ and

$$\sum_{i=1}^{n}\mathbb{E}\left[|w_iY_i|^{2+p}\right] = O(n).$$

Moreover, for $x \in (0,1)$, we have

$$v_n^2 = \sum_{i=1}^{n} \mathrm{Var}\left[w_iY_i\right] = \frac{Var[Y|X=x]}{f(x)}\psi(x)nm_n^{1/2} + o\left(nm_n^{1/2}\right),$$

then, we have

$$\begin{aligned} \frac{1}{v_n^{2+p}}\sum_{i=1}^{n}\mathbb{E}\left[|w_iY_i|^{2+p}\right] &= O\left(\frac{n}{n^{\frac{2+p}{2}}m_n^{\frac{2+p}{4}}}\right), \\ &= O\left(n^{-\frac{p}{2}}m_n^{-\frac{2+p}{4}}\right) = o(1). \end{aligned}$$

Then the convergence in (22) follows from the application of Lyapounov's theorem.

Now to prove the convergence, For $x \in \{0,1\}$, we use the fact that

$$\sqrt{\frac{n}{m_n}}\left(\widehat{r}_n(x)-\mathbb{E}\left[\widehat{r}_n(x)\right]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\mathrm{Var}\left[Y|X=x\right]}{f(x)}\right), \tag{23}$$

which will be proved later. We have

$$\begin{aligned}
\sqrt{\frac{n}{m_n}}\left(\widehat{r}_n(x)-r(x)\right) &= \sqrt{\frac{n}{m_n}}\left(\widehat{r}_n(x)-\mathbb{E}\left[\widehat{r}_n(x)\right]+\sqrt{\frac{n}{m_n}}\left(\mathbb{E}\left[\widehat{r}_n(x)\right)\right]-r(x)\right), \\
&= \sqrt{\frac{n}{m_n}}\left(\widehat{r}_n(x)-\mathbb{E}\left[\widehat{r}_n(x)\right]\right)+n^{1/2}m_n^{-3/2}\Delta(x)[1+o(1)],
\end{aligned}$$

we infer that, when $nm_n^{-3} \to c$ for some constant $c \geq 0$, then Part 3 of Theorem 3.1 follows.

Now, if $nm_n^{-3} \to \infty$, we have

$$\begin{aligned}
m_n\left(\widehat{r}_n(x)-r(x)\right) &= m_n\left(\widehat{r}_n(x)-\mathbb{E}\left[\widehat{r}_n(x)\right]\right)+m_n\left(\mathbb{E}\left[\widehat{r}_n(x)\right]-r(x)\right), \\
&= \left(n^{1/2}m_n^{3/2}\right)\sqrt{\frac{n}{m_n}}\left(\widehat{r}_n(x)-\mathbb{E}\left[\widehat{r}_n(x)\right]\right)+\Delta(x)[1+o(1)],
\end{aligned}$$

and then Part 4 of Theorem 3.1 follows from (23) and the fact that $n^{1/2}m_n^{5/4} \to 0$.

To prove (23), For $x \in \{0,1\}$, we have

$$v_n^2 = \sum_{i=1}^n \mathrm{Var}\left[\left|w_iY_i-\mathbb{E}\left[w_iY_i\right]\right|^{2+p}\right] = \frac{Var[Y|X=x]}{f(x)}nm_n + o\left(nm_n\right),$$

hence

$$\begin{aligned}
\frac{1}{v_n^{2+p}}\sum_{i=1}^n \mathbb{E}\left[\left|w_iY_i-\mathbb{E}\left[w_iY_i\right]\right|^{2+p}\right] &= O\left(\frac{n}{n^{\frac{2+p}{2}}m_n^{\frac{2+p}{2}}}\right), \\
&= O\left(n^{-\frac{p}{2}}m_n^{-\frac{2+p}{2}}\right) = o(1).
\end{aligned}$$

Then the convergence in (23) follows from the application of Lyapounov's theorem.

A.4. **Proof of the results obtained for r_n.** First, we set $n_0 \geq 3$ such that $\forall k \geq n_0, \quad \gamma_k \leq \left(2\left\|f\right\|_\infty\right)^{-1}$ and $\gamma_k m_k \leq 1$. Moreover, we introduce the following notations:

$$\begin{aligned}
s_n &= \sum_{k=n_0}^n \gamma_k \\
\Pi_n(s) &= \prod_{j=n_0}^n \left(1-s\gamma_j\right) \quad \text{for} \quad s>0, \\
U_{k,n}(s) &= \Pi_n(s)\Pi_k^{-1}(s) \quad \text{for} \quad s>0.
\end{aligned}$$

Further, we define the sequences (λ_n), $(\widetilde{\lambda}_n)$, (β_n) and $(\widetilde{\beta}_n)$ by setting

$$(\lambda_n) = \begin{cases} \left(\sqrt{\gamma_n m_n^{1/2}}\right) & \text{if } \lim_{n\to\infty}\left(\gamma_n m_n^{5/2}\right)=\infty, \\ \left(m_n^{-1}\right) & \text{otherwise.}\end{cases}$$

$$(\widetilde{\lambda}_n) = \begin{cases} \left(\sqrt{\gamma_n m_n^{1/2}}\ln n\right) & \text{if } \lim_{n\to\infty}\left(\gamma_n m_n^{5/2}\ln n\right)=\infty, \\ \left(m_n^{-1}\right) & \text{otherwise.}\end{cases}$$

$$(\beta_n) = \begin{cases} \left(\sqrt{\gamma_n m_n}\right) & \text{if } \lim_{n\to\infty}\left(\gamma_n m_n^{3}\right)=\infty, \\ \left(m_n^{-1}\right) & \text{otherwise.}\end{cases} \tag{24}$$

$$(\widetilde{\beta}_n) = \begin{cases} \left(\sqrt{\gamma_n m_n}\ln n\right) & \text{if } \lim_{n\to\infty}\left(\gamma_n m_n^3 \ln n\right) = \infty, \\ \left(m_n^{-1}\right) & \text{otherwise.} \end{cases}$$

Note that (λ_n) and $(\widetilde{\lambda}_n)$ belong to $\mathcal{GS}(-\lambda^*)$ with $\lambda^* = \min\left\{\frac{2\alpha-a}{4}, a\right\}$, and (β_n) and $(\widetilde{\beta}_n)$ belong to $\mathcal{GS}(-\beta^*)$ with $\beta^* = \min\left\{\frac{\alpha-a}{2}, a\right\}$.

To establish the characteristics of our recursive estimator r_n defined by (5), we state the following technical lemma, which is proved in Mokkadem et al. [21], and which will be used throughout the demonstrations.

Lemma A.2. *Let* $(v_n) \in \mathcal{GS}(v^*)$, $(\gamma_n) \in \mathcal{GS}(-\alpha)$ *with* $\alpha > 0$, *and set* $l > 0$. *If* $ls - v^*\xi > 0$ *(where* $\xi = \lim_{n\to\infty}(n\gamma_n)^{-1}$*), then*

$$\lim_{n\to\infty} v_n \Pi_n^l(s) \sum_{k=n_0}^{n} \Pi_k^{-l}(s) \frac{\gamma_k}{v_k} = \frac{1}{ls - v^*\xi}.$$

Moreover, for all positive sequence (α_n) *such that* $\lim_{n\to\infty}\alpha_n = 0$, *and all* C,

$$\lim_{n\to\infty} v_n \Pi_n^l(s) \left[\sum_{k=n_0}^{n} \Pi_k^{-l}(s) \frac{\gamma_k}{v_k}\alpha_k + C\right] = 0.$$

We first give the asymptotic behavior of (ρ_n) defined in (7). Then, we follow similar steps as Mokkadem et al. [21] to show how the asymptotic behavior of (r_n) (6) can be deduced from that of (ρ_n).

A.4.1. *Asymptotic behavior of* ρ_n. The following Lemma gives the bias and the variance of the estimator ρ_n defined in (6).

Lemma A.3. *(Bias and Variance of* ρ_n*)*

(1) *Assume that Assumptions* $(A1)-(A4)$ *hold. For* $x \in (0,1)$, *such that* $f(x) > 0$, *we have*

$$\mathbb{E}\left[\rho_n(x)\right] - r(x) = \frac{f(x)\Delta(x)}{f(x) - a\xi}\mathbb{1}_{\left\{a\in\left(\frac{1-\alpha}{4},\frac{2}{5}\alpha\right]\right\}} m_n^{-1} + \mathbb{1}_{\left\{a\in\left(\frac{2}{5}\alpha,\frac{2}{3}\alpha\right]\right\}} o\left(\sqrt{\gamma_n m_n^{1/2}}\right) + o\left(m_n^{-1}\right), \tag{25}$$

$$Var\left[\rho_n(x)\right] = \frac{2f(x)\psi(x)Var\left[Y|X=x\right]}{4f(x) - (2\alpha - a)}\mathbb{1}_{\left\{a\in\left(\frac{2}{5}\alpha,\frac{2}{3}\alpha\right]\right\}}\gamma_n m_n^{1/2} + \mathbb{1}_{\left\{a\in\left(\frac{1-\alpha}{4},\frac{2}{5}\alpha\right)\right\}} o\left(m_n^{-2}\right) + o\left(\gamma_n m_n^{1/2}\right). \tag{26}$$

(2) *Assume that Assumptions* $(A1)-(A'4)$ *hold. For* $x \in \{0,1\}$, *such that* $f(x) > 0$, *we have*

$$\mathbb{E}\left[\rho_n(x)\right] - r(x) = \frac{f(x)\Delta(x)}{f(x) - a\xi}\mathbb{1}_{\left\{a\in\left(\frac{1-\alpha}{4},\frac{\alpha}{3}\right]\right\}} m_n^{-1} + \mathbb{1}_{\left\{a\in\left(\frac{\alpha}{3},\frac{2}{3}\alpha\right)\right\}} o\left(\sqrt{\gamma_n m_n}\right) + o\left(m_n^{-1}\right), \tag{27}$$

$$Var\left[\rho_n(x)\right] = \frac{f(x)\psi(x)Var\left[Y|X=x\right]}{2f(x) - (\alpha - a)}\mathbb{1}_{\left\{a\in\left(\frac{\alpha}{3},\frac{2}{3}\alpha\right]\right\}}\gamma_n m_n + \mathbb{1}_{\left\{a\in\left(\frac{1-\alpha}{4},\frac{\alpha}{3}\right)\right\}} o\left(m_n^{-2}\right) + o\left(\gamma_n m_n\right). \tag{28}$$

A.4.2. *Proof of Lemma A.3.* We have, for $n \geq n_0$,

$$\begin{aligned}
\rho_n(x) - r(x) &= (1 - \gamma_n f(x))(\rho_{n-1}(x) - r(x)) + \gamma_n (W_n(x) - r(x)Z_n(x)), \\
&= \Pi_n(f(x)) \sum_{k=n_0}^{n} \Pi_k^{-1}(f(x))\gamma_k (W_k(x) - r(x)Z_k(x)) \\
&\quad + \Pi_n(f(x))(\rho_{n_0-1}(x) - r(x)), \\
&= T_n(x) + R_n(x).
\end{aligned}$$

Remark A.1. (1) Since $\rho_{n_0-1}(x) = r_{n_0-1}(x)$, we have

$$\begin{aligned}
T_n(x) &= \sum_{k=n_0}^{n} U_{k,n}(f(x))\gamma_k (W_k(x) - r(x)Z_k(x)), \\
R_n(x) &= \Pi_n(f(x))(r_{n_0-1}(x) - r(x)).
\end{aligned}$$

(2) Since $|r_{n_0-1}(x) - r(x)| = O(1)$ a.s. The application of Lemma A.2, ensures that

(29)
$$\begin{aligned}
R_n(x) &= O(\Pi_n(f(x))) \quad a.s. \\
&= \begin{cases} o(\lambda_n) & \text{for } x \in (0,1), \\ o(\beta_n) & \text{for } x \in \{0,1\}. \end{cases} \quad a.s.
\end{aligned}$$

We infer that Lemma A.3 hold when $\rho_n(x)$ is replaced by $T_n(x)$. Then, for $x \in [0,1]$ such that $f(x) > 0$, we have

$$\mathbb{E}[T_n(x)] = \sum_{k=n_0}^{n} U_{k,n}(f(x))\gamma_k (\mathbb{E}[W_k(x)] - r(x)\mathbb{E}[Z_k(x)]),$$

where

$$\begin{aligned}
\mathbb{E}[Z_k(x)] &= m_k \sum_{k=0}^{m_k-1} \left[F\left(\frac{k+1}{m_k}\right) - F\left(\frac{k}{m_k}\right)\right] b_k(m-1,x), \\
&= f(x) + \Delta_1(x)m_k^{-1} + o\left(m_k^{-1}\right),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[W_k(x)] &= m_k \mathbb{E}\left[Y \sum_{k=0}^{m_k-1} \mathbb{1}_{\left\{\frac{k}{m_k} < X \leq \frac{k+1}{m_k}\right\}} b_k(m_k-1,x)\right], \\
&= r(x)f(x) + \Delta_2(x)m_k^{-1} + o\left(m_k^{-1}\right).
\end{aligned}$$

Then, we obtain

$$\mathbb{E}[T_n(x)] = f(x)\Delta(x) \sum_{k=n_0}^{n} U_{k,n}(f(x))\gamma_k \left[m_k^{-1} + o\left(m_k^{-1}\right)\right].$$

Moreover, we have

$$\begin{aligned}
Var[T_n(x)] = \sum_{k=n_0}^{n} U_{k,n}^2(f(x))\gamma_k^2 \{&\mathrm{Var}[W_k(x)] \\
&+ r^2(x)\mathrm{Var}[Z_k(x)] - 2r(x)\mathrm{Cov}(W_k(x), Z_k(x))\},
\end{aligned}$$

where

$$\mathrm{Var}[W_k(x)] = \begin{cases} \mathbb{E}[Y^2|X=x]f(x)\psi(x)m_k^{1/2} + o_x\left(m_k^{1/2}\right) & \text{for } x \in (0,1), \\ \mathbb{E}[Y^2|X=x]f(x)m_k + o_x(m_k) & \text{for } x = 0,1. \end{cases}$$

$$\begin{aligned}
Var\left[Z_k(x)\right] &= \begin{cases} f(x)\psi(x)m_k^{1/2} + o_x\left(m_k^{1/2}\right) & \text{for } x \in (0,1), \\ f(x)m_k + o\left(m_k\right) & \text{for } x = 0,1. \end{cases} \\
Cov\left(W_k(x), Z_k(x)\right) &= \begin{cases} r(x)f(x)\psi(x)m_k^{1/2} + o_x\left(m_k^{1/2}\right) & \text{for } x \in (0,1), \\ r(x)f(x)m_k + o_x\left(m_k\right) & \text{for } x = 0,1. \end{cases}
\end{aligned}$$

which gives

$$\begin{aligned}
&Var\left[T_n(x)\right] \\
&= \begin{cases} f(x)Var\left[Y|X=x\right]\psi(x)\displaystyle\sum_{k=n_0}^{n} U_{k,n}^2(f(x))\gamma_k^2\left[m_k^{1/2} + o_x\left(m_k^{1/2}\right)\right] & \text{for } x \in (0,1), \\ f(x)Var\left[Y|X=x\right]\displaystyle\sum_{k=n_0}^{n} U_{k,n}^2(f(x))\gamma_k^2\left[m_k + o_x\left(m_k\right)\right] & \text{for } x = 0,1. \end{cases}
\end{aligned}$$

Remark A.2. (1) For $x \in (0,1)$, the application of Lemma A.2 gives (25) and (26).
(2) For $x \in \{0,1\}$, the application of Lemma A.2 gives (27) and (28).

The following lemma gives the weak convergence rate of the estimator ρ_n defined in (7), for $x \in [0,1]$ such that $f(x) > 0$.

Lemma A.4. *(Weak convergence rate of ρ_n)*
Let Assumption $(A1)-(A4)$ hold. For $x \in (0,1)$, we have

(1) *If $\gamma_n^{-1}m_n^{-5/2} \to c$ for some constant $c \geq 0$, then*

$$\gamma_n^{-1/2}m_n^{-1/4}\left(\rho_n(x) - r(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\sqrt{c}\frac{f(x)\Delta(x)}{f(x)-a\xi}, \frac{2f(x)Var\left[Y|X=x\right]\psi(x)}{4f(x)-(2\alpha-a)\xi}\right),$$

(2) *If $\gamma_n^{-1}m_n^{-5/2} \to \infty$, then*

$$m_n\left(\rho_n(x) - r(x)\right) \xrightarrow{\mathbb{P}} \frac{f(x)\Delta(x)}{f(x)-a\xi}.$$

Let Assumption $(A1)$-$(A'4)$ hold. For $x \in \{0,1\}$, we have

(1) *If $\gamma_n^{-1}m_n^{-3} \to c$ for some constant $c \geq 0$, then*

$$\gamma_n^{-1/2}m_n^{-1/2}\left(\rho_n(x) - r(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\sqrt{c}\frac{f(x)\Delta(x)}{f(x)-a\xi}, \frac{f(x)Var\left[Y|X=x\right]\psi(x)}{2f(x)-(\alpha-a)\xi}\right),$$

(2) *If $\gamma_n^{-1}m_n^{-3} \to \infty$, then*

$$m_n\left(\rho_n(x) - r(x)\right) \xrightarrow{\mathbb{P}} \frac{f(x)\Delta(x)}{f(x)-a\xi}.$$

A.4.3. *Proof of Lemma A.4.* To prove Lemma A.4, for $x \in (0,1)$, we use the fact that if $a \in [\frac{2}{5}\alpha, \frac{2}{3}\alpha)$, we have

$$\gamma_n^{-1/2}m_n^{-1/4}\left(\rho_n(x) - \mathbb{E}\left[\rho_n(x)\right]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{2f(x)Var\left[Y|X=x\right]\psi(x)}{4f(x)-(2\alpha-a)\xi}\right), \tag{30}$$

which will be proved later.

Remark A.3. (1) The result in (30) hold if we replace $\rho_n(x)$ by $T_n(x)$.
(2) Part 1 of Lemma A.4 follows from the combination (25) and (30).
(3) Part 1 and 2 of Lemma A.4 follows from the combination of (27) and (30)

Now, we set

$$\eta_k(x) = \Pi_k^{-1}(f(x))\gamma_k\left(W_k(x) - r(x)Z_k(x)\right). \tag{31}$$

Then,

$$T_n(x) - \mathbb{E}\left[T_n(x)\right] = \Pi_n(f(x))\sum_{k=n_0}^{n}\left(\eta_k(x) - \mathbb{E}\left[\eta_k(x)\right]\right).$$

Moreover, for $x \in (0,1)$, we have

$$Var\left[\eta_k(x)\right] = \Pi_k^{-2}(f(x))\gamma_k^2 m_k^{1/2}\left[f(x)\psi(x)Var\left[Y|X=x\right] + o(1)\right],$$

and, since $\lim_{n\to\infty}(n\gamma_n) > (2\alpha - a)/(4f(x))$, Lemma A.2 ensures that

$$\begin{aligned}
v_n^2 &= \sum_{k=n_0}^{n} Var\left[\eta_k(x)\right] \\
&= \sum_{k=n_0}^{n} \Pi_k^{-2}(f(x))\gamma_k^2 m_k^{1/2}\left[f(x)\psi(x)Var\left[Y|X=x\right] + o(1)\right] \\
&= \frac{2\Pi_n^{-2}(f(x))\gamma_n m_n^{1/2}}{4f(x) - (2\alpha - a)\xi}\left[f(x)\psi(x)Var\left[Y|X=x\right] + o(1)\right]. \qquad (32)
\end{aligned}$$

Further, for all $p > 0$ and $x \in [0,1]$, we make use of Lemma A.1 (ii) to ensure that

$$\begin{aligned}
&\mathbb{E}\left(|Y_k - r(x)|^{2+p}\left\{\sum_{i=0}^{m_k-1}\mathbb{1}_{\left\{\frac{i}{m_i} < X_n \le \frac{i+1}{m_i}\right\}} b_k(m_i - 1, x)\right\}^{2+p}\right) \\
&\le \left\{\sum_{i=0}^{m_k-1} b_k^2(m_i - 1, x)\right\}^{(2+p)/2}\int_0^1\int_{\mathbb{R}}|y - r(x)|^{2+p}g(z,y)dzdy \\
&\le m_n^{-(2+p)/4}\int_0^1\left\{\int_{\mathbb{R}}|y|^{2+p}g(z,y)dy + |r(x)|^{2+p}\int_{\mathbb{R}}g(z,y)dy\right\}dz \\
&= O\left(m_n^{-(2+p)/4}\right). \qquad (33)
\end{aligned}$$

Moreover, since $\lim_{n\to\infty}(n\gamma_n) > (\alpha - a/2)/(2f(x))$, there exists a $p > 0$ such that $\lim_{n\to\infty}(n\gamma_n) > (1+p)(\alpha - a/2)/(2+p)(f(x)) > \frac{(1+p)\alpha - (3(2+p)/4)a}{2+p}$, then the application of Lemma A.2 gives

$$\begin{aligned}
\sum_{k=n_0}^{n}\mathbb{E}\left[|\eta_k(x)|^{2+p}\right] &= O\left(\sum_{k=n_0}^{n}\Pi_k^{-2-p}(f(x))\gamma_k^{2+p}m_k^{2+p}\right. \\
&\quad\left.\times\mathbb{E}\left(|Y_k - r(x)|^{2+p}\left\{\sum_{i=0}^{m_k-1}\mathbb{1}_{\left\{\frac{i}{m_i} < X_n \le \frac{i+1}{m_i}\right\}} b_k(m_i - 1, x)\right\}^{2+p}\right)\right) \\
&= O\left(\sum_{k=n_0}^{n}\Pi_k^{-2-p}(f(x))\gamma_k^{2+p}m_k^{3(2+p)/4}\right) = O\left(\frac{\gamma_n^{1+p}m_n^{3(2+p)/4}}{\Pi_n^{2+p}(f(x))}\right),
\end{aligned}$$

we infer that

$$\frac{1}{v_n^{2+p}}\sum_{k=n_0}^{n}\mathbb{E}\left[|\eta_k(x)|^{2+p}\right] = O\left(m_n\left(\gamma_n m_n\right)^{p/2}\right),$$

and the convergence in (30) follows from the application of Lyapounov's Theorem.

Now, to prove Lemma A.4, For $x \in \{0,1\}$, we use the fact that when $a \in [\frac{\alpha}{3}, \frac{2}{3}\alpha)$, we have

$$\gamma_n^{-1/2} m_n^{-1/2} \left(\rho_n(x) - \mathbb{E}\left[\rho_n(x)\right]\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(0, \frac{f(x)\mathrm{Var}\left[Y|X=x\right]}{2f(x) - (\alpha - a)\xi}\right), \tag{34}$$

which will be proved later.

Remark A.4. (1) Part 3 and 4 of Lemma A.4 follows from the combination of (26) and (34).

(2) The result in (34) hold if we replace $\rho_n(x)$ by $T_n(x)$.

In the case when $x = 0, 1$, we have

$$\mathrm{Var}\left[\eta_k(x)\right] = \Pi_k^{-2}(f(x))\gamma_k^2 m_k \left[f(x)\mathrm{Var}\left[Y|X=x\right] + o(1)\right].$$

Since $\lim_{n\to\infty}(n\gamma_n) > (\alpha - a)/(2f(x))$, we make use of Lemma A.2 to ensure that

$$\begin{aligned} v_n^2 &= \sum_{k=n_0}^{n} \mathrm{Var}\left[\eta_k(x)\right] \\ &= \sum_{k=n_0}^{n} \Pi_k^{-2}(f(x))\gamma_k^2 m_k \left[f(x)\mathrm{Var}\left[Y|X=x\right] + o(1)\right] \\ &= \frac{2\Pi_n^{-2}(f(x))\gamma_n m_n}{4f(x) - (2\alpha - a)\xi}\left[f(x)\mathrm{Var}\left[Y|X=x\right] + o(1)\right]. \end{aligned}$$

Moreover, there exists a $p > 0$ such that

$$\sum_{k=n_0}^{n} \mathbb{E}\left[\left|\eta_k(x)\right|^{2+p}\right] = O\left(\frac{\gamma_n^{1+p} m_n^{3(2+p)/4}}{\Pi_n^{2+p}(f(x))}\right),$$

then,

$$\frac{1}{v_n^{2+p}} \sum_{k=n_0}^{n} \mathbb{E}\left[\left|\eta_k(x)\right|^{2+p}\right] = O\left(m_n^{1/2}\left(\gamma_n m_n^{1/2}\right)^{p/2}\right),$$

and the convergence in (34) follows from the application of Lyapounov's Theorem.

The following lemma gives the strong pointwise convergence rate of ρ_n, for $x \in [0,1]$ such that $f(x) > 0$.

Lemma A.5. *(Strong pointwise convergence rate of ρ_n) Let Assumption (A1) – (A4) hold. For $x \in (0,1)$, we have*

(1) *If $\gamma_n^{-1} m_n^{-5/2}/\ln(s_n) \to c$ for some constant $c \geq 0$, then, with probability one, the sequence*

$$\left(\sqrt{\frac{\gamma_n^{-1} m_n^{-1/2}}{2\ln(s_n)}}\left(\rho_n(x) - r(x)\right)\right)$$

is relatively compact and its limit set is the interval

$$\left[\sqrt{\frac{c}{2}}\frac{f(x)\Delta(x)}{f(x) - a\xi} - \sqrt{\frac{2f(x)\mathrm{Var}\left[Y|X=x\right]\psi(x)}{4f(x) - (2\alpha - a)\xi}}, \right.$$
$$\left.\sqrt{\frac{c}{2}}\frac{f(x)\Delta(x)}{f(x) - a\xi} + \sqrt{\frac{2f(x)\mathrm{Var}\left[Y|X=x\right]\psi(x)}{4f(x) - (2\alpha - a)\xi}}\right].$$

(2) *If $\gamma_n^{-1} m_n^{-5/2}/\ln(s_n) \to \infty$, then with probability one*

$$\lim_{n\to\infty} m_n\left(\rho_n(x) - r(x)\right) = \frac{f(x)\Delta(x)}{f(x) - a\xi}.$$

Let Assumption (A1)-(A′4) hold. For $x \in \{0,1\}$, we have

(1) *If $\gamma_n^{-1}m_n^{-3}/\ln(s_n) \to c$ for some constant $c \geq 0$, then, with probability one, the sequence*

$$\left(\sqrt{\frac{\gamma_n^{-1}m_n^{-1}}{2\ln(s_n)}}\left(\rho_n\left(x\right)-r\left(x\right)\right)\right)$$

is relatively compact and its limit set is the interval

$$\left[\sqrt{\frac{c}{2}}\frac{f(x)\Delta(x)}{f(x)-a\xi}-\sqrt{\frac{f(x)\mathrm{Var}\left[Y|X=x\right]\psi(x)}{2f(x)-(\alpha-a)\xi}},\right.$$
$$\left.\sqrt{\frac{c}{2}}\frac{f(x)\Delta(x)}{f(x)-a\xi}+\sqrt{\frac{f(x)\mathrm{Var}\left[Y|X=x\right]\psi(x)}{2f(x)-(\alpha-a)\xi}}\right].$$

(2) *If $\gamma_n^{-1}m_n^{-3}/\ln(s_n) \to \infty$, then, with probability one*

$$\lim_{n\to\infty} m_n\left(\rho_n\left(x\right)-r\left(x\right)\right)=\frac{f(x)\Delta(x)}{f(x)-a\xi}.$$

A.4.4. *Proof of Lemma A.5.*

Remark A.5. (1) In view of (30), Lemma A.5 hold when $\rho_n\left(x\right)-r\left(x\right)$ is replaced by $T_n(x)$

(2) We give the proof in the case when $x \in (0,1)$, the case $x=0,1$ can be proven by following similar steps.

First, we set

$$B_n(x)=\sum_{k=n_0}^{n}\left(\eta_k(x)-\mathbb{E}\left[\eta_k(x)\right]\right),$$

where η_k is defined in (31).

- We consider the case $a \geq \frac{2}{5}\alpha$ (in which $\lim_{n\to\infty}(n\gamma_n) > (\alpha-a/2)/(2f(x))$). We set $H_n^2(f(x)) = \Pi_n^2(f(x))\gamma_n^{-1}m_n^{-1/2}$, and note that, since $\left(\gamma_n^{-1}m_n^{-1/2}\right) \in \mathcal{GS}(\alpha-a/2)$, we have

$$\begin{aligned}
&\ln\left(H_n^{-2}(f(x))\right)\\
&=-2\ln\left(\Pi_n(f(x))\right)+\ln\left(\prod_{k=n_0}^{n}\frac{\gamma_{k-1}^{-1}m_{k-1}^{-1/2}}{\gamma_k^{-1}m_k^{-1/2}}\right)+\ln\left(\gamma_{n_0-1}m_{n_0-1}^{1/2}\right)\\
&=-2\sum_{k=n_0}^{n}\ln\left(1-f(x)\gamma_k\right)+\sum_{k=n_0}^{n}\ln\left(1-\frac{\alpha-a/2}{k}+o\left(\frac{1}{k}\right)\right)+\ln\left(\gamma_{n_0-1}m_{n_0-1}^{1/2}\right)\\
&=\sum_{k=n_0}^{n}\left(2f(x)\gamma_k+o\left(\gamma_k\right)\right)-\sum_{k=n_0}^{n}\left(\left(\alpha-a/2\right)\xi\gamma_k+o\left(\gamma_k\right)\right)+\ln\left(\gamma_{n_0-1}m_{n_0-1}^{1/2}\right)
\end{aligned}$$

$$=\left(2f(x)-\left(\alpha-a/2\right)\xi\right)s_n+o\left(s_n\right). \tag{35}$$

Since $2f(x)-(\alpha-a/2)\xi>0$, it follows in particular that $\lim_{n\to\infty}H_n^{-2}(f(x))=\infty$. Moreover, we have $\lim_{n\to\infty}H_n^2(f(x))/H_{n-1}^2(f(x))=1$, and the application of Lemma A.2 together with (32), ensures that

$$\lim_{n\to\infty}H_n^2(f(x))\sum_{k=n_0}^{n}Var\left[\eta_k(x)\right] = \frac{f(x)\psi(x)\mathrm{Var}\left[Y|X=x\right]}{\left(2f(x)-\left(\alpha-a/2\right)\xi\right)}.$$

Moreover, in view of (33), we have

$$\mathbb{E}\left[|\eta_k(x)|^3\right] = O\left(\Pi_n^{-3}(f(x))\gamma_n^3 m_n^{9/4}\right).$$

Now, since $\left(\gamma_n^{-1}m_n^{1/2}\right) \in \mathcal{GS}(\alpha - a/2)$, the application of Lemma A.2 together with (35), ensures that

$$\begin{aligned}
&\frac{1}{n\sqrt{n}}\sum_{k=n_0}^{n}\mathbb{E}\left(|H_n(f(x))\eta_k(x)|^3\right)\\
&= O\left(\frac{H_n^3(f(x))}{n\sqrt{n}}\left(\sum_{k=n_0}^{n}\Pi_k^{-3}(f(x))\gamma_k^3 m_k^{9/4}\right)\right)\\
&= O\left(\frac{\Pi_n^3(f(x))\gamma_n^{-3/2}m_n^{-3/4}}{n\sqrt{n}}\left(\sum_{k=n_0}^{n}\Pi_k^{-3}(f(x))\gamma_k m_k o\left(\left(\gamma_k m_k^{1/2}\right)^{3/2}\right)\right)\right)\\
&= o\left(\frac{m_n}{n\sqrt{n}}\right).
\end{aligned}$$

The application of Theorem 1 in Mokkadem and Pelletier [19] then ensures that, with probability one, the sequence

$$\left(\frac{H_n(f(x))B_n(x)}{\sqrt{2\ln\ln\left(H_n^{-2}(f(x))\right)}}\right) = \left(\frac{\sqrt{\gamma_n^{-1}m_n^{1/2}}\left(T_n(x) - \mathbb{E}\left[T_n(x)\right]\right)}{\sqrt{2\ln\ln\left(H_n^{-2}(f(x))\right)}}\right)$$

is relatively compact and its limit set is the interval

$$\left[-\sqrt{\frac{2f(x)\mathrm{Var}\left[Y|X=x\right]\psi(x)}{4f(x)-(2\alpha-a)\xi}}, \sqrt{\frac{2f(x)\mathrm{Var}\left[Y|X=x\right]\psi(x)}{4f(x)-(2\alpha-a)\xi}}\right]. \tag{36}$$

In view (35), we have $\lim_{n\to\infty}\ln\ln\left(H_n^{-2}(f(x))\right)/\ln(s_n) = 1$. It, follows that, with probability one, the sequence $\left(\sqrt{\gamma_n^{-1}m_n^{1/2}}\left(T_n(x) - \mathbb{E}\left[T_n(x)\right]\right)/\sqrt{2\ln(s_n)}\right)$ is relatively compact and its limit set is the interval given in (36). The application of (25) concludes the proof of Lemma A.5 in the cases $a \geq 2\alpha/5$.

- We consider the case $a < \frac{2}{5}\alpha$ (in which $\lim_{n\to\infty}(n\gamma_n) > a/f(x)$). We set $H_n^{-2}(f(x)) = \Pi_n^{-2}(f(x))m_n^{-2}\left(\ln\ln\left(\Pi_n^{-2}(f(x))m_n^{-2}\right)\right)^{-1}$, and note that, since $\left(m_n^2\right) \in \mathcal{GS}(2a)$, we have

$$\begin{aligned}
&\ln\left(\Pi_n^{-2}(f(x))m_n^{-2}\right)\\
&= -2\ln\left(\Pi_n(f(x))\right) + \ln\left(\prod_{k=n_0}^{n}\frac{m_{k-1}^2}{m_k^2}\right) + \ln\left(m_{n_0-1}^{-2}\right)\\
&= -2\sum_{k=n_0}^{n}\ln\left(1 - f(x)\gamma_k\right) + \sum_{k=n_0}^{n}\ln\left(1 - \frac{2a}{k} + o\left(\frac{1}{k}\right)\right) + \ln\left(\gamma_{n_0-1}m_{n_0-1}^{1/2}\right)\\
&= \sum_{k=n_0}^{n}\left(2f(x)\gamma_k + o\left(\gamma_k\right)\right) - \sum_{k=n_0}^{n}\left(2a\xi\gamma_k + o\left(\gamma_k\right)\right) + \ln\left(\gamma_{n_0-1}m_{n_0-1}^{1/2}\right)\\
&= \left(2f(x) - 2a\xi\right)s_n + o\left(s_n\right).
\end{aligned} \tag{37}$$

Since $2f(x) - 2a\xi > 0$, it follows in particular that $\lim_{n\to\infty}\Pi_n^{-2}(f(x))m_n^{-2} = \infty$ and thus $\lim_{n\to\infty}H_n^{-2}(f(x)) = \infty$. Moreover, we have $\lim_{n\to\infty}H_n^2(f(x))/H_{n-1}^2(f(x)) = 1$. Now, set $\varepsilon \in (0, \alpha - 5a/2)$ such that $\lim_{n\to\infty}(n\gamma_n) > a/f(x) + \varepsilon/2$. Then, the application of Lemma A.2 together with (32), ensures that

$$\begin{aligned}
&H_n^2(f(x)) \sum_{k=n_0}^{n} Var\left[\eta_k(x)\right] \\
&\quad = O\left(\Pi_n^2(f(x)) m_n^2 \left(\ln\ln\left(\Pi_n^{-2}(f(x)) m_n^{-2}\right)\right) \sum_{k=n_0}^{n} \Pi_k^{-2}(f(x)) \gamma_k m_k^{1/2}\right) \\
&\quad = O\left(\Pi_n^2(f(x)) m_n^2 \left(\ln\ln\left(\Pi_n^{-2}(f(x)) m_n^{-2}\right)\right) \sum_{k=n_0}^{n} \Pi_k^{-2}(f(x)) \gamma_k o\left(m_k^{-2} k^{-\varepsilon}\right)\right) \\
&\quad = o(1).
\end{aligned}$$

Moreover, in view of (33)

$$\mathbb{E}\left[|\eta_k(x)|^3\right] = O\left(\Pi_n^{-3}(f(x)) \gamma_n^3 m_n^{9/4}\right),$$

and thus in view of (37), we get

$$\begin{aligned}
&\frac{1}{n\sqrt{n}} \sum_{k=n_0}^{n} \mathbb{E}\left(|H_n(f(x))\eta_k(x)|^3\right) \\
= \quad & O\left(\frac{H_n^3(f(x))}{n\sqrt{n}} \left(\ln\ln\left(\Pi_n^{-2}(f(x)) m_n^{-2}\right)\right)^{3/2} \left(\sum_{k=n_0}^{n} \Pi_k^{-3}(f(x)) \gamma_k^3 m_k^{9/4}\right)\right) \\
= \quad & O\left(\frac{\Pi_n^3(f(x)) m_n^3}{n\sqrt{n}} \left(\ln\ln\left(\Pi_n^{-2}(f(x)) m_n^{-2}\right)\right)^{3/2} \left(\sum_{k=n_0}^{n} \Pi_k^{-3}(f(x)) \gamma_k o\left(m_k^{-3}\right)\right)\right) \\
= \quad & O\left(\frac{\left(\ln\ln\left(\Pi_n^{-2}(f(x)) m_n^{-2}\right)\right)^{3/2}}{n\sqrt{n}}\right) \\
= \quad & o\left(\left[\ln\left(H_n^{-2}(f(x))\right)\right]^{-1}\right).
\end{aligned}$$

The application of Theorem 1 in Mokkadem and Pelletier [19] then ensures that, with probability one

$$\begin{aligned}
&\lim_{n\to\infty} \frac{H_n(f(x)) B_n(x)}{\sqrt{2\ln\ln\left(H_n^{-2}(f(x))\right)}} \\
&\qquad = \lim_{n\to\infty} m_n^{-1} \frac{\sqrt{\ln\ln\left(\Pi_n^{-2}(f(x)) m_n^{-2}\right)}}{\sqrt{2\ln\ln\left(H_n^{-2}(f(x))\right)}} \left(T_n(x) - \mathbb{E}\left[T_n(x)\right]\right) = 0.
\end{aligned}$$

Noting that (37) ensures that

$$\lim_{n\to\infty} \ln\ln\left(H_n^{-2}(f(x))\right) / \ln\ln\left(\Pi_n^{-2}(f(x)) m_n^{-2}\right) = 1,$$

then,

$$\lim_{n\to\infty} m_n^{-1}\left(T_n(x) - \mathbb{E}\left[T_n(x)\right]\right) = 0 \quad a.s.$$

and Lemma A.5 in the case when $a < 2\alpha/5$ follows from (25).

A.4.5. *MISE of ρ_n.* The following Lemma give the $MISE$ of ρ_n.

Lemma A.6. *(MISE of ρ_n)*
Let Assumptions $(A1)-(A''4)$ hold, we have

$$MISE\left(\rho_n\right) = K_1 \mathbb{1}_{\left\{a\in\left(\frac{1-\alpha}{4},\frac{2}{5}\alpha\right]\right\}} m_n^{-2} + K_2 \mathbb{1}_{\left\{a\in\left[\frac{2}{5}\alpha,\frac{2}{3}\alpha\right)\right\}} \gamma_n m_n^{1/2} + o\left(m_n^{-2}+\gamma_n m_n^{1/2}\right).$$

A.4.6. *Proof of Lemma A.6.* We have

$$MISE\left(\rho_n\right) = \int_0^1 Bias^2\left[\rho_n(x)\right]dx + \int_0^1 \mathrm{V}ar\left[\rho_n(x)\right]dx.$$

- In the case when $a\in\left(\frac{1-\alpha}{4},\frac{2}{5}\alpha\right]$, we use the fact that for all $x\in(0,1)$,

$$\lim_{n\to\infty}\left(n\gamma_n\right) > a/\varphi > a/f(x),$$

we have

$$\begin{aligned}\int_0^1 Bias^2\left[\rho_n(x)\right]dx &= \int_0^1\left[m_n^{-1}\frac{f(x)}{f(x)-a\xi}\Delta(x)+o\left(m_n^{-1}\right)\right]^2 dx\\ &= K_1 m_n^{-2}+o\left(m_n^{-2}\right).\end{aligned} \tag{38}$$

- In the case when $a\in\left(\frac{2}{5}\alpha,\frac{2}{3}\alpha\right)$, we have $m_n^{-1}=o\left(\gamma_n m_n^{1/2}\right)$, then Lemma A.2 gives

$$\begin{aligned}\int_0^1 Bias^2\left[\rho_n(x)\right]dx &= \int_0^1 \Pi_n(f(x))\sum_{k=n_0}^{n}\Pi_k^{-1}(f(x))\gamma_k o\left(\gamma_k m_k^{1/2}\right)dx\\ &= o\left(\gamma_n m_n^{1/2}\right).\end{aligned} \tag{39}$$

On the other hand, we note that

$$\begin{aligned}\int_0^1 \mathrm{V}ar\left[\rho_n(x)\right]dx &= \int_0^1 \mathrm{V}ar\left[T_n(x)\right]dx\\ &= \int_0^1\sum_{k=n_0}^{n}U_{k,n}^2(f(x))\gamma_k^2\Bigg\{\mathrm{V}ar\left[W_k(x)\right]+r^2(x)\mathrm{V}ar\left[Z_k(x)\right]\\ &\qquad -2r(x)\mathrm{Cov}\left(W_k(x),Z_k(x)\right)\Bigg\}dx.\end{aligned}$$

Using the same argument as in the proof of Proposition 3.2, we obtain

- In the case when $a\in\left[\frac{2}{5}\alpha,\frac{2}{3}\alpha\right)$, since, for all $x\in(0,1)$, $\lim_{n\to\infty}(n\gamma_n) > (2\alpha-a)/(4\varphi) > (2\alpha-a)/(4f(x))$, Lemma A.2 gives

$$\int_0^1 \mathrm{V}ar\left[\rho_n(x)\right]dx = K_2\gamma_n^2 m_n^{1/2}+o\left(\gamma_n m_n^{1/2}\right). \tag{40}$$

- In the case when $a\in\left(\frac{1-\alpha}{4},\frac{2}{5}\alpha\right)$, we have $\gamma_n m_n^{1/2}=o\left(m_n^{-2}\right)$ and Lemma A.2 gives

$$\begin{aligned}\int_0^1 \mathrm{V}ar\left[\rho_n(x)\right]dx &= \int_0^1 \Pi_n^2(f(x))\sum_{k=n_0}^{n}\Pi_k^{-2}(f(x))o\left(m_k^{-2}\right)dx\\ &= o\left(m_n^{-2}\right).\end{aligned} \tag{41}$$

Then, Part 1 of Lemma A.6 follows from the combination of (38) and (41), Part 2 from that of (38) and (40) and Part 3 from (39) and (40).

A.4.7. *Asymptotic behaviour of r_n.* We show in this section how to deduce the asymptotic behaviour of r_n from that of ρ_n. To do so, we set

$$\delta_n(x) = r_n(x) - \rho_n(x),$$

and we prove that δ_n is negligible in front of ρ_n. Note that, in view of (5) and (6), and since $\rho_{n_0-1} = r_{n_0-1}$, we have, for $n \geq n_0$

$$\begin{aligned} \delta_n(x) &= (1-\gamma_n f(x))\,\delta_{n-1}(x) + \gamma_n\,(f(x) - Z_n(x))\,(r_{n-1}(x) - r(x)) \\ &= \sum_{k=n_0}^{n} U_{k,n}(f(x))\gamma_k\,(f(x) - Z_k(x))\,(r_{k-1}(x) - r(x))\,. \end{aligned} \tag{42}$$

To obtain an upper bound of δ_n, we must have an upper bound of $r_n - r$. To do so, we use the following property given by Mokkadem et al. [21]
$(\mathcal{P})$: if $(r_n - r)$ is known to be bounded almost surely by a sequence (w_n), then it can be shown that (δ_n) is bounded almost surely by a sequence (w'_n) such that $\lim_{n\to\infty} w'_n w_n = 0$, which may allow to upper bound $r_n - r$ by a sequence smaller than (w_n).

We thus proceed as follows. We first establish an upper bound of $(r_n - r)$. Then, we apply the Property $(\mathcal{P})$ several times until we obtain an upper bound which allows to prove that δ_n is negligible in front of ρ_n.

The proof of the results given in Section 4 relies on the repeated application of the following lemma.

Lemma A.7. *Let Assumptions (A1) − (A3) hold, and assume that there exists $(w_n) \in \mathcal{GS}(w^*)$ such that $|r_n(x) - r(x)| = O(w_n)$ a.s. For $x \in (0,1)$, we have*

(1) *If the sequence $(n\gamma_n)$ is bounded, if*

$$\lim_{n\to\infty}(n\gamma_n) > \min\left\{a/f(x), (2\alpha - a)/(4f(x))\right\},$$

and if $w^* > 0$, then, for all $\delta > 0$,

$$|\delta_n(x)| = O\left(\lambda_n w_n (\ln n)^{\frac{1+\delta}{2}}\right) + o(\lambda_n) \quad a.s.$$

(2) *If $\lim_{n\to\infty}(n\gamma_n) = \infty$, then, for all $\delta > 0$,*

$$|\delta_n(x)| = O\left(\lambda_n w_n \left(n^{1+\delta}\gamma_n\right)^{\frac{1+\delta}{2}}\right) \quad a.s.$$

For $x \in \{0,1\}$, we have

(1) *If the sequence $(n\gamma_n)$ is bounded, if*

$$\lim_{n\to\infty}(n\gamma_n) > \min\left\{a/f(x), (\alpha - a)/(2f(x))\right\},$$

and if $w^* > 0$, then, for all $\delta > 0$,

$$|\delta_n(x)| = O\left(\beta_n w_n (\ln n)^{\frac{1+\delta}{2}}\right) + o(\beta_n) \quad a.s.$$

(2) *If $\lim_{n\to\infty}(n\gamma_n) = \infty$, then, for all $\delta > 0$,*

$$|\delta_n(x)| = O\left(\beta_n w_n \left(n^{1+\delta}\gamma_n\right)^{\frac{1+\delta}{2}}\right) \quad a.s.$$

A.4.8. *Proof of Lemma A.7.* To prove Lemma A.7, we use the following decomposition, which can be deduced from (42)

$$\delta_n(x) = \delta_n^{(1)}(x) + \delta_n^{(2)}(x),$$

with

$$\delta_n^{(1)}(x) = \sum_{k=n_0}^{n} U_{k,n}(f(x))\gamma_k\,(\mathbb{E}\left[Z_k(x)\right] - Z_k(x))\,(r_{k-1}(x) - r(x))\,,$$

$$\delta_n^{(2)}(x) = \sum_{k=n_0}^{n} U_{k,n}(f(x))\gamma_k \left(f(x) - \mathbb{E}\left[Z_k(x)\right]\right)\left(r_{k-1}(x) - r(x)\right).$$

Moreover, we have for $x \in [0,1]$

$$\mathbb{E}\left[Z_k(x)\right] = f(x) + \Delta_1(x)m_k^{-1} + o\left(m_k^{-1}\right).$$

Then, the application of Lemma A.2, ensures that

$$\begin{aligned}\left|\delta_n^{(2)}(x)\right| &= O\left(\Pi_n(f(x))\sum_{k=n_0}^{n}\Pi_k^{-1}(f(x))\gamma_k m_k^{-1}w_k\right) \quad a.s.\\ &= O\left(\Pi_n(f(x))\sum_{k=n_0}^{n}\Pi_k^{-1}(f(x))\gamma_k O\left(\lambda_k\right)w_k\right) \quad a.s.\\ &= O\left(\lambda_n w_n\right) \quad a.s.\end{aligned}$$

Further, we set

$$\begin{aligned}\varepsilon_k(x) &= \mathbb{E}\left[Z_k(x)\right] - Z_k(x),\\ G_k(x) &= r_k(x) - r(x),\\ Q_n(x) &= \sum_{k=n_0}^{n}\Pi_k^{-1}(f(x))\gamma_k\varepsilon_k(x)G_{k-1}(x),\end{aligned}$$

and $\mathcal{F}_k = \sigma\left(\left(X_1, Y_1\right), \ldots, \left(X_k, Y_k\right)\right)$. Since, we have

$$\mathrm{V}ar\left[Z_k(x)\right] = f(x)\psi(x)m_k^{1/2} + o\left(m_k^{1/2}\right), \quad \forall x \in (0,1),$$

and of Lemma A.2, the increasing process of the martingale $(Q_n(x))$ satisfies

$$\begin{aligned}< Q >_n (x) &= \sum_{k=n_0}^{n}\mathbb{E}\left[\Pi_k^{-2}(f(x))\gamma_k^2\varepsilon_k^2(x)G_{k-1}^2(x)|\mathcal{F}_{k-1}\right]\\ &= \sum_{k=n_0}^{n}\Pi_k^{-2}(f(x))\gamma_k^2 G_{k-1}^2(x)\mathrm{V}ar\left[Z_k(x)\right]\\ &= O\left(\sum_{k=n_0}^{n}\Pi_k^{-2}(f(x))\gamma_k^2 w_k^2 m_k^{1/2}\right) \quad a.s.\\ &= O\left(\sum_{k=n_0}^{n}\Pi_k^{-2}(f(x))\gamma_k\lambda_k^2 w_k^2\right) \quad a.s.\\ &= O\left(\Pi_n^{-2}(f(x))\lambda_n^2 w_n^2\right) \quad a.s.\end{aligned}$$

- Let us first consider the case when the sequence $(n\gamma_n)$ is bounded. In this case we have $\left(\Pi_n^{-1}(f(x))\right) \in \mathcal{GS}(\xi^{-1}f(x))$, and thus $\ln\left(< Q >_n (x)\right) = O\left(\ln n\right)$ a.s. Theorem 1.3.15 in Duflo [6] then ensures that, for any $\delta > 0$,

$$\begin{aligned}|Q_n(x)| &= o\left(< Q >_n^{\frac{1}{2}} (x)\left(\ln < Q >_n (x)\right)^{\frac{1+\delta}{2}}\right) + O(1) \quad a.s.\\ &= o\left(\Pi_n^{-1}(f(x))\lambda_n w_n\left(\ln n\right)^{\frac{1+\delta}{2}}\right) + O(1) \quad a.s.\end{aligned}$$

It follows that, for any $\delta > 0$,

$$\begin{aligned}\left|\delta_n^{(1)}(x)\right| &= o\left(\lambda_n w_n\left(\ln n\right)^{\frac{1+\delta}{2}}\right) + O\left(\Pi_n(f(x))\right) \quad a.s.\\ &= o\left(\lambda_n w_n\left(\ln n\right)^{\frac{1+\delta}{2}}\right) + o\left(\lambda_n\right) \quad a.s.\end{aligned}$$

which concludes the proof of Lemma A.7 in this case.

- Let us now consider the case $\lim_{n\to\infty}(n\gamma_n) = \infty$. In this case, for all $\delta > 0$, we have

$$\begin{aligned} \ln\left(\Pi_n^{-2}(f(x))\right) &= \sum_{k=n_0}^{n} \ln\left(1-\gamma_k f(x)\right)^{-2} \\ &= \sum_{k=n_0}^{n} \left(2\gamma_k f(x) + o(\gamma_k)\right) \\ &= O\left(\sum_{k=1}^{n} \gamma_k k^{\delta}\right). \end{aligned}$$

Moreover, since $\left(\gamma_n n^{\delta}\right) \in \mathcal{GS}(-(\alpha-\delta))$ with $(\alpha-\delta) < 1$, we have

$$\lim_{n\to\infty} \frac{n\left(\gamma_n n^{\delta}\right)}{\sum_{k=1}^{n} \gamma_k k^{\delta}} = 1-(\alpha-\delta).$$

It follows that $\ln\left(\Pi_n^{-2}(f(x))\right) = O\left(n^{1+\delta}\gamma_n\right)$. The sequence $(\lambda_n w_n)$ being in $\mathcal{GS}(-\lambda^* + w*)$, we deduce that, for all $\delta > 0$, we have

$$\ln\left(< Q >_n (x)\right) = O\left(n^{1+\delta}\gamma_n\right) \quad a.s.$$

Theorem 1.3.15 in Duflo [6] then ensures that, for any $\delta > 0$,

$$\begin{aligned} |Q_n(x)| &= o\left(< Q >_n^{\frac{1}{2}} (x) \left(\ln < Q >_n (x)\right)^{\frac{1+\delta}{2}}\right) + O(1) \quad a.s. \\ &= o\left(\Pi_n^{-1}(f(x))\lambda_n w_n \left(n^{1+\delta}\gamma_n\right)^{\frac{1+\delta}{2}}\right) + O(1) \quad a.s. \end{aligned}$$

The application of Lemma A.2 then ensures that, for any $\delta > 0$,

$$\begin{aligned} \left|\delta_n^{(1)}(x)\right| &= o\left(\lambda_n w_n \left(n^{1+\delta}\gamma_n\right)^{\frac{1+\delta}{2}}\right) + O\left(\Pi_n(f(x))\right) \quad a.s. \\ &= o\left(\lambda_n w_n \left(n^{1+\delta}\gamma_n\right)^{\frac{1+\delta}{2}}\right) \quad a.s. \end{aligned}$$

which concludes the proof of Lemma A.7 for $x \in (0,1)$.

Remark A.6. (1) For $x \in \{0,1\}$, we use the same steps as in the case $x \in (0,1)$ with the sequence (β_n) defined in (24).
(2) To use Lemma A.7, we must establish an upper bound for $r_n(x) - r(x)$. For this purpose, we follow similar steps as Mokkadem et al. [21] to prove Proposition 4.1, Theorem 4.1 and Proposition 4.3 in the case $(n\gamma_n)$ is bounded and in the case $\lim_{n\to\infty}(n\gamma_n) = \infty$ (see the supplementary material).
(3) We use the same idea to prove Proposition 4.2 and Theorem 4.2 which correspond to the case $x = 0, 1$.

Acknowledgements

The authors would like to thank the Editor, the associate Editor and the referee for their very helpful comments, which led to a considerable improvement of the original version of the paper and a more sharply focused presentation. This work benefited from the financial support of the GDR 3477 GeoSto.

References

1. G. J. Babu, A. J. Canty and Y. P. Chaubey, *Application of Bernstein polynomials for smooth estimation of a distribution and density function*, J. Statist. Plann. Inference **105** (2002), 377–392. https://doi.org/10.1016/S0378-3758(01)00265-8
2. G. J. Babu and Y. P. Chaubey, *Smooth estimation of a distribution function and density function on a hypercube using Bernstein polynomials for dependent random vectors*, Statist. Probab. Lett. **76** (2006), 959–969. https://doi.org/10.1016/j.spl.2005.10.031
3. R. Bojanic and E. Seneta, *A unified theory of regularly varying sequences*, Math. Z. **134** (1973), 91–106. https://doi.org/10.1007/BF01214468
4. S. Boukabour and A. Masmoudi, *Semiparametric Bayesian networks for continuous data*, Comm. Statist. Theory Methods (2020). https://doi.org/10.1080/03610926.2020.1738486
5. R. A. Djojosugito and P. L. Speckman, *Boundary bias correction in nonparametric density estimation*, Comm. Statist. Theory Methods **21** (1992), 69–88. https://doi.org/10.1080/03610929208830765
6. M. Duflo, *Random Iterative Models, in Applications of Mathematics*, Springer, Berlin, 1997.
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. T. Gasser and H.-G. Müller, *Kernel estimation of regression functions*, Smoothing Techniques for Curve Estimation, Lecture Notes in Mathematics **757** (1979), 23–68. https://doi.org/10.1007/BFb0098489
9. T. Gasser, H.-G. Müller and V. Mammitzsch, *Kernels for nonparametric curve estimation*, J. Roy. Statist. Soc. Ser. B **47** (1985), 238–252. https://doi.org/10.1111/j.2517-6161.1985.tb01350.x
10. S. Ghosal, *Convergence rates for density estimation with Bernstein polynomials*, Ann. Statist. **29** (2000), 1264–1280. https://doi.org/10.1214/aos/1013203453
11. B. L. Granovsky and H.-G. Müller, *Optimizing kernel methods: a unifying variational principle*, Internat. Statist. Rev. **59** (1991), 373–388. https://doi.org/10.2307/1403693
12. I. Igarashi and Y. Kakizawa, *On improving convergence rate of Bernstein polynomial density estimator*, J. Nonparametr. Stat. **26** (2014), 61–84. https://doi.org/10.1080/10485252.2013.827195
13. A. Jmaei, Y. Slaoui and W. Dellagi, *Recursive distribution estimator defined by stochastic approximation method using Bernstein polynomials*, J. Nonparametr. Stat. **29** (2017), 792–805. https://doi.org/10.1080/10485252.2017.1369538
14. Y. Kakizawa, *Bernstein polynomial probability density estimation*, J. Nonparametr. Stat. **16** (2004), 709–729. https://doi.org/10.1080/1048525042000191486
15. Y. Kakizawa, *A note on generalized Bernstein polynomial density estimators*, Statistical Methodology **8** (2011), 136–153. https://doi.org/10.1016/j.stamet.2010.08.004
16. A. Leblanc, *A bias-reduced approach to density estimation using Bernstein polynomials*, J. Nonparametr. Stat. **22** (2010), 459–475. https://doi.org/10.1080/10485250903318107
17. A. Leblanc, *On estimating distribution function using Bernstein polynomials*, Ann. Inst. Statist. Math., **64** (2012), 919–943. https://doi.org/10.1007/s10463-011-0339-4
18. A. Leblanc, *On the boundary properties of Bernstein polynomial estimators of density and distribution functions*, J. Statist. Plann. Inference **142** (2012), 2762–2778. https://doi.org/10.1016/j.jspi.2012.03.016
19. A. Mokkadem and M. Pelletier, *Compact law of the iterated logarithm for matrix normalized sums of random vectors*, Theory Probab. Appl. **52** (2008), 636–650. https://doi.org/10.1137/S0040585X97983298
20. A. Mokkadem, M. Pelletier and Y. Slaoui, *The stochastic approximation method for the estimation of a multivariate probability density*, J. Statist. Plann. Inference **139** (2009), 2459–2478. https://doi.org/10.1016/j.jspi.2008.11.012
21. A. Mokkadem, M. Pelletier and Y. Slaoui, *Revisiting Révész's stochastic approximation method for the estimation of a regression function*, ALEA. Lat. Am. J. Probab. Math. Stat. **6** (2009), 63–114.
22. H.-G. Müller, *Smooth optimal kernel estimators near endpoints*, Biometrika **78** (1991), 521–530. https://doi.org/10.1093/biomet/78.3.521
23. E. A. Nadaraya, *On Estimating Regression*, *Theory Probab. Appl.* **9** (1964), no. 1, 141–142. https://doi.org/10.1137/1109020
24. M. B. Priestley and M. T. Chao, *Non-parametric function fitting*, J. R. Stat. Soc. Ser. B Stat. Methodol. **34** (1972), 385–392. https://doi.org/10.1111/j.2517-6161.1972.tb00916.x
25. B. L. S. P. Rao, *Estimation of distribution and density functions by generalized Bernstein polynomials*, Indian J. Pure Appl. Math. **36** (2005), 63–88.

26. P. Révész, *Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I*, Studia Sci. Math. Hung. **8** (1973), 391–398.
27. 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
28. H. Robbins and S. Monro, *A Stochastic Approximation Method*, Ann. Math. Stat. **22** (1951), 400–407. https://doi.org/10.1214/aoms/1177729586
29. Y. Slaoui, *Large and moderate principles for recursive kernel density estimators defined by stochastic approximation method*, Serdica Math. J. **39** (2013), 53–82.
30. Y. Slaoui, *Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method*, J. Probab. Stat. **2014** (2014), ID 739640. https://doi.org/10.1155/2014/739640
31. Y. Slaoui, *The stochastic approximation method for the estimation of a distribution function*, Math. Methods Statist. **23** (2014), 306–325. https://doi.org/10.3103/S1066530714040048
32. Y. Slaoui, *Moderate deviation principles for recursive regression estimators defined by stochastic approximation method*, Int. J. Math. Stat. **16** (2015), 51–60.
33. Y. Slaoui, *Large and moderate deviation principles for recursive regression estimators defined by stochastic approximation method*, Serdica Math. J. **41** (2015), 307–328.
34. 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
35. 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
36. Y. Slaoui and A. Jmaei, *Recursive density estimators based on Robbins-Monro's scheme and using Bernstein polynomials*, Stat. Interface **12** (2019), 439–455. https://doi.org/10.4310/19-SII561
37. Y. Slaoui, *Wild Bootstrap Bandwidth Selection of Recursive Nonparametric Relative Regression for Independent Functional Data*, J. Multivariate Anal. **173** (2019), 494–511. https://doi.org/10.1016/j.jmva.2019.04.009
38. Y. Slaoui, *Recursive nonparametric regression estimation for independent functional data*, Statist. Sinica **30** (2020), 417–437.
39. A. Tenbusch, *Two-dimensional Bernstein polynomial density estimation*, Metrika **41** (1994), 233–253. https://doi.org/10.1007/BF01895321
40. A. Tenbusch, *Nonparametric Curve Estimation with Bernstein estimates*, Metrika (1997) **45**, 1–30. https://doi.org/10.1007/BF02717090
41. A. B. Tsybakov, *Recurrent estimation of the mode of a multidimensional distribution*, Probl. Inf. Transm. **8** (1990), 119–126.
42. R. A. Vitale, *A Bernstein polynomial approach to density function estimation*, Stat. Inference Related Topics **2** (1975), 87–99. https://doi.org/10.1016/B978-0-12-568002-8.50011-2
43. G. S. Watson, *Smooth regression analysis*, Sankhya A. Mathematical Statistics and Probability **26** (1975), 359–372.

Université de Poitiers, Laboratoire de Mathématiques et applications
Current address: UMR 7348, 11 Boulevard Marie et Pierre Curie, 86962 Futuroscope Chasseneuil, France
E-mail address: yousri.slaoui@math.univ-poitiers.fr

Faculté des sciences de Bizerte, 7021 Jarzouna, Tunisie
E-mail address: asma.jmaei@math.univ-poitiers.fr