STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **10**, September 2022, pp 1021–1043.
Published online in International Academic Press (www.IAPress.org)

Non-parametric Multivariate Kernel Regression Estimation to Describe Cognitive Processes and Mental Representations

Sahar Slama [1,2], Yousri Slaoui [2,*], Gwendoline Le Du [3], Cyril Perret [2,4]

[1] *Laboratoire de Mathématiques Modélisation Déterministe et Aléatoire,*
Sousse University and ESST Hammam Sousse, Tunisia
[2] *Laboratoire de Mathématiques et Applications, University of Poitiers, Poitiers, France*
[3] *UMR-S INSERM 1237- Physiopathology & Imaging of Neurological Disorders (PhIND), Caen Normandie University, France*
[4] *Centre de Recherches sur la Cognition et l'Apprentissage (CeRCa), Poitiers University, France*

Abstract
In this research paper, we set forward a non-parametric multivariate recursive kernel regression estimator under missing data using the propensity score approach in order to describe writing word production. Our main objective is to explore cognitive processes and mental representations mobilized when a human being prepares to write a word according to the idea developed in [21]. We investigate the asymptotic properties of the proposed recursive estimator and compare them to the well known Nadaraya-Watson's regression estimator. We calculate the bias and the variance of the proposed estimator which depend on the choice of some parameters such as the stepsize and the bandwidth. We examine some data-driven procedures to select these parameters. Thus, we demonstrate that, under some optimal choices of these parameters, the MSE (Mean Squared Error) of the proposed estimator can be smaller than the one obtained by using Nadaraya Watson's regression estimator. The elaborated estimator is then applied to the behavioral data to classify some participants in groups. This classification may stand for a departure point to tackle written behavior variations.

Keywords regression function estimation, stochastic approximation algorithm, classification and cluster analysis, plug-in method, handwritten and cognitive psychology, propensity score matching.

AMS 2010 subject classifications 62G08, 62L20, 62H30, 65D10, 91E10, 93A30.

DOI: 10.19139/soic-2310-5070-1507

1. Introduction

Research on the handwritten word production aims to describe the cognitive processes and mental representations mobilized when a human being prepares to hand-write a word from an idea (see [21]). The most frequently used method to explore this issue relies on relating a behavioral variable, reaction time and a set of factors aiming at predicting different cognitive treatments (e.g., [3], [20]). It is possible to imagine some variations in the cognitive treatments performed by participants. This could result in variations in the relationship between the behavioral variables and the explanatory factors. The intrinsic target lies in being able to group participants with similar degrees of variation.

In order to achieve our purpose, we resort to regression analysis, which corresponds to the study of how a response variable depends on one or more predictors. In fact, it is a reliable method for identifying which variables have impact on a topic of interest. The process of performing a regression allows us to confidently determine which

[*] Correspondence to: Yousri Slaoui (Email: yousri.slaoui@math.univ-poitiers). Laboratoire de Mathématiques et Applications, University of Poitiers, Poitiers, France

ISSN 2310-5070 (online) ISSN 2311-004X (print)

factors matter most, which factors can be ignored, and how these factors influence each other. Regression problems can be usefully summarized using non-parametric regression methods which represent a category of regression analysis in which the predictor does not take a predetermined form but is constructed according to information derived from the data. Since we ignore the behavior of our data, and we don't have the normality (see [14], [13] and [12]), we resorted to non-parametric approach. In this paper, we shall focus on Kernel regression which is a non-parametric technique in statistics to estimate the conditional expectation of a random variable. The main objective is to find a non-linear relation between a pair of random variables X and T.

In addition to the non-parametric fact, we introduce the recursive approach of estimation using stochastic approximation method. The use of stochastic approximation algorithms for regression function estimation was first introduced by [24] and [11] and then extended with [18,36], [23], [8], [16] for univariate framework. Subsequently, a generalization in a multivariate case of this estimator was carried out by [17].

More recently, [29] developed a new recursive kernel estimator to estimate a regression function. The missing data question is a former problem in psychology, which can contaminate the results and disrupt them. In order to settle the missing data problem, multiple 'naive' methods have been incorporated to solve this problem, such as the replacement of the missing value by the mean/median or complete outliers detection and treatment (see [5]). Recently, [30] used the propensity score probability technique and constructed an estimator of the density function under missing data. Our central focus resides in building up a multivariate kernel regression estimator under missing data.

Presentation of the method

Consider a couple (X, T) of random variables defined in $\mathbb{R}^d \times \mathbb{R}$ to be independent random vectors identically distributed as (X, T) with joint density function $g(x, t)$ and let f denote the probability density of X. Assuming that $T_1, \ldots, T_n$ are subjects to missing data, the observed random variables are then Y_i and δ_i, where

$$\delta_i = \mathbb{1}_{\{T_i \text{ is observed}\}} \text{ and } Y_i = T_i * \delta_i, \forall i \in \{1, \ldots, n\}.$$

Accordingly, when some T_i are missing, we introduce the propensity score, a probability elaborated by Rosenbaum and Rubin (1983) [25] and defined as followed

$$\psi_i = \mathbb{P}[\delta_i = 1 | T_i], \forall i \in \{1, \ldots, n\}.$$

In the remainder, Y is considered as the response variable of interest and X its associated regressor vector variable. Our basic purpose in this paper is to propose a recursive estimator to estimate recursively the regression function $p(x) = \mathbb{E}[T|X = x]$ under censoring data. Our aim then resides in building up a stochastic algorithm, which approaches the regression function $m : x \longmapsto \mathbb{E}[T|X = x]f(x) = \int_{\mathbb{R}} yg(x, y)dy$ at a given vector x. For this reason, we define an algorithm of search of the zero function $\phi : y \longmapsto m(x) - y$. We therefore proceed as follows: we fix $m_0(x) \in \mathbb{R}$, and then we set for all $n \geq 1, m_n(x) = m_{n-1}(x) + \beta_n U_n(x)$, where (β_n) is a positive sequence of real numbers decreasing towards zero and $U_n(x)$ is an observation of the function ϕ at the point $m_{n-1}(x)$. In order to construct $U_n(x)$, we adopt the approach considered first by [23], [34] and more recently by [31] and we introduce a multivariate kernel $\mathbf{K}$, which is a function satisfying $\int_{\mathbb{R}^d} \mathbf{K}(t)dt = 1$, and a bandwidth (h_n), which is a sequence of positive real numbers that tends to zero. By assuming $U_n(x) = Y_n \psi_n^{-1} h_n^{-d} \mathbf{K}\left(\frac{x - X_n}{h_n}\right) - m_{n-1}(x)$, the stochastic approximation algorithm that we consider to estimate recursively the regression function m at a vector x can be expressed as follows :

$$m_n(x) = (1 - \beta_n)m_{n-1}(x) + \beta_n Y_n \psi_n^{-1} h_n^{-d} \mathbf{K}\left(\frac{x - X_n}{h_n}\right). \tag{1}$$

Throughout this paper, we consider that $m_0(x) = 0$ and we set $Q_n = \prod_{j=1}^{n}(1-\beta_j)$. It follows that

$$m_n(x) = Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k\, Y_k \psi_k^{-1} h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right). \tag{2}$$

Moreover, we use the recursive multivariate probability density estimator of the density function f defined in [15] which was constructed with the same tools of stochastic approximation algorithm and under the condition that $f_0(x) = 0$, it follows that

$$f_n(x) = Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k\, h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right). \tag{3}$$

In this paper, we consider the following recursive estimator of the regression function $p : x \longmapsto \mathbb{E}[T|X = x]$ at the vector x

$$p_n(x) = \begin{cases} \dfrac{m_n(x)}{f_n(x)} & \text{if} \quad f_n(x) \neq 0 \\ 0 & \text{if} \quad f_n(x) = 0 \end{cases}. \tag{4}$$

We explore the asymptotic properties of our proposed multivariate recursive kernel regression estimator. Afterwards, we compare our proposed estimator to the multivariate non-recursive Nadaraya-Watson's regression estimator (see [18] and [36]) $\widetilde{p}_n$ indicated by

$$\widetilde{p}_n(x) = \begin{cases} \dfrac{\widetilde{m}_n(x)}{\widetilde{f}_n(x)} & \text{if} \quad \widetilde{f}_n(x) \neq 0 \\ 0 & \text{if} \quad \widetilde{f}_n(x) = 0. \end{cases}, \tag{5}$$

with

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

2. Notations and assumptions

Throughout this paper, we invest the following useful notations:

$$\xi_\beta = \lim_{n\to+\infty}(n\beta_n)^{-1}, \quad \psi = \lim_{n\to+\infty}\psi_n, \quad R(\mathbf{K}) = \int_{\mathbb{R}^d}\mathbf{K}^2(z)\,dz, \quad \mu_i(\mathbf{K}) = \int_{\mathbb{R}^d} z_i^2\mathbf{K}(z)dz,$$

$$I_1 = \int_{\mathbb{R}^d}\left(\sum_{j=1}^{d}\mu_j(\mathbf{K})m_{jj}^{(2)}(x)\right)^2 f(x)dx, \quad I_2 = \int_{\mathbb{R}^d}\left(\sum_{j=1}^{d}\mu_j(\mathbf{K})m_{jj}^{(2)}(x)\right)\left(\sum_{j=1}^{d}\mu_j(\mathbf{K})f_{jj}^{(2)}(x)\right)p(x)f(x)dx,$$

$$I_3 = \int_{\mathbb{R}^d}\left(\sum_{j=1}^{d}\mu_j(\mathbf{K})f_{jj}^{(2)}(x)\right)^2 p^2(x)f(x)dx, \quad I_4 = \int_{\mathbb{R}^d}\mathbb{E}[T^2|X=x]f^2(x)dx, \quad I_5 = \int_{\mathbb{R}^d}p^2(x)f^2(x)dx.$$

Before stating our assumptions, let us recall the definition of class of regularly varying sequences introduced by Galambos and Seneta in [7].

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

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

In what follows, we exhibit a lemma that will be widely used for the study of our estimator p_n. The proof of this lemma was introduced in [15].

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

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

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

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

The assumptions upon which we shall rely are the following.

Assumptions:

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

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

(A_2) (i) $(\beta_n)_{n\geq 1} \in \mathcal{GS}(-\beta)$, with $\beta \in \left(\frac{1}{2}, 1\right]$.

(ii) $(h_n)_{n\geq 1} \in \mathcal{GS}(-a)$, with $a \in (0, 1)$.

(iii) $\lim\limits_{n\to+\infty} (n\beta_n) \in \left(\min\{2a, \frac{\beta - ad}{2}\}, \infty\right]$.

(A_3) (i) f is bounded, twice differentiable and $\forall i, j \in \{1, \ldots, d\}, f_{ij}^{(2)} := \dfrac{\partial^2 f}{\partial x_i \partial x_j}$ is bounded.

(ii) m is bounded, twice differentiable and $\forall i, j \in \{1, \ldots, d\}, m_{ij}^{(2)} := \dfrac{\partial^2 m}{\partial x_i \partial x_j}$ is bounded.

(iii) $g(s, t)$ is twice continuously differentiable with respect to s.

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

(v) For $q \in \{0, 1\}$, $i, j \in \{1, \ldots, d\}$, $\displaystyle\int_{\mathbb{R}} |t|^q \left| \frac{\partial g}{\partial x_i}(x, t) \right| dt < \infty$, and $s \longmapsto \displaystyle\int_{\mathbb{R}} t^q \frac{\partial^2 g}{\partial s_i \partial s_j}(s, t)dt$ is a bounded function continuous at $s = x$.

3. Main results

In order to investigate the asymptotic properties of our estimator p_n, we first need to introduce the following two propositions which provide the bias and the variance of m_n as well as those of f_n.

3.1. Bias and variance of f_n

Proposition 1
Let assumptions $(A_1)-(A_3)$ hold and assume that, for all $i,j\in\{1,\dots,d\}$, $f_{ij}^{(2)}$ is continuous at x. We therefore get

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

$$\mathbb{E}[f_n(x)]-f(x)=\frac{h_n^2}{2(1-2a\xi_\beta)}\sum_{j=1}^{d}\mu_j(\mathbf{K})f_{jj}^{(2)}(x)+o\left(h_n^2\right). \tag{6}$$

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

$$\mathbb{E}[f_n(x)]-f(x)=o\left(\sqrt{\beta_n h_n^{-d}}\right). \tag{7}$$

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

$$Var[f_n(x)]=o\left(h_n^4\right). \tag{8}$$

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

$$Var[f_n(x)]=\frac{\beta_n}{h_n^d}\frac{1}{2-(\beta-ad)\xi_\beta}f(x)R(\mathbf{K})+o\left(\beta_n h_n^{-d}\right). \tag{9}$$

3.2. Bias and variance of m_n

Proposition 2
Let assumptions $(A_1)-(A_3)$ hold and assume that, for all $i,j\in\{1,\dots,d\}$, $m_{ij}^{(2)}$ is continuous at x, we hence obtain

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

$$\mathbb{E}[m_n(x)]-m\left(x\right)=\frac{h_n^2}{2(1-2a\xi_\beta)}\sum_{j=1}^{d}\mu_j(\mathbf{K})m_{jj}^{(2)}(x)+o\left(h_n^2\right). \tag{10}$$

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

$$\mathbb{E}[m_n(x)]-m\left(x\right)=o\left(\sqrt{\beta_n h_n^{-d}}\right). \tag{11}$$

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

$$Var[m_n(x)]=o\left(h_n^4\right). \tag{12}$$

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

$$Var[m_n(x)]=\frac{\beta_n}{h_n^d}\psi_n^{-1}\frac{\mathbb{E}[T^2|X=x]}{2-(\beta-ad)\xi_\beta}f(x)R(\mathbf{K})+o\left(\beta_n h_n^{-d}\right). \tag{13}$$

Proof
Throughout this proof, we use the following notations:

$$\mathcal{Z}_n(x)=h_n^{-d}Y_n\psi_n^{-1}\mathbf{K}\left(\frac{x-X_n}{h_n}\right)\qquad\text{and}\qquad\mathcal{W}_n(x)=h_n^{-d}\mathbf{K}\left(\frac{x-X_n}{h_n}\right).$$

We have

$$\begin{aligned}
m_n(x) - m\,(x) &= (1-\beta_n)m_{n-1}(x) + \beta_n \mathcal{Z}_n(x) - m\,(x) \\
&= (1-\beta_n)[m_{n-1}(x) - m\,(x)] + \beta_n[\mathcal{Z}_n(x) - m\,(x)] \\
&= \prod_{i=1}^{n}(1-\beta_i)[m_0(x) - m\,(x)] + \sum_{k=1}^{n-1}\prod_{i=k+1}^{n}(1-\beta_i)\beta_k(\mathcal{Z}_k(x) - m\,(x)) + \beta_n(\mathcal{Z}_n(x) - m\,(x)) \\
&= Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k(\mathcal{Z}_k(x) - m\,(x)) + Q_n[m_0(x) - m\,(x)].
\end{aligned}$$

It follows that,

$$\mathbb{E}[m_n(x)] - m\,(x) = Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k(\mathbb{E}[\mathcal{Z}_k(x)] - m\,(x)) + Q_n[m_0(x) - m\,(x)]. \tag{14}$$

Moreover, we have

$$\begin{aligned}
\mathbb{E}[\mathcal{Z}_k(x)] &= h_k^{-d}\psi_k^{-1}\mathbb{E}\left[Y_k \mathbf{K}\left(\frac{x - X_k}{h_k}\right)\right] \\
&= h_k^{-d}\psi_k^{-1}\mathbb{E}\left[T_k \mathbb{1}_{\{T_k = Y_k\}} \mathbf{K}\left(\frac{x - X_k}{h_k}\right)\right] \\
&= h_k^{-d}\psi_k^{-1}\mathbb{E}[\mathbb{1}_{\{T_k = Y_k\}}] \int_{\mathbb{R}^d} \mathbb{E}[T|X=y]\mathbf{K}\left(\frac{x-y}{h_k}\right) f(y)dy \\
&= h_k^{-d}\int_{\mathbb{R}^d} \mathbf{K}\left(\frac{x-y}{h_k}\right) m\,(y)\,dy.
\end{aligned}$$

Since we have $\int_{\mathbb{R}^d} \mathbf{K}(z)dz = 1$, we infer that

$$\begin{aligned}
\mathbb{E}[\mathcal{Z}_k(x)] - m\,(x) &= \int_{\mathbb{R}^d} h_k^{-d}\mathbf{K}\left(\frac{x-y}{h_k}\right) m\,(y)\,dy - \int_{\mathbb{R}^d} \mathbf{K}\,(y)\,m\,(x)\,dy \\
&= \int_{\mathbb{R}^d} \mathbf{K}\,(z)\,[m\,(x - zh_k) - m\,(x)]\,dz.
\end{aligned}$$

A Taylor expansion of m around x ensures that

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

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

Owing to the fact that $m_{ij}^{(2)}$ is bounded and continuous at x for all $i, j \in \{1, \dots, d\}$, we thus get

$$\begin{aligned}
\mathbb{E}[m_n(x)] - m\,(x) &= Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k (\mathbb{E}[\mathcal{Z}_k(x)] - m\,(x)) + Q_n[m_0(x) - m\,(x)] \\
&= Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k \left(\frac{h_k^2}{2} \sum_{j=1}^{d} \mu_j(\mathbf{K}) m_{jj}^{(2)}(x) + h_k^2 \eta_k(x) \right) + Q_n[m_0(x) - m\,(x)] \\
&= \frac{1}{2} \sum_{j=1}^{d} \mu_j(\mathbf{K}) m_{jj}^{(2)}(x) Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k h_k^2 + Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k h_k^2 \eta_k(x) + Q_n[m_0(x) - m\,(x)].
\end{aligned}$$

For the case $a \leq \beta/(d+4)$, we have $\lim_{n\to\infty} (n\beta_n) > 2a$ and then $1 - 2a\xi_\beta > 0$. The application of lemma 1 ensures that

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

We infer that

$$\mathbb{E}[m_n(x)] - m\,(x) = \frac{1}{2(1 - 2a\xi_\beta)} h_n^2 \sum_{j=1}^{d} \mu_j(\mathbf{K}) m_{jj}^{(2)}(x) + o\left(h_n^2\right).$$

For the case $a > \beta/(d+4)$, we have $\lim_{n\to\infty} (n\beta_n) > \frac{\beta - a}{2}$, which yields $h_n^2 = o\left(\sqrt{\beta_n h_n^{-d}}\right)$. Hence, the application of lemma 1 ensures that

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

As a matter of fact, the result can be expressed as

$$\mathbb{E}[m_n(x)] - m\,(x) = o\left(\sqrt{\beta_n h_n^{-d}}\right).$$

Let us now compute the variance of $m_n\,(x)$. We state

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

Moreover, we have

$$\begin{aligned}\mathbb{E}[\mathcal{Z}_k^2(x)] &= \int_{\mathbb{R}^d} h_k^{-2d}\psi_k^{-2}\mathbb{E}[T^2|X=y]\psi_k\mathbf{K}^2\left(\frac{x-y}{h_k}\right)f(y)dy\\ &= \int_{\mathbb{R}^d} h_k^{-d}\psi_k^{-1}\mathbf{K}^2\left(z\right)\mathbb{E}[T^2|X=x-zh_k]f(x-zh_k)dz.\end{aligned}$$

Hence, the Taylor's expansion for $h : x \longmapsto \mathbb{E}[T^2|X=x]f(x) = \int_{\mathbb{R}} y^2 g(x,y)dy$ ensures that

$$\mathbb{E}[\mathcal{Z}_k^2(x)] = h_k^{-d}\psi_k^{-1}\left[\mathbb{E}[T^2|X=x]f(x)\int_{\mathbb{R}^d}\mathbf{K}^2\left(z\right)dz + \nu_k(x)\right].$$

Thus,

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

where

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

For the case $a \geqslant \beta/(d+4)$, we have $\lim_{n\to\infty}(n\beta_n) > \frac{\beta-ad}{2}$ and then $1-2a\xi_\beta > 0$. Since we have $\lim_{k\to+\infty}\nu_k(x) = 0$ and $\lim_{k\to+\infty} h_k\eta_k(x) = 0$, then the application of lemma 1 ensures that

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

Therefore, the result is provided by

$$Var[m_n(x)] = \frac{\mathbb{E}[T^2|X=x]}{2-(\alpha-a)\xi_\beta}\frac{\beta_n}{h_n}\psi_n^{-1}f(x)R(\mathbf{K}) + o\left(\frac{\beta_n}{h_n}\right).$$

For the case $a < \beta/(d+4)$, we have $\lim_{n\to\infty}(n\beta_n) > 2a$ which yields $\beta_n h_n^{-d} = o\left(h_n^4\right)$. Then, the application of lemma 1 ensures that

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

□

Our main result rests on the following theorem, which provides us the bias and the variance of p_n.

3.3. Bias and variance of p_n

Theorem 1
Let assumptions $(A_1) - (A_3)$ hold and assume that, for all $i, j \in \{1, \dots, d\}$, $m_{ij}^{(2)}$ and $f_{ij}^{(2)}$ are continuous at x, we obtain

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

$$\mathbb{E}[p_n(x)] - p(x) = \frac{1}{2(1-2a\xi_\beta)} \frac{h_n^2}{f(x)} \sum_{j=1}^{d} \mu_j(\mathbf{K}) \left(m_{jj}^{(2)}(x) - p(x) f_{jj}^{(2)}(x) \right) + o\left(h_n^2\right). \tag{15}$$

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

$$\mathbb{E}[p_n(x)] - p(x) = o\left(\sqrt{\beta_n h_n^{-d}}\right). \tag{16}$$

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

$$Var[p_n(x)] = o\left(h_n^4\right). \tag{17}$$

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

$$Var[p_n(x)] = \frac{\beta_n}{h_n^d} \frac{\psi_n^{-1}}{2 - (\beta - ad)\xi_\beta} \frac{R(\mathbf{K})}{f(x)} \left(\mathbb{E}[T^2 | X = x] - \psi p^2(x) \right) + o\left(\frac{\beta_n}{h_n^d}\right). \tag{18}$$

The bias and the variance of the estimator p_n defined by the stochastic approximation algorithm (4) then heavily depend on the choice of the stepsizes (β_n).

Proof
For this proof, let us note that for $f_n \neq 0$, we have

$$p_n(x) - p(x) = A_n(x) \frac{f(x)}{f_n(x)}, \tag{19}$$

with

$$A_n(x) = \frac{1}{f(x)} \left(m_n(x) - m(x) \right) - \frac{p(x)}{f(x)} \left(f_n(x) - f(x) \right). \tag{20}$$

It follows from (19) that the asymptotic behavior of $p_n(x) - p(x)$ can be deduced from the one of $A_n(x)$. Hence, we can state

$$\mathbb{E}[A_n(x)] = \frac{1}{f(x)} \left(\mathbb{E}[m_n(x)] - m(x) \right) - \frac{p(x)}{f(x)} \left(\mathbb{E}[f_n(x)] - f(x) \right).$$

Since we already have the *bias* of $m_n(x)$ as well as that of $f_n(x)$, and considering the fact that $m(x) = p(x) f(x)$, then we just need to combine the results (10), (11), (6) and (7) in order to obtain (15) and (16). Now, we have

$$Var[A_n(x)] = \frac{1}{(f(x))^2} Var[m_n(x)] - \frac{(p(x))^2}{(f(x))^2} Var[f_n(x)] - 2\frac{p(x)}{(f(x))^2} Cov(m_n(x), f_n(x)).$$

Let us now compute the covariance between $m_n(x)$ and $f_n(x)$. Indeed, we have

$$\begin{aligned}
Cov(m_n(x), f_n(x)) &= Cov\left(Q_n\sum_{k=1}^{n} Q_k^{-1}\beta_k\, Y_k\psi_k^{-1}\, h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right), Q_n\sum_{i=1}^{n} Q_i^{-1}\beta_i h_i^{-d}\mathbf{K}\left(\frac{x-X_i}{h_i}\right)\right)\\
&= Q_n\sum_{k=1}^{n} Q_k^{-1}\beta_k Q_n\sum_{i=1}^{n} Q_i^{-1}\beta_i\, Cov\left(Y_k\psi_k^{-1}\, h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right), h_i^{-d}\mathbf{K}\left(\frac{x-X_i}{h_i}\right)\right)\\
&= Q_n^2\sum_{k=1}^{n} Q_k^{-2}\beta_k^2 Cov\left(Y_k\psi_k^{-1}\, h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right), h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right)\right)\\
&= Q_n^2\sum_{k=1}^{n} Q_k^{-2}\beta_k^2 \left(\mathbb{E}\left[Y_k\psi_k^{-1}\, h_k^{-2d}\mathbf{K}^2\left(\frac{x-X_k}{h_k}\right)\right]\right.\\
&\quad \left. -\,\mathbb{E}\left[Y_k\psi_k^{-1}\, h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right)\right]\mathbb{E}\left[h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right)\right]\right)\\
&= Q_n^2\sum_{k=1}^{n} Q_k^{-2}\beta_k^2\left(\mathbb{E}[T|X=x]f(x)R(\mathbf{K})h_k^{-d} - \mathbb{E}[T|X=x]f^2(x)\right) + o\left(h_k^{-d}\right)\\
&= Q_n^2\sum_{k=1}^{n} Q_k^{-2}\beta_k^2 h_k^{-d}\left(p(x)f(x)R(\mathbf{K}) + o(1)\right)\\
&= \frac{\beta_n h_n^{-d}}{2-(\beta-ad)\xi_\beta}p(x)f(x)R(\mathbf{K}) + o\left(\beta_n h_n^{-d}\right). \qquad (21)
\end{aligned}$$

Consequently, (17) and (18) follow from the combination of (12), (13), (8), (9) and (21). For the case $a \geq \beta/(d+4)$, we can deduce

$$\begin{aligned}
Var[p_n(x)] &= \frac{1}{f(x)}\frac{\beta_n}{h_n^d}\psi_n^{-1}\frac{\mathbb{E}[T^2|X=x]}{2-(\beta-ad)\xi_\beta}R(\mathbf{K}) + \frac{p(x)^2}{f(x)}\frac{\beta_n}{h_n^d}\frac{1}{2-(\beta-ad)\xi_\beta}R(\mathbf{K})\\
&\quad - 2\frac{p(x)}{f(x)^2}\frac{\beta_n h_n^{-d}}{2-(\beta-ad)\xi_\beta}p(x)f(x)R(\mathbf{K}) + o\left(\beta_n h_n^{-d}\right)\\
&= \frac{\beta_n}{h_n^d}\frac{\psi_n^{-1}}{2-(\beta-ad)\xi_\beta}\frac{R(\mathbf{K})}{f(x)}\left(\mathbb{E}[T^2|X=x] - \psi p(x)^2\right) + o\left(\beta_n h_n^{-d}\right).
\end{aligned}$$

□

Now, let us state the following theorem which yields the asymptotic normality of the proposed multivariate recursive regression estimator under missing data p_n denoted in (4).

3.4. Asymptotic normality of p_n

Theorem 2
Let assumptions $(A_1)-(A_3)$ hold and assume that, for all $i,j \in \{1,\dots,d\}$, $m_{ij}^{(2)}$ and $f_{ij}^{(2)}$ are continuous at x. We therefore have:
If there exists $c \geq 0$ such that $\beta_n^{-1}h_n^{d+4} \underset{n\to+\infty}{\longrightarrow} c$, then

$$\sqrt{\beta_n^{-1}h_n^d\psi_n}\left(p_n(x)-p(x)\right) \underset{n\to+\infty}{\overset{\mathcal{D}}{\longrightarrow}} \mathcal{N}\left(\sqrt{c}M(x)\ ,\ \Sigma(x)\right). \qquad (22)$$

with

$$M(x) = \frac{1}{2(1-2a\xi_\beta)f(x)}\sum_{j=1}^{d}\mu_j(\mathbf{K})\left(m_{jj}^{(2)}(x) - p(x)f_{jj}^{(2)}(x)\right)$$

and

$$\Sigma(x) = \frac{1}{2-(\beta - ad)\xi_\beta}\frac{R(\mathbf{K})}{f(x)}\Big(\mathbb{E}[T^2|X=x] - \psi p^2(x)\Big),$$

where $\underset{n\to+\infty}{\xrightarrow{\mathcal{D}}}$ represents convergence in distribution and $\mathcal{N}$ denotes the Gaussian distribution.

Proof
We have

$$\begin{aligned}A_n(x) - \mathbb{E}[A_n(x)] =& \frac{1}{f(x)}[m_n(x) - \mathbb{E}[m_n(x)]] - \frac{p(x)}{f(x)}[f_n - \mathbb{E}[f_n]]\\ =& \frac{1}{f(x)}Q_n\sum_{k=1}^{n}\left(L_k(x) - \mathbb{E}[L_k(x)]\right),\end{aligned}$$

with

$$L_k(x) = Q_k^{-1}\beta_k\left(\mathcal{Z}_k(x) - p(x)\mathcal{W}_k(x)\right).$$

In this proof, we note

$$S_k(x) = L_k(x) - \mathbb{E}[L_k(x)].$$

On the one hand, it's obvious that

$$p_n(x) - \mathbb{E}[p_n(x)] = \frac{1}{f(x)}Q_n\sum_{k=1}^{n}S_k(x). \tag{23}$$

On the other hand, we attempt to apply Lyapunov's theorem for $S_k(x)$. For this reason, we consider

$$\begin{aligned}\upsilon_n^2 &= \sum_{k=1}^{n}Var[S_k(x)]\\ &= \sum_{k=1}^{n}Var[L_k(x)]\\ &= \sum_{k=1}^{n}Q_k^{-2}\beta_k^2\left(Var\left[\mathcal{Z}_k(x)\right] + p(x)^2Var\left[\mathcal{W}_k(x)\right] - 2p(x)cov\left(\mathcal{Z}_k(x),\mathcal{W}_k(x)\right)\right).\end{aligned}$$

Moreover, we have

$$Var\left[\mathcal{Z}_k(x)\right] = h_k^{-d}\psi_k^{-1}\Big(\mathbb{E}[T^2|X=x]f(x)R(\mathbf{K}) + o(1)\Big).$$

$$Var\left[\mathcal{W}_k(x)\right] = h_k^{-d}\Big(f(x)R(\mathbf{K}) + o(1)\Big).$$

$$cov\left(\mathcal{Z}_k(x),\mathcal{W}_k(x)\right) = h_k^{-d}\Big(p(x)f(x)R(\mathbf{K}) + o(1)\Big).$$

Hence, by applying lemma 1 , it can be inferred that

$$\begin{aligned}\upsilon_n^2 \quad =& \quad \sum_{k=1}^{n}Q_k^{-2}\beta_k^2h_k^{-d}\psi_k^{-1}\Big(\mathbb{E}[T^2|X=x]f(x)R(\mathbf{K}) + o(1)\Big)\\ & +p(x)^2\sum_{k=1}^{n}Q_k^{-2}\beta_k^2h_k^{-d}\Big(f(x)R(\mathbf{K}) + o(1)\Big) - 2p(x)\sum_{k=1}^{n}Q_k^{-2}\beta_k^2h_k^{-d}\Big(p(x)f(x)R(\mathbf{K}) + o(1)\Big)\\ =& \quad \frac{\beta_n}{h_n^d}\psi_n^{-1}\frac{f(x)^2}{Q_n^2}[\Sigma + o\,(1)].\end{aligned} \tag{24}$$

In addition, we have

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

Therefore,

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

Hence,

$$\mathbb{E}\left[|S_k(x)|^{2+p}\right] = O\left(Q_k^{-2-p}\beta_k^{2+p}\mathbb{E}\left[|L_k(x)|^{2+p}\right]\right).$$

We then deduce that

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

In the following, let us assume that there is $p > 0$, such that

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

By applying lemma 1 , we obtain

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

Thus,

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

Then, it follows that

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

Moreover, since we have

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

by applying the Lyapunov theorem, we get

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

which implies

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

Moreover, (23) ensures that

$$f(x)Q_n^{-1}\upsilon_n^{-1}\left(p_n(x)-\mathbb{E}[p_n(x)]\right)\underset{n\to+\infty}{\xrightarrow{\mathcal{D}}}\mathcal{N}\left(0\,,1\right). \tag{25}$$

Then, the combination of (24) and (25) ensures that

$$\sqrt{\beta_n^{-1}h_n^d\psi_n}\left(p_n(x)-\mathbb{E}[p_n(x)]\right)\underset{n\to+\infty}{\xrightarrow{\mathcal{D}}}\mathcal{N}\left(0\,,\Sigma\right). \tag{26}$$

Hence, the application of Lyapounov's Theorem coupled with the combination of (15), (16) and (26) ensures the convergence in (22). □

In order to measure the asymptotic performance of the proposed recursive kernel regression estimator under missing data p_n and to be able to use a data-driven bandwidth selection procedure, through proposing an asymptotic unbiased estimators of the unknown quantities, we consider the Mean Weighted Integrated Squared Error ($MWISE$), where the weight function is selected to be equal to $f^3\left(x\right)$. This choice was motivated by the fact that we can propose an asymptotic unbiased kernel estimator for the unknown quantities, which will appear in the $MWISE$ as reported previously in [29], and which shall be detailed later.

3.5. Asymptotic expressions of $MWISE$ of p_n

The $MWISE$ of the estimator p_n is determined by,

$$MWISE[p_n]=\int_{\mathbb{R}^d}\left(\mathbb{E}[p_n(x)]-p(x)\right)^2f^3(x)dx+\int_{\mathbb{R}^d}Var[p_n(x)]f^3(x)dx. \tag{27}$$

For simplicity, we set

$$C_1=\frac{I_1-2I_2+I_3}{\left(1-2a\xi_\beta\right)^2}\quad\text{and}\quad C_2=\frac{I_4-\psi I_5}{2-(\beta-ad)\xi_\beta}.$$

It follows that

$$MWISE[p_n]=\begin{cases}\dfrac{1}{4}C_1h_n^4+o(h_n^4) & \text{if}\quad a\in\left(0,\frac{\beta}{d+4}\right)\\ C_2R(\mathbf{K})\beta_nh_n^{-d}\psi_n^{-1}+\dfrac{1}{4}C_1h_n^4+o(h_n^4) & \text{if}\quad a=\frac{\beta}{d+4}\\ C_2R(\mathbf{K})\beta_nh_n^{-d}\psi_n^{-1}+o\left(\beta_nh_n^{-d}\right) & \text{if}\quad a\in\left(\frac{\beta}{d+4},1\right)\end{cases}.$$

The corollary bellow ensures that the bandwidth which minimizes the $MWISE$ of p_n depends on the choice of the stepsizes (β_n) and then the corresponding $MWISE$ depends in turn on (β_n).

Corollary 1
Let assumptions $(A_1)-(A_3)$ hold. To minimize the $MWISE$ of p_n, the bandwidth (h_n) must be equal to

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

Then, the corresponding $MWISE$ is estimated in terms of

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

The following corollary is presented in the special case, where (β_n) is chosen as $(\beta_n)=(\beta_0n^{-1})$. We can check easily that the optimal choice of β_0 is obtained by getting β_0 equal to 1.

Corollary 2
Let assumptions $(A_1) - (A_3)$ hold. To minimize the $MWISE$ of p_n, we must choose the stepsize (β_n) in $\mathcal{GS}(-1)$ such that $\lim_{n\to\infty} (n\beta_n) = 1$. Consequently, the optimal bandwidth (h_n) must be equal to

$$\left(\left(\frac{d(d+2)}{2(d+4)}\right)^{\frac{1}{d+4}} \left(\frac{I_4 - \psi I_5}{I_1 - 2I_2 + I_3}\right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{\frac{-1}{d+4}} \psi_n^{\frac{-1}{d+4}}\right). \tag{28}$$

Thus, the corresponding $MWISE$ is provided by

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

3.6. Asymptotic properties of $\widetilde{p}_n$

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

Proposition 3
Let assumptions (A_1) and (A_3) hold and assume that, for all $i, j \in \{1, \dots, d\}$, $m_{ij}^{(2)}$ and $f_{ij}^{(2)}$ are continuous at x. Therefore, the bias and variance of Nadaraya-Watson's regression estimator are equal to:

$$\mathbb{E}[\widetilde{p}_n(x)] - p(x) = \frac{1}{2f(x)} h_n^2 \left(\sum_{j=1}^{d} \mu_j(\mathbf{K}) m_{jj}^{(2)}(x) - p(x) \sum_{j=1}^{d} \mu_j(\mathbf{K}) f_{jj}^{(2)}(x)\right) + o\left(h_n^2\right).$$

$$Var[\widetilde{p}_n(x)] = \frac{1}{nh_n^d} \psi_n^{-1} \frac{1}{f(x)} Var[T|X = x] R(\mathbf{K}) + o\left(\frac{1}{nh_n^d}\right).$$

It is inferred that

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

Corollary 3
Let assumptions (A_1) and (A_3) hold. To minimize the $MWISE$ of $\widetilde{p}_n$, the bandwidth (h_n) must be equal to

$$\left(d^{\frac{1}{d+4}} \left(\frac{I_4 - \psi I_5}{I_1 - 2I_2 + I_3}\right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{-\frac{1}{d+4}} \psi_n^{-\frac{1}{d+4}}\right). \tag{29}$$

Then, the corresponding $MWISE$ is specified by

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

Clearly, the use of such bandwidth (29), is not possible when we use real data. From this perspective, the next section is devoted to build up a data-driven bandwidth procedure, which will be helpful in practice.

4. Bandwidth selection

Within the framework of non-parametric kernel estimation, the choice of the smoothing parameter is crucial for the effective performance of the estimators. There are a myriad of data-driven bandwidth selection methods recorded in literature which can be divided into three broad classes: cross-validation techniques, plug-in methods, and the bootstrap approach. A detailed comparison of the three techniques is exhibited in [6]. In this paper, based on the previous work conducted by [27–29] for unidimensional data, we propose a second generation Plug-in bandwidth data-driven procedures in the multivariate data for regression estimation.

4.1. Plug-in bandwidth selection method

A widely used criterion stands for selecting a bandwidth that minimizes the estimate of the mean squared error, using the density function as a weight function. [2] proposed an efficient method of bandwidth selection, a plug-in estimate. Since the $MWISE$ depends on unknown quantities I_j, $j = 1 \dots 5$, we suggest elaborating an asymptotic unbiased estimator of those quantities.
As a matter of fact, we adopt the approach proposed in [2], called the second generation Plug-in estimation. For this purpose, we should introduce the so called pilot bandwidth $(b_n)_{n \geq 1} \in \mathcal{GS}(-\delta)$, $\delta \in (0, 1)$.
In practice, we take $b_n = n^{-\delta} \min \left\{ \widehat{s}, \frac{Q_3 - Q_1}{1.349} \right\}$, with, $\widehat{s}$ is the sample standard deviation and Q_1, Q_3 are the first and third quartiles. In order to select the parameter δ, we follow the work of [27–29].
First of all, for the sake of simplicity, the kernel $\mathbf{K}$ is considered as a product of univariate kernels K satisfying $\int_{\mathbb{R}} K(x)dx = 1$. For this purpose, we let $\mu(K) = \int_{\mathbb{R}} z^2 K(z)dz$,

$$I_j = \mu^2(K) I'_j, \; j = 1 \dots 3,$$

where

$$I'_1 = \int_{\mathbb{R}^d} \left(\sum_{j=1}^{d} m_{jj}^{(2)}(x) \right)^2 f(x)dx,$$

$$I'_2 = \int_{\mathbb{R}^d} \left(\sum_{j=1}^{d} m_{jj}^{(2)}(x) \right) \left(\sum_{j=1}^{d} f_{jj}^{(2)}(x) \right) p(x) f(x)dx,$$

$$I'_3 = \int_{\mathbb{R}^d} \left(\sum_{j=1}^{d} f_{jj}^{(2)}(x) \right)^2 p^2(x) f(x)dx.$$

4.1.1. Multivariate recursive kernel regression estimator under missing data p_n

Here, we can state

$$m_n(x) = Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k \, h_k^{-d} Y_k \psi_k^{-1} \mathbf{K} \left(\frac{x - X_k}{h_k} \right) = Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k Y_k \psi_k^{-1} h_k^{-d} \prod_{i=1}^{d} K \left(\frac{x_i - X_{k_i}}{h_k} \right)$$

and

$$f_n(x) = Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k \, h_k^{-d} \mathbf{K} \left(\frac{x - X_k}{h_k} \right) = Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k h_k^{-d} \prod_{i=1}^{d} K \left(\frac{x_i - X_{k_i}}{h_k} \right).$$

At this stage of analysis, in order to estimate the optimal bandwidth (28), we need to estimate I_j, $j = 1 \dots 5$.

Estimation of I_1, I_2 and I_3: We consider the following kernel estimators to estimate respectively I_1, I_2 and I_3:

$$\widehat{I'}_1 = \frac{Q_n^2}{n} \sum_{\substack{i,j,k=1\\ i\neq j\neq k}}^{n} Q_j^{-1} Q_k^{-1} \beta_j \beta_k {b'}_j^{-(d+2)} {b'}_k^{-(d+2)} \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{i_t} - X_{j_t}}{b_j^{'}} \right) \prod_{\substack{l=1\\ l\neq t}}^{d} K_b \left(\frac{X_{i_l} - X_{j_l}}{b_j} \right) \right]$$
$$\times \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{i_t} - X_{k_t}}{b_k^{'}} \right) \prod_{\substack{l=1\\ l\neq t}}^{d} K_b \left(\frac{X_{i_l} - X_{k_l}}{b_k} \right) \right] Y_j \psi_j^{-1} Y_k \psi_k^{-1},$$

$$\widehat{I'}_2 = \frac{Q_n \Pi_n}{n} \sum_{\substack{i,j,k=1\\ i\neq j\neq k}}^{n} Q_j^{-1} \Pi_k^{-1} \beta_j \gamma_k {b'}_j^{-(d+2)} {b'}_k^{-(d+2)} \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{i_t} - X_{j_t}}{b_j^{'}} \right) \prod_{\substack{l=1\\ l\neq t}}^{d} K_b \left(\frac{X_{i_l} - X_{j_l}}{b_j} \right) \right]$$
$$\times \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{i_t} - X_{k_t}}{b_k^{'}} \right) \prod_{\substack{l=1\\ l\neq t}}^{d} K_b \left(\frac{X_{i_l} - X_{k_l}}{b_k} \right) \right] Y_j \psi_j^{-1} Y_i \psi_i^{-1},$$

$$\widehat{I'}_3 = \frac{\Pi_n^2}{n} \sum_{\substack{i,j,k,m=1\\ i\neq j\neq k\neq m}}^{n} \Pi_j^{-1} \Pi_k^{-1} \gamma_j \gamma_k {b'}_j^{-(d+2)} {b'}_k^{-(d+2)} \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{i_t} - X_{j_t}}{b_j^{'}} \right) \prod_{\substack{l=1\\ l\neq t}}^{d} K_b \left(\frac{X_{i_l} - X_{j_l}}{b_j} \right) \right]$$
$$\times \left[\sum_{t=1}^{d} K_{b'}^{(2)} \left(\frac{X_{i_t} - X_{k_t}}{b_k^{'}} \right) \prod_{\substack{l=1\\ l\neq t}}^{d} K_b \left(\frac{X_{i_l} - X_{k_l}}{b_k} \right) \right] Y_i \psi_i^{-1} Y_m \psi_m^{-1},$$

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

Estimation of I_4 and I_5: We consider the following kernel estimators to estimate respectively I_4 and I_5:

$$\widehat{I_4} = \frac{\Pi_n}{n} \sum_{\substack{i,k=1\\ i\neq k}}^{n} \Pi_k^{-1} \gamma_k b_k^{-d} \prod_{l=1}^{d} K_b \left(\frac{X_{i_l} - X_{k_l}}{b_k} \right) Y_i^2 \psi_i^{-2},$$

and

$$\widehat{I_5} = \frac{Q_n}{n} \sum_{\substack{i,k=1\\ i\neq k}}^{n} Q_k^{-1} \beta_k b_k^{-d} \prod_{l=1}^{d} K_b \left(\frac{X_{i_l} - X_{k_l}}{b_k} \right) Y_i \psi_i^{-1} Y_k \psi_k^{-1},$$

where, K_b is a kernel and b_n is the associated bandwidth, such that $\delta = -2/5$. It follows that, the plug-in bandwidth selection estimator of (28) is expressed by

$$(h_n) = \left(\left(\frac{d(d+2)}{2(d+4)} \right)^{\frac{1}{d+4}} \left(\frac{\widehat{I_4} - \psi \widehat{I_5}}{\widehat{I_1} - 2\widehat{I_2} + \widehat{I_3}} \right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{\frac{-1}{d+4}} \psi_n^{-\frac{1}{d+4}} \right), \tag{30}$$

with, $\widehat{I_i} = \mu^2(\mathbf{K}) \widehat{I'}_i$, $i = 1 \ldots 3$.
Then, the plug-in estimator of $MWISE[p_n]$ is equal to

$$\widehat{MWISE}[p_n] = \frac{(d+4)^{\frac{3d+8}{d+4}}}{4^{\frac{d+6}{d+4}} d^{\frac{d}{d+4}} (d+2)^{\frac{d+6}{d+4}}} \left(\widehat{I_1} - 2\widehat{I_2} + \widehat{I_3} \right)^{\frac{d}{d+4}} \left(\widehat{I_4} - \psi \widehat{I_5} \right)^{\frac{4}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{\frac{-4}{d+4}} \psi_n^{-\frac{4}{d+4}} + o\left(n^{\frac{-4}{d+4}} \psi_n^{-\frac{4}{d+4}} \right).$$

Now, let us examine the asymptotic properties of the multivariate non-recursive Nadaraya-Watson's regression estimator under missing data $\widetilde{p}_n$. First we display the multivariate bias and variance of $\widetilde{p}_n$.

4.1.2. Multivariate non-recursive kernel regression estimator under missing data $\widetilde{p}_n$

Here, we can state

$$\widetilde{m}_n(x) = \frac{1}{nh_n^d}\sum_{k=1}^{n} Y_k\psi_k^{-1}\,\mathbf{K}\left(\frac{x-X_k}{h_n}\right) = \frac{1}{nh_n^d}\sum_{k=1}^{n} Y_k\psi_k^{-1}\prod_{i=1}^{d} K\left(\frac{x_i-X_{k_i}}{h_k}\right)$$

and

$$\widetilde{f}_n(x) = \frac{1}{nh_n^d}\sum_{k=1}^{n}\mathbf{K}\left(\frac{x-X_k}{h_n}\right) = \frac{1}{nh_n^d}\sum_{k=1}^{n}\prod_{i=1}^{d} K\left(\frac{x_i-X_{k_i}}{h_k}\right).$$

In order to estimate the optimal bandwidth (29), we need to estimate I_j, $j = 1\dots 5$.

Estimation of I_1, I_2 and I_3: We consider the following kernel estimators to estimate respectively I_1, I_2 and I_3:

$$\widetilde{I'}_1 = \frac{1}{n^3 {b'}_n^{2(d+2)}}\sum_{\substack{i,j,k=1\\ i\neq j\neq k}}^{n}\left[\sum_{t=1}^{d} K_{b'}^{(2)}\left(\frac{X_{i_t}-X_{j_t}}{b_n^{'}}\right)\prod_{\substack{l=1\\ l\neq t}}^{d} K_b\left(\frac{X_{i_l}-X_{j_l}}{b_n}\right)\right] \times\left[\sum_{t=1}^{d} K_{b'}^{(2)}\left(\frac{X_{i_t}-X_{k_t}}{b_n^{'}}\right)\prod_{\substack{l=1\\ l\neq t}}^{d} K_b\left(\frac{X_{i_l}-X_{k_l}}{b_n}\right)\right] Y_j\psi_j^{-1}Y_k\psi_k^{-1},$$

$$\widetilde{I'}_2 = \frac{1}{n^3 {b'}_n^{2(d+2)}}\sum_{\substack{i,j,k=1\\ i\neq j\neq k}}^{n}\left[\sum_{t=1}^{d} K_{b'}^{(2)}\left(\frac{X_{i_t}-X_{j_t}}{b_n^{'}}\right)\prod_{\substack{l=1\\ l\neq t}}^{d} K_b\left(\frac{X_{i_l}-X_{j_l}}{b_n}\right)\right] \times\left[\sum_{t=1}^{d} K_{b'}^{(2)}\left(\frac{X_{i_t}-X_{k_t}}{b_n^{'}}\right)\prod_{\substack{l=1\\ l\neq t}}^{d} K_b\left(\frac{X_{i_l}-X_{k_l}}{b_n}\right)\right] Y_j\psi_j^{-1}Y_i\psi_i^{-1},$$

$$\widetilde{I'}_3 = \frac{1}{n^4 {b'}_n^{2(d+2)}}\sum_{\substack{i,j,k,m=1\\ i\neq j\neq k\neq m}}^{n}\left[\sum_{t=1}^{d} K_{b'}^{(2)}\left(\frac{X_{i_t}-X_{j_t}}{b_n^{'}}\right)\prod_{\substack{l=1\\ l\neq t}}^{d} K_b\left(\frac{X_{i_l}-X_{j_l}}{b_n}\right)\right] \times\left[\sum_{t=1}^{d} K_{b'}^{(2)}\left(\frac{X_{i_t}-X_{k_t}}{b_n^{'}}\right)\prod_{\substack{l=1\\ l\neq t}}^{d} K_b\left(\frac{X_{i_l}-X_{k_l}}{b_n}\right)\right] Y_i\psi_i^{-1}Y_m\psi_m^{-1},$$

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

Estimation of I_4 and I_5: We consider the following kernel estimators to estimate respectively I_4 and I_5:

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

and

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

where, K_b is a kernel and b_n is the associated bandwidth, such that $\delta = -2/5$. Hence, the plug-in bandwidth selection estimator of (29) is indicated by

$$(h_n) = \left(d^{\frac{1}{d+4}} \left(\frac{\widetilde{I}_4 - \psi \widetilde{I}_5}{\widetilde{I}_1 - 2\widetilde{I}_2 + \widetilde{I}_3} \right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{-\frac{1}{d+4}} \psi_n^{-\frac{1}{d+4}} \right), \tag{31}$$

with, $\widetilde{I}_i = \mu^2(\mathbf{K}) \widetilde{I'}_i$, $i = 1 \ldots 3$.
It follows that, the plug-in non-recursive estimator of $MWISE[p_n]$ is equal to

$$\begin{aligned} \widetilde{MWISE}[\widetilde{p}_n] &= \frac{(d+4)}{4d^{\frac{d}{d+4}}} \frac{5}{4} \left(\widetilde{I}_4 - \psi \widetilde{I}_5 \right)^{\frac{4}{d+4}} \left(\widetilde{I}_1 - 2\widetilde{I}_2 + \widetilde{I}_3 \right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{-\frac{4}{d+4}} \psi_n^{-\frac{4}{d+4}} \\ &\quad + o\left(n^{-\frac{4}{d+4}} \psi_n^{-\frac{4}{d+4}} \right). \end{aligned}$$

5. Application to the handwritten word production

Research on the handwritten word production aims to describe the cognitive processes and mental representations mobilized when a human being prepares to handwrite a word from an idea of [21]. One of the most widely used tasks to experimentally explore these issues is object naming. Participants have to produce words corresponding to the names of a set of drawings in handwriting as quickly as possible. It is generally accepted that the handwritten objects naming involves four levels of processing [19]. First, a perceptual analysis of the visual input is performed, which results in activation of stored structural knowledge about the object. A second processing level corresponds to the retrieval of semantic/conceptual information. The lexical selection level makes orthographic word form information available. Eventually, the motoric programming level allows the access to motoric codes corresponding to each produced letter.

These theoretical propositions concerning the cognitive processes and representations involved in the handwritten object naming stem from studies aiming at finding predictors of reaction times (RTs hereafter), i.e., the time between the presentation of the image and the first graphic movement (e.g., [3]; [20]). Four factors have been reported to significantly influence RTs, each of which allows indexing a specific processing level. Image Agreement (IA) captures the similarity between structural representations stored in memory and the visual characteristics of an object's drawing. This factor has extensive influence in terms of the perceptual analysis. The IA is measured on a Like rt scale, generally in five points, from '1 - weakly similar' to '5 - strongly similar'. A negative linear relationship is observed between this variable and the RTs (see [3]; [20]). Image variability (Ivar or Image ability) is designed to index the 'richness' of semantic representations. Like AI, it is rated on a 5-point scale, from 1 = few images to 5 = many images. A negative linear relationship is reported between handwritten RTs and Ivar (see [3]; [20]).

Name agreement (NA) refers to the degrees of agreement on the use of a specific label for an image, measured using an entropy measure (h-index). A positive linear relationship is reported between RTs and the h-index (see [3]; [20]). NA indexes the influence of the number of correct alternative names existing for an image (e.g., couch => sofa). Latencies would be more or less impacted by the time needed to manage the competition between the higher or lower number of alternatives during lexical access. Finally, the influence of age-limited learning (Age of Acquisition, AoA) has been systematically emphasized in studies on the predictors of handwritten RTs (see [3]; [20]). AoA is usually measured using a Like rt scale (from 1 = learned at 0-3 years to 5 = learned at age +12, with 3-year bands in between), with a population of young adults who are asked to estimate the age at which they learned the proposed word. A positive linear relationship is observed between the RTs and the rated values of AoA (see [3]; [20]). Experimental work [22] suggests that this variable influences the orthographic wordform encoding processes.

The major target of this work is to classify the participants in groups of clusters. From this perspective, we first have to predict the regression function, i.e the relation between the variable $T = RTs$ and the four covariates $X_1 = H$, $X_2 = IA$, $X_3 = Ivar$ and $X_4 = AoA$. Since the response variable RTs is subject of missing data, we should introduce a correction variable $Y := CRTs$ defined as $Y_i = T_i * \mathbb{1}_{\{T_i \text{ is observed}\}}$.

Here, we have Np individual estimators of each participant $\widehat{Y}_1, \ldots, \widehat{Y}_{Np}$ ($\widetilde{Y}_1, \ldots, \widetilde{Y}_{Np}$) and a general estimator $\widehat{Y}^g$ ($\widetilde{Y}^g$) which estimates the whole database of Np participants.

It's worth noting that, for each participant/covariate behavior test, we invested a different method for bandwidth selection, namely the plug-in univariate selection for multivariate data.

This implies that, instead of opting for a single value of bandwidth h_n, we considered a vector $h_{n1}, \ldots, h_{nd}$, an individual choice of bandwidth for each covariate. Then, for the recursive case, we have a matrix of bandwidths:

$$H = \begin{pmatrix} h_{11} & \cdots & h_{1d} \\ \vdots & \ddots & \vdots \\ h_{n1} & \cdots & h_{nd} \end{pmatrix}.$$

We denote by p_i^* the reference regression vector and by p_i the test regression. Thus, we calculate the two following measures: the Mean Squared Error: $MSE = \dfrac{1}{n}\sum_{i=1}^{n}(p_i - p_i^*)^2$, and the Mean Relative Error: $MRE = \dfrac{1}{n}\sum_{i=1}^{n}\left|\frac{p_i - p_i^*}{p_i^*}\right|$.

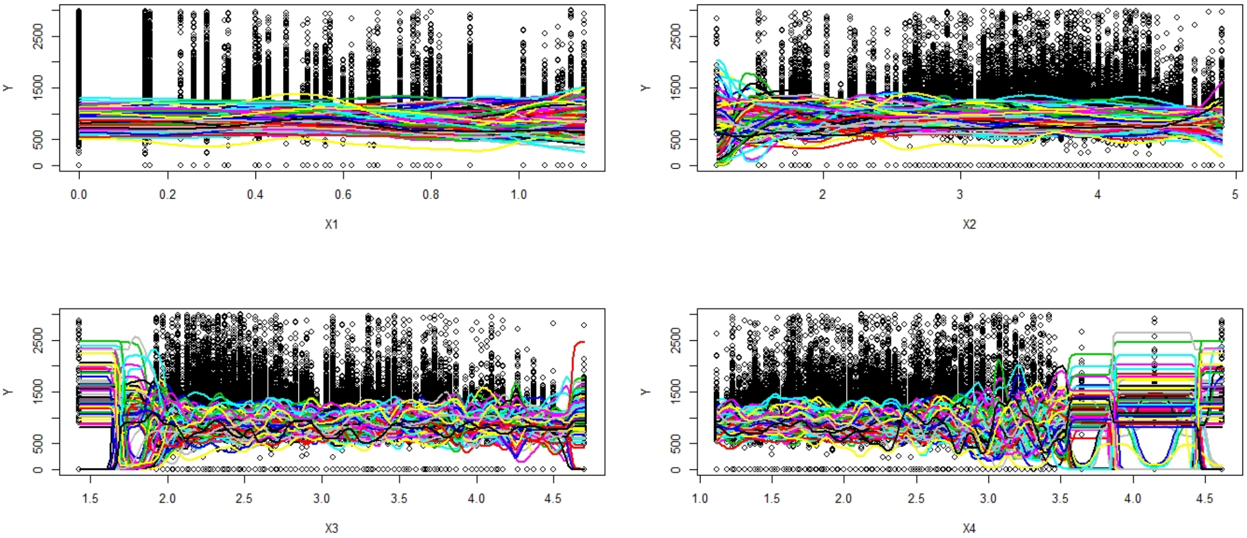

Figure 1. Participants' behavior representations, the regression between the reactions time variable $Y = CRTs$ and each covariate ($X_1 = H$, $X_2 = IA$, $X_3 = Ivar$ and $X_4 = AoA$) with the entire database (a total of 137 participants).

Algorithm 1 $X_1, \dots, X_4$ are the covariates such that $X_1 = H$, $X_2 = IA$, $X_3 = Ivar$ and $X_4 = AoA$, Y is the response variable with $Y = CRTs$, $\mathbf{K}$ is the Gaussian kernel n the number of items and Np is the number of participants.

```
Input: $Y$, $X_1, \dots, X_4$, $\mathbf{K}$, $n$ and $Np$.
  A choice value for the recursive bandwidth vectors $h_1, \dots, h_n$. (resp. the non-recursive bandwidth values $h_n$) // using the plug-in approach provided in (28) (resp. (31)).
  The choice of the stepsize $(\beta_n) = \left(n^{-1}\right)$ (then, $(Q_n) = \left(n^{-1}\right)$).
  An arbitrary sampling vectors $x_1, \dots, x_4$.
  The estimation of $\psi$ is carried out according to the algorithm proposed in [30]
  for $l = 1, \dots, Np$ do
      $\widehat{Y}_l = \frac{\sum_{k=1+(l-1)n}^{ln} k\beta_k Y_k \psi_k^{-1} \prod_{i=1}^{4} h_{k_i}^{-1} \prod_{i=1}^{4} K\left(\frac{x_i - X_{k_i}}{h_{k_i}}\right)}{\sum_{k=1+(l-1)n}^{ln} k\beta_k \prod_{i=1}^{4} h_{k_i}^{-1} \prod_{i=1}^{4} K\left(\frac{x_i - X_{k_i}}{h_{k_i}}\right)}$ for the multivariate recursive kernel regression estimator (resp. $\widetilde{Y}_l = \frac{\sum_{k=1+(l-1)n}^{ln} Y_k \psi_k^{-1} \prod_{i=1}^{4} K\left(\frac{x_i - X_{k_i}}{h_{n_i}}\right)}{\sum_{k=1+(l-1)n}^{ln} \prod_{i=1}^{4} K\left(\frac{x_i - X_{k_i}}{h_{n_i}}\right)}$ for the multivariate non-recursive kernel regression estimator).
  end for
output: $\widehat{Y}_1, \dots, \widehat{Y}_{Np}$ and $\widetilde{Y}_1, \dots, \widetilde{Y}_{Np}$.
```

Mean Relative Error	Min	1st Qu	Median	Mean	3rd Qu	Max
`Recursive`	0.0000014	0.1208883	0.2643889	0.3837469	0.5322441	2.4902917
`Non-recursive`	0.0000103	0.1412407	0.3301375	0.4748705	0.9999976	2.4525833

Table 1. Quantitative comparison between the mean relative error of the multivariate non-recursive Nadaraya-Watson's regression estimator (`Non-recursive`) and the proposed multivariate recursive kernel regression estimator (`Recursive`) with stepsize $(\beta_n) = (n^{-1})$ through a plug-in method.

Let us underline that in order to classify participants in groups, we use the MSE as a reference vector. Thus, we use the k-means method to specify the maximum number of needed clusters.

Algorithm 2 Participants classification algorithm:
Y is the response variable with $Y = \texttt{CRTs}$, $\widehat{Y}_1, \dots, \widehat{Y}_{Np}$ are the predicted multivariate recursive kernel regression estimators and $\widetilde{Y}_1, \dots, \widetilde{Y}_{Np}$ are the predicted multivariate non-recursive kernel regression estimators.

```
Input: $Y$, $\widehat{Y}_1, \dots, \widehat{Y}_{Np}$ and $\widetilde{Y}_1, \dots, \widetilde{Y}_{Np}$.
  Start with writing $Y$ in a matrix form participant per participant.
  for $l = 1, \dots, Np$ do
      $MSER_l = \frac{1}{n}\sum_{i=1}^{n}(\widehat{Y}_i - Y_i)^2$. for the recursive estimation (resp. $MSET_l = \frac{1}{n}\sum_{i=1}^{n}(\widetilde{Y}_i - Y_i)^2$. for the non-recursive estimation).
  end for
  A classification of the remote distance // through kmeans package in R.
output: The classification list using both considered estimators.
```

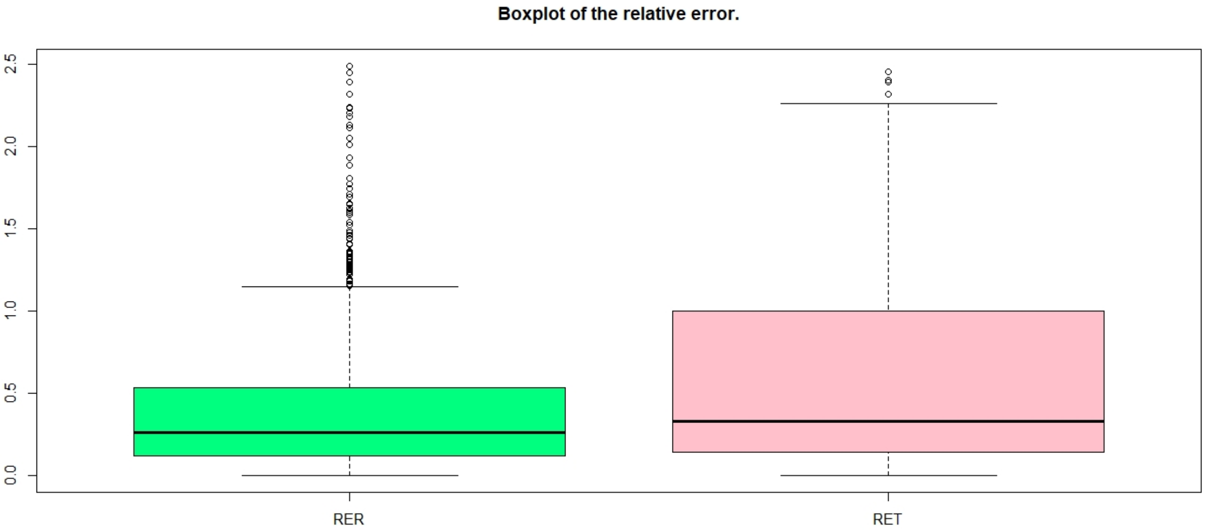

Figure 2. Box-plot of the relative error estimation of both considered estimators, the recursive one on the left and the non-recursive one on the right.

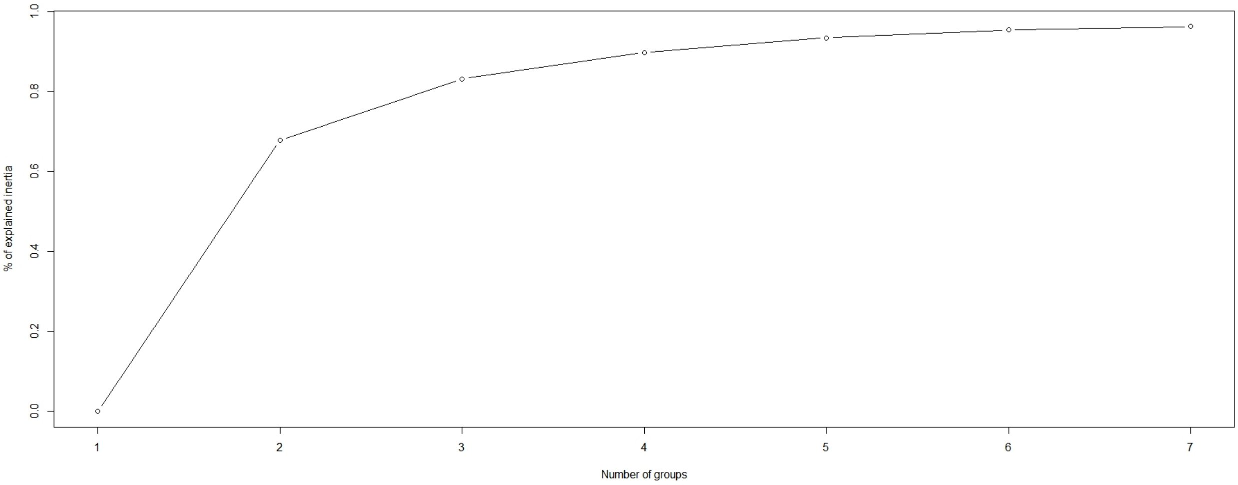

Figure 3. The `elbow` method of selecting the optimal number of clusters ($k = 3$) for K-means clustering on the MSE vector.

Result Analysis:

Departing form figure 2 and table 1, we deduce that the proposed recursive estimator outperforms the non-recursive one in terms of mean relative error estimation. Meanwhile, figures 1 and 3 indicate that it is advisable to consider three clusters. As far as written production behavior is concerned, this implies that the classification procedure suggests three clusters to measure the distance of each participant from the reference. In other words, three forms of variation can be observed when participants have to write the label of a drawing. Further exploration of the available characteristics of the participants suggests that such anthropological factors as the age and gender do not account for the result of clustering. Descriptive analysis of executive function task data suggests that there are differences between the three groups of participants. This indicates that the variations would be interpreted in part by the participants' cognitive processing ability and by differences in the mobilization of participants' executive functions. Studies based upon procedures for fitting reaction time distributions with ex-Gaussian-type probability density distributions (convolution of a normal and exponential law) have corroborated the role of these executive functions in simple tasks (e.g., [26]; [35]). Our analyses yield that this result can be extended to more complex activities such as written production. Eventually, this work confirms the significance of the use of non-parametric regressions for modeling behavior in experimental psychology area.

6. Conclusion

In this research paper, we elaborated a multivariate recursive regression estimator under missing data. We first investigated the asymptotic properties of the proposed estimator by providing the bias as well as the variance in order to demonstrate that our estimator asymptotically follows a normal distribution. Subsequently, we compared our recursive estimator with the non-recursive multivariate Nadaraya-Watson's regression estimator using the plug-in bandwidth selection approach. In our application of real dataset, and for all the cases, the proposed estimator (4) with stepsize $(\beta_n) = \left(n^{-1}\right)$ yielded smaller MSE and MRE compared to the non-recursive Nadaraya Watson's estimator.

As part of the application, it was possible to estimate the response variable RTs (Reaction Times) according to the other covariates through classifying the participants into clusters of membership according to their approximation to the real value of RTs.

To conclude, the use of the multivariate recursive kernel regression estimator under missing data enabled us to obtain better results compared to the multivariate non-recursive kernel regression estimator under missing data. With an appropriate choice of the bandwidth, we depicted that our proposed estimator is closer to the true regression function than the non-recursive one.

A future research direction would be to extend our findings to the case of functional data like in [32] and [33]. We can also consider the k nearest neighbours smoothing with functional regressor, see [1] in finite dimensional data and [10] in the case of functional data. Another direction is to consider same estimators grounded on bias reduction technique (see [9], [31]), which requires non trivial mathematics and goes therefore beyond the scope of the present paper. Finally, we can also explore the idea developed in the recent work of [4] through considering some semi-parametric Bayesian networks approaches based on the current work.

Acknowledgments

This work has received support from the Nouvelle Aquitaine Region and the European Union within the framework of the "FEDER/FSE Operational Programme 2014-2020" (CPER "E-data" and CPER "insect"). This work benefited also from the financial support of the PHC-Utique project no.17G1505.

References

1. I. M. Almanjahie, K. A. Aissiri, A. Laksaci, and Z. Chikr Elmezouar, *The k nearest neighbors smoothing of the relative-error regression with functional regressor*, Comm. Statist. Theory Methods., vol. 51, no. 12, pp. 4196–4209, 2022.
2. N. Altman and C. Léger, *Bandwidth selection for kernel distribution function estimation*, Journal of Statistical Planning and Inference, vol. 46, no. 2, pp. 195–214, 1995.

3. P. Bonin, M. Chalard, A. Meot, and M. Fayol, *The determinants of spoken and written picture naming latencies,* British Journal of Psychology, vol. 93, pp. 89–114, 2002.
4. S. Boukabour, and A. Masmoudi, *Semiparametric Bayesian networks for continuous data*, Comm. Statist. Theory Methods, vol. 50, no. 24, pp. 5974–5996, 2021.
5. D. Cousineau, and S. Chartier, *Outliers detection and treatment: a review,* International Journal of Psychological Research, vol. 3, pp. 58–67, 2010.
6. A. Delaigle, and I. Gijbels, *Practical bandwidth selection in deconvolution kernel density estimation*, Comput. Statist. Data Anal., vol. 45, no. 2, pp. 249–267, 2004.
7. J. Galambos, and E. Seneta, *Regularly varying sequences*, Proceedings of the American Mathematical Society, vol. 41, pp. 110–116, 1973.
8. W. Hardle, and J. S. Marron, *Optimal Bandwidth Selection in Nonparametric Regression Function Estimation*, The Ann. Statist., vol. 13, no. 4, pp. 1465–1481, 1985.
9. A. Jmaei, Y. Slaoui, and W. Dellagi, *Recursive distribution estimators defined by stochastic approximation method using Bernstein polynomials*, J. Nonparametr. Stat., vol. 29, no. 4, pp. 792–805, 2017.
10. L. Z. Kara, A. Laksaci, M. Rachdi, and P. Vieu, *Data-driven kNN estimation in nonparametric functional data analysis*, J. Multivariate Anal., vol. 153, no. C, pp. 176–188, 2017.
11. J. Kiefer and J. Wolfowitz, *Stochastic estimation of the maximum of a regression function*, Annals of Mathematical Statistics, vol. 23, pp. 462–466, 1952.
12. R. D. Luce, *Response times: their role in inferring elementary mental organization*, Handbook of Mathematical Psychology, New York: Oxford, 1986.
13. P. F. McCormack, and N. M. Wright, *The positive skew observed in reaction time distributions*, Canadian Journal of Psychology, vol. 18, pp. 43–51, 1964.
14. W. J. McGill, *Stochastic latency mechanisms.*, Handbook of Mathematical Psychology, vol. 19, no. 1, pp. 309–360, 1963.
15. A. Mokkadem, M. Pelletier, and Y. Slaoui, *The stochastic approximation method for the estimation of a multivariate probability density*, Journal of Statistical Planning and Inference, vol. 139, no. 7, pp. 2459–2478, 2009.
16. A. Mokkadem, M. Pelletier, and Y. Slaoui, *Revisiting Révész's stochastic approximation method for the estimation of a regression function*, ALEA. Latin American Journal of Probability and Mathematical Statistics, vol. 6, pp. 63–114, 2009.
17. A. Mokkadem, and M. Pelletier, *The Multivariate Revesz's Online Estimator of a Regression Function and Its Averaging.*, *M. Math. Meth. Stat.*, vol. 25, no. 3, pp. 151–167, 2016.
18. E. A. Nadaraya, *On estimating regression*, Theory of Probability and Its Applications, vol. 9, pp. 141–142, 1964.
19. C. Perret, and P. Bonin, *Which variables should be controlled for to investigate picture naming in adults? A Bayesian meta-analysis*, Behavior Research Methods, vol. 51, pp. 2533–2545, 2019.
20. C. Perret, and M. Laganaro, *Why are written naming latencies (not) longer than spoken naming? Reading and Writing*, An Interdisciplinary Journal, vol. 26, pp. 225–239, 2013.
21. C. Perret, and T. Olive, *Spelling and Writing Words: Theoretical and Methodological Advances*, Brills Edition, 2019.
22. C. Perret, P. Bonin, and M. Laganaro, *Exploring the multiple-level hypothesis of AoA effects in spoken and written picture naming using a topographic ERP analysis*, Brain and Language, vol. 135, pp. 20–31, 2014.
23. P. Révész, *How to apply the method of stochastic approximation in the non-parametric estimation of a regression function*, Math. Operationsforsch. Statist. Ser. Statist., vol. 8, no. 1, pp. 119–126, 1977.
24. H. Robbins and S. Monro, *A stochastic approximation method*, Annals of Mathematical Statistics, vol. 22, pp. 400–407, 1951.
25. P. R. Rosenbaum, and D. B. Rubin, *The central role of the propensity score in observational studies for causal effects*, Biometrika, vol. 70, no. 1, pp. 41–55, 1983.
26. F. Schmiedek, K. Oberauer, O. Wilhelm, H. M Süβ, and W. W. Wittmann, *Individual differences in components of reaction time distributions and their relations to working memory and intelligence*, Journal of Experimental Psychology: General, vol. 136, pp. 414–429, 2007.
27. Y. Slaoui, *Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method*, Journal of Probability and Statistics, ID 739640, pp. 1–11, 2014.
28. Y. Slaoui, *The stochastic approximation method for estimation of a distribution function*, Mathematical Methods of Statistics, vol. 23, no. 4, pp. 306–325, 2014.
29. Y. Slaoui, Optimal bandwidth selection for semi-recursive kernel regression estimators, Stat. Interface, vol. 9, no. 3, pp. 375–388, 2016.
30. Y. Slaoui, *Recursive kernel density estimators under missing data*, Comm. Statist. Theory Methods. vol. 46, no. 18, pp. 9101–9125, 2017.
31. Y. Slaoui, *Bias reduction in kernel density estimation*. J. Nonparametr. Stat., vol. 30, no. 2, pp. 505–522, 2018.
32. Y. Slaoui, *Wild Bootstrap Bandwidth Selection of Recursive Nonparametric Relative Regression for Independent Functional Data*, J. Multivariate Anal., vol. 173, pp. 494–511, 2019.
33. Y. Slaoui, *Recursive non-parametric regression estimation for independent functional data*, Statist. Sinica, vol. 30, pp. 417–437, 2020.
34. A. B. Tsybakov, *Recurrent estimation of the mode of a multidimensional distribution*, Probl. Inf. Transm., vol. 26, no. 1, pp. 31–37, 1990.
35. N. Unsworth, T.S. Redick, C.E. Lakey and D.L. Young, *Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation*, Intelligence, vol. 38, pp. 111–122, 2010.
36. G. S. Watson, *Smooth regression analysis*, Sankhyā (Statistics). The Indian Journal of Statistics. Series A, vol. 26, pp. 359–372, 1964.