Stochastic approximation algorithms for multivariate functionals estimation with medical and cognitive fields applications.

Sahar Slama

Republic of Tunisia
Ministry of Higher Education and Scientific Research

Thesis

Presented at

High School of Sciences and Technologie of Hammam Sousse
Mathematics Laboratory LAMMDA
University of Sousse

And in collaboration with the Laboratory of Mathematics and Applications
University of Poitiers

to obtain the degree of

DOCTOR OF MATHEMATICS

Mention: Applied Mathematics
Specialty: Statistics

by

Sahar SLAMA

Stochastic approximation algorithms for multivariate functionals estimation with medical and cognitive fields applications.

Defended on 31 October 2022 front the committee composed of :

Pr. Adel DAOUAS	Sousse University	President
Pr. Sophie DABO	Lille University	Referee
Pr. Sofiane GASMI	Tunis University	Referee
Pr. Ali GANNOUN	Montpelier University	Member
Pr. Moez KHENISSI	Sousse University	Member
Pr. Cyril PERRET	Poitiers University	Member (Invited)
Pr. Khalifa EL MABROUK	Sousse University	Supervisor
Pr. Yousri SLAOUI	Poitiers University	Supervisor
Pr. Hamdi FATHALLAH	Sousse University	Co-Supervisor

Acknowledgments

As a preamble to this thesis, I thank **ALLAH** who helped me and granted me patience and courage over these years of study.

My highest acknowledgments go first of all to the members of the jury for having agreed to asses this work and for all their pertinent remarks and outstanding reviews. I have the honor to warmly thank the referees Pr. Sophie DABO and Pr. Sofiane GASMI, for having agreed to participate in the assessment of this work.
I have the great pleasure to specifically thank the chairman Pr. Adel DAOUAS and the members Pr. Moez KHENISSI, Pr. Ali GANNOUN and Pr. Cyril PERRET.

I am deeply indebted to my supervisor in Poitiers Pr. Yousri SLAOUI and my supervisors in Sousse, Pr. Khalifa EL MABROUK and Pr. Hamdi FATHALLAH. I thank them infinitely for having framed me and advised me.

I also wish to address my most sincere thanks to the professorial and administrative staff of the Higher School of Sciences and Technology of Hammam Sousse, of the Laboratory *LAMMDA* and of the Laboratory of Mathematics and Applications of Poitiers, for not only their helpful assistance but also their active contribution to the development of this thesis as well as to the success of these long years of preparation.

My special thanks go to my dearest parents, Salem and Sihem, who have always been there for me, "You have sacrificed everything for your only daughter, sparing neither health nor efforts. I am indebted to you for a life of which I am proud."

I am equally grateful to my fiance Chokri. "You are the only one who knows the magic trick to make me smile even when I am going through tough time. I am truly blessed to have you as part of my life. Thank you for being my lover, my friend and confident.

I especially extend my honorable thanks to my colleague Carlos for his constant support and his judicious advice.

Eventually, I would like to express my endless gratitude to my grandmother Habiba as well as all my family members and friends who have provided me with moral and intellectual support throughout my academic journey.

My sincere apology goes to those whom I didn't mention personally one by one.

Résumé

L'estimation récursive multivariée est le point central de cette thèse. Notre objectif fondamental est de construire des estimateurs des fonctionnelles multivariées en utilisant des méthodes d'approximations stochastiques. Dans la section d'ouverture, nous fournissons une introduction générale du sujet de l'estimation non-paramétrique et de l'algorithme original d'approximation stochastique récursive. Pour le premier chapitre, nous introduisons un estimateur récursif multivarié pour la fonction de répartition. Nous étudions les propriétés asymptotiques de cet estimateur généralisé et nous le comparons avec l'estimateur multivarié non récursif de Nadaraya. Il s'avère qu'avec un choix adéquat de la taille des pas et un choix approprié de la fenêtre, l'erreur quadratique moyenne MSE (Mean Squared Error en anglais) de l'estimateur multivarié avec plug-in comme méthode de sélection de la fenêtre est plus petite que celle des deux autres estimateurs, à savoir l'estimateur multivarié récursif avec sélection par validation croisée et l'estimateur non récursif de Nadaraya. Le deuxième chapitre traite le problème de l'estimation non paramétrique d'une fonction de répartition cumulative conditionnelle $\pi : (y|x) \longmapsto \mathbb{P}[Y \leqslant y|X = x]$. En utilisant la même approche récursive, nous proposons un estimateur récursif multivarié défini par un algorithme d'approximation stochastique. Nous étudions les inférences statistiques de notre estimateur et les comparons avec celles de l'estimateur non récursif de Nadaraya-Watson. Étant donné l'idée d'estimation conditionnelle, et pour le troisième chapitre, nous construisons un estimateur semi-récursif généralisé de type noyau de la fonction de régression $r_\varphi : x \longmapsto \mathbb{E}[\varphi(Y)|X = x]$, pour une fonction mesurable choisie φ et $x \in \mathbb{R}^d$. Afin d'examiner ces propriétés asymptotiques, nous calculons d'abord le biais et la variance de notre estimateur proposé qui dépendent fortement du choix de trois paramètres qui sont les pas et la fenêtre. De plus, nous sommes intéressés par l'étude de la convergence forte de notre estimateur. Il s'avère que pour l'estimation par intervalles de confiance, l'estimateur proposé est meilleur que celui de l'estimateur de Nadaraya Watson. En ce qui concerne le quatrième chapitre, nous tentons d'explorer les processus cognitif et les représentations mentales mobilisées lorsqu'un être humain se prépare à écrire un mot selon l'idée développée dans Perret and Olive(2019). Dans cette perspective, nous proposons une estimation non paramétrique à noyau récursif d'une régression multivariée avec données manquantes pour décrire la production de mots d'écriture. Nous étudions les propriétés asymptotiques de notre estimateur récursif et ses performances par rapport à l'estimateur non récursif. L'estimateur proposé est ensuite appliqué aux données comportementales pour classer certains participants dans des groupes. Cette classification peut être un début pour comprendre les variations de comportement écrites. La dernière section est consacré à la partie conclusion ainsi qu'à quelques perspectives de future recherches.

Mots-clés: estimation multivariée, algorithme d'approximation stochastique, méthode d'estimation du noyau, estimation par plug-in, lissage et ajustement de courbe, champ manuscrit et psychologie.

Abstract

Multivariate recursive estimation is the central focus of this thesis. Our basic objective is to construct multivariate functional estimators using stochastic approximations methods. In the opening section, we provide a general introduction of non-parametric estimation topic and the original recursive stochastic approximation algorithm. In the first chapter, we introduce a multivariate recursive estimator for the distribution function. We study the asymptotic properties of this generalized estimator and we compare it with non-recursive Nadaraya's multivariate distribution estimator. It turns out that, with an adequate choice of the stepsize and an appropriate choice of the bandwidth, the MSE (Mean Squared Error) of the multivariate estimator with plug-in bandwidth selection method can be smaller than the two other estimators, namely the multivariate recursive one with cross-validation selection and the non-recursive one of Nadaraya's estimator. The second chapter deals with non-parametric estimation of a conditional cumulative distribution function (CCDF) $\pi : (y|x) \longmapsto \mathbb{P}\left[Y \leqslant y | X = x\right]$. Using the same recursive approach, we suggest a multivariate recursive estimator defined by stochastic approximation algorithm. We investigate the statistical inferences of our estimator and compare them with those of non-recursive Nadaraya-Watson's estimator. Given the idea of conditional estimation, and for the third chapter, we construct a generalized semi-recursive kernel-type estimator of the regression function $r_\varphi : x \longmapsto \mathbb{E}[\varphi(Y)|X = x]$, for a chosen measurable function φ and $x \in \mathbb{R}^d$. In order to examine the asymptotic properties of this estimator, we first calculate the bias and the variance of our proposed estimator which strongly depend on the choice of three parameters which are the stepsizes and the bandwidth. Moreover, we are interested in studying the strong pointwise convergence rate of our estimator. It turns out that under the estimation by confidence intervals, the proposed estimator is better than that of the non-recursive Nadaraya Watson regression one. As far as the fourth chapter is concerned, we attempt to explore cognitive processes and mental representations enacted when a human being prepares to write a word according to the idea developed in Perret and Olive(2019). For this objective, we set forward a non-parametric multivariate recursive kernel regression estimation under missing data to describe writing word production. We examine the asymptotic properties of our recursive estimator and its performance against the non-recursive one. As application, the proposed estimator is then applied to the behavioral data to classify some participants in groups. This classification can be a departure point to understand written behavior variations. The closing section is devoted for the conclusion part as well as some perspectives for future researches.

Keywords: multivariate estimation, stochastic approximation algorithm, kernel method, plug-in estimate, smoothing and curve fitting, handwritten and psychology field.

Contents

List of Figures

List of Tables

Introduction en français

Analyse Statistiques

De nos jours, les données sont partout. Elles sont à la disposition de tous. Les données elles-mêmes ne sont que des faits et des chiffres qui doivent être explorés pour obtenir des informations significatives. L'analyse des données est donc cruciale. Il s'agit du processus d'application de techniques statistiques et logiques pour visualiser, réduire, décrire et évaluer les données en informations utiles qui fournissent un meilleur contexte pour les données.
L'analyse des données joue un rôle-clé dans la recherche d'informations significatives qui aideront les entreprises à prendre de meilleures décisions sur la base des résultats.
La science des données est tout aussi importante que l'analyse des données. La science des données correspond à un domaine combinant de multiples méthodes de méthodologie scientifique, des processus, des algorithmes et des outils pour extraire des informations, en particulier d'énormes ensembles de données pour obtenir des informations sur des données structurées et non structurées. Différents termes liés à l'extraction, au nettoyage, à l'analyse et à l'interprétation des données sont souvent utilisés de manière interchangeable dans la science des données.
L'analyse des données consiste à examiner divers facteurs ou variables pour déterminer leur impact sur certaines situations et résultats. Cela nous aide à comprendre pourquoi certains résultats se produisent, ce qui nous permet de faire des prévisions et de prendre des décisions éclairées pour l'avenir. Lorsque nous élaborons des données impliquant plus de deux variables, nous utilisons l'analyse multivariée qui n'est pas une méthode spécifique, mais qui englobe plutôt l'ensemble des techniques statistiques utilisées pour analyser plus de deux variables à la fois. Ces approches nous permettent d'obtenir une vision plus approfondie des données en relation avec des scénarios spécifiques de l'entreprise ou du monde réel. En fait, si vous êtes un analyste ou un scientifique des données ambitieux, l'analyse multivariée est un concept intrinsèque à aborder. Cette technique sera le point central de cette thèse.

Introduction à la théorie d'estimation

La statistique d'estimation, ou simplement l'estimation, correspond à un cadre d'analyse des données qui repose sur une combinaison de taille d'effet, d'intervalles de confiance, de planification de précision pour analyser les données et interpréter les résultats. Elle se distingue du test de signification de l'hypothèse nulle (TSN) qui est moins informatif.
En statistiques, un estimateur est une fonction permettant d'estimer un paramètre inconnu lié à une distribution de probabilité. Il peut être investi pour déterminer certaines caractéristiques d'une population totale à partir des données observées. Par exemple, dans une enquête, la moyenne de l'échantillon est l'estimateur le plus couramment utilisée de la moyenne de la population.
L'approche de l'estimation est l'une des branches les plus remarquables de la statistique, qui s'intéresse aux propriétés des estimateurs exprimées en terme de convergence, de biais et d'efficacité. De nombreuses approches peuvent être élaborées afin de proposer des estimateurs de qualités multiples pour une même quantité et sur une même base de données.

Par ailleurs, la théorie de l'estimation peut être divisée en deux composantes principales, l'estimation paramétrique et l'estimation non paramétrique.
La statistique paramétrique, qui remonte à Fisher en 1920, représente le cadre classique de la statistique. Le modèle statistique est caractérisé par un nombre fini de paramètres. On prend $\mathcal{M} = \{\mathbb{P}_\theta : \theta \in \mathbb{R}^p\}$ comme modèle statistique typique qui décrit la distribution des variables aléatoires observées.
Cependant, la statistique non paramétrique étudie des problèmes statistiques où la paramétrisation n'est pas fixe, mais où il existe différents choix de paramètres afin de trouver ceux qui entraînent les procédures les plus efficaces. En fait, la loi est entièrement inconnue. Pour cette raison, nous avons recours à l'estimation des fonctionnelles décrivant le modèle.
Nous nous intéressons essentiellement à la question de l'estimation non-paramétrique d'une fonction de répartition F, d'une densité f et d'une fonction de régression r.
Par conséquent, afin de résoudre un nombre de problèmes statistiques, il suffit d'avoir une estimation convenable.

Estimation non paramétrique

L'estimation non paramétrique joue un rôle important dans l'exploration d'une infinité de phénomènes de nature aléatoire. Elle se trouve au cœur de la modélisation de problèmes réels relevant de multiples domaines scientifiques, à savoir les sciences de l'environnement, la sismologie et la psychologie, l'imagerie médicale et les neurosciences.
Un large éventail de méthodes d'estimation non paramétrique de la densité a été élaboré, comme l'estimateur par histogramme, l'estimateur simple et l'estimateur par noyau.
L'estimateur de l'histogramme, introduit par John Graunt (1662), est une fonction en escalier, et donc discontinue, ce qui constitue une lacune pour l'estimation d'une densité. Pour surmonter ce problème, la fonction indicatrice est remplacée par une fonction réelle, appelée estimateur à noyau. En effet, alors qu'un histogramme compte le nombre de points de données dans des intervalles assez arbitraires, l'estimation de la densité par noyau est une fonction définie comme la somme d'une fonction noyau sur chaque point de données. Par conséquent, l'estimation de la densité du noyau est une question essentielle de lissage des données qui permet de faire des inférences sur la population à partir d'un échantillon fini de données.
La technique d'estimation par noyau constitue le point central du paragraphe suivant et de l'ensemble de l'ouvrage.

Estimation par noyau

En statistique, l'estimation par noyau désigne un processus non paramétrique d'estimation d'une fonction de probabilité inconnue d'une variable aléatoire à l'aide d'une fonction noyau.
Fondamentalement, on suppose que le noyau présente généralement les propriétés suivantes :

i) $\displaystyle\int_{\mathbb{R}} \mathbf{K}(x)dx = 1.$

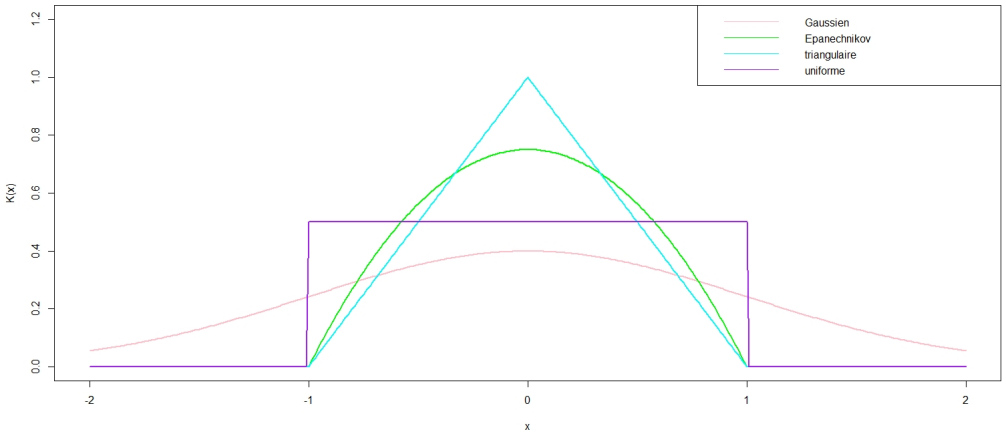
ii) $\displaystyle\int_{\mathbb{R}} \mathbf{K}^2(x)dx < +\infty.$

iii) $\displaystyle\int_{\mathbb{R}} x\mathbf{K}(x)dx = 0.$

iv) $\displaystyle\int_{\mathbb{R}} x^2\mathbf{K}(x)dx < +\infty.$

Examples 0.1. *Les noyaux les plus couramment utilisés sont :*

* *Le noyau uniforme :* $\mathbf{K}(x) = \frac{1}{2}\mathbb{1}_{[-1\,,\,1]}(x)$.
* *Le noyau triangulaire :* $\mathbf{K}(x) = (1-|x|)\,\mathbb{1}_{[-1\,,\,1]}(x)$.
* *Le noyau d'Epanechnikov :* $\mathbf{K}(x) = \frac{3}{4}\left(1-x^2\right)\mathbb{1}_{[-1\,,\,1]}(x)$.
* *Le noyau gaussien :* $\mathbf{K}(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$.



Représentation de quelques noyaux classiques.

Nous pouvons distinguer trois estimateurs de noyaux. Nous observons l'estimateur de Parzen-Rosenblatt pour la densité de probabilité, l'estimateur de Nadaraya pour la fonction de répartition et l'estimateur de Nadaraya-Watson pour la fonction de régression.
Soient $(X,Y) \in \mathbb{R}\times\mathbb{R}$ et $(X_1,Y_1),\ldots,(X_n,Y_n)$ des vecteurs aléatoires indépendants identiquement distribués comme (X,Y) de densité jointe $g(x,y)$ et soient f et F la densité de probabilité et la fonction de répartition de X, qui sont inconnues.

L'estimateur de densité à noyau : Rosenblatt (1956) et Parzen (1962)

L'estimateur de densité à noyau étudié par Murray Rosenblatt (1956) et Emanuel Parzen (1962) et appelé estimateur de Parzen-Rosenblatt est indiqué par

$$\widetilde{f}_n(x) = \frac{1}{nh_n}\sum_{k=1}^{n}\mathbf{K}\left(\frac{x-X_k}{h_n}\right),$$

où, $\mathbf{K}$ désigne le noyau et (h_n) représente le paramètre de lissage (la fenêtre) qui est une suite de nombres réels positifs qui tendent vers zéro.

L'estimateur de la fonction de répartition à noyau : Nadaraya (1964)

L'estimateur de la fonction de répartition à noyau élaboré par Nadaraya (1964) et appelé estimateur de Nadaraya s'exprime comme suit

$$\widetilde{F}_n(x) = \frac{1}{n}\sum_{k=1}^{n}\mathcal{K}\left(\frac{x-X_k}{h_n}\right),$$

avec $\mathcal{K}(z) = \displaystyle\int_{-\infty}^{z}\mathbf{K}(u)du.$

L'estimateur de régression à noyau : Nadaraya-Watson (1964)

L'estimateur de régression à noyau identifié par Nadaraya (1964) et Watson (1964) et appelé estimateur de Nadaraya-Watson est noté par

$$\widetilde{r}_n(x) = \begin{cases} \dfrac{\widetilde{a}_n(x)}{\widetilde{f}_n(x)} & \text{si} \quad \widetilde{f}_n(x) \neq 0 \\ \\ 0 & \text{sinon} \end{cases},$$

avec $\widetilde{a}_n(x) = \dfrac{1}{nh_n} \sum_{k=1}^{n} Y_k \mathbf{K}\left(\dfrac{x - X_k}{h_n}\right)$.

Pour la construction adéquate de ces estimateurs, deux choix pertinents doivent être faits : la fonction noyau K et le paramètre de lissage h_n. Comme le noyau est peu sensible à la forme de l'estimateur, le choix de la fonction noyau n'est pas critique et a moins d'importance que celui de la fenêtre.
Bien que l'étude théorique asymptotique permette d'obtenir la fenêtre optimale, le fait de ne pas connaître la fonction de densité rend son interprétation en pratique assez difficile. De ce point de vue, le lissage par noyau en statistique non paramétrique nécessite le choix d'un paramètre de fenêtre qui est crucial pour la performance effective des estimateurs. Une fenêtre appropriée peut aider à obtenir une fonction estimée proche de la vraie. Cependant, une fenêtre mal choisie peut sérieusement déformer les véritables caractéristiques sous-jacentes de la fonctionnelle. En fait, une petite fenêtre entraîne un sous-lissage et une grande fenêtre entraîne un sur-lissage. Ainsi, un choix judicieux de la fenêtre est fortement recommandé.

Sélection du paramètre de lissage:

Afin de déterminer le choix optimal du paramètre de lissage, il faut trouver le paramètre qui minimise le risque. Les critères les plus utilisés pour le choix de h_n sont l'Erreur quadratique moyenne MSE (Mean Squared Error), lErreur quadratique moyenne intégrée $MISE$ (Mean Integrated Squared Error) et l'Erreur quadratique moyenne intégrée et pondéréee $MWISE$ (Mean Weighted Integrated Squared Error). Les $MWISE$ et $MISE$ correspondent au choix du paramètre de lissage constant et le MSE se réfère au paramètre de lissage variable. Tout au long de la thèse, la valeur optimale de la fenêtre h_n est obtenue en minimisant la $MWISE$ asymptotique.
Bien que l'étude théorique asymptotique permette d'obtenir la fenêtre optimale, le fait de ne pas connaître la fonction de densité rend son interprétation en pratique assez difficile. De ce point de vue, le lissage par noyau dans les statistiques non paramétriques nécessite le choix d'un paramètre de fenêtre qui est crucial pour la performance effective des estimateurs. Une fenêtre appropriée peut aider à obtenir une fonction estimée proche de la vraie fonction cible. Cependant, une fenêtre mal choisie peut sérieusement déformer les véritables caractéristiques sous-jacentes de la fonctionnelle. Ainsi, un choix judicieux de la fenêtre est fortement recommandé. Il existe une myriade de méthodes de sélection de la fenêtre basées sur les données recensées dans la littérature, que l'on peut diviser en trois grandes catégories : les techniques de validation croisée, les méthodes du plug-in de deuxième génération et l'approche bootstrap.
La méthode de sélection par validation croisée introduite par Sarda (1993) est une technique de ré-échantillonnage qui consiste à minimiser un estimateur approprié de l'erreur quadratique moyenne. Cependant, cette méthode présente certains inconvénients, comme l'a souligné Altman and Leger (1995) qui a développé une autre méthode efficace, une estimation plug-in qui minimise une estimation de l'erreur quadratique intégrée pondérée moyenne, en utilisant la fonction de densité comme fonction de poids. Une autre méthode est le bootstrap sauvage, introduit dans Hardle and Marron (1991) qui consiste à ré-échantillonner à partir des résidus estimés. La méthode plug-in et la méthode bootstrap sont toutes deux indiscernables et il a été largement

prouvé qu'elles se comportent de manière similaire. Une comparaison détaillée des trois techniques pratiques de sélection de la fenêtre est présentée dans Delaigle and Gijbels (2004).
Tout au long de la thèse, les approches non-paramétriques qui nous intéressent reposent sur la Méthode d'Approximation Stochastique notée SAM. Cette dernière est étudiée dans ce qui suit.

Méthode d'approximation stochastique

Au cours de la dernière décennie, les flux de données sont devenus un domaine de recherche de plus en plus important. Certains des flux de données les plus courants incluent les données de paquets Internet, l'activité de Twitter, le fil d'actualité de Facebook, les transactions par carte de crédit et, plus récemment, les données relatives à l'épidémie COVID-19. Les algorithmes stochastiques ont été couramment utilisés dans de nombreuses applications de recherche, à savoir l'identification de systèmes, le contrôle adaptatif, les systèmes de transmission et la détection de changements séquentiels.
Dans ces situations, les données arrivent régulièrement de sorte qu'il est impossible de les stocker dans une base de données traditionnelle. Dans un tel contexte, il est très intéressant de construire un estimateur récursif qui n'a pas besoin de stocker toutes les données en mémoire et qui peut être facilement mis à jour pour gérer les données en ligne.
Le mérite fondamental des estimateurs récursifs réside dans le fait que l'estimation peut être mise à jour à chaque nouvelle observation. Par conséquent, au lieu de ré-exécuter les données à chaque fois, il est possible de réécrire notre estimateur considéré comme une combinaison de deux (ou plus) estimateurs, où chaque estimateur est basé sur des ensembles de données distincts, ce qui peut être très intéressant pour maintenir le coût de calcul raisonnablement bas. Il est à noter que tous les calculs et simulations ont été effectués à l'aide du logiciel statistique **R**.

Cadre historique

Les algorithmes d'approximation stochastiques sont des versions stochastiques d'algorithmes déterministes comme l'algorithme de Newton ou l'algorithme du gradient. Nous nous intéressons au problème suivant : trouver le zéro x^* d'une fonction à valeur réelle S. Nous supposons avoir une valeur très approximative x_0 de cette racine. L'idée naturelle de l'algorithme de Newton est de remplacer la courbe représentative de la fonction S par sa tangente au point x_n. L'abscisse x_{n+1} du point d'intersection de cette tangente avec l'axe des x est alors donnée pour $n \geq 1$ par

$$x_{n+1} = x_n - \frac{S(x_n)}{S'(x_n)}.$$

Lorsque la dérivée de S n'est pas facilement calculable, nous pouvons considérer une version déterministe de l'algorithme de Robbins-Monro qui consiste à remplacer le calcul de la dérivée par une suite positive décroissante tendant vers 0 de pas (γ_n). Sous réserve que S ait de bonnes propriétés de régularité, la suite définie par

$$x_{n+1} = x_n - \gamma_n S(x_n)$$

converge vers le zéro de la fonction S, noté x^* , pour toute valeur initiale x_0.
Dans de nombreuses situations, la fonction dont le zéro est recherché n'est connue que pour une perturbation proche du zéro. La recherche des zéros par des méthodes d'optimisation déterministes devient alors plus périlleuse et donc nous avons recours à des algorithmes stochastiques comme celui introduit par Robbins and Monro (1951). Les estimateurs récursifs que nous proposons ont été construits à partir de SAM. En effet, l'incorporation d'algorithmes d'approximation stochastique dans le contexte de la statistique non-paramétrique remonte aux papiers de Robbins and Monro (1951) et Kiefer and Wolfowitz (1952) pour un cadre unidimensionnel.

L'algorithme stochastique général, consacré essentiellement à l'approximation du mode d'une fonction de régression, a pour forme :

$$\theta_n = \theta_{n-1} + \gamma_n \Phi\left(\theta_{n-1}, W_n\right) + \gamma_n^2 \mu_n\left(\theta_{n-1}, W_n\right), \tag{1}$$

où :

- (γ_n) est une suite positive de nombres réels qui tend vers zéro,
- (θ_n) est la suite à mettre à jour récursivement,
- (W_n) est une suite de variables aléatoires représentant les observations en ligne,
- $\Phi\left(\theta, W\right)$ est la fonction qui détermine essentiellement comment le paramètre est mis à jour en fonction d'une nouvelle observation,
- $\mu_n\left(\theta_{n-1}, W_n\right)$ définit une petite perturbation sur l'algorithme.

Par la suite, Blum (1954) a fourni une version multidimensionnelle de cet algorithme. Ces travaux de recherche ont été étendus dans plusieurs directions. En nous inspirant des plus importants, nous pouvons présenter l'algorithme suivant, analysé par Kushner and Clark (1978), Ruppert (1982).

$$\theta_n = \theta_{n-1} + \gamma_n \left(\phi(\theta_{n-1}) - W_n + \beta_n\right), \tag{2}$$

où (β_n) désigne une suite de variables aléatoires convergeant vers 0 presque sûrement et ϕ désigne une fonction mesurable inconnue. Ils ont démontré que (2) coïncide avec l'algorithme (1) et inclut les processus d'approximation stochastique de Robbins and Monro (1951) et Kiefer and Wolfowitz (1952), qui permettent la recherche du zéro θ^* de la fonction ϕ.
Certaines modifications de base ont été incorporées par Nadaraya (1964), Polyak and Tsybakov (1990), Dippon and Renz (1997) et Dippon (2003). Plus tard, Duflo (1997) a corroboré que, sous des conditions standard sur la fonction ϕ et sur la suite (γ_n), (θ_n) tend vers θ^* presque sûrement. Afin de construire un algorithme d'approximation stochastique, Révész (1973) a introduit une application de la procédure de Robbins-Monro et Mokaddem and Pelletier (2007b) ont développé l'application de la procédure Robbins-Monro-Blum. De plus, un algorithme d'estimation d'une fonction de régression a été élaboré par Révész (1977) et a ensuite été utilisé par Tsybakov (1990) pour approximer le mode d'une densité de probabilité. , Slaoui (2006) a signalé un estimateur lisse de la fonction de densité dans un cas unidimensionnel en utilisant la méthode d'approximation stochastique. Ensuite, dans l'article de Mokkadem *et al.* (2009a), le cas multidimensionnel a été étudié afin d'estimer une densité de probabilité multivariée en utilisant l'estimation par intervalles de confiance. De plus, Slaoui (2014b) a réutilisé des méthodes d'approximation stochastique pour améliorer les qualités de l'estimateur de la fonction de répartition. Le principe de déviation large et modérée de cet estimateur a été prouvé dans Slaoui (2019).

L'estimateur classique de régression récursive a été abordé dans Mokaddem *et al.* (2009b) pour un cadre univarié et une extension multivariée de cet estimateur a été réalisée par Mokaddem and Pelletier(2016). Par la suite, Slaoui (2016) a établi le cas semi-récursif et a introduit un nouvel estimateur qui est la fraction d'une régression récursive par une fonction de densité récursive. De plus, Slaoui (2017) a adopté la technique de probabilité du score de propension et a construit un estimateur de la fonction de densité sous données manquantes. Récemment, un estimateur de la densité conditionnelle a été proposé dans Slaoui and Khardani (2020). L'objectif fondamental de ce chapitre est d'utiliser des algorithmes d'approximation stochastique pour définir des estimateurs d'une densité de probabilité en un point donné.

L'objectif principal de ce chapitre est d'utiliser des algorithmes d'approximation stochastique pour définir des estimateurs d'une densité de probabilité en un point donné.
Rappelons que les algorithmes d'approximation stochastique utilisés pour la recherche du zéro d'une fonction inconnue $\phi : y \longmapsto S(x) - y$ sont construits comme suit.

(i) On fixe $S_0(x) \in \mathbb{R}$ arbitrairement.

(ii) Pour tout $n \geq 1$, la suite (S_n) est définie récursivement par

$$S_n(x) = S_{n-1}(x) + \gamma_n T_n(x),$$

où

- (T_n) est une suite de fonctions $T_n : \mathbb{R} \to \mathbb{R}$ définie par

$$T_n(x) = \phi(S_{n-1}) - W_n + \beta_n. \tag{3}$$

  En d'autres termes, $T_n(x)$ est une observation de la fonction ϕ au point $S_{n-1}(x)$.

- (γ_n) est une suite de nombres réels positifs qui va jusqu'à zéro telle que

$$\sum \gamma_n = +\infty \quad \text{et} \quad \sum \gamma_n^2 < +\infty. \tag{4}$$

Il convient de noter que, sous la condition $\mathbb{E}\left[W_n|\mathcal{F}_{n-1}\right] = 0$ (où $\mathcal{F}_{n-1}$ représente la σ-algèbre des événements se produisant au temps $n-1$), nous avons

$$\mathbb{E}[T_n(x)] = \phi(S_{n-1}) + \beta_n = S(x) - S_{n-1}(x) + \beta_n. \tag{5}$$

Estimation multivariée récursive:

Les statistiques multivariées sont axées sur l'exploration des relations entre les variables et leur adéquation au problème en question, ce qui implique plusieurs types d'analyses univariées et multivariées. Fondamentalement, cette dernière repose sur une procédure statistique comprenant les mesures et observations simultanées de données incluant plus d'un facteur de variables indépendantes qui ont un impact sur la variabilité des variables dépendantes. Par conséquent, le principal mérite de l'analyse multivariée réside dans le fait que, comme elle prend en compte plus d'une variable de résultat où diverses quantités différentes présentent un intérêt pour la même analyse, les conclusions obtenues sont plus précises et authentiques par rapport à la situation réelle. De plus, l'analyse multivariée récursive serait une procédure révolutionnaire pour résoudre de nombreux sujets non paramétriques reposant sur des ensembles de données complexes.

L'objectif principal de cette thèse est de proposer une large classe d'estimateurs récursifs à noyau de différentes fonctionnelles multivariées basés sur la méthode d'approximation stochastique.
Tout d'abord, nous introduisons l'estimateur récursif de la densité de probabilité multivariée noté f_n et défini dans Mokkadem *et al.* (2009a).
Soit $X \in \mathbb{R}^d, d \geq 1$ et soient $X_1, \ldots, X_n$ des vecteurs aléatoires indépendants, identiquement distribués dans $\mathbb{R}^d$, et soient f et F la densité de probabilité et la fonction de distribution de X.

L'estimateur récursif multivarié de la densité:

Pour construire un algorithme stochastique, qui approxime la fonction f en un point donné x, nous devons définir un algorithme de recherche du zéro de la fonction

$$\phi : y \longmapsto f(x) - y.$$

Nous procédons donc de la manière suivante :

(i) On fixe $f_0(x) \in \mathbb{R}$.

(ii) Pour tout $n \geq 1$, on pose $f_n(x) = f_{n-1}(x) + \gamma_n T_n(x)$,

où $T_n(x)$ est une observation de la fonction ϕ au point $f_{n-1}(x)$ vérifiant (17) et (γ_n) est une suite positive satisfaisant (18).
Pour identifier $T_n(x)$, nous adoptons l'approche de Révész (1977) et de Tsybakov (1990), et insérons un noyau multivarié $\mathbf{K}$ (une fonction satisfaisant $\int_{\mathbb{R}^d} \mathbf{K}(t)dt = 1$), et une fenêtre (h_n) (une suite de nombres réels positifs qui vont jusqu'à zéro). Par conséquent, sous certaines conditions de régularité sur la fonction de densité de X, on a $\mathbb{E}\left[h_n^{-d} K\left(\frac{x - X_n}{h_n}\right)\right] = f(x) + \varepsilon_n(x)$, où $\varepsilon_n(x)$ devient nul lorsque n va à l'infini. Puis suivant (19), on pose

$$T_n(x) = h_n^{-d}\mathbf{K}\left(\frac{x - X_n}{h_n}\right) - f_{n-1}(x).$$

L'algorithme d'approximation stochastique que nous intégrons pour estimer récursivement la densité f au vecteur x peut donc être exprimé comme suit

$$f_n(x) = (1 - \gamma_n) f_{n-1}(x) + \gamma_n h_n^{-d}\mathbf{K}\left(\frac{x - X_n}{h_n}\right). \tag{6}$$

Une forme générale utilisant la récurrence, et sous la condition que $f_0(x) = 0$, a été proposée par Slaoui (2006) dans une version univariée et Mokkadem *et al.* (2009a) pour une version multivariée. Par conséquent, l'estimateur récursif de la densité est donné par

$$f_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \, h_k^{-d}\, \mathbf{K}\left(\frac{x - X_k}{h_k}\right) \quad \text{avec} \quad \Pi_n = \prod_{j=1}^{n}(1 - \gamma_j). \tag{7}$$

Remarque 0.1. *La relation (21) définit une classe entière d'estimateurs récursifs à noyau d'une densité de probabilité. Il est intéressant de noter que cette classe inclut la sous-classe suivante introduite dans Hall et Patil (1994). Étant donné (u_n) une suite positive non croissante telle que $\sum u_n = +\infty$ et lorsque le pas (γ_n) est choisi égal à $\left(u_n(\sum_{i=1}^{n} u_i)^{-1}\right)$, l'estimateur f_n (21) est alors exprimé comme suit*

$$f_n(x) = \frac{1}{\sum_{i=1}^{n} u_i} \sum_{k=1}^{n} u_k \, h_k^{-d}\, \mathbf{K}\left(\frac{x - X_k}{h_k}\right). \tag{8}$$

Nous pouvons mentionner quelques choix particuliers de (u_n). Plus tard, Amiri (2010) a exposé l'estimateur récursif de densité à noyau donné par le choix $(u_n) = \left(h_n^{d(1-l)}\right)$, $l \in [0, 1]$, qui englobe les trois exemples ci-dessous.

- (E_1) *Le choix $l = 1$, c'est-à-dire $(u_n) = 1$, qui correspond au cas $(\gamma_n) = \left(\frac{1}{n}\right)$, produit l'estimateur proposé par Wolverton and Wagner (1969).*
- (E_2) *Le choix $l = 1/2$, c'est-à-dire $(u_n) = (h_n^{d/2})$ produit l'estimateur considéré par Wegman and Davies (1979).*
- (E_3) *Le choix $l = 0$, c'est-à-dire $(u_n) = (h_n^d)$ donne l'estimateur considéré par Deheuvels (1973) et Duflo (1997).*

Comme premier résultat de recherche, nous présentons dans ce qui suit notre estimateur de fonction de répartition multivariée.

L'estimateur récursif multivariée de la fonction de répartition:

Pour construire un algorithme stochastique qui approche la fonction F d'un vecteur donné x, nous définissons un algorithme de recherche de la fonction zéro $\phi : y \longmapsto F(x) - y$ et nous fixons :

$$(i)\ F_0(x) \in [0,1] \qquad (ii)\ \text{pour tout } n \geq 1, F_n(x) = F_{n-1}(x) + \gamma_n T_n(x),$$

Afin de définir $T_n(x)$, nous adoptons l'approche de Slaoui (2014b) et nous introduisons un noyau multivarié modifié

$$\mathcal{K} :\ \mathbb{R}^d \longrightarrow [0,1]\,,\ x \longmapsto \int_{\prod_{i=1}^{d}(-\infty,x_i)} \mathbf{K}(t)dt.$$

En fixant $T_n(x) = \mathcal{K}\left(\frac{x - X_n}{h_n}\right) - F_{n-1}(x)$, l'algorithme d'approximation stochastique que nous considérons pour estimer récursivement la fonction de distribution F au vecteur x peut être indiqué comme suit

$$F_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \, \mathcal{K}\left(\frac{x - X_k}{h_k}\right). \tag{9}$$

Pour notre deuxième sujet, reposant sur l'estimation de la fonction de distribution, nous présenterons notre version conditionnelle multivariée notée $CCDF$, l'estimateur récursif de la fonction de distribution cumulative conditionnelle multivariée.

L'estimateur récursif de la fonction de distribution cumulative conditionnelle multivariée:

Soit (X, Y) un vecteur aléatoire à valeurs dans $\mathbb{R}^d \times \mathbb{R}^q$, $q \geq 1$, avec densité jointe $f_{(X,Y)}$ et soit f_X la densité de probabilité marginale de X. De plus, soit $(X_1, Y_1), \ldots, (X_n, Y_n)$ des vecteurs aléatoires indépendants identiquement distribués suivant (X, Y) . L'algorithme d'approximation stochastique qui est consacré à l'estimation récursive de la fonction

$$a : (x, y) \longmapsto \int_{\mathbb{R}^q} \mathbb{1}_{\{u \leqslant y\}} f_{(X,Y)}(x, u)du$$

à un couple de vecteurs (x, y) peut être énoncé comme suit :

$$a_n(x, y) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \, \chi_k(y) h_k^{-d} \, \mathbf{K}\left(\frac{x - X_k}{h_k}\right), \tag{10}$$

où χ est une fonction indicatrice multivariée identifiée par $\chi_k : \mathbb{R}^q \longrightarrow \mathbb{R},\ ; y \longmapsto \mathbb{1}_{\{Y_k \leqslant y\}}$.
Nous nous concentrons sur le problème de l'estimation de la CCDF de Y étant donné $X = x$, fournie par

$$\begin{aligned} \pi :\quad & \mathbb{R}^q \times \mathbb{R}^d \longrightarrow \mathbb{R} \\ & (y|x) \quad \longmapsto \mathbb{P}[Y \leqslant y | X = x] = \frac{a(x, y)}{f_X(x)}, \end{aligned}$$

Un estimateur récursif de π a été identifié dans Slama *et al.* (2021) et spécifié par

$$\pi_n(y|x) = \begin{cases} \dfrac{a_n(x, y)}{f_n(x)} & \text{si} \quad f_n(x) \neq 0 \\ 0 & \text{sinon} \end{cases}. \tag{11}$$

Dans une autre vision de l'estimation conditionnelle, nous étudions la régression multivariée dans un cas de type noyau.

L'estimateur récursif de la régression multivariée de type noyau:

Soit (X, Y) un vecteur aléatoire à valeurs dans $\mathbb{R}^d \times \mathbb{R}$ avec densité jointe $g(x, y)$ et soit f la densité de probabilité de X. De plus, soit $(X_1, Y_1), \ldots, (X_n, Y_n)$ des vecteurs aléatoires indépendants identiquement distribués suivant (X, Y). L'algorithme d'approximation stochastique, qui estime récursivement la fonction de régression

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

à un vecteur donné x, pour une fonction mesurable φ et $x \in \mathbb{R}^d$, peut être noté comme suit :

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

où (β_n) est une suite positive de nombres réels décroissant vers zéro satisfaisant (18).
Tout au long de ce travail, nous considérons l'estimateur général multivarié de type noyau pour la fonction de régression $r : x \longmapsto \mathbb{E}[\varphi(Y)|X = x]$ au niveau du vecteur x

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

Exemples particuliers sur l'estimation de la régression multivariée: Supposons que nous ayons une fonction mesurable $\varphi : \mathbb{R} \to \mathbb{R}$, on distingue les exemples particuliers suivants.

1. Pour $\varphi(y) := \mathbb{I}(y) = y$, nous avons la fonction de régression classique

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

   Un estimateur récursif de $r_{\mathbb{I}}$ a été traité dans Slaoui (2015).

2. Pour $\varphi(y) := \mathbb{I}(y) = y^m$, $m \in \mathbb{N}$, on a les moments conditionnels

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

3. Pour $\varphi(y) := \chi_t(y) = \mathbb{1}_{\{y \leqslant t\}}$, $; t \in \mathbb{R}$, nous avons la fonction de distribution cumulative conditionnelle

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

   Un estimateur récursif de r_{χ_t} a été identifié dans Slama *et al.* (2021).

En suivant l'esprit de l'estimation par régression, nous construisons une classe particulière d'estimateurs pour les régressions sous données manquantes. Cette dernière est donnée par l'approche du score de propension.

Application : estimation récursive d'une fonction de régression multivariée avec données manquantes :

Soit (X, T) un vecteur aléatoire à valeurs dans $\mathbb{R}^d \times \mathbb{R}$ avec une densité jointe $h(x, t)$ et soit f la densité de probabilité de X. De plus, soit $(X_1, T_1), \ldots, (X_n, T_n)$ des vecteurs aléatoires indépendants identiquement distribués suivant (X, T). En supposant que $T_1, \ldots, T_n$ sont sujets à des données manquantes, les variables aléatoires observées sont alors Y_i et δ_i, où

$$\delta_i = \mathbb{1}_{\{T_i \text{ est observé}\}} \text{ et } Y_i = T_i \times \delta_i, \text{ pour tout} i \in \{1, \ldots, n\}.$$

En conséquence, lorsque certains T_i sont manquants, nous introduisons le score de propension, une probabilité élaborée par Rosenbaum and Rubin (1983) et définie comme suit

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

Notre tâche principal dans ce chapitre est de proposer un estimateur récursif permettant d'estimer récursivement la fonction de régression $p(x) = \mathbb{E}[T|X = x]$ sous censure des données. Notre objectif réside alors dans la construction d'un algorithme stochastique, qui approche la fonction de régression

$$m : x \longmapsto \mathbb{E}[T|X = x]f(x) = \int_{\mathbb{R}} th(x, t)dt$$

à un vecteur donné x. L'algorithme d'approximation stochastique que nous considérons pour estimer récursivement la fonction de régression m à un vecteur x peut être exprimé comme suit

$$m_n(x) = Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k Y_k \psi_k^{-1} h_k^{-d} \mathbf{K}\left(\frac{x - X_k}{h_k}\right). \tag{12}$$

Un estimateur récursif de p a été proposé dans Slama *et al.* (2021). Ce dernier est identifié par

$$p_n(x) = \begin{cases} \dfrac{m_n(x)}{f_n(x)} & \text{si} \quad f_n(x) \neq 0 \\ 0 & \text{sinon} \end{cases}, \tag{13}$$

où

$$m_n(x) = Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k Y_k \psi_k^{-1} h_k^{-d} \mathbf{K}\left(\frac{x - X_k}{h_k}\right). \tag{14}$$

A ce niveau de l'analyse, nous pouvons affirmer que le point central de cette thèse réside dans l'extension des approches existantes au cas multidimensionnel, ce qui est une tâche épineuse, que ce soit d'un point de vue théorique ou pratique, afin d'obtenir des compétences de programmation appropriées pour l'exécution d'algorithmes de simulation ainsi que pour l'analyse d'ensembles de données réelles. Dans cette optique, nous avons tenté d'élargir les domaines de recherche non seulement en construisant des estimateurs récursifs fonctionnels mais aussi en automatisant le choix de la fenêtre où nous avons utilisé les trois méthodes les plus efficaces de sélection de la fenêtre, à savoir la méthode plug-in, la validation croisée et la procédure bootstrap.

La première section est investi comme une partie introductive. En effet, les techniques d'estimation non paramétrique multivariée ont été élaborées. Ainsi, nous avons mis en avant la fameuse estimation par la méthode des noyaux qui est un outil de base pour l'approche d'estimation récursive basée sur des méthodes d'approximation stochastique.

Cette thèse est composée de quatre chapitres principaux qui s'organisent comme suit.

Dans le premier chapitre, qui est une extension de mon projet de master, nous essayons de nous appuyer sur le travail de Slaoui (2014b) et de l'étendre au cas des données multivariées. Nous examinons les propriétés asymptotiques de cet estimateur généralisé et nous les comparons à celles de l'estimateur non récursif de la distribution multivariée de Nadaraya. Il s'avère que, avec un choix adéquat du pas (γ_n) et un choix approprié de la fenêtre (h_n), en investissant l'une des deux méthodes de sélection de la fenêtre, la procédure de validation croisée ainsi que la méthode du plug-in de deuxième génération, le MSE des estimateurs généralisés peut être plus petit que celui de l'estimateur de Nadaraya. Nous corroborons nos résultats théoriques par des études de simulation et en considérant quelques ensembles de données réelles telles qu'une application médicale unidimensionnelle des données de l'`hormone lutéinisante` dans des échantillons de sang féminin, une application bidimensionnelle des données des tremblements de terre

se produisant dans le nord-ouest de la péninsule ibérique `nwip` ainsi qu'une application multidimensionnelle des données de l'`iris` de Fisher ou d'Anderson.

Dans le même esprit que l'estimation d'une fonction de répartition multivariée mais dans un cadre conditionnel et en utilisant une approche récursive, notre deuxième chapitre aborde l'estimation non paramétrique d'une fonction de répartition cumulative conditionnelle (CCDF). En utilisant une approche récursive, nous présentons un estimateur récursif multivarié défini par un algorithme d'approximation stochastique. Notre objectif principal est d'étudier l'inférence statistique de notre estimateur et de la comparer à celle de l'estimateur non récursif de Nadaraya-Watson. Dans cette perspective, nous dérivons d'abord les propriétés asymptotiques de l'estimateur proposé qui dépendent fortement du choix de deux paramètres, le pas (γ_n) ainsi que la fenêtre (h_n). La méthode plug-in de deuxième génération implique le choix optimal de la fenêtre et maintient donc une sélection appropriée du paramètre pas. Fondamentalement, nous démontrons que, sous certaines conditions, le MSE de l'estimateur proposé peut être plus petit que celui de l'estimateur de Nadaraya Watson. Nous corroborons nos résultats théoriques par des études de simulation et deux applications de jeux de données réelles, à savoir le jeu de données de référence des compagnies d'assurance (COIL 2000) ainsi que les données hospitalières françaises de l'épidémie COVID-19.

En ce qui concerne le troisième chapitre, étant donné l'idée de l'estimation conditionnelle et en considérant un concept général, nous élaborons une extension de l'estimateur semi-récursif de la fonction de régression de type noyau. Nous étudions les propriétés asymptotiques de cet estimateur et nous les comparons à celles de l'estimateur de régression non récursif de Nadaraya Watson. Dans cette perspective, nous calculons d'abord le biais et la variance de l'estimateur proposé qui dépendent fortement du choix de trois paramètres, à savoir les pas (β_n) et (γ_n) ainsi que la fenêtre (h_n) choisie en utilisant l'une des meilleures méthodes de sélection de fenêtre, l'approche bootstrap combinée avec la méthode plug-in. Un choix judicieux de ces paramètres permet de constater que, sous certaines conditions, l'MSE de l'estimateur proposé peut être inférieur à celui de l'estimateur de Nadaraya Watson. Nous confirmons nos résultats théoriques par des études de simulation et en considérant deux applications de jeux de données réelles, à savoir les données hospitalières françaises de l'épidémie COVID-19 ainsi que la charge parasitaire de Plasmodium Falciparum (PL).

Enfin, en ce qui concerne le quatrième chapitre, notre objectif central est d'explorer les processus cognitifs et les représentations mentales mobilisés lorsqu'un être humain se prépare à écrire un mot selon l'idée développée dans Perret and Olive(2019). Pour ce faire, nous mettons en avant un estimateur multivarié non paramétrique de régression récursive à noyau sous données manquantes utilisant l'approche du score de propension afin de caractériser la production de mots écrits. Nous examinons les propriétés asymptotiques de l'estimateur récursif proposé et les comparons à l'estimateur de régression bien connu de Nadaraya-Watson. Nous calculons le biais et la variance de l'estimateur proposé qui dépendent du choix de certains paramètres tels que le pas et la fenêtre. Nous utilisons des procédures basées sur les données pour sélectionner ces paramètres. Ainsi, nous démontrons que, sous certains choix optimaux de ces paramètres, le MSE de l'estimateur proposé peut être plus petit que celui obtenu en utilisant l'estimateur de régression de Nadaraya Watson. L'estimateur développé est ensuite appliqué aux données comportementales afin de classer certains participants en groupes. Cette classification peut constituer un point de départ pour aborder les variations de comportement écrites.
Enfin, la dernière section résume la conclusion, fournit des remarques finales et offre de nouvelles perspectives pour les futur travaux de recherche.

Introduction

Statistical Analysis

Nowadays, data are everywhere. They are available to everybody. Data themselves are just facts and figures that need to be explored to get meaningful information. Therefore, data analysis is crucial. It stands for the process of applying statistical and logical techniques to visualize, reduce, describe and assess data into useful information that provides a better context for the data. Data analysis plays a key role in finding meaningful information which will help business take better decision basis the output. Along with Data analysis, Data science is equally significant. Data science corresponds to an area combining multiple methods of scientific methodology, processes, algorithms, and tools to extract information from, particularly huge datasets for insights on structured and unstructured data. A different range of terms related to mining, cleaning, analyzing, and interpreting data are often used interchangeably in data science.

Data analytics is concerned with examining various factors or variables to trace how they might impact certain situations and outcomes. This helps us understand why certain outcomes occur, which in turn allows us to make informed predictions and decisions for the future.
When elaborating data involving more than two variables, we shall use multivariate analysis which isn't just one specific method, it rather encompasses whole statistical techniques that are used to analyze more than two variables at once. These approaches allow us to gain a deeper insight of data in relation to specific business or real-world scenarios. As a matter of fact, if one is an ambitious data analyst or data scientist, multivariate analysis is an intrinsic concept to address. This technique would be the central focus of this thesis.

Introduction to estimation principle

Estimation statistics, or simply estimation corresponds to a data analysis framework that rests upon a combination of effect sizes, confidence intervals, precision planning to analyze data and interpret results. It differs from null hypothesis significance testing (NHST) which is less informative.
In statistics, an estimator is a function for estimating an unknown parameter related to a probability distribution. It can be invested to determine certain characteristics of a total population from observed data. For instance, in a survey, the sample mean is the most commonly used estimator of the population mean. Estimation approach is one of the most outstanding branches of statistics, concerned with the properties of estimators expressed in terms of their convergence, bias and efficiency. Numerous approaches can be elaborated in order to set forward estimators of multiple qualities for the same quantity and on the same database.
Furthermore, estimation theory can be divided into two main components, parametric and non-parametric estimation. Parametric statistics, which dates back to Fisher in 1920, stands for the classical framework of statistics. The statistical model is characterized by a finite number of parameters. We take $\mathcal{M} = \{\mathbb{P}_\theta : \theta \in \mathbb{R}^p\}$ as a typical statistical model that describes the distribution of observed random variables.

However, non-parametric statistics investigates statistical problems where the parametrization is not fixed, but there are various choices of parameters in order to find those entailing to the most efficient procedures. In fact, the law is entirely unknown. For this reason, we resort to the estimation of the functionals describing the model. We are basically interested in the issue of the non-parametric estimation of a distribution function F, a density f and a regression function r. Therefore, in order to settle a number of statistical problems, it is sufficient to have a suitable estimate.

Non-parametric estimation

Non-parametric estimation plays a significant role in exploring a myriad of phenomena with a random nature. It lies at the heart of modeling topic in real problems pertaining to multiple scientific areas, namely environmental sciences, seismology and psychology, medical imaging and neuroscience.

A wide range of methods for non parametric density estimation were elaborated, such as the Histogram estimator, the simple estimator and the kernel estimator. The histogram estimator, introduced by John Graunt (1662), is a stepped function, and therefore discontinuous, which is a deficiency for estimating a density. To overcome this issue, the indicator function is replaced by a real one, called the kernel estimator. Indeed, while a histogram counts the number of data points in quite arbitrary intervals, a kernel density estimate is a function defined as the sum of a kernel function on every data point. Therefore, Kernel density estimation is a critical data smoothing concern that allows inferences to be enacted about the population from a finite sample of data. The kernel estimation technique stands for the central focus of the following paragraph and the whole research work.

Kernel estimation

In statistics, kernel estimation refers to a non-parametric process of estimating an unknown probability functional of a random variable using a kernel function.
Basically, it is assumed that the kernel function typically exhibits the following properties:

i) $\displaystyle\int_{\mathbb{R}} \mathbf{K}(x)dx = 1.$

ii) $\displaystyle\int_{\mathbb{R}} \mathbf{K}^2(x)dx < +\infty.$

iii) $\displaystyle\int_{\mathbb{R}} x\mathbf{K}(x)dx = 0.$

iv) $\displaystyle\int_{\mathbb{R}} x^2\mathbf{K}(x)dx < +\infty.$

Examples 0.2. *The most commonly used kernels are:*

* *The uniform kernel:* $\mathbf{K}(x) = \dfrac{1}{2}\mathbb{1}_{[-1\,,1]}(x).$
* *The triangular kernel:* $\mathbf{K}(x) = (1-|x|)\,\mathbb{1}_{[-1\,,1]}(x)$.
* *The Epanechnikov kernel:* $\mathbf{K}(x) = \dfrac{3}{4}\left(1-x^2\right)\mathbb{1}_{[-1\,,1]}(x).$
* *The Gaussian kernel:* $\mathbf{K}(x) = \dfrac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}.$

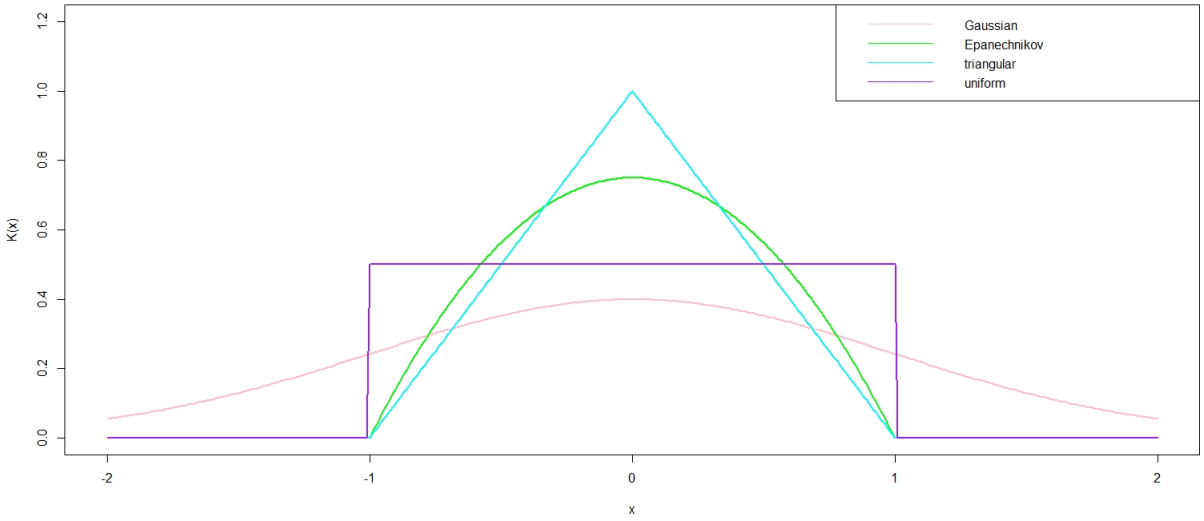


Figure 1: Representation of some classical kernels.

We can distinguish three kernel estimators. We observe the Parzen-Rosenblatt estimator for the probability density, the Nadaraya estimator for the distribution function and the Nadaraya-Watson estimator for the regression function.

Let (X, Y) be a random vector with values in $\mathbb{R} \times \mathbb{R}$ with a joint density function $g(x, y)$ and let f and F denote the probability density and the distribution function of X which are unknown. Moreover, let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent random vectors identically distributed as (X, Y).

The kernel density estimator: Rosenblatt (1956) and Parzen (1962)

The kernel density estimator of the density function f investigated by Murray Rosenblatt (1956) and Emanuel Parzen (1962) and called the Parzen-Rosenblatt's estimator is indicated by

$$\widetilde{f}_n(x) = \frac{1}{nh_n} \sum_{k=1}^{n} \mathbf{K}\left(\frac{x - X_k}{h_n}\right),$$

where $\mathbf{K}$ refers to the kernel function and (h_n) denotes the smoothing parameter (the bandwidth), a positive deterministic sequence tending to zero.

The kernel distribution function estimator: Nadaraya (1964)

The kernel distribution function estimator of the distribution function F elaborated by Nadaraya (1964) and called Nadaraya's estimator is expressed by

$$\widetilde{F}_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathcal{K}\left(\frac{x - X_k}{h_n}\right),$$

with $\mathcal{K}(z) = \displaystyle\int_{-\infty}^{z} \mathbf{K}(u)du.$

The kernel regression estimator: Nadaraya-Watson (1964)

The kernel regression estimator of the regression function $r : x \mapsto \mathbb{E}[Y|X = x]$ identified by Nadaraya (1964) and Watson (1964) and called Nadaraya-Watson's estimator is denoted by

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

with $\widetilde{a}_n(x) = \dfrac{1}{nh_n} \sum_{k=1}^{n} Y_k \mathbf{K}\left(\dfrac{x - X_k}{h_n}\right)$.

For the adequate construction of these estimators, two pertinent choices shall be made: the kernel function K and the smoothing parameter h_n. Since the kernel is barely sensitive to the shape of the estimator, the choice of the kernel function is not critical and of less importance than that of the bandwidth.
Although theoretical asymptotic study yields the optimal bandwidth, not knowing the density function makes it quite difficult to interpret it in practice. From this perspective, kernel smoothing in non-parametric statistics requires the choice of a bandwidth parameter which is crucial for the effective performance of the estimators. An appropriate bandwidth can help obtain an estimated functional close to the true one. However, a poorly chosen bandwidth can seriously distort the true underlying characteristics of the functional. In fact, a small bandwidth leads to undersmoothing and a large bandwidth leads to oversmoothing. Thus, judicious choice of bandwidth should be highly recommended.

Bandwidth selection:

In order to determine the optimal choice of the smoothing parameter, one should find the parameter that minimises the risk. The most widely used criteria for the choice of h_n are the MSE (Mean Squared Error), $MISE$ (Mean Integrated Squared Error) and the $MWISE$ (Mean Weighted Integrated Squared Error). The $MWISE$ and $MISE$ correspond to the choice of the constant smoothing parameter and the MSE refers to the variable smoothing parameter. Throughout the whole thesis work, the optimal value of the bandwidth h_n is obtained through minimizing the asymptotic $MWISE$.
There are a myriad of data-driven bandwidth selection methods recorded in literature which can be divided into three broad classes: cross-validation techniques, second generation plug-in methods, and the bootstrap approach. The cross-validation selection method introduced by Sarda (1993) stands for a resampling technique that consists of minimising a suitable estimator of the mean squared error. Yet, this method has certain short comings as highlighted by Altman and Leger (1995) who developed another efficient method, a plug-in estimate which minimizes an estimate of the mean weighted integrated squared error, using the density function as a weight function. An alternative method is the wild bootstrap, introduced in Hardle and Marron (1991) which lies in resampling from the estimated residuals. Both, the plug-in method and the bootstrap one are indistinguishable and it has been widely proven that they behave similarly. A detailed comparison of the three practical bandwidth selection techniques is exhibited in Delaigle and Gijbels (2004).

Throughout the entire thesis, the non-parametric approaches we are concerned with rest upon the Stochastic Approximation Method denoted SAM. The latter is investigated in what follows.

Stochastic approximation method

Over the past decade, data streams have become an increasingly important area of research. Some of the most common data streams include Internet packet data, Twitter activity, Facebook newsfeed, credit card transactions and more recently COVID-19 epidemic data. Stochastic algorithms have been commonly used in many research applications, namely system identification, adaptive control, transmission systems and sequential change detection.
Recursivity can be crucial when one seeks to infer these kind of phenomena that evolve over time and that require constant updating of the estimates made. In such situations, the data arrives regularly so that it is impossible to store it in a traditional database. In such a context, it is very interesting to build a recursive estimator that does not need to store all the data in memory and that can be easily updated to handle the online data.
Suppose we have a big number of data, we need to store a lot of data in order to recalculate them and then the time of execution would be enormous, since adding a new observation means that non-recursive estimators have to be completely recalculated. For reasons of time calculation optimization and the nature of the data studied, we have chosen to study recursive estimators which can be updated with each new observation added to the database. Therefore, instead of re-running the data each time, it is possible to rewrite our considered estimator as a combination of two (or more) estimators, where each estimator is based on separate datasets, which can be very interesting to keep the computational cost reasonably low. Moreover, recursive estimators can be preferable to non-recursive versions because of their lower asymptotic variance.
It is noteworthy that all computation and simulation have been done using the **R** statistical software.

Historical framework

Stochastic approximation algorithms correspond to stochastic versions of deterministic algorithms like Newton's algorithm or the gradient algorithm. We are basically interested in the following problem: finding the zero x^* of a real-valued function S. We suppose to have a very approximate value x_0 of this root. The natural idea of Newton's algorithm is to replace the representative curve of the function S by its tangent at the point x_n. The abscissa x_{n+1} of the intersection point of this tangent with the x-axis is then provided for $n \geq 1$ by

$$x_{n+1} = x_n - \frac{S(x_n)}{S'(x_n)}.$$

When the derivative of S is not easily computable, one can consider a deterministic version of the Robbins-Monro algorithm consisting of replacing the computation of the derivative with a decreasing positive sequence tending to 0 of steps (γ_n). Provided that S has good regularity properties, the sequence expressed by

$$x_{n+1} = x_n - \gamma_n S(x_n)$$

converges to the zero of the function S, denoted x^* , for any initial value x_0. In many situations the function whose zero is sought is known only at a perturbation close to zero. The search for zeros by deterministic optimization methods therefore becomes more perilous and one needs to resort to stochastic algorithms like the one introduced by Robbins and Monro (1951).

The recursive estimators that we propose were constructed based on dint of SAM. Indeed, incorporating stochastic approximation algorithms in the context of non-parametric statistics dates back to the papers Robbins and Monro (1951) and Kiefer and Wolfowitz (1952) for a unidimensional framework.
The general stochastic algorithm, devoted essentially to the approximation of the mode of a regression function, has the form:

$$\theta_n = \theta_{n-1} + \gamma_n \Phi\left(\theta_{n-1}, W_n\right) + \gamma_n^2 \mu_n\left(\theta_{n-1}, W_n\right), \tag{15}$$

where :

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

. (θ_n) is the sequence to be updated recursively,

. (W_n) is a sequence of random variables representing the online observations,

. $\Phi(\theta, W)$ is the function which essentially determines how the parameter is updated according to a new observation,

. $\mu_n(\theta_{n-1}, W_n)$ defines a small perturbation on the algorithm.

Subsequently, Blum (1954) provided a multidimensional version of this algorithm. These research works have been extended in several directions. Inspired by the most prominent ones, we can introduce the following algorithm analyzed by Kushner and Clark (1978), Ruppert (1982).

$$\theta_n = \theta_{n-1} + \gamma_n \left(\phi(\theta_{n-1}) - W_n + \beta_n\right), \tag{16}$$

where (β_n) refers to a random variable converging to 0 almost surely and ϕ denotes an unknown measurable function. They demonstrated that (16) coincides with the (15) algorithm and includes the stochastic approximation processes of Robbins and Monro (1951) and Kiefer and Wolfowitz (1952), which allow the search for the zero θ^* of the ϕ function.
Some basic modifications were incorporated by Nadaraya (1964), Polyak and Tsybakov (1990), Dippon and Renz (1997) and Dippon (2003). Later, Duflo (1997) corroborated that, under standard conditions on the function ϕ and on the sequence (γ_n), (θ_n) tends to θ^* almost surely.
In order to build up a stochastic approximation algorithm, Révész (1973) introduced an application of the Robbins-Monro procedure and Mokaddem and Pelletier (2007b) developed the application of the Robbins-Monro-Blum procedure. In addition, an algorithm for estimating a regression function was elaborated by Révész (1977) and was afterwards used by Tsybakov (1990) to approximate the mode of a probability density.
Latterly, Slaoui (2006) reported a smooth estimator of the density function in a unidimensional case using the SAM. Next, in the paper of Mokkadem *et al.* (2009a), the multidimensional case was investigated in order to estimate a multivariate probability density using the estimation by confidence intervals. Additionally, Slaoui (2014b) reused SAM to enhance the qualities of the distribution function estimator. The large and moderate deviation principle of this estimator was proven in Slaoui (2019).
The classical recursive regression estimator was addressed in Mokaddem *et al.* (2009b) for univariate framework and a multivariate extension of this estimator was carried out by Mokaddem and Pelletier(2016). Subsequently, Slaoui (2016) established the semi-recursive case and introduced a new estimator which is the fraction of a recursive regression by a recursive density function. Afterward, Slaoui (2017) adopted the propensity score probability technique and constructed an estimator of the density function under missing data. More recently, a conditional density estimator was set forward in Slaoui and Khardani (2020).
The basic target of this chapter is to use stochastic approximation algorithms to define estimators of a probability density at a given point.

Let us recall that stochastic approximation algorithms used for the search of the zero of an unknown function $\phi : y \longmapsto S(x) - y$ are built up as follows.

(i) We set $S_0(x) \in \mathbb{R}$ arbitrarily.

(ii) For all $n \geq 1$, the sequence (S_n) is recursively defined by

$$S_n(x) = S_{n-1}(x) + \gamma_n T_n(x),$$

where

- (T_n) is a sequence of functions $T_n : \mathbb{R} \to \mathbb{R}$ defined by

$$T_n(x) = \phi(S_{n-1}) - W_n + \beta_n. \tag{17}$$

  In other words, $T_n(x)$ is an observation of the function ϕ at the point $S_{n-1}(x)$.

- (γ_n) is a sequence of positive real numbers that goes to zero such that

$$\sum \gamma_n = +\infty \quad \text{and} \quad \sum \gamma_n^2 < +\infty. \tag{18}$$

It's noteworthy that, under the condition $\mathbb{E}\left[W_n|\mathcal{F}_{n-1}\right] = 0$ (where $\mathcal{F}_{n-1}$ stands for the σ-algebra of the events occurring at the time $n-1$), we have

$$\mathbb{E}[T_n(x)] = \phi(S_{n-1}) + \beta_n = S(x) - S_{n-1}(x) + \beta_n. \tag{19}$$

Recursive multivariate estimation:

Multivariate statistics centers around exploring the relationships between variables and their goodness of fit to the problem in question involving several types of univariate and multivariate analyses. Basically, the latter relies upon statistical procedure comprising the simultaneous measurements and observations of data including more than one factor of independent variables that impact the variability of dependent variables. Therefore, the major merit of multivariate analysis lies in the fact that as it takes into account more than one outcome variable where various different quantities are of interest to the same analysis, the obtained conclusions are more accurate and authentic to the real-life situation. Further more, recursive multivariate analysis would be a revolutionary procedure in terms of fixing many non-parametric topics relying on complex sets of data.

The major objective of this thesis is to set forward a large class of recursive kernel estimators of different multivariate functionals based on the stochastic approximation method.

First of all, we introduce the recursive multivariate probability density estimator noted f_n which is defined in Mokkadem *et al.* (2009a).
Let $X \in \mathbb{R}^d, d \geq 1$ and let $X_1, \ldots, X_n$ be independent, identically distributed $\mathbb{R}^d$-valued random vectors, and let f and F denote the probability density and the distribution function of X.

The multivariate recursive density estimator:

To construct a stochastic algorithm, which approximates the function f at a given point x, we need to define an algorithm of search of the zero of the function

$$\phi : y \longmapsto f(x) - y.$$

We thus proceed in the following way :

(i) We set $f_0(x) \in \mathbb{R}$.

(ii) For all $n \geq 1$, we set $f_n(x) = f_{n-1}(x) + \gamma_n T_n(x)$,

where $T_n(x)$ is an observation of the function ϕ at the point $f_{n-1}(x)$ verifying (17) and (γ_n) is a positive sequence satisfying (18).
To identify $T_n(x)$, we adopt the approach of Révész (1977) and of Tsybakov (1990), and insert a

multivariate kernel $\mathbf{K}$ (a function satisfying $\int_{\mathbb{R}^d} \mathbf{K}(t)dt = 1$), and a bandwidth (h_n) (a sequence of positive real numbers that goes to zero). Therefore, under some regularity conditions on the density function of X, we have $\mathbb{E}\left[h_n^{-d} K\left(\frac{x - X_n}{h_n}\right)\right] = f(x) + \varepsilon_n(x)$, where $\varepsilon_n(x)$ goes to zero as n goes to infinity. Then following (19), we set

$$T_n(x) = h_n^{-d}\mathbf{K}\left(\frac{x - X_n}{h_n}\right) - f_{n-1}(x).$$

The stochastic approximation algorithm we integrate to recursively estimate the density f at the vector x can thus be expressed as

$$f_n(x) = (1 - \gamma_n) f_{n-1}(x) + \gamma_n h_n^{-d}\mathbf{K}\left(\frac{x - X_n}{h_n}\right). \tag{20}$$

A general form using recurrence, and under the condition that $f_0(x) = 0$, was proposed by Slaoui (2006) in a univariate version and Mokkadem *et al.* (2009a) for a multivariate one. Therefore, the recursive density estimator is given by

$$f_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \, h_k^{-d}\, \mathbf{K}\left(\frac{x - X_k}{h_k}\right) \quad \text{with} \quad \Pi_n = \prod_{j=1}^{n}(1 - \gamma_j). \tag{21}$$

Remark 0.3. *The relation (21) defines a whole class of recursive kernel estimators of a probability density. It's noteworthy that this class includes the following subclass introduced in Hall and Patil (1994). Given (u_n) a nonincreasing positive sequence such that $\sum u_n = +\infty$ and when the stepsize (γ_n) is chosen equal to $\left(u_n(\sum_{i=1}^{n} u_i)^{-1}\right)$, the f_n estimator (21) is then expressed as*

$$f_n(x) = \frac{1}{\sum_{i=1}^{n} u_i} \sum_{k=1}^{n} u_k \, h_k^{-d}\, \mathbf{K}\left(\frac{x - X_k}{h_k}\right). \tag{22}$$

Later Amiri (2010) exhibited the recursive kernel density estimator given by the choice $(u_n) = \left(h_n^{d(1-l)}\right)$, $l \in [0, 1]$, which englobes the three examples below.

- (E_1) *The choice $l = 1$, i.e $(u_n) = 1$, which corresponds to the case $(\gamma_n) = \left(\frac{1}{n}\right)$, produces the estimator proposed by Wolverton and Wagner (1969).*
- (E_2) *The choice $l = 1/2$, i.e $(u_n) = (h_n^{d/2})$ yields the estimator considered by Wegman and Davies (1979).*
- (E_3) *The choice $l = 0$, i.e $(u_n) = (h_n^d)$ gives the estimator considered by Deheuvels (1973) and Duflo (1997).*

As a first research result, we introduce in the following our proposed multivariate distribution function estimator.

The multivariate recursive distribution function estimator:

To build up a stochastic algorithm which approaches the function F to a given vector x, we define an algorithm of search of the zero function $\phi : y \longmapsto F(x) - y$ and we set:

$$(i)\ F_0(x) \in [0, 1] \qquad (ii) \text{ for all } n \geq 1, F_n(x) = F_{n-1}(x) + \gamma_n T_n(x),$$

In order to define $T_n(x)$, we adopt the approach of Slaoui (2014b) and we introduce a modified multivariate kernel

$$\mathcal{K}: \mathbb{R}^d \longrightarrow [0,1], \; x \longmapsto \int_{\prod_{i=1}^{d}(-\infty,x_i)} \mathbf{K}(t)dt.$$

By setting $T_n(x) = \mathcal{K}\left(\dfrac{x-X_n}{h_n}\right) - F_{n-1}(x)$, the stochastic approximation algorithm that we consider to estimate recursively the distribution function F at the vector x can be indicated as follows

$$F_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \, \mathcal{K}\left(\frac{x-X_k}{h_k}\right). \tag{23}$$

For our second topic, resting upon the distribution function estimation, we shall present our multivariate conditional version denoted $CCDF$, the multivariate recursive conditional cumulative distribution function estimator.

The multivariate recursive conditional cumulative distribution function estimator:

Let (X,Y) be a random vector with values in $\mathbb{R}^d \times \mathbb{R}^q$ with a joint density function $f_{(X,Y)}$ and let f_X denote the marginal probability density of X. Moreover, let $(X_1,Y_1),\ldots,(X_n,Y_n)$ be independent random vectors identically distributed as (X,Y).
The stochastic approximation algorithm that is devoted to estimate recursively the function

$$a: (x,y) \longmapsto \int_{\mathbb{R}^q} \mathbb{1}_{\{u \leqslant y\}} f_{(X,Y)}(x,u)du$$

at a couple of vectors (x,y) can be stated as follows

$$a_n(x,y) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \, \chi_k(y) h_k^{-d} \, \mathbf{K}\left(\frac{x-X_k}{h_k}\right), \tag{24}$$

where χ is a multivariate indicator function identified by $\chi_k : \mathbb{R}^q \longrightarrow \mathbb{R}, \; y \longmapsto \mathbb{1}_{\{Y_k \leqslant y\}}$.
Our central focus is upon the problem of estimating the CCDF of Y given $X = x$ provided by

$$\begin{aligned} \pi: \quad & \mathbb{R}^q \times \mathbb{R}^d \longrightarrow \mathbb{R} \\ & (y|x) \quad \longmapsto \mathbb{P}[Y \leqslant y | X = x] = \frac{a(x,y)}{f_X(x)}, \end{aligned}$$

A recursive estimator of π was identified in Slama *et al.* (2021) and specified by

$$\pi_n(y|x) = \begin{cases} \dfrac{a_n(x,y)}{f_n(x)} & \text{if} \quad f_n(x) \neq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{25}$$

In an other vision of conditional estimation, we investigate the multivariate regression in a kernel-type case.

The multivariate recursive kernel-type regression estimator:

Let (X,Y) be a random vector with values in $\mathbb{R}^d \times \mathbb{R}$ with a joint density function $g(x,y)$ and let f denote the probability density of X. Moreover, let $(X_1,Y_1),\ldots,(X_n,Y_n)$ be independent random vectors identically distributed as (X,Y).
The stochastic approximation algorithm, which estimates recursively the regression function

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

at a given vector x, for a chosen measurable function φ and $x \in \mathbb{R}^d$, can be denoted as follows

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

where (β_n) is a positive sequence of real numbers decreasing towards zero satisfying (18). Throughout this work, we consider the general multivariate kernel-type estimator for the regression function $r_\varphi : x \longmapsto \mathbb{E}[\varphi(Y)|X = x]$ at the vector x

$$r_{\varphi_n}(x) = \begin{cases} \dfrac{a_{\varphi_n}(x)}{f_n(x)} & \text{if} \quad f_n(x) \neq 0 \\ 0 & \text{if} \quad f_n(x) = 0 \end{cases} . \tag{26}$$

Particular regression estimation cases: Suppose we have a measurable function $\varphi : \mathbb{R} \to \mathbb{R}$, we distinguish the following particular examples.

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

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

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

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

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

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

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

   A recursive estimator of r_{χ_t} was identified in Slama *et al.* (2021).

Following the spirit of regression estimation, we construct a particular class of estimators for regressions under missing data. This latter is given by the propensity score approach.

Application: recursive estimation of multivariate regression function under missing data:

Let (X, T) be a random vector with values in $\mathbb{R}^d \times \mathbb{R}$ with a joint density function $h(x, t)$ and let f denote the probability density of X. Moreover, let $(X_1, T_1), \ldots, (X_n, T_n)$ be independent random vectors identically distributed as (X, T). Assuming that $T_1, \ldots, T_n$ are subjects to missing data, the observed random variables are then Y_i and δ_i, where

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

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

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

Our basic purpose in this chapter 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 $m : x \longmapsto \mathbb{E}[T|X = x]f(x) =$

$\int_{\mathbb{R}} th(x,t)dt$ at a given vector x. The stochastic approximation algorithm that we consider to estimate recursively the regression function m at a vector x can be expressed as follows

$$m_n(x) = Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k \, Y_k \psi_k^{-1} h_k^{-d} \mathbf{K}\left(\frac{x - X_k}{h_k}\right). \tag{27}$$

A recursive estimator of p was recorded in Slama *et al.* (2021) and identified by

$$p_n(x) = \begin{cases} \dfrac{m_n(x)}{f_n(x)} & \text{if} \quad f_n(x) \neq 0 \\ 0 & \text{if} \quad f_n(x) = 0 \end{cases}, \tag{28}$$

Thesis contributions

At this stage of analysis, we would assert that the central focus in this thesis resides in extending existing approaches to the multidimensional case, which is a thorny task, either from a theoretical or a practical point of view, in order to achieve fitting programming skills for running simulation algorithms as well as real datasets analysis. From this perspective, we attempted to broaden the research areas by not only building functional recursive estimators but also by automating the choice of the bandwidth where we used the three most effective methods of bandwidth selection, namely the plug-in method, the cross validation and the bootstrap procedure.

This section was invested as an introductory part. Indeed, the techniques for multivariate non-parametric estimation were elaborated. Therefore, we set forward the famous kernel method estimate which is a basic tool for the recursive estimation approach based on SAM.

This thesis is composed of four main chapters which are laid out as follows.

In the first chapter, being an extension to my master project, we attempt to build upon the work of Slaoui (2014b) and extend it to the case of multivariate data. We examine the asymptotic properties of this generalized estimator and we compare them with those of the non-recursive Nadaraya's multivariate distribution estimator. It turns out that, with an adequate choice of the stepsize (γ_n) and an appropriate choice of the bandwidth (h_n), investing one of both methods of bandwidth selection, the cross-validation procedure as well as the second generation plug-in method, the MSE of the generalized estimators can be smaller than the one of Nadaraya's estimator. We corroborate our theoretical results through simulation studies and by considering some real datasets such that a medical unidimensional application of the `luteinizing hormone` in female blood samples data, a bidimensional application of the earthquakes occurring in the Northwest of the Iberian Peninsula `nwip` data as well as a multidimensional application of Fisher's or Anderson's `iris` data.

Proceeding with the same spirit in terms of estimating a multivariate distribution function but in a conditional framework, our second chapter tackles non-parametric estimation of a conditional cumulative distribution function (CCDF). Using a recursive approach, we set forward a multivariate recursive estimator defined by stochastic approximation algorithm. Our basic objective is to investigate the statistical inference of our estimator and compare it with that of non-recursive Nadaraya-Watson's estimator. From this perspective, we first derive the asymptotic properties of the proposed estimator which highly depend on the choice of two parameters, the stepsize (γ_n) as well as the bandwidth (h_n). The second generation plug-in method entails the optimal choice of the bandwidth and therefore maintains an appropriate selection of the stepsize parameter.

Basically, we demonstrate that, under some conditions, the MSE of the proposed estimator can be smaller than the one of Nadaraya Watson's estimator. We corroborate our theoretical results through simulation studies and two real dataset applications, namely the Insurance Company Benchmark (COIL 2000) dataset as well as the French Hospital Data of COVID-19 epidemic.

As far as the third chapter is concerned, given the idea of conditional estimation and considering a general concept, we elaborate an extension of the semi-recursive kernel-type regression function estimator. We investigate the asymptotic properties of this estimator and compare them with non-recursive Nadaraya Watson regression estimator ones. From this perspective, we first calculate the bias and the variance of the proposed estimator which strongly depend on the choice of three parameters, namely the stepsizes (β_n) and (γ_n) as well as the bandwidth (h_n) chosen using one of the best methods of bandwidth selection, the bootstrap approach combined with the plug-in method. A convenient choice of those parameters yields that, under some conditions, the MSE of the proposed estimator can be smaller than that of Nadaraya Watson's estimator. We confirm our theoretical results through simulation studies and by considering two real dataset applications, namely the French Hospital Data of COVID-19 epidemic as well as the Plasmodium Falciparum Parasite Load (PL).

Last but not least, in terms of the fourth chapter, our central 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 Perret and Olive(2019). For this purpose, we foreground a non-parametric multivariate recursive kernel regression estimator under missing data using the propensity score approach so as to characterize writing word production. We examine 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 handle some data-driven procedures to select these parameters. Thus, we demonstrate that, under some optimal choices of these parameters, the MSE of the proposed estimator can be smaller than the one obtained by using Nadaraya Watson's regression estimator. The developed estimator is then applied to the behavioral data so as to classify some participants in groups. This classification may stand for a departure point to tackle written behavior variations.

Eventually, the closing section wraps up the conclusion, provides outstanding concluding remarks and offers new perspectives for future research works.

Chapter 1

Multivariate distribution function estimation using stochastic approximation method

1.1 Introduction

The estimation of a distribution function stands for an intrinsic tool in the study of multiple random phenomena. In fact, it's at the heart of modeling questions in real problems of such scientific fields, such as environmental sciences, seismology and particularly in biology, medical imaging and neuroscience. Numerous methods were set forward and explored in order to elaborate an effective estimator compared to the empirical distribution function. Our basic method applied in this study will be the non-parametric adaptive kernel density estimator for recursive multivariate distribution function estimation.
Inspired by the most prominent expansions, Slaoui (2014b) reused stochastic approximation methods to set forward an univariate recursive estimator of the distribution function. Thereafter, he provided a special stepsize and an adequate plug-in bandwidth selection to achieve a better estimation compared to Nadaraya's unidimensional data estimation. The large and moderate deviation principle of this estimator was proven in Slaoui (2019). The central objective of this chapter lies in extending this latter estimator to a multidimensional case. We demonstrated that, under some suitable conditions, essentially on the stepsize and on the bandwidth, the generalized estimators are closer to the true distribution function compared to the non-recursive Nadaraya's multidimensional distribution estimator.

1.1.1 Presentation of the method

Let $X, X_1, \ldots, X_n$ be independent, identically distributed $\mathbb{R}^d$-valued random vectors, $d \geq 1$, and let f and F denote the probability density and the distribution function of X.
To build up a stochastic algorithm which approaches the function F to a given vector $x \in \mathbb{R}^d$, we define an algorithm of search of the zero function $\phi : y \longmapsto F(x) - y$ and we set:

$$(i)\ F_0(x) \in [0,1] \qquad (ii)\ \text{for all } n \geq 1, F_n(x) = F_{n-1}(x) + \gamma_n T_n(x),$$

where (γ_n) is a positive sequence of real numbers decreasing towards zero satisfying (18) and $T_n(x)$ is an observation of the function ϕ at the point $F_{n-1}(x)$ verifying (17).
In order to define $T_n(x)$, we first introduce a modified multivariate kernel

$$\mathcal{K} :\ \mathbb{R}^d \longrightarrow [0,1]\,,\ x \longmapsto \int_{\prod_{i=1}^{d}(-\infty, x_i)} \mathbf{K}(t)dt.$$

By setting $T_n(x) = \mathcal{K}\left(\dfrac{x - X_n}{h_n}\right) - F_{n-1}(x)$, the stochastic approximation algorithm that we consider to estimate recursively the distribution function F at the vector x can be expressed as follows :

$$F_n(x) = (1 - \gamma_n)F_{n-1}(x) + \gamma_n \mathcal{K}\left(\frac{x - X_n}{h_n}\right). \tag{1.1}$$

Throughout this chapter, we consider that $F_0(x) = 0$. Then, our estimator (4.1) can be rewritten

$$F_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \mathcal{K}\left(\frac{x - X_k}{h_k}\right), \text{ with } \Pi_n = \prod_{j=1}^{n}(1 - \gamma_j). \tag{1.2}$$

Our major aim in this part is to explore the asymptotic properties of our generalized recursive kernel estimator of a distribution in the case of multivariate data and subsequently confirm its performances. We are basically interested also in comparing our generalized estimator to the generalized non-recursive multivariate distribution estimator of Nadaraya $\widetilde{F}_n$ indicated by:

$$\widetilde{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}\left(\frac{x - X_i}{h_n}\right). \tag{1.3}$$

For this reason, we attempt first to calculate the bias and the variance of the generalized estimator F_n which largely depends on the choice of two parameters, namely the stepsize (γ_n) and the bandwidth (h_n). Moreover, we introduce the asymptotic properties of multivariate Nadaraya's estimator F_n. It turns out that, by using an adequate choice of the stepsize of the proposed algorithm, the Mean Weighted Integrated Squared Error of the generalized estimators is smaller than the one of Nadaraya's estimator. Furthermore, we depict that with an appropriate choice of the bandwidth (h_n), using one of two data-driven bandwidth selections; the cross-validation procedure as well as the second generation plug-in method, the MSE of the generalized estimators can be smaller than the one of Nadaraya's multidimensional estimator. Likewise, we confirm that in our context the plug-in method can be more efficient than the cross-validation. We corroborate these theoretical results with some simulations as well as real datasets. We consider as a first application, a real dataset in a unidimensional case, the `lh` (Luteinizing Hormone in Blood Samples data) available on the `datasets` package for a biological concept. As a bivariate example, we consider the `nwip` data available on the `kerdiest` package for a seismology concept. Moreover, for a multidimensional application, we consider the `iris` data available on the `datasets` package for an environmental sciences concept.

First of all, let us recall the following definition of the class of regularly varying sequences.

Definition 1.1. *Let $(v_n)_{n\geq 1}$ be a nonrandom positive sequence and $\gamma \in \mathbb{R}$. We state that*

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

This définition was introduced by Galambos and Seneta (1973) to characterize regularly varying sequences and by Mokaddem and Pelletier (2007a) in the context of stochastic approximation algorithms. Note that the acronym GS stands for (Galambos and Seneta).
Typical sequences in $\mathcal{GS}(\gamma)$ are, for $b \in \mathbb{R}, n^\gamma, n^\gamma(\log n)^b, n^\gamma(\log\log n)^b$, and so on.

At this stage of analysis, it is worth mentioning that the following assumptions will be useful for the whole thesis investigation.

(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) $(\gamma_n)_{n\geq 1} \in \mathcal{GS}(-\alpha)$, with $\alpha \in \left(\frac{1}{2}, 1\right]$.
(ii) $(h_n)_{n\geq 1} \in \mathcal{GS}(-a)$, with $a \in (0,1)$.
(iii) $\lim_{n\to+\infty} n\gamma_n \in \left(\min\{2a, \frac{a+\alpha}{2}\}, +\infty\right]$.

Discussion of the assumptions:
All these assumptions are standard and are generally assumed within the context of non-parametric estimation. Classical assumption (A_1) provides regularity conditions on the kernel density estimator introduced by Rosenblat (1956) and Parzen (1962). It is widely used in the non-parametric framework for the functional estimation. Assumption (A_2) on the stepsize and the bandwidth was used in the recursive framework for the estimation of the density function in Mokkadem *et al.* (2009a), Slaoui (2014a) and for the distribution function estimation in Slaoui (2014b). Furthermore, it is notable that the assumption $(A_2)(iii)$ regarding the limit of $(n\gamma_n)$ as n goes to infinity is quite common within the context of stochastic approximation algorithms. More specifically, the limit $\xi := \lim_{n\to+\infty} (n\gamma_n)^{-1}$ is implied to be finite.

In what follows, we introduce two lemmas that will be widely invested throughout the theoretical studies of our recursive estimators. The proof of the first lemma was provided in Mokkadem *et al.* (2009a).

Lemma 1.2. *Let $(v_n)_{n\geq 1} \in \mathcal{GS}(v^*), v^* \in \mathbb{R}$, $(\gamma_n)_{n\geq 1} \in \mathcal{GS}(-\alpha)$ and let $m > 0$ such that $m - v^*\xi > 0$. Therefore,*

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

Moreover, for any positive sequence $(\alpha_n)_{n\geq 1}$ such that $\lim_{n\to+\infty} \alpha_n = 0$ 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 second lemma is a direct consequence of the definition 1.1 and is exhibited so as to help the reader simplify the assessment of proofs details.

Lemma 1.3. *Let $(a_n)_{n\geq 1} \in \mathcal{GS}(a)$ and $(b_n)_{n\geq 1} \in \mathcal{GS}(b)$, $a, b \in \mathbb{R}$. Hence,*

$$\left(\frac{a_n^k}{b_n^z}\right)_{n\geq 1} \in \mathcal{GS}(ka - zb), \text{ with } k, z \in \mathbb{R}.$$

In particular, we get $(a_n^{-1})_{n\geq 1} \in \mathcal{GS}(-a)$.

1.1.2 Notations and assumptions

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

(A_3) (i) $F : \mathbb{R}^d \longrightarrow [0,1]$ is twice continuously differentiable.
(ii) For all $i, j \in \{1, \ldots, d\}$, $F_i^{(1)} := \dfrac{\partial F}{\partial y_i}$ and $F_{ij}^{(2)} := \dfrac{\partial^2 F}{\partial y_i \partial y_j}$ are bounded.
(iii) For all $k \in \{1, \ldots, d\}$, $\dfrac{\partial^{d-k+1} F}{\partial y_k \ldots \partial y_d}$ exists and are continuous.
(iv) For all $i \in \{1, \ldots, d\}$, there exists $f_i : \mathbb{R} \longrightarrow \mathbb{R}_+$ bounded and integrable functions such that

$$f(x_1, \ldots, x_d) \leq \prod_{i=1}^{d} f_i(x_i).$$

1.2 Main results

Our first result rests upon the following propositions highlighting respectively the bias and the variance of F_n. For the following, we note

$$\text{For all } i,j \in \{1,\dots,d\}, \qquad \mu_j(\mathbf{K}) := \int_{\mathbb{R}^d} z_j^2 \mathbf{K}(z)dz, \qquad \phi_j(\mathbf{K}) := 2^d \int_{\mathbb{R}^d} z_j \mathbf{K}(z)\mathcal{K}(z)dz.$$

1.2.1 Bias and variance of F_n

Proposition 1.4. *Under the assumptions $(A_1)-(A_3)$, we obtain*

1. *If $a \in \left(0, \frac{\alpha}{3}\right]$, then*

$$\mathbb{E}[F_n(x)] - F(x) = \frac{h_n^2}{2(1-2a\xi)} \sum_{j=1}^{d} \mu_j(\boldsymbol{K}) F_{jj}^{(2)}(x) + o\left(h_n^2\right). \tag{1.4}$$

If $a \in \left(\frac{\alpha}{3}, 1\right)$, then

$$\mathbb{E}[F_n(x)] - F(x) = o\left(\sqrt{\gamma_n h_n}\right). \tag{1.5}$$

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

$$Var[F_n(x)] = o\left(h_n^4\right). \tag{1.6}$$

If $a \in \left[\frac{\alpha}{4}, \frac{\alpha}{3}\right)$, then

$$Var[F_n(x)] = \frac{\gamma_n}{2-\alpha\xi} F(x)(1-F(x)) + o\left(\gamma_n\right). \tag{1.7}$$

If $a \in \left[\frac{\alpha}{3}, 1\right)$, then

$$Var[F_n(x)] = \frac{\gamma_n}{2-\alpha\xi} F(x)(1-F(x)) - \frac{\gamma_n h_n}{2-(a+\alpha)\xi} \sum_{i=1}^{d} \phi_i(\mathbf{K}) F_i^{(1)}(x) + o\left(\gamma_n h_n\right). \tag{1.8}$$

The bias and the variance of the estimator F_n defined by the stochastic approximation algorithm (1.2) then largely depend on the choice of the stepsize (γ_n).

At this level, let us display the following theorem which provides the weak pointwise convergence rate of the generalized recursive estimator F_n.
It is noteworthy that $\xrightarrow[n\to+\infty]{\mathcal{D}}$ denotes the convergence in distribution, $\mathcal{N}$ corresponds to the Gaussian distribution and $\xrightarrow[n\to+\infty]{\mathbb{P}}$ stands for the convergence in probability.

1.2.2 Weak pointwise convergence rate of F_n

Theorem 1.5. *Under the assumptions $(A_1)-(A_3)$, we have:*

1. *If there exists $c \geq 0$ such that $\gamma_n^{-1} h_n^4 \xrightarrow[n\to+\infty]{} c$, then*

$$\sqrt{\gamma_n^{-1}}(F_n(x) - F(x)) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(\frac{\sqrt{c}}{2(1-2a\xi)} \sum_{j=1}^{d} \mu_j(\boldsymbol{K}) F_{jj}^{(2)}(x)\ ,\ \frac{1}{2-\alpha\xi} F(x)\left(1-F(x)\right)\right).$$

2. *If $\gamma_n^{-1} h_n^4 \xrightarrow[n\to+\infty]{} +\infty$, then*

$$\frac{1}{h_n^2}\left(F_n(x) - F(x)\right) \xrightarrow[n\to+\infty]{\mathbb{P}} \frac{1}{2(1-2a\xi)} \sum_{j=1}^{d} \mu_j(\boldsymbol{K}) F_{jj}^{(2)}(x).$$

1.3 Optimal choice of the stepsizes

In order to measure the asymptotic quality of the recursive estimator F_n, we need to use the Mean Weighted Integrated Squared Error ($MWISE$).

1.3.1 Asymptotic expressions of $MWISE[F_n]$

The $MWISE$ of the estimator F_n is determined by

$$MWISE[F_n] = \int_{\mathbb{R}^d} \left(\mathbb{E}[F_n(x)] - F(x)\right)^2 f(x)dx + \int_{\mathbb{R}^d} Var[F_n(x)]f(x)dx. \tag{1.9}$$

We first note,

$$J_1 := \int_{\mathbb{R}^d} \sum_{i=1}^{d} F_i^{(1)}(x)\phi_i(\mathbf{K})f(x)dx, \qquad J_2 := \int_{\mathbb{R}^d} \left(\sum_{j=1}^{d} \mu_j(\mathbf{K})F_{jj}^{(2)}(x)\right)^2 f(x)dx,$$

$$V_F := \int_{\mathbb{R}^d} F(x)(1-F(x))f(x)dx.$$

Proposition 1.6. *Under the assumptions $(A_1)-(A_3)$, we have*

$$MWISE[F_n] = \begin{cases} \frac{J_2}{4(1-2a\xi)^2}h_n^4 + o(h_n^4) & \text{if} \quad a \in \left(0, \frac{\alpha}{4}\right) \\ \frac{V_F}{2-\alpha\xi}\gamma_n + \frac{J_2}{4(1-2a\xi)^2}h_n^4 + o(h_n^4) & \text{if} \quad a \in \left[\frac{\alpha}{4}, \frac{\alpha}{3}\right) \\ \frac{V_F}{2-\alpha\xi}\gamma_n - \frac{J_1}{2-(\alpha+a)\xi}\gamma_n h_n + \frac{J_2}{4(1-2a\xi)^2}h_n^4 + o(h_n^4) & \text{if} \quad a = \frac{\alpha}{3} \\ \frac{V_F}{2-\alpha\xi}\gamma_n - \frac{J_1}{2-(\alpha+a)\xi}\gamma_n h_n + o(\gamma_n h_n) & \text{if} \quad a \in \left(\frac{\alpha}{3}, 1\right). \end{cases}$$

Proof. By distinguishing the different possible cases according to the expressions of the Bias and Variance, one can prove this proposition and find the required result. □

The following corollary ensures that, for a special choice of the stepsize $(\gamma_n) = \left(\gamma_0 n^{-1}\right)$, the optimal value for the bandwidth (h_n) depends on γ_0 and as a matter of fact the corresponding $MWISE$ depends also on γ_0.

Corollary 1.7. *Let assumptions $(A_1)-(A_3)$ hold. To minimize the $MWISE$ of F_n, we need to choose the stepsize (γ_n) in $\mathcal{GS}(-1)$ such that $\lim_{n\to+\infty} n\gamma_n = \gamma_0$, and the bandwidth (h_n) needs to be equal to*

$$\left(2^{-\frac{1}{3}}\left(\gamma_0 - \frac{2}{3}\right)^{\frac{1}{3}}\left(\frac{J_1}{J_2}\right)^{\frac{1}{3}} n^{-\frac{1}{3}}\right). \tag{1.10}$$

Therefore, the corresponding $MWISE$ is specified by

$$MWISE[F_n] = \frac{\gamma_0^2}{2\gamma_0 - 1}n^{-1}V_F - \frac{3}{4}\frac{1}{2^{4/3}}\frac{\gamma_0^2}{(\gamma_0 - 2/3)^{2/3}}J_1^{\frac{4}{3}}J_2^{-\frac{1}{3}}n^{-\frac{4}{3}} + o\left(n^{-\frac{4}{3}}\right). \tag{1.11}$$

The following corollary follows immediately from theorem 1.5 and exhibits the asymptotic normality of F_n.

Corollary 1.8. *Under the assumptions* $(A_1) - (A_3)$*, we have:*

If there exists $s \geq 0$ *such that* $\gamma_n^{-1} h_n^3 \underset{n\to+\infty}{\longrightarrow} s$*, so that* $\gamma_n^{-1} h_n^4 \underset{n\to+\infty}{\longrightarrow} 0$*, then*

$$\sqrt{\gamma_n^{-1}}(F_n(x) - F(x)) \underset{n\to+\infty}{\xrightarrow{\mathcal{D}}} \mathcal{N}\left(0\ ,\ \frac{1}{2-\alpha\xi} F(x)\left(1 - F(x)\right)\right).$$

Now, let us examine the asymptotic properties of the generalized Nadaraya's non-recursive distribution estimator $\widetilde{F}_n$.

1.4 Asymptotic properties of $\widetilde{F}_n$

The main properties of the generalized non-recursive estimator $\widetilde{F}_n$ are identified in the following proposition. (see Nadaraya (1964), Reiss (1981) and Hill (1985) for further details in the univariate case.)

Proposition 1.9. *Let assumptions* (A_1) *and* (A_3) *hold. Hence, the asymptotic properties of Nadaraya's estimator are displayed as follows*

- *The bias of* $\widetilde{F}_n$*:*

$$\mathbb{E}[\widetilde{F}_n(x)] - F(x) = \frac{h_n^2}{2} \sum_{j=1}^{d} \mu_j(\boldsymbol{K}) F_{jj}^{(2)}(x) + o\left(h_n^2\right).$$

- *The variance of* $\widetilde{F}_n$*:*

$$Var[\widetilde{F}_n(x)] = \frac{1}{n} F(x)(1 - F(x)) - \frac{h_n}{n} \sum_{i=1}^{d} F_i^{(1)}(x) \phi_i(\mathbf{K}) + o\left(h_n n^{-1}\right).$$

- *The asymptotic normality of* $\widetilde{F}_n$*:*
  If we have $nh_n^4 \underset{n\to+\infty}{\longrightarrow} 0$*, then*

$$\sqrt{n}(\widetilde{F}_n(x) - F(x)) \underset{n\to+\infty}{\xrightarrow{\mathcal{D}}} \mathcal{N}\left(0\ ,\ F(x)\left(1 - F(x)\right)\right).$$

It follows from this proposition that

$$MWISE[\widetilde{F}_n] = \frac{h_n^4}{4} J_2 + o\left(h_n^4\right) + \frac{1}{n} V_F - \frac{h_n}{n} J_1 + o\left(h_n^4\right).$$

Corollary 1.10. *Let assumptions* (A_1) *and* (A_3) *hold. To minimize the* $MWISE$ *of* $\widetilde{F}_n$*, the bandwidth* (h_n) *must be equal to*

$$\left(\left(\frac{J_1}{J_2}\right)^{\frac{1}{3}} n^{-\frac{1}{3}}\right). \tag{1.12}$$

Thus, the corresponding $MWISE$ *is expressed in terms of*

$$MWISE[\widetilde{F}_n] = n^{-1} V_F - \frac{3}{4} J_1^{\frac{4}{3}} J_2^{-\frac{1}{3}} n^{-\frac{4}{3}} + o\left(n^{-\frac{4}{3}}\right). \tag{1.13}$$

Remark 1.11. *In practice, the Robbins-Monro algorithm is very sensitive to parameters like the starting point or the calibration of the step sequence* (γ_n). *The optimal speed of the algorithm is reached for a step of the form* $\gamma_n = cn^{-\alpha}$ *but the choice of the constants* c *and* α *is extremely significant and determining it in practice proves to be very tricky.*

The following theorem establishes that, for a special choice of the stepsize (γ_n), the proposed recursive distribution estimator F_n can dominate the generalized non-recursive Nadaraya's estimator $\widetilde{F}_n$ in terms of the MWISE.

Theorem 1.12. *Under assumptions* $(A_1) - (A_3)$, *and assuming that* $(\gamma_n) = (\gamma_0 n^{-1})$ *with* $\gamma_0 = \frac{2}{3} + \varepsilon$ *where* $\varepsilon > 0$ *(very close to zero), we consider the generalized estimators* (1.2) *with the bandwidth*

$$(h_n) = \left(2^{-\frac{1}{3}} \left(\gamma_0 - \frac{2}{3} \right)^{\frac{1}{3}} \left(\frac{J_1}{J_2} \right)^{\frac{1}{3}} n^{-\frac{1}{3}} \right)$$

and the generalized Nadaraya's estimator (1.3) *with the bandwidth*

$$(h_n) = \left(\left(\frac{J_1}{J_2} \right)^{\frac{1}{3}} n^{-\frac{1}{3}} \right).$$

For $n \in \left(n_\varepsilon^{Lower}, n_\varepsilon^{Upper} \right)$, *where*

$$n_\varepsilon^{Lower} = \frac{3^3}{2^{10}} \frac{J_1^4}{J_2^1 {V_F}^3} \frac{(\frac{1}{3} + 2\varepsilon)^3}{\varepsilon^2},$$

and

$$n_\varepsilon^{Upper} = \frac{3^3}{2^{10}} \frac{J_1^4}{J_2 V_F^3} \frac{(\frac{1}{3} + 2\varepsilon)^3 \left(\frac{2}{3} + \varepsilon - 2^{\frac{2}{3}} \varepsilon^{\frac{1}{3}} \right)^3 \left(\frac{2}{3} + \varepsilon + 2^{\frac{2}{3}} \varepsilon^{\frac{1}{3}} \right)^3}{(\varepsilon - \frac{1}{3})^6 \varepsilon^2},$$

we have

$$MWISE[F_n] < MWISE[\widetilde{F}_n].$$

Remark 1.13. *Note that since the optimal* $MWISE$ *formulas* (1.11) *and* (1.13) *must be positive, we need to opt for* n *such that* $n > n_\varepsilon^{Lower}$. *In addition, in order to get a smaller* $MWISE$ *of the proposed recursive estimator compared to non-recursive estimator, we need to choose* $n < n_\varepsilon^{Upper}$; *for supplementary details one can consult the proof given in Slaoui (2014b) in the case of unidimensionnel case, see also Jmaei* et al. *(2017) in the case of using the Bernstein polynomials rather than kernels. We can determine for each selected law and for a fixed sample size* n *the upper and the lower bounds of the interval, in which the proposed recursive estimator can be better than the generalized classic Nadaraya's estimator.*

1.5 Bandwidth selection

Kernel smoothing in non-parametric statistics requires the choice of a bandwidth parameter. This choice is critical and can substantially reduce precision. An appropriate bandwidth can help obtain an estimated distribution function close to the true distribution function. However, a poorly selected bandwidth can seriously distort the true underlying characteristics of the distribution function. Thus, wise choice of bandwidth is highly recommended. There are a myriad of methods for bandwidth selection. As far as our research is concerned, we are mainly interested in

two of them, namely the cross-validation procedure and the second generation plug-in approach. The cross-validation selection method was proposed by Sarda (1993), but this method has deficiencies as revealed by Altman and Leger (1995) who elaborated another efficient method, a plug-in approach which minimizes the estimate of the $MWISE$, using the density function as a weight function.
First of all, for the sake of simplicity, the kernel $\mathbf{K}$ is considered as a product of univariate kernels K satisfying

$$\mathbf{K} = \otimes_1^d K \text{ and } \int_{\mathbb{R}} K(x)dx = 1. \tag{1.14}$$

Hence, we introduce a function $\mathscr{K}$ defined by

$$\mathscr{K}(x) = \int_{-\infty}^{x} K(t)dt \text{ such that } \mathcal{K} = \otimes_1^d \mathscr{K}.$$

Beside, it is noteworthy that using the fact that X is a an $n \times d$ matrix, we will adopt the following notation in the whole thesis

$$X_{ki} := X_{k,i}, \quad \text{the entry in the k-th row and i-th column of the matrix X.}$$

Hence, we infer that

$$F_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \mathcal{K}\left(\frac{x - X_k}{h_k}\right) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \prod_{i=1}^{d} \mathscr{K}\left(\frac{x_i - X_{ki}}{h_k}\right)$$

and

$$\widetilde{F}_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathcal{K}\left(\frac{x - X_k}{h_n}\right) = \frac{1}{n} \sum_{k=1}^{n} \prod_{i=1}^{d} \mathscr{K}\left(\frac{x_i - X_{ki}}{h_n}\right).$$

Let us start by introducing the first bandwidth selection method.

1.5.1 Cross-Validation

The cross-validation procedure corresponds to a method of selecting the smoothing parameter. It's noteworthy that (see Sarda (1993)):

$$CV(h) = \frac{1}{n} \sum_{i=1}^{n} (F_e - F_{-i})^2,$$

with F_e is the empirical distribution function and F_{-i} is the leave-one-out version of the considered estimator.
Thus, an estimator of the cross-validation criterion based on the generalized recursive estimator is identified by

$$\widehat{CV}(h) = \frac{1}{n} \sum_{i=1}^{n} \left(F_e - \Pi_{n-1} \sum_{\substack{k=1 \\ i \neq k}}^{n} \Pi_k^{-1} \gamma_k \prod_{t=1}^{d} \mathscr{K}\left(\frac{X_{it} - X_{kt}}{h_k}\right) \right)^2,$$

and the considered cross-validation bandwidth selection adapted to our generalized recursive estimator is indicated by

$$\widehat{h}_{\text{opt}} = \arg\min_{h \in H} \widehat{CV}(h).$$

Likewise, an estimator of the cross-validation criterium based on the generalized Nadaraya's estimator is determined by

$$\widetilde{CV}(h) = \frac{1}{n}\sum_{i=1}^{n}\left(F_e - \frac{1}{n-1}\sum_{\substack{k=1\\ i\neq k}}^{n}\prod_{t=1}^{d}\mathscr{K}\left(\frac{X_{it}-X_{kt}}{h_n}\right)\right)^2,$$

and the considered cross-validation bandwidth selection adapted to the generalized Nadaraya's estimator is defined by

$$\widetilde{h}_{\text{opt}} = \arg\min_{h\in H}\widetilde{CV}(h).$$

In the next subsection, we tackle the second considered bandwidth selection method, which is called the second generation plug-in approach.

1.5.2 Plug-in method

In statistics, the plug-in principle stands for approximating a functional of a given population distribution by the same functional at the empirical distribution. Plug-in bandwidth selectors are a major class of bandwidth selectors which are derived from the $MWISE$ expansion. Since the $MWISE$ depends on the unknown quantities J_1 and J_2, we suggest constructing asymptotic unbiased estimators of J_1 and J_2.
Here, we notice $J_1 = \phi(K)I_1$ and $J_2 = \mu^2(K)I_2$, where

$$\mu(K) = \int_{\mathbb{R}} z^2K(z)dz, \qquad \phi(K) = 2\int_{\mathbb{R}} zK(z)\mathscr{K}(z)dz,$$

$$I_1 = \int_{\mathbb{R}^d}\sum_{i=1}^{d}F_i^{(1)}(x)f(x)dx \qquad \text{and} \qquad I_2 = \int_{\mathbb{R}^d}\left(\sum_{j=1}^{d}F_{jj}^{(2)}(x)\right)^2 f(x)dx.$$

In order to estimate the optimal bandwidth (1.10), we need to estimate I_1 and I_2, by using the Plug-in estimate approach of Altman and Leger (1995). From this perspective, we introduce $(b_n)_{n\geq 1} \in \mathcal{GS}(-\delta)$, $\delta \in (0,1)$. In practice, we set

$$b_n = n^{-\delta}\min\left\{\widehat{s}, \frac{Q_3 - Q_1}{1.349}\right\}, \tag{1.15}$$

with $\widehat{s}$ being the sample standard deviation, and Q_1, Q_3 denoting the first and third quartiles of the sample X, respectively.
Moreover, we assume that K_b stands for a kernel and b_n is the associated bandwidth selected to be equal to (1.15) such that $\delta = \frac{2}{5}$, $K_{b'}^{(1)}$ denotes the first derivative of a kernel $K_{b'}$ with the associated bandwidth $b_n^{'}$ such that $\delta = \frac{3}{10}$ and $\mathscr{K}_{b''}$ corresponds to the distribution function of a kernel $K_{b''}$ with the associated bandwidth $b_n^{''}$ such that $\delta = \frac{1}{3}$.
For additional details concerning our choice for the parameter δ, we recommend the reader to consult the work of Slaoui (2014a) for recursive kernel density estimation and Slaoui (2014b) for the distribution function estimation. In order to clarify this notion for the reader, we develop in the next subsection the proposed bandwidth selection based on the generalized recursive distribution estimator.

Recursive estimator F_n:

To estimate the optimal bandwidth (1.10), we need to estimate I_1 and I_2.
Estimation of I_1:

$$\widehat{I}_1 = \frac{\Pi_n}{n} \sum_{\substack{i,k=1 \\ i\neq k}}^{n} \Pi_k^{-1} \gamma_k b_k^{-1} \left[\sum_{t=1}^{d} K_b \left(\frac{X_{it} - X_{kt}}{b_k} \right) \prod_{\substack{l=1 \\ l\neq t}}^{d} \mathscr{K}_{b''} \left(\frac{X_{il} - X_{kl}}{b_k''} \right) \right].$$

Estimation of I_2:

$$\widehat{I}_2 = \frac{\Pi_n^2}{n} \sum_{\substack{i,j,k=1 \\ i\neq j\neq k}}^{n} \Pi_j^{-1} \Pi_k^{-1} \gamma_j \gamma_k b_j'^{-2} b_k'^{-2} \left[\sum_{t=1}^{d} K_{b'}^{(1)} \left(\frac{X_{it} - X_{jt}}{b_j'} \right) \prod_{\substack{l=1 \\ l\neq t}}^{d} \mathscr{K}_{b''} \left(\frac{X_{il} - X_{jl}}{b_j''} \right) \right]$$
$$\times \left[\sum_{t=1}^{d} K_{b'}^{(1)} \left(\frac{X_{it} - X_{kt}}{b_k'} \right) \prod_{\substack{l=1 \\ l\neq t}}^{d} \mathscr{K}_{b''} \left(\frac{X_{il} - X_{kl}}{b_k''} \right) \right],$$

At this stage, we obtain

$$\widehat{J}_1 = \phi(K) \widehat{I}_1 \text{ and } \widehat{J}_2 = \mu^2(K) \widehat{I}_2.$$

As a matter of fact, our considered plug-in bandwidth selection procedure to estimate (1.10) is specified by

$$\left(2^{-\frac{1}{3}} \left(\gamma_0 - \frac{2}{3} \right)^{\frac{1}{3}} \left\{ \frac{\widehat{J}_1}{\widehat{J}_2} \right\}^{\frac{1}{3}} n^{-\frac{1}{3}} \right), \tag{1.16}$$

Estimation of V_F:
Furthermore, in order to compute the $MWISE$ of F_n, we need to estimate V_F as follows

$$\widehat{V}_F = \frac{\Pi_n}{n} \sum_{\substack{i,k=1 \\ i\neq k}}^{n} \Pi_k^{-1} \gamma_k \left[\prod_{l=1}^{d} \mathscr{K}_{b''} \left(\frac{X_{il} - X_{kl}}{b_k''} \right) \right]$$
$$- \frac{\Pi_n^2}{n} \sum_{\substack{i,j,k=1 \\ i\neq j\neq k}}^{n} \Pi_j^{-1} \Pi_k^{-1} \gamma_j \gamma_k \left[\prod_{l=1}^{d} \mathscr{K}_{b''} \left(\frac{X_{il} - X_{jl}}{b_j''} \right) \right] \times \left[\prod_{l=1}^{d} \mathscr{K}_{b''} \left(\frac{X_{il} - X_{kl}}{b_k''} \right) \right].$$

Therefore, the plug-in estimator of $MWISE[F_n]$ is provided by

$$\widehat{MWISE}[F_n] = \frac{\gamma_0^2}{2\gamma_0 - 1} n^{-1} \widehat{V}_F - \frac{3}{4} \frac{1}{2^{4/3}} \frac{\gamma_0^2}{(\gamma_0 - 2/3)^{2/3}} \widehat{J}_1^{\frac{4}{3}} \widehat{J}_2^{-\frac{1}{3}} n^{-\frac{4}{3}} + o\left(n^{-\frac{4}{3}} \right). \tag{1.17}$$

In the next subsection, we handle the generalized Nadaraya's distribution estimator in order to be able to compare our generalized recursive distribution estimator to the non-recursive one.

Non-recursive estimator $\widetilde{F}_n$:

To estimate the optimal bandwidth (1.12), we need to estimate I_1 and I_2.
Estimation of I_1:

$$\widetilde{I}_1 = \frac{1}{n(n-1)b_n} \sum_{\substack{i,k=1 \\ i\neq k}}^{n} \left[\sum_{t=1}^{d} K_b \left(\frac{X_{it} - X_{kt}}{b_n} \right) \prod_{\substack{l=1 \\ l\neq t}}^{d} \mathscr{K}_{b''} \left(\frac{X_{il} - X_{kl}}{b_n''} \right) \right],$$

Estimation of I_2:

$$\widetilde{I}_2 = \frac{1}{n^3 b_n^4} \sum_{\substack{i,j,k=1 \\ i\neq j\neq k}}^{n} \left[\sum_{t=1}^{d} K_{b'}^{(1)} \left(\frac{X_{it} - X_{jt}}{b'_n} \right) \prod_{\substack{l=1 \\ l\neq t}}^{d} \mathscr{K}_{b''} \left(\frac{X_{il} - X_{jl}}{b''_n} \right) \right] \times \left[\sum_{t=1}^{d} K_{b'}^{(1)} \left(\frac{X_{it} - X_{kt}}{b'_n} \right) \prod_{\substack{l=1 \\ l\neq t}}^{d} \mathscr{K}_{b''} \left(\frac{X_{il} - X_{kl}}{b''_n} \right) \right],$$

At this stage, we obtain

$$\widetilde{J}_1 = \phi(K)\widetilde{I}_1 \text{ and } \widetilde{J}_2 = \mu^2(K)\widetilde{I}_2.$$

Hence, the considered plug-in bandwidth selection estimator of (1.12) is denoted by

$$\left(\left\{ \frac{\widetilde{J}_1}{\widetilde{J}_2} \right\}^{\frac{1}{3}} n^{-\frac{1}{3}} \right), \tag{1.18}$$

Estimation of V_F:
Furthermore, in order to compute the associated $MWISE$ of $\widetilde{F_n}$ using the previous plug-in bandwidth estimator of (1.18), we need to estimate V_F as follows

$$\widetilde{V}_F = \frac{1}{n(n-1)} \sum_{\substack{i,k=1 \\ i\neq k}}^{n} \left[\prod_{l=1}^{d} \mathscr{K}_{b''} \left(\frac{X_{il} - X_{kl}}{b''_n} \right) \right] - \frac{1}{n(n-1)^2} \sum_{\substack{i,j,k=1 \\ i\neq j\neq k}}^{n} \left[\prod_{l=1}^{d} \mathscr{K}_{b''} \left(\frac{X_{il} - X_{jl}}{b''_n} \right) \right] \times \left[\prod_{l=1}^{d} \mathscr{K}_{b''} \left(\frac{X_{il} - X_{kl}}{b''_n} \right) \right].$$

Thus, the associated plug-in estimator of $MWISE[\widetilde{F_n}]$ is expressed by

$$\widetilde{MWISE}[\widetilde{F}_n] = n^{-1}\widetilde{V}_F - \frac{3}{4}\widetilde{J}_1^{\frac{4}{3}}\widetilde{J}_2^{-\frac{1}{3}} n^{-\frac{4}{3}} + o\left(n^{-\frac{4}{3}}\right). \tag{1.19}$$

The equations incorporated in (1.17) and (1.19) will be very useful in the next section at the level of comparing the performance of both estimators in terms of estimation error.

1.6 Numerical applications

The target underlying our numerical studies lies in comparing the performance of our generalized recursive estimator (1.2) with that of Nadaraya (1.3).

1.6.1 Simulation studies

When applying our generalized recursive estimator F_n, we need to choose three quantities:

- The function K, we use the Epanechnikov kernel.
- The stepsize $(\gamma_n) = (\gamma_0 n^{-1})$, where $\gamma_0 = 2/3 + \varepsilon$.
- The bandwidth (h_n) is chosen to be equal to (1.16).

When applying the non-recursive estimator $\widetilde{F}_n$, we need to choose two quantities:

- The function K, we use the Epanechnikov kernel.
- The bandwidth (h_n) is chosen to be equal to (1.18).

We denote by F_i^* the reference distribution and by F_i the test distribution. Therefore, we calculate the following measures:

- Mean squared error: $MSE = \frac{1}{n}\sum_{i=1}^{n}(F_i - F_i^*)^2$.
- The linear correlation: $Cor = Cov(F_i, F_i^*)\sigma(F_i)^{-1}\sigma(F_i^*)^{-1}$.

Here are certain properties of the used Epanechnikov kernel :

Kernel	$K(x)$	Support	$\phi(K)$	$\mu(K)$
Epanechnikov	$\frac{3}{4}(1-x^2)$	$[-1,1]$	$\frac{9}{35}$	$\frac{1}{5}$

Table 1.1: Epanechnikov proporties.

The unidimensional case: d=1

In order to compare the considered generalized recursive estimator and the Nadaraya's generalized distribution one, we consider two sample sizes; $n = 50$ and 100 using $N = 500$ trials of sample n and the following four models:

- Model 1: The standard normal distribution $\mathcal{N}(0,1)$.
- Model 2: The normal distribution $\mathcal{N}\left(\frac{1}{2},1\right)$.
- Model 3: The mixture normal distribution $\frac{1}{2}\mathcal{N}\left(\frac{1}{2},1\right)+\frac{1}{2}\mathcal{N}\left(-\frac{1}{2},1\right)$.
- Model 4: The mixture normal distribution $\frac{2}{3}\mathcal{N}\left(\frac{1}{3},1\right)+\frac{1}{3}\mathcal{N}\left(-\frac{1}{3},1\right)$.

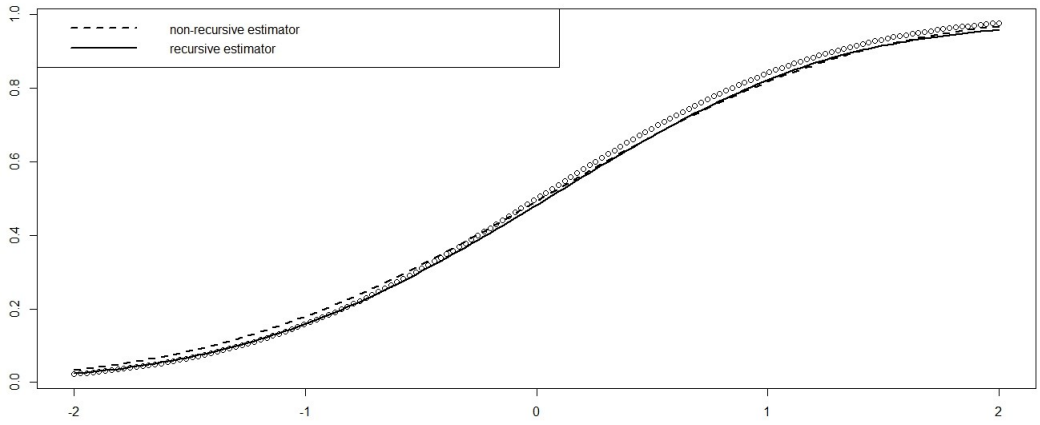


Figure 1.1: Qualitative comparison between Nadaraya's distribution estimator and the generalized recursive estimator for Model 1 with n=50 and $(\gamma_n) = \left((2/3+0.05)n^{-1}\right)$.

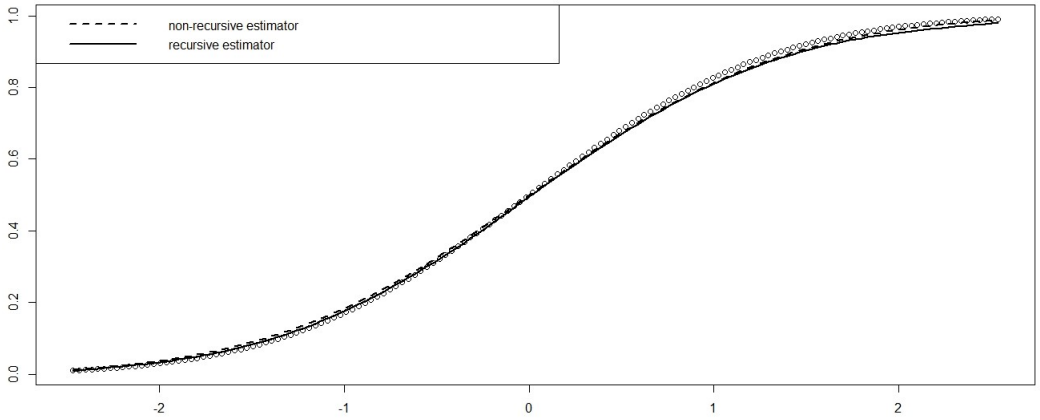


Figure 1.2: Qualitative comparison between Nadaraya's distribution estimator with the generalized recursive estimator using Model 1, n=100 and $(\gamma_n) = \left((2/3 + 0.05)n^{-1}\right)$.

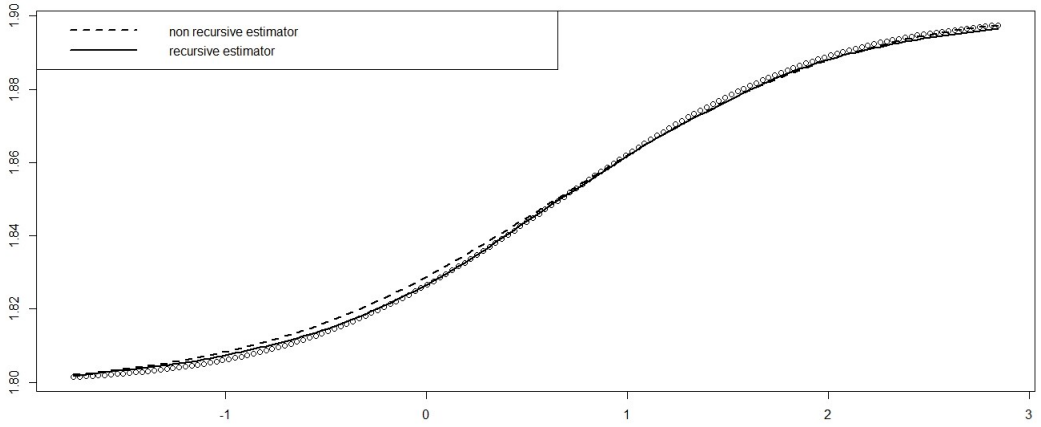


Figure 1.3: Qualitative comparison between Nadaraya's distribution estimator with the generalized recursive estimator using Model 4, n=50 and $(\gamma_n) = \left((2/3 + 0.05)n^{-1}\right)$.

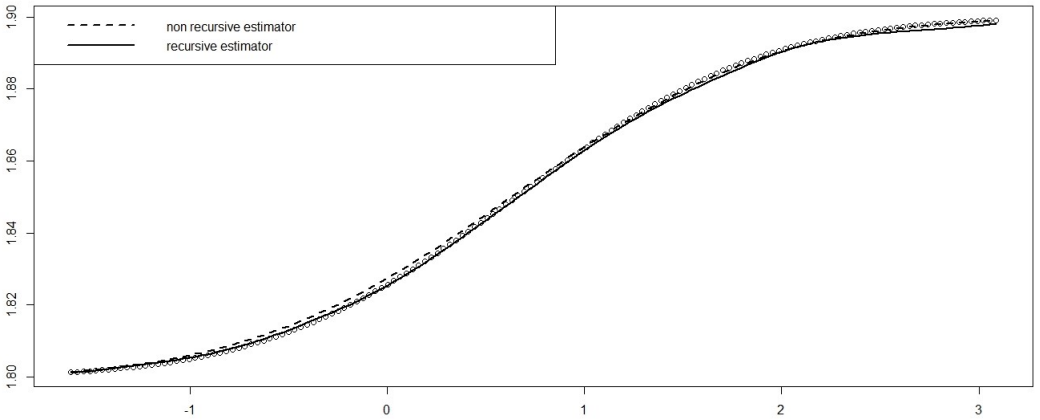


Figure 1.4: Qualitative comparison between Nadaraya's distribution estimator with the generalized recursive estimator using Model 4, n=100 and $(\gamma_n) = \left((2/3 + 0.05)n^{-1}\right)$.

			Plug-in		Cross-Validation	
Model	MSE/Cor	n	Nadaraya's estimator	Recursive estimator $\gamma_n = (\frac{2}{3} + 0.05)n^{-1}$	Nadaraya's estimator	Recursive estimator $\gamma_n = (\frac{2}{3} + 0.05)n^{-1}$
Model 1	MSE	50	0.0008760976	**0.0006684492**	0.0008719240	0.0007052983
		100	0.0003355597	**0.0002681363**	0.0003507760	0.0002798545
	Cor	50	0.9991972854	**0.9993061033**	0.9985833690	0.9992424589
		100	0.9996993596	**0.9997911018**	0.9996216460	0.9997260820
Model 2	MSE	50	0.0003243463	**0.0002961442**	0.0003424373	0.0003047189
		100	0.0002010728	**0.0001532752**	0.0001743897	0.0001540263
	Cor	50	0.9997654430	**0.9996854602**	0.9996520644	0.9996463968
		100	0.9998041657	**0.9998307559**	0.9997582078	0.9997787056
Model 3	MSE	50	0.0006376542	**0.0005857468**	0.0006901210	0.0006039373
		100	0.0005868853	**0.0004730782**	0.0006321977	0.0004962125
	Cor	50	0.9989832200	**0.9988033588**	0.9987992520	0.9987239326
		100	0.9990871155	**0.9990431684**	0.9989331049	0.9989420103
Model 4	MSE	50	0.0005038682	**0.0004442070**	0.0006349498	0.0004793215
		100	0.0000701021	**0.0000310626**	0.0001136351	0.0000382802
	Cor	50	0.9993125926	**0.9989769800**	0.9988661040	0.9988299239
		100	0.9999081901	**0.9999493515**	0.9998286569	0.9999350136

Table 1.2: Quantitative comparison between Nadaraya's distribution estimator and the generalized recursive distribution estimator with the stepsize $(\gamma_n) = ([2/3+0.05]n^{-1})$ through a plug-in method as well as a cross-validation one in the unidimensional case.

For the following, it is worthy to denote that I_d stands for the identity matrix of size d.

The bidimensional case: d=2

In order to compare the generalized recursive distribution estimator with Nadaraya's distribution estimator, we consider two sample sizes, $n = 50$ and 100 using $N = 500$ trials of sample n and four models:

- Model 1: The standard normal distribution $\mathcal{N}\left(\binom{0}{0}, I_2\right)$.
- Model 2: The normal distribution $\mathcal{N}\left(\binom{1/2}{1/2}, I_2\right)$.
- Model 3: The mixture normal distribution $\frac{1}{2}\mathcal{N}\left(\binom{1/2}{1/2}, I_2\right) + \frac{1}{2}\mathcal{N}\left(\binom{-1/2}{-1/2}, I_2\right)$.
- Model 4: The mixture normal distribution $\frac{2}{3}\mathcal{N}\left(\binom{1/3}{1/3}, I_2\right) + \frac{1}{3}\mathcal{N}\left(\binom{-1/3}{-1/3}, I_2\right)$.

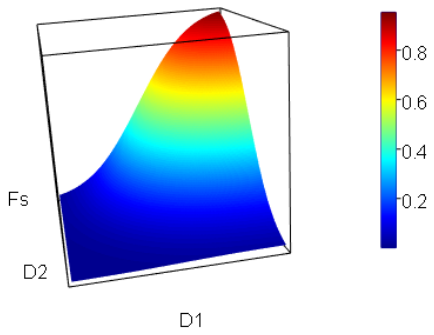


Figure 1.5: The reference distribution function F_s using Model 1.

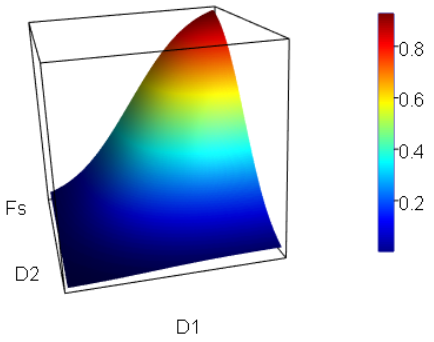


Figure 1.6: The reference distribution function F_s using Model 3.

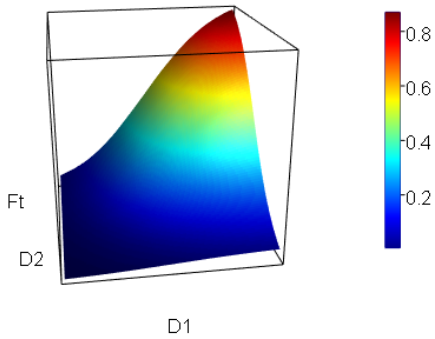


Figure 1.7: Nadaraya's estimator $\widetilde{F}$ using Model 1 with n=50.

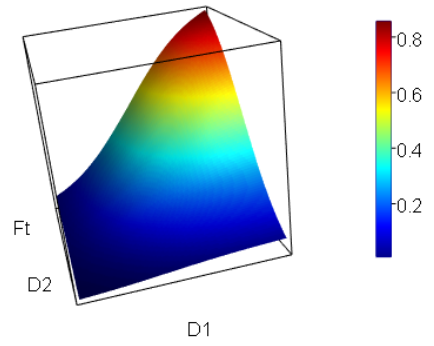


Figure 1.8: Nadaraya's estimator $\widetilde{F}$ using Model 3 with n=50.

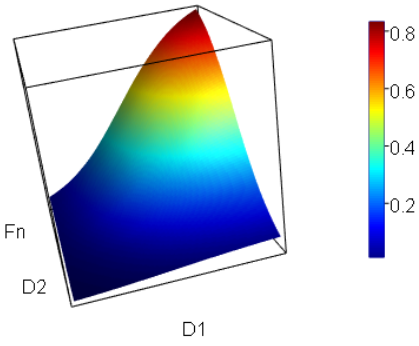


Figure 1.9: The recursive estimator F_n using Model 1 with n=50.

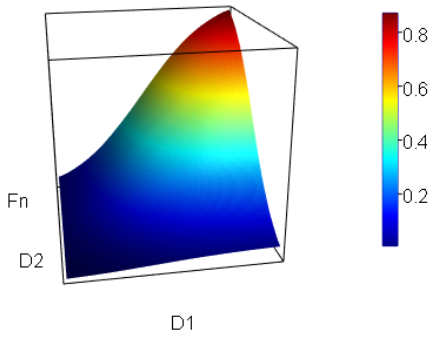


Figure 1.10: The recursive estimator F_n using Model 3 with n=50.

			Plug-in		Cross-Validation	
Model	MSE/Cor	n	Nadaraya's estimator	Recursive estimator $\gamma_n = (\frac{2}{3} + 0.05)n^{-1}$	Nadaraya's estimator	Recursive estimator $\gamma_n = (\frac{2}{3} + 0.05)n^{-1}$
Model 1	MSE	50	0.0003157969	**0.0002118529**	0.0003309580	0.0003047189
		100	0.0000876097	**0.0000668449**	0.0000871924	0.0000705320
	Cor	50	0.9994881093	**0.9992621212**	0.9989031200	0.9989383580
		100	0.9999197285	**0.9999306103**	0.9998583369	0.9999242405
Model 2	MSE	50	0.0008261247	**0.0006548646**	0.0008385428	0.0006632885
		100	0.0000908300	**0.0000600300**	0.0000977200	0.0000713295
	Cor	50	0.9993041754	**0.9996147794**	0.9993122882	0.9996248268
		100	0.9998238000	**0.9998288000**	0.9998037000	0.9998278900
Model 3	MSE	50	0.0008668519	**0.0008134314**	0.0010215940	0.0009445246
		100	0.0005842987	**0.0005468488**	0.0007121525	0.0006637816
	Cor	50	0.9976693040	**0.9977774985**	0.9967051420	0.9974661303
		100	0.9986429796	**0.9987608168**	0.9982977372	0.9983191907
Model 4	MSE	50	0.0006459259	**0.0004847252**	0.0006801415	0.0005460630
		100	0.0002643474	**0.0001391675**	0.0002736224	0.0001653388
	Cor	50	0.9986615779	**0.9989035867**	0.9985514195	0.9986705950
		100	0.9997714558	**0.9998242084**	0.9997033162	0.9998010731

Table 1.3: Quantitative comparison between Nadaraya's distribution estimator and the proposed distribution estimator with stepsize $(\gamma_n) = ([2/3 + 0.05]n^{-1})$ through a plug-in method as well as a cross-validation one in the bidimensional case.

From Tables 1.2 and 1.3, we conclude that:

1 - The MSE of the generalized recursive estimator with stepsize $(\gamma_n) = ([2/3 + 0.05]n^{-1})$ through a plug-in method is smaller than that of the generalized Nadaraya's estimator with the second generation considered plug-in method as well as the generalized recursive estimator and the generalized Nadaraya's estimator with the cross-validation approach.

2 - The MSE decreases as the simple size increases.

3 - The Cor increases as the sample size increases.

1.6.2 Real Datasets

In this section, we exhibit three real data applications. The first one is in an unidimensional case, the second one is in a bidimensional case while the third one is in a multidimensional case. For all of the considered cases, we report the values of the plug-in I_1, I_2, V_F and $MWISE$ using the generalized Nadaraya's estimator as well as the generalized recursive estimator using a specific choice of the stepsize. Moreover, we calculate the error defined by

$$PSE = \frac{1}{n}\sum_{i=1}^{n}(F_i - F_i^e)^2,$$

where F_i^e is the empirical distribution function and F_i is the test distribution function.

An unidimensional application

At this level, we use `lh` data which appears in the `R` package `datasets`. These data are a sort of a regular time series giving the `luteinizing hormone` in blood samples at 10 mins intervals from a human female. Totally, we got 48 samples.

	I_1	I_2	V_F	$MWISE$	PSE
Nadaraya's estimator	0.42214350	0.64455170	0.16382990	0.00265951	0.00048493
Recursive estimator	**0.45531620**	**0.33022020**	**0.16532920**	**0.00251792**	**0.00043248**

Table 1.4: Quantitative comparison between the I_1, I_2, V_F, $MWISE$ and PSE of Nadaraya's distribution estimator as well as the proposed distribution estimator with stepsize $(\gamma_n) = ([2/3 + 0.05]n^{-1})$ via `lh` data of the package datasets and through a plug-in method.

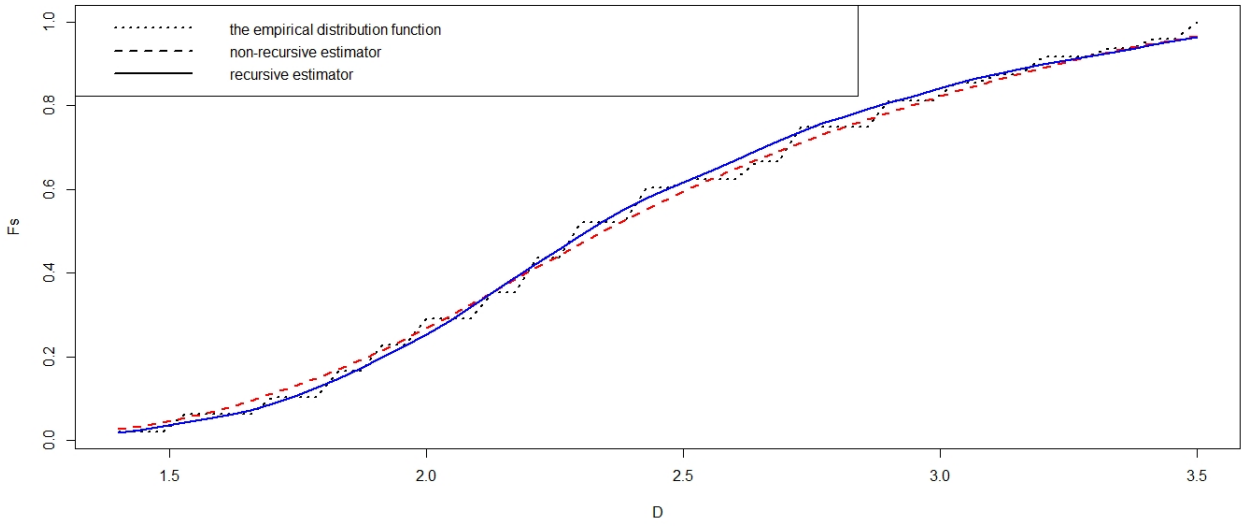


Figure 1.11: Qualitative comparison between Nadaraya's distribution estimator and the proposed distribution estimator with stepsize $(\gamma_n) = ([2/3 + 0.05]n^{-1})$ via lh data of the package datasets and through a plug-in method.

A bidimensional application

Within this framework, we use the `nwip` data which appears in the `R` package `kerdiest`. This data set corresponds to the earthquakes occurring in the Northwest of the Iberian Peninsula, from 25/November/1924 to 31/July/2010. The area is limited by the coordinates 41 N – 44 N and 6 W – 10 W, involving the autonomic region of Galicia (Spain) and northern Portugal.
The data catalog was obtained from the National Geographic Institute (IGN) of Spain. These data are available online at the web page `http://www.ign.es`. Within this data frame, we have 3491 observations on 10 variables, corresponding to the earthquake epicenters and time of occurrence. Within this application, we are interested in two variables, namely the longitude in degrees as well as the magnitude in Richter Scale. For the latter, we consider the date with the magnitude values > 3.1, to end up with a sample of 326 observations.

	I_1	I_2	V_F	$MWISE$	PSE
Nadaraya's estimator	0.44454490	0.09554306	0.16091480	0.00037498	0.00170011
Recursive estimator	**0.33153120**	**0.01013669**	**0.16707990**	**0.00035297**	**0.00089828**

Table 1.5: Quantitative comparison between the I_1, I_2, V_F, $MWISE$ and PSE of Nadaraya's distribution estimator as well as the proposed distribution estimator with stepsize $(\gamma_n) = ([2/3 + 0.05]n^{-1})$ via nwip data of the package kerdiest and through a plug-in method.

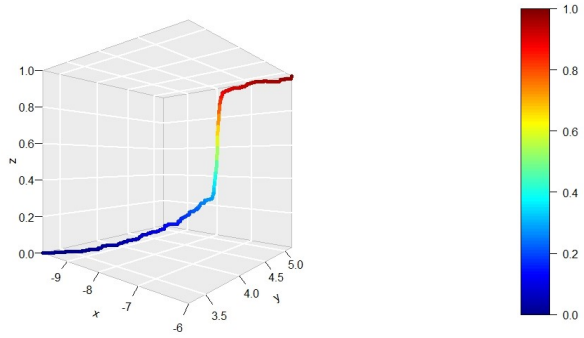


Figure 1.12: The empirical distribution function.

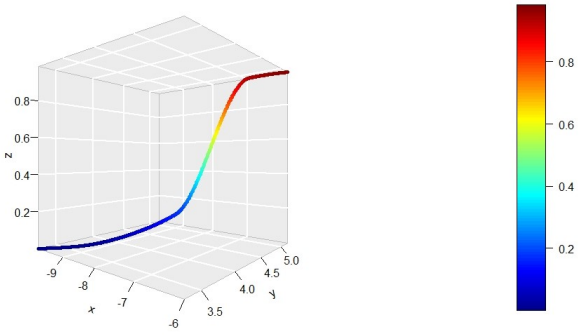


Figure 1.13: Nadaraya's estimator.

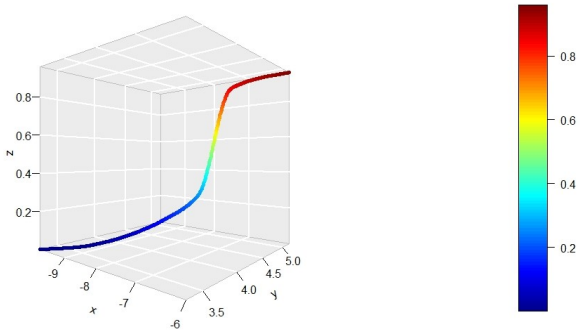


Figure 1.14: The recursive estimator.

A multidimensional application

In this application, we use the `iris` data which appears in the `R` package `datasets`. This famous (Fisher's or Anderson's) iris data set provides the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. These species are `Iris setosa`, `versicolor`, and `virginica`. As far as this application is concerned, we consider data with the petal width values < 2 to end up with a sample of 121 observations. Therefore, the following results are obtained.

	I_1	I_2	V_F	$MWISE$	PSE
Nadaraya's estimator	0.44976830	0.22750983	0.08894939	0.00039688	0.00086480
Recursive estimator	**0.39024930**	**0.07616996**	**0.09344776**	**0.00030997**	**0.00075062**

Table 1.6: Quantitative comparison between the I_1, I_2, V_F, $MWISE$ and PSE of Nadaraya's distribution estimator as well as the proposed distribution estimator with stepsize $(\gamma_n) = ([2/3+0.05]n^{-1})$ via iris data of the package datasets and through a plug-in method.

1.7 Conclusion

In this research work, we set forward a smooth estimator of the multivariate distribution function. We first studied the asymptotic properties of the generalized estimator. We computed the bias as well as the variance in order to demonstrate that our estimator asymptotically follows a normal distribution. Afterwards, we compared our generalized recursive estimator to non-recursive Nadaraya's multivariate distribution estimator using two bandwidth selection approaches, namely the cross-validation method as well as the plug-in technique. In the simulation studies, and for all the cases, the generalized recursive estimator (1.2) with stepsize

$(\gamma_n) = \left([2/3 + 0.05]n^{-1}\right)$ ensures a better performance in terms of estimation error compared to the generalized non-recursive Nadaraya's estimator.

To sum up, the use of the generalized recursive distribution estimators enables us to obtain better results compared to the non-recursive estimator. With an appropriate choice of the bandwidth, we can demonstrate that our generalized recursive estimator is closer to the true distribution function than the generalized non-recursive Nadaraya's estimator.

1.8 Proofs

Throughout this section, devoted to the proofs of our main results, we use the following notation:

$$\mathcal{Z}_n(x) = \mathcal{K}\left(\frac{x - X_n}{h_n}\right). \tag{1.20}$$

Proof of Proposition 1.4.
Let x be in $\mathbb{R}^d$.
In the general framework, using the relation (1.1) and without assuming that $F_0(x) = 0$, we have:

$$\begin{aligned} F_n(x) - F(x) &= (1-\gamma_n)F_{n-1}(x) + \gamma_n \mathcal{Z}_n(x) - F(x) \\ &= (1-\gamma_n)[F_{n-1}(x) - F(x)] + \gamma_n[\mathcal{Z}_n(x) - F(x)]. \end{aligned}$$

By a simple recurrence, we obtain

$$\begin{aligned} F_n(x) - F(x) &= \prod_{i=1}^{n}(1-\gamma_i)[F_0(x) - F(x)] + \sum_{k=1}^{n-1}\prod_{i=k+1}^{n}(1-\gamma_i)\gamma_k(\mathcal{Z}_k(x) - F(x)) + \gamma_n(\mathcal{Z}_n(x) - F(x)) \\ &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k(\mathcal{Z}_k(x) - F(x)) + \Pi_n[F_0(x) - F(x)]. \end{aligned}$$

Thus,

$$\mathbb{E}[F_n(x)] - F(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k(\mathbb{E}[\mathcal{Z}_k(x)] - F(x)) + \Pi_n[F_0(x) - F(x)]. \tag{1.21}$$

1. **Bias of F_n:**

Our intrinsic goal at this level is to calculate the quantity:

$$\mathbb{E}[\mathcal{Z}_k(x)] = \int_{\mathbb{R}^d} \mathcal{K}\left(\frac{x-y}{h_k}\right) f(y)dy.$$

Notably, under the conditions in (A_1), we assume that

$$f(y) = \frac{\partial^d F}{\partial y_1 \dots \partial y_d}(y),\ y = (y_1, \dots, y_d).$$

Correspondingly, under (A_1) and (1.14), we infer that

$$\mathbf{K}(y) = \frac{\partial^d \mathcal{K}}{\partial y_1 \dots \partial y_d}(y) = \frac{\partial^d \mathcal{K}}{\partial y_d \dots \partial y_1}(y).$$

At this level, we aim to apply the following result, an extension of the integration by parts formula.

Under the assumptions (A_1), (A_3) and (1.14), we infer that

$$\begin{aligned}\int_{\mathbb{R}^d} \mathcal{K}\left(\frac{x-y}{h_k}\right) f(y)dy &= \int_{\mathbb{R}^d} \mathcal{K}\left(\frac{x-y}{h_k}\right) \frac{\partial^d F}{\partial y_1 \dots \partial y_d}(y)dy \\ &= (-1)^d \int_{\mathbb{R}^d} \left(\frac{-1}{h_k}\right)^d \frac{\partial^d \mathcal{K}}{\partial y_1 \dots \partial y_d}\left(\frac{x-y}{h_k}\right) F(y)dy \\ &= \int_{\mathbb{R}^d} h_k^{-d}\mathbf{K}\left(\frac{x-y}{h_k}\right) F(y)dy.\end{aligned}$$

Hence, a change of variables ensures

$$\mathbb{E}[\mathcal{Z}_k(x)] = \int_{\mathbb{R}^d} \mathbf{K}(z)F(x-zh_k)dz, \quad z=(z_1,\dots,z_d).$$

Moreover, using (A_3) and by applying the well known Taylor's development with integral remainder formula for F, we obtain

$$F(x-zh_k) = F(x) - \sum_{i=1}^d F_i^{(1)}(x)z_i h_k + \int_0^1 (1-t)\sum_{i,j=1}^d F_{ij}^{(2)}(x-tzh_k)z_i z_j h_k^2 dt.$$

Then, it will be obvious that

$$\mathbb{E}[\mathcal{Z}_k(x)] - F(x) = \int_{\mathbb{R}^d} \mathbf{K}(z)\left[F(x-zh_k) - F(x)\right]dz.$$

Hence,

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

where

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

Therefore, thanks to the relation (1.21), we have

$$\begin{aligned}\mathbb{E}[F_n(x)] - F(x) &= \Pi_n\sum_{k=1}^n \Pi_k^{-1}\gamma_k\left(\frac{h_k^2}{2}\sum_{j=1}^d \mu_j(\mathbf{K})F_{jj}^{(2)}(x) + h_k^2\eta_k(x)\right) + \Pi_n[F_0(x) - F(x)] \\ &= \frac{1}{2}\sum_{j=1}^d \mu_j(\mathbf{K})F_{jj}^{(2)}(x)\Pi_n\sum_{k=1}^n \Pi_k^{-1}\gamma_k h_k^2 + \Pi_n\sum_{k=1}^n \Pi_k^{-1}\gamma_k h_k^2\eta_k(x) + \Pi_n[F_0(x) - F(x)] \\ &= S_{1,n}(x) + S_{2,n}(x) + \Pi_n[F_0(x) - F(x)]\end{aligned}$$

where,

$$S_{1,n}(x) = \frac{1}{2}\sum_{j=1}^d \mu_j(\mathbf{K})F_{jj}^{(2)}(x)\Pi_n\sum_{k=1}^n \Pi_k^{-1}\gamma_k h_k^2 \text{ and } S_{2,n}(x) = \Pi_n\sum_{k=1}^n \Pi_k^{-1}\gamma_k h_k^2\eta_k(x).$$

Here, we distinguish two cases:

1. For the case $a \leqslant \alpha/3$, we have $\lim\limits_{n\to+\infty} n\gamma_n > \min\left\{\dfrac{a+\alpha}{2}, 2a\right\} = 2a$.

Asymptotic behaviour of $S_{1,n}(x)$:

Here we opt to verify the conditions of lemma 1.2. We have $(v_n) := \left(h_n^{-2}\right) \in \mathcal{GS}(2a)$ and $m = 1$, since $\xi^{-1} > 2a$, then $1 - 2a\xi > 0$. The application of lemma 1.2 ensures that

$$\lim_{n\to+\infty} h_n^{-2}\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^2 = \frac{1}{1-2a\xi}.$$

Hence,

$$S_{1,n}(x) = \frac{h_n^2}{2(1-2a\xi)} \sum_{j=1}^{d} \mu_j(\mathbf{K}) F_{jj}^{(2)}(x) + o\left(h_n^2\right).$$

Asymptotic behaviour of $S_{2,n}(x)$:

Owing to the fact that $F_{ij}^{(2)}$ is bounded and continuous at x for all $i, j \in \{1, \ldots, d\}$, Lebesgue's convergence theorem ensures that $\lim\limits_{k\to+\infty} \eta_k(x) = 0$, which ensures that, $\eta_k(x) = o(1)$.

Therefore, the second part of lemma 1.2 ensures that

$$\lim_{n\to+\infty} h_n^{-2}\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^2 \eta_k = 0.$$

Thus,

$$S_{2,n}(x) = o\left(h_n^2\right).$$

As a matter of fact, we infer that,

$$\mathbb{E}[F_n(x)] - F(x) = \frac{h_n^2}{2(1-2a\xi)} \sum_{j=1}^{d} \mu_j(\mathbf{K}) F_{jj}^{(2)}(x) + o\left(h_n^2\right).$$

2. For the case $a > \alpha/3$, we have $h_n^2 = o\left(\sqrt{\gamma_n h_n}\right)$. Consequently, we obtain

$$\begin{aligned}
\mathbb{E}[F_n(x)] - F(x) &= \frac{1}{2}\sum_{j=1}^{d} \mu_j(\mathbf{K}) F_{jj}^{(2)}(x) \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^2 + \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^2 o(1) + \Pi_n[F_0(x) - F(x)] \\
&= \frac{1}{2}\sum_{j=1}^{d} \mu_j(\mathbf{K}) F_{jj}^{(2)}(x) \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k\, o\left(\sqrt{\gamma_k h_k}\right) + \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k\, o\left(\sqrt{\gamma_k h_k}\right) + O\left(\Pi_n\right) \\
&= o\left(\sqrt{\gamma_n h_n}\right).
\end{aligned}$$

As a result, we get

$$\mathbb{E}[F_n(x)] - F(x) = o\left(\sqrt{\gamma_n h_n}\right).$$

Variance of F_n:

In view of the independence of X_i's, for $i = 1, \ldots, n$, we deduce that

$$\begin{aligned}
Var[F_n(x)] &= Var[\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k \mathcal{Z}_k(x)] \\
&= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 Var[\mathcal{Z}_k(x)] \\
&= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 \left(\mathbb{E}[\mathcal{Z}_k^2(x)] - \mathbb{E}[\mathcal{Z}_k(x)]^2\right).
\end{aligned}$$

Hence, following (1.14) and owing to the fact that $2^d \int_{\mathbb{R}^d} \mathbf{K}(z)\mathcal{K}(z)dz = 1$, we infer that

$$\mathcal{K}^2(y_1,\dots,y_d) = \prod_{i=1}^{d} \mathscr{K}^2(y_i) \text{ and } \frac{\partial^d}{\partial y_1 \dots \partial y_d}(\mathcal{K}^2(y)) = \frac{\partial^d}{\partial y_d \dots \partial y_1}(\mathcal{K}^2(y)) = 2^d \mathbf{K}(y)\mathcal{K}(y).$$

Then, we get

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

From the application of taylor's theorem with integral remainder, we obtain

$$F(x-zh_k) = F(x) - \int_0^1 \sum_{i=1}^{d} F_i^{(1)}(x-tzh_k)z_i h_k dt.$$

Consequently, we find

$$\begin{aligned}
\mathbb{E}[\mathcal{Z}_k^2(x)] &= 2^d \int_{\mathbb{R}^d} \mathbf{K}(z)\mathcal{K}(z) \left[F(x) - \int_0^1 \sum_{i=1}^{d} F_i^{(1)}(x-tzh_k)z_i h_k dt\right] dz \\
&= F(x) - h_k \sum_{i=1}^{d} \phi_i(\mathbf{K}) F_i^{(1)}(x) - \nu_k(x),
\end{aligned}$$

with

$$\phi_i(\mathbf{K}) = 2^d \int_{\mathbb{R}^d} z_i \mathbf{K}(z)\mathcal{K}(z)dz \text{ and } \nu_k(x) = 2^d \sum_{i=1}^{d} \int_{\mathbb{R}^d} \int_0^1 \mathbf{K}(z)\mathcal{K}(z)\left[F_i^{(1)}(x-tzh_k) - F_i^{(1)}(x)\right] z_i h_k dt dz.$$

Moreover, we have

$$\mathbb{E}[\mathcal{Z}_k(x)] = F(x) + \widetilde{\nu}_k(x),$$

where $\widetilde{\nu}_k(x) = \int_{\mathbb{R}^d} \mathbf{K}(z)\left[F(x-zh_k) - F(x)\right]dz.$

Then, combining the previous results we state

$$\begin{aligned}
Var[F_n(x)] &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 \left(\mathbb{E}[\mathcal{Z}_k^2(x)] - \mathbb{E}[\mathcal{Z}_k(x)]^2\right) \\
&= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 \left[F(x) - h_k \sum_{i=1}^{d} \phi_i(\mathbf{K}) F_i^{(1)}(x) - \nu_k(x) - (F(x) + \widetilde{\nu}_k(x))^2\right] \\
&= [F(x)(1-F(x))]\Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 - \sum_{i=1}^{d} \phi_i(\mathbf{K}) F_i^{(1)}(x) \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 h_k \\
&\quad - \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 [\nu_k(x) + \widetilde{\nu}_k(x)^2 + 2F(x)\widetilde{\nu}_k(x)].
\end{aligned}$$

Likewise,

$$Var[F_n(x)] = S_{3,n}(x) + S_{4,n}(x) - S_{5,n}(x). \tag{1.22}$$

$$S_{3,n}(x) = [F(x)(1-F(x))]\Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2, \qquad S_{4,n}(x) = \sum_{i=1}^{d} \phi_i(\mathbf{K}) F_i^{(1)}(x) \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 h_k$$

$$\text{and } S_{5,n}(x) = \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2[\nu_k(x) + \widetilde{\nu}_k(x)^2 + 2F(x)\widetilde{\nu}_k(x)].$$

1. For the case where $a \in \left[\frac{\alpha}{3}, 1\right)$, we have $\lim\limits_{n\to+\infty} n\gamma_n > \min\left\{\frac{a+\alpha}{2}, 2a\right\} = \frac{a+\alpha}{2}$.

Asymptotic behavior of $S_{3,n}(x)$:
The application of lemma 1.2 ensures that

$$\lim_{n\to+\infty} \frac{1}{\gamma_n}\Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 = \frac{1}{2-\alpha\xi}.$$

Consequently, we get

$$S_{3,n}(x) = \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2[F(x)(1-F(x))] = F(x)(1-F(x))\frac{\gamma_n}{2-\alpha\xi} + o(\gamma_n).$$

Asymptotic behavior of $S_{4,n}(x)$:
Now, since $(\gamma_n h_n^{-1})_{n\geq 1} \in \mathcal{GS}(a+\alpha)$, the application of lemma 1.2 ensures that

$$\lim_{n\to+\infty} \frac{1}{\gamma_n h_n}\Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k\gamma_k h_k = \frac{1}{2-(a+\alpha)\xi}.$$

Therefore, we get

$$S_{4,n}(x) = \frac{\gamma_n h_n}{2-(a+\alpha)\xi} \sum_{i=1}^{d} \phi_i(\mathbf{K}) F_i^{(1)}(x) + o(\gamma_n h_n).$$

Asymptotic behaviour of $S_{5,n}(x)$:
Owing to the fact that F and $F_i^{(1)}$ are bounded and continuous at x for all $i \in \{1,\dots,d\}$, we have $\lim\limits_{k\to+\infty} \nu_k(x) = 0$ and $\lim\limits_{k\to+\infty} \widetilde{\nu}_k(x) = 0$. Thus, $\lim\limits_{k\to+\infty} \left(\nu_k(x) + \widetilde{\nu}_k(x)^2 + 2F(x)\widetilde{\nu}_k(x)\right) = 0$.
The application of lemma 1.2 ensures that

$$S_{5,n}(x) = \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2[\nu_k(x) + \widetilde{\nu}_k(x)^2 + 2F(x)\widetilde{\nu}_k(x)] = o(\gamma_n).$$

Therefore,

$$Var[F_n(x)] = \frac{\gamma_n}{2-\alpha\xi}F(x)(1-F(x)) - \frac{\gamma_n h_n}{2-(a+\alpha)\xi} \sum_{i=1}^{d} \phi_i(\mathbf{K}) F_i^{(1)}(x) + o(\gamma_n h_n).$$

2. In the case where $a \in [\frac{\alpha}{4}, \frac{\alpha}{3})$, we have $\gamma_n h_n = o(h_n^4)$ thus $\gamma_n h_n^{-3} \underset{n\to+\infty}{\longrightarrow} 0$. Then, by applying the second part of lemma 1.2 along with the fact that $\alpha_k = o(h_k^4) \underset{n\to+\infty}{\longrightarrow} 0$, it follows that

$$\Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k o(h_k^4) = o(1).$$

Hence,

$$\begin{aligned} Var[F_n(x)] &= \frac{\gamma_n}{2-\alpha\xi}F(x)(1-F(x)) + o(\gamma_n) - \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k o(h_k^4) + o(\gamma_n) \\ &= \frac{\gamma_n}{2-\alpha\xi}F(x)(1-F(x)) + o(\gamma_n). \end{aligned}$$

3. For the case where $a \in \left(0, \frac{\alpha}{4}\right)$, we have $\gamma_n = o(h_n^4)$ thus $\gamma_n h_n^{-4} \underset{n\to+\infty}{\longrightarrow} 0$. The application of lemma 1.2 ensures that

$$\begin{aligned} Var[F_n(x)] &= [F(x)(1-F(x))]\Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k o\left(h_k^4\right) - \sum_{i=1}^{d} \phi_i(\mathbf{K}) F_i^{(1)}(x) \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k o\left(h_k^4\right) + o\left(\gamma_n\right) \\ &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k o\left(h_k^4\right) + o\left(\gamma_n\right) \\ &= o\left(h_n^4\right). \end{aligned}$$

□

Now, let's recall the precise statement of Lyapunov's theorem, which shall be invested in the next proof.

Theorem 1.14. *Let (X_n) be a sequence of independent random variables, centered all with a finite moment of order $2+p$, $p>0$. We note, $u_n^2 = Var\left[\sum_{i=1}^{n} X_i\right] > 0$.*
Under the following Lyapunov condition,

$$\lim_{n\to+\infty} \frac{1}{u_n^{2+p}} \sum_{i=1}^{n} \mathbb{E}\left[|X_i|^{2+p}\right] = 0,$$

we obtain

$$\frac{1}{u_n} \sum_{i=1}^{n} X_i \underset{n\to+\infty}{\overset{\mathcal{D}}{\longrightarrow}} \mathcal{N}(0,1).$$

Proof of Theorem 1.5.
First, by the relation (1.1), we have

$$F_n(x) = (1-\gamma_n)F_{n-1}(x) + \gamma_n \mathcal{Z}_n(x),$$

It follows that

$$F_n(x) - \mathbb{E}[F_n(x)] = (1-\gamma_n)\left(F_{n-1}(x) - \mathbb{E}[F_{n-1}(x)]\right) + \gamma_n\left(\mathcal{Z}_n(x) - \mathbb{E}[\mathcal{Z}_n(x)]\right).$$

Using a simple recurrence, we obtain

$$\begin{aligned} F_n(x) - \mathbb{E}[F_n(x)] &= \prod_{i=1}^{n}(1-\gamma_i)(F_0(x) - \mathbb{E}[F_0(x)]) + \sum_{k=1}^{n}\prod_{j=k+1}^{n}(1-\gamma_j)\gamma_k(\mathcal{Z}_k(x) - \mathbb{E}[\mathcal{Z}_k(x)]) \\ &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k(\mathcal{Z}_k(x) - \mathbb{E}[\mathcal{Z}_k(x)]). \end{aligned}$$

Thus,

$$F_n(x) - \mathbb{E}[F_n(x)] = \Pi_n \sum_{k=1}^{n} Y_k(x),$$

where,

$$Y_k(x) = \Pi_k^{-1}\gamma_k(\mathcal{Z}_k(x) - \mathbb{E}[\mathcal{Z}_k(x)]). \tag{1.23}$$

Now, in order to apply Lyapunov's theorem for $Y_k(x)$, we mention that $\mathbb{E}[Y_k(x)] = 0$ and we state

$$\begin{aligned}
\upsilon_n^2 &= \sum_{k=1}^{n} Var[Y_k(x)] \\
&= \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2\, Var[\mathcal{Z}_k(x)] \\
&= \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 \left(\mathbb{E}[\mathcal{Z}_k^2(x)] - \mathbb{E}[\mathcal{Z}_k(x)]^2\right) \\
&= \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 \Big(F(x) + \nu_k'(x) - F^2(x) - \tilde{\nu}_k(x)^2 - 2F(x)\tilde{\nu}_k(x)\Big),
\end{aligned}$$

where $\nu_k'(x) = 2^d \displaystyle\int_{\mathbb{R}^d} \mathbf{K}(z)\mathcal{K}(z)\left[F(x - zh_k) - F(x)\right] dz$ and $\widetilde{\nu}_k(x) = \displaystyle\int_{\mathbb{R}^d} \mathbf{K}(z)\left[F(x - zh_k) - F(x)\right] dz$. It is obvious that $\lim\limits_{k\to+\infty} \nu_k'(x) = 0$ and $\lim\limits_{k\to+\infty} \widetilde{\nu}_k(x) = 0$. Thus, $\lim\limits_{k\to+\infty} \left(\nu_k'(x) - \widetilde{\nu}_k(x)^2 - 2F(x)\widetilde{\nu}_k(x)\right) = 0$. Hence, we get

$$\upsilon_n^2 = \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2[F(x)(1 - F(x)) + o(1)].$$

The application of lemma 1.2 ensures that

$$\upsilon_n^2 = \frac{\gamma_n}{\Pi_n^2}\left(\frac{1}{2 - \alpha\xi}F(x)(1 - F(x)) + o(1)\right).$$

Additionally, using (1.23) we have

$$\begin{aligned}
\mathbb{E}\left[|Y_k(x)|^{2+p}\right] &= \mathbb{E}\left[|\Pi_k^{-1}\gamma_k(\mathcal{Z}_k(x) - \mathbb{E}[\mathcal{Z}_k(x)])|^{2+p}\right] \\
&= \Pi_k^{-2-p}\gamma_k^{2+p}\mathbb{E}\left[|\mathcal{Z}_k(x) - \mathbb{E}[\mathcal{Z}_k(x)]|^{2+p}\right].
\end{aligned}$$

Thus, we get

$$\begin{aligned}
\mathbb{E}\left[|Y_k(x)|^{2+p}\right] &\leq 2\Pi_k^{-2-p}\gamma_k^{2+p}\mathbb{E}\left[|\mathcal{Z}_k(x)|^{2+p}\right]. \\
&= O\left(\Pi_k^{-2-p}\gamma_k^{2+p}\mathbb{E}\left[|\mathcal{Z}_k(x)|^{2+p}\right]\right).
\end{aligned}$$

Moreover, since we have $\mathcal{Z}_k(x) \leq 1$, which gives,

$$\forall p > 0,\ \mathbb{E}[|\mathcal{Z}_k(x)|^{2+p}] = O(1),$$

we infer that

$$\begin{aligned}
\sum_{k=1}^{n} \mathbb{E}[|Y_k(x)|^{2+p}] &= O\left(\sum_{k=1}^{n} \Pi_k^{-2-p}\gamma_k^{2+p}\mathbb{E}\left[|\mathcal{Z}_k(x)|^{2+p}\right]\right) \\
&= O\left(\sum_{k=1}^{n} \Pi_k^{-2-p}\gamma_k^{2+p}\right).
\end{aligned}$$

Here we suppose that $a \geq \alpha/4$. Departing from the fact that $\lim\limits_{n\to+\infty}(n\gamma_n) > \alpha/2$, which implies that there exists $p > 0$ such that

$$\xi^{-1} = \lim_{n\to+\infty} n\gamma_n > \frac{1+p}{2+p}\alpha.$$

Thus $(2+p)-(1+p)\alpha\xi > 0$ and the application of lemma 1.2 yields

$$\sum_{k=1}^{n} \mathbb{E}[|Y_k(x)|^{2+p}] = O\left(\frac{\gamma_n^{1+p}}{\Pi_n^{2+p}}\right).$$

Thus,

$$\frac{1}{\upsilon_n^{2+p}} \sum_{k=1}^{n} \mathbb{E}[|Y_k(x)|^{2+p}] \quad = \quad O\left(\gamma_n^{\frac{p}{2}}\right) = o\left(1\right).$$

Furthermore, since we have

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

Then, the application of the Lyapunov's theorem ensures that

$$\sqrt{\gamma_n^{-1}}\left(F_n(x) - \mathbb{E}[F_n(x)]\right) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{2-\alpha\xi}F(x)(1-F(x))\right). \tag{1.24}$$

In what follows, we distinguish the two cases displayed below:
First case: $a > \alpha/3$**:**
We have

$$\mathbb{E}[F_n(x)] - F(x) = o\left(\sqrt{\gamma_n h_n}\right).$$

By replacing the latter in (1.24), we obtain

$$\sqrt{\gamma_n^{-1}}\left(F_n(x) - F(x)\right) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{2-\alpha\xi}F(x)(1-F(x))\right).$$

Second case: $\alpha/4 \leq a \leq \alpha/3$**:**
We have

$$\mathbb{E}[F_n(x)] - F(x) = \frac{1}{2(1-2a\xi)} h_n^2 \sum_{j=1}^{d} \mu_j(\mathbf{K}) F_{jj}^{(2)}(x) + o\left(h_n^2\right).$$

As $\gamma_n^{-\frac{1}{2}} h_n^2 \underset{n\to+\infty}{\longrightarrow} c^{\frac{1}{2}}$, we obtain

$$\sqrt{\gamma_n^{-1}}\left(F_n(x) - F(x)\right) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(\frac{c^{\frac{1}{2}}}{2(1-2a\xi)} \sum_{j=1}^{d} \mu_j(\mathbf{K}) F_{jj}^{(2)}(x), \frac{1}{2-\alpha\xi}F(x)(1-F(x))\right).$$

The case where $a < \alpha/4$ implies the convergence in probability in the second part of the theorem. By applying the Bienaymé–Chebyshev inequality, we get

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

Since we have $\gamma_n^{-1} h_n^4 \underset{n\to+\infty}{\longrightarrow} +\infty$, then $\frac{Var[F_n(x)]}{h_n^4\epsilon^2} \underset{n\to+\infty}{\longrightarrow} 0$. Hence, we deduce that

$$\frac{1}{h_n^2}\left(F_n(x) - F(x)\right) \xrightarrow[n\to+\infty]{\mathbb{P}} \frac{1}{2(1-2a\xi)} \sum_{j=1}^{d} \mu_j(\mathbf{K}) F_{jj}^{(2)}(x).$$

□

Proof of Theorem 1.12.
Following similar steps as Slaoui (2014b), and by using (1.11) and (1.13), we obtain

$$MWISE[F_n] - MWISE[\widetilde{F}_n] = \frac{n^{-1}V_F}{(2\gamma_0-1)(\gamma_0-\frac{2}{3})^{\frac{2}{3}}}\Big\{ (\gamma_0-1)^2\left(\gamma_0-\frac{2}{3}\right)^{\frac{2}{3}} - 2^{-\frac{4}{3}}\frac{3}{4}(2\gamma_0-1)\left(\gamma_0^2-2^{\frac{4}{3}}\left(\gamma_0-\frac{2}{3}\right)^{\frac{2}{3}}\right)V_F^{-1}J_1^{\frac{4}{3}}J_2^{-\frac{1}{3}}n^{-\frac{1}{3}}+o\left(n^{-\frac{1}{3}}\right)\Big\}.$$

Which is no other than

$$n < \left(\frac{2^{-\frac{4}{3}}\frac{3}{4}(2\gamma_0-1)\left(\gamma_0^2-2^{\frac{4}{3}}\left(\gamma_0-\frac{2}{3}\right)^{\frac{2}{3}}\right)V_F^{-1}J_1^{\frac{4}{3}}J_2^{-\frac{1}{3}}}{(\gamma_0-1)^2(\gamma_0-\frac{2}{3})^{\frac{2}{3}}}\right)^3.$$

The fact that $\gamma_0 = \frac{2}{3}+\varepsilon$ ensures that

$$n < \frac{3^3}{2^{10}}\frac{J_1^4}{J_2V_F^3}\frac{(\frac{1}{3}+2\varepsilon)^3\left(\frac{2}{3}+\varepsilon-2^{\frac{2}{3}}\varepsilon^{\frac{1}{3}}\right)^3\left(\frac{2}{3}+\varepsilon+2^{\frac{2}{3}}\varepsilon^{\frac{1}{3}}\right)^3}{(\varepsilon-\frac{1}{3})^6\varepsilon^2}.$$

Thus, the result of the theorem is derived by taking into account the last inequality. □

Chapter 2

Statistical inferences for multivariate conditional cumulative distribution function estimation by stochastic approximation method

2.1 Introduction

Assume that we observe independent identically distributed vectors $(X_1, Y_1), \ldots, (X_n, Y_n)$ of a bivariate random variable $(X, Y) \in \mathbb{R} \times \mathbb{R}$ with common cumulative distribution function $\pi(x, y)$ where one is interested in modeling the functional dependence of the observation Y on the covariable X by the conditional cumulative distribution function (CCDF) of Y given $X = x$, denoted by, for all real y and x,

$$\pi(y|x) := \mathbb{P}\left[Y \leqslant y | X = x\right].$$

We shall also assume that the bivariate random variable (X, Y) (resp. the random variable X) has a density function $f_{(X,Y)}$ (resp. f_X) with respect to the Lebesgue measure. Recall that for all real y and x such that $f_X(x) \neq 0$, the CCDF of Y given $X = x$ is expressed by

$$\pi(y|x) = \frac{\int_{\mathbb{R}} \mathbb{1}_{\{u \leqslant y\}} f_{(X,Y)}(x, u) du}{f_X(x)}.$$

In a variety of non-parametric statistical problems, the estimation of a CCDF is a key aspect of inference. Remember that the CCDF has the merit of characterizing the conditional law of the considered bivariate random variables. Notably, the CCDF is often useful in reliability or survival analysis.
More specifically, the conditional survival function $S(y, x)$ defined by, for all real y and x, $S(y, x) := 1 - \pi(y|x)$ is of extreme interest, either by itself, or by its independence with the conditional hazard function $H(y, x)$ indicated by, for all real y and x, $H(y, x) := \dfrac{f(y, x)}{S(y, x)}$ where $f(y, x)$ denotes the conditional density of Y given $X = x$. Furthermore, conditional quantiles can also be deduced from the CCDF π by (pseudo)-inversion given x of the function $y \to \pi(y|x)$ and the same procedure may be applied to the estimator of CCDF to find conditional quantile estimators.
Several non-parametric estimators have been elaborated to estimate the CCDF. Many of them rely initially on estimating the $\int_{\mathbb{R}} \mathbb{1}_{\{u \leqslant y\}} f_{(X,Y)(x,u)} du$. The conditional cumulative distribution function was first extensively explored by Stute (1986) using a nearest-neighbor-type conditional empirical process. Subsequently, Hall *et al.* (1999), motivated by the problem of setting prediction intervals in time series analysis, developed a new non-parametric method for CCDF

estimation resting on an adjusted form of Nadaraya–Watson estimator. Afterwards, Ferrigno *et al.* (2014) established uniform asymptotic certainty bands for the CCDF using the same strategy.
For a general non-parametric regression model, Kiwitt and Neumeyer (2012) set up two estimators using a kernel approch, where the distribution of the error given the covariate is modeled by a CCDF provided by $\mathbb{P}(\epsilon \leq y|X = x)$.
On a given compact set, Brunel *et al.* (2010) constructed a minimax estimator of the CCDF. Thereafter, Veraverbeke *et al.* (2014) built up a new estimator of CCDF investing a method of pre-adjusting the original observations non-parametrically. Recently, Bouanani *et al.* (2020) introduced a new method to settle CCDF estimation problem based on local polynomial technique.
Many functional estimations are grounded on estimating the CCDF see e.g Laksaci and Maref (2009), Laksaci and Hachemi (2012) and Almanjahie *et al.* (2018). The CCDF is involved in a wide range of applications, for instance, in medicine see Gannoun *et al.* (2002), econometrics see Li *et al.* (2013) or machine learning domain see the recent work of Chilinski and Silva (2020). In a broader context, extensive state of art works including various non-parametric approaches tackled the conditional estimation. We can state for example Yu and Jones (1998), Fan *et al.* (1996), Berlinet *et al.* (1998b),Berlinet *et al.* (1998a), Honda (2000b),Honda (2000a) and Plancade (2013). For recent references see Benziadi *et al.* (2016), Choudhury *et al.* (2018), Al-Awadhi *et al.* (2019a) , Chikr-Elmezouar *et al.* (2019b) and Slaoui and Khardani (2020).
Over the past decade, data streams have become an increasingly important area of research. Some of the most common data streams include Internet packet data, Twitter activity, Facebook newsfeed, credit card transactions and more recently COVID-19 epidemic data. In these situations, the data arrives regularly so that it is impossible to store it in a traditional database. In such a context, it is very interesting to build a recursive multivariate conditional cumulative distribution estimator that does not need to store all the data in memory and that can be easily updated to handle the online data. The basic target of this chapter is to provide a non-parametric strategy to recursively estimate the CCDF.

2.1.1 Presentation of the method

Let (X, Y) be a random vector with values in $\mathbb{R}^d \times \mathbb{R}^q$, $q \geq 1$, with a joint density function $f_{(X,Y)}$ and let f_X denote the marginal probability density of X given by $f_X(x) = \int_{\mathbb{R}^q} f_{(X,Y)}(x, u)du$. Moreover, let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent random vectors identically distributed as (X, Y). In this work, our central focus is upon the problem of estimating the CCDF of Y given $X = x$ provided by

$$\begin{aligned} \pi : \quad & \mathbb{R}^q \times \mathbb{R}^d \longrightarrow \mathbb{R} \\ & (y|x) \quad \longmapsto \mathbb{P}[Y \leqslant y|X = x] = \frac{a(x, y)}{f_X(x)}, \end{aligned}$$

where

$$a(x, y) = \int_{\mathbb{R}^q} \mathbb{1}_{\{u \leqslant y\}} f_{(X,Y)}(x, u)du \quad \text{and} \quad f_X(x) = \int_{\mathbb{R}^q} f_{(X,Y)}(x, u)du.$$

The recursive estimator was constructed based on dint of stochastic approximation method. Since, Slaoui (2014b) reused stochastic approximation methods to enhance the qualities of the univariate distribution function estimator and the previous chapter elaborated the multivariate one and following the same recursive approach, we intend to establish a multivariate conditional cumulative distribution function estimator.
To build up a stochastic algorithm, which approaches the function a at a given couple of vectors (x, y), we define an algorithm of search of the zero function $\phi : z \longmapsto a(x, y) - z$ and we set:

$$(i)\ a_0(x, y) \in \mathbb{R} \qquad (ii)\ \text{for all } n \geq 1,\ a_n(x, y) = a_{n-1}(x, y) + \gamma_n U_n(x, y),$$

where $U_n(x,y)$ corresponds to an observation of the function ϕ at the point $a_{n-1}(x,y)$.
In the following, we set a multivariate indicator function denoted χ and identified by

$$\chi_k : \mathbb{R}^q \longrightarrow \mathbb{R},\ y \longmapsto \mathbb{1}_{\{Y_k \leqslant y\}}.$$

By considering $U_n(x,y) = \chi_n(y)\, h_n^{-d}\, \mathbf{K}\left(\frac{x - X_n}{h_n}\right) - a_{n-1}(x,y)$, the stochastic approximation algorithm that is devoted to estimate recursively the function a at a couple of vectors (x,y) can be stated as follows :

$$a_n(x,y) = (1-\gamma_n) a_{n-1}(x,y) + \gamma_n\, \chi_n(y)\, h_n^{-d}\, \mathbf{K}\left(\frac{x - X_n}{h_n}\right). \tag{2.1}$$

Throughout this section, we consider that $a_0(x,y) = 0$. Therefore, by recurrence, we get

$$a_n(x,y) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k\, \chi_k(y) h_k^{-d}\, \mathbf{K}\left(\frac{x - X_k}{h_k}\right). \tag{2.2}$$

Within this framework, we recall the recursive multivariate probability density estimator of the density function noted f_n and defined in Mokkadem *et al.* (2009a). It was constructed with the same tools of stochastic approximation algorithm and under the condition that $f_0(x) = 0$, we have:

$$f_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k\, h_k^{-d}\, \mathbf{K}\left(\frac{x - X_k}{h_k}\right). \tag{2.3}$$

Therefore, we introduce our recursive estimator π_n specified by

$$\pi_n(y|x) = \begin{cases} \dfrac{a_n(x,y)}{f_n(x)} & \text{if} \quad f_n(x) \neq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{2.4}$$

Our main purpose is to examine the asymptotic properties of the proposed multivariate estimator of the CCDF and to corroborate its performances.
Moreover, we set forward the non-recursive estimator of the function a given by

$$\widetilde{a}_n(x,y) = \frac{1}{n h_n^d} \sum_{k=1}^{n} \chi_k(y)\, \mathbf{K}\left(\frac{x - X_k}{h_n}\right)$$

and the non-recursive estimator of the multivariate density function f defined by

$$\widetilde{f}_n(x) = \frac{1}{n h_n^d} \sum_{k=1}^{n} \mathbf{K}\left(\frac{x - X_k}{h_n}\right). \tag{2.5}$$

Hence, we shall compare our estimator to the generalized kernel CCDF estimator of Nadaraya-Watson Nadaraya (1964) and Watson (1964) $\widetilde{\pi}_n$ expressed by

$$\widetilde{\pi}_n(y|x) = \begin{cases} \dfrac{\widetilde{a}_n(x,y)}{\widetilde{f}_n(x)} & \text{if} \quad \widetilde{f}_n(x) \neq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{2.6}$$

2.1.2 Notations and assumptions

For this section and under (A_1) and (A_2) given in 1.1.1, we provide the following notations and assumptions which will be intensively used for our theoretical main results.

(A_4) (i) The functions f_X and a are bounded and twice differentiable.

(ii) For all $i,j \in \{1,\dots,d\}$, $f^{(2)}_{X_{ij}} := \dfrac{\partial^2 f_X}{\partial x_i \partial x_j}$ and $a^{(2)}_{ij} := \dfrac{\partial^2 a}{\partial x_i \partial x_j}$ are bounded and continuous at x.

First of all, we need to recall the following proposition which introduces the bias and the variance of f_n. The proof of this result was depicted in Mokkadem *et al.* (2009a).

2.1.3 Bias and variance of f_n

Proposition 2.1. *Under assumptions* (A_1), (A_2) *and* (A_4), *we obtain*

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

$$\mathbb{E}[f_n(x)] - f_X(x) = \frac{1}{2(1-2a\xi)} \left(\sum_{j=1}^{d} \mu_j(\mathbf{K}) f^{(2)}_{X_{jj}}(x) \right) h_n^2 + \mathbf{o}\left(h_n^2\right). \tag{2.7}$$

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

$$\mathbb{E}[f_n(x)] - f_X(x) = \mathbf{o}\left(\sqrt{\gamma_n h_n^{-d}}\right). \tag{2.8}$$

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

$$Var[f_n(x)] = \mathbf{o}\left(h_n^4\right). \tag{2.9}$$

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

$$Var[f_n(x)] = \frac{1}{2-(\alpha-ad)\xi} f_X(x) R(\mathbf{K}) \frac{\gamma_n}{h_n^d} + \mathbf{o}\left(\gamma_n h_n^{-d}\right). \tag{2.10}$$

2.2 Main results

In order to explore the asymptotic properties of our estimator π_n, we need first to introduce the following proposition which provides the bias and the variance of a_n.

2.2.1 Bias and variance of a_n

Proposition 2.2. *Under assumptions* (A_1), (A_2) *and* (A_4), *we obtain*

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

$$\mathbb{E}[a_n(x,y)] - a(x,y) = \frac{1}{2(1-2a\xi)} \left(\sum_{j=1}^{d} \mu_j(\mathbf{K}) a^{(2)}_{jj}(x,y) \right) h_n^2 + \mathbf{o}\left(h_n^2\right). \tag{2.11}$$

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

$$\mathbb{E}[a_n(x,y)] - a(x,y) = \mathbf{o}\left(\sqrt{\gamma_n h_n^{-d}}\right). \tag{2.12}$$

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

$$Var[a_n(x,y)] = \mathbf{o}\left(h_n^4\right). \tag{2.13}$$

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

$$Var[a_n(x,y)] = \frac{1}{2-(\alpha-ad)\xi} a(x,y) R(\mathbf{K}) \frac{\gamma_n}{h_n^d} + \mathbf{o}\left(\gamma_n h_n^{-d}\right). \tag{2.14}$$

In the following theorem, we introduce our main result which provides the bias and the variance of our CCDF multivariate estimator π_n. The following notations are highly useful as they are invested throughout the whole chapter.

$$\text{For all } i,j \in \{1,\dots,d\}, \qquad \pi_i^{(1)}(y,.) := \frac{\partial \pi}{\partial x_i}(y,.) \qquad \pi_{ij}^{(2)}(y,.) := \frac{\partial^2 \pi}{\partial x_i \partial x_j}(y,.),$$
$$a_i^{(1)}(.,y) := \frac{\partial a}{\partial x_i}(.,y), \qquad f_{X_i}^{(1)}(.) := \frac{\partial f_X}{\partial x_i}(.).$$

2.2.2 Bias and variance of π_n

Theorem 2.3. *Let assumptions* (A_1), (A_2) *and* (A_4) *hold and note* $R(\mathbf{K}) := \int_{\mathbb{R}^d} \mathbf{K}^2(z)\, dz$, *we obtain*

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

$$\mathbb{E}[\pi_n(y|x)] - \pi(y|x) = \frac{1}{2(1-2a\xi)} \frac{1}{f_X(x)} \sum_{j=1}^{d} \mu_j(\mathbf{K}) \left[\pi_{jj}^{(2)}(y|x) f_X(x) + 2\pi_j^{(1)}(y|x) f_{X_j}^{(1)}(x)\right] h_n^2 + \mathbf{o}\left(h_n^2\right). \tag{2.15}$$

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

$$\mathbb{E}[\pi_n(y|x)] - \pi(y|x) = \mathbf{o}\left(\sqrt{\gamma_n h_n^{-d}}\right). \tag{2.16}$$

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

$$Var[\pi_n(y|x)] = \mathbf{o}\left(h_n^4\right). \tag{2.17}$$

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

$$Var[\pi_n(y|x)] = \frac{R(\mathbf{K})}{2-(\alpha-ad)\xi} \frac{\pi(y|x)(1-\pi(y|x))}{f_X(x)} \frac{\gamma_n}{h_n^d} + \mathbf{o}\left(\frac{\gamma_n}{h_n^d}\right). \tag{2.18}$$

In the sequel, let us present the following theorem which identifies the asymptotic normality of our recursive estimator π_n.

2.2.3 Weak pointwise convergence rate of π_n

Theorem 2.4. *Let assumptions* (A_1), (A_2) *and* (A_4) *hold.*

1. *If there exists a non-negative real c such that $\gamma_n^{-1} h_n^{d+4} \underset{n\to+\infty}{\longrightarrow} c$, then*

$$\sqrt{\gamma_n^{-1} h_n^d}\,(\pi_n(y|x) - \pi(y|x)) \underset{n\to+\infty}{\xrightarrow{\mathcal{D}}} \mathcal{N}\Big(\sqrt{c}\, M(x,y)\ ,\ \sigma^2(x,y)\Big). \tag{2.19}$$

with

$$M(x,y) = \frac{1}{2(1-2a\xi)} \frac{1}{f_X(x)} \sum_{j=1}^{d} \mu_j(\mathbf{K}) \left[\pi_{jj}^{(2)}(y|x) f_X(x) + 2\pi_j^{(1)}(y|x) f_{X_j}^{(1)}(x)\right]$$

and

$$\sigma^2(x,y) = \frac{R(\mathbf{K})}{2-(\alpha-ad)\xi} \frac{\pi(y|x)(1-\pi(y|x))}{f_X(x)}.$$

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

$$\frac{1}{h_n^2}\,(\pi_n(y|x) - \pi(y|x)) \underset{n\to+\infty}{\xrightarrow{\mathbb{P}}} M(x,y).$$

2.3 Optimal choice of the stepsizes

In order to assess the asymptotic quality of the CCDF recursive estimator π_n, we set up the Mean Weighted Integrated Squared Error ($MWISE$). We first introduce the $MWISE$ expression:

$$MWISE[\pi_n] = \int_{\mathbb{R}^{d+q}} \left[(\mathbb{E}[\pi_n(y|x)] - \pi(y|x))^2 + Var[\pi_n(y|x)]\right] {f_X}^2(x) f_{X,Y}(x,y)dxdy. \tag{2.20}$$

2.3.1 Asymptotic expressions of $MWISE[\pi_n]$

First of all, let us set the following notations

$$I_1 := \int_{\mathbb{R}^{d+q}} \left(\sum_{j=1}^{d} \mu_j(\mathbf{K}) \left[\pi_{jj}^{(2)}(y|x) f_X(x) + 2\pi_j^{(1)}(y|x) f_{X_j}^{(1)}(x)\right]\right)^2 f_{X,Y}(x,y)dxdy$$

and

$$I_2 := \int_{\mathbb{R}^{d+q}} \pi(y|x)\,(1-\pi(y|x))\, f_X(x) f_{X,Y}(x,y)dxdy.$$

Proposition 2.5. *The $MWISE$ of the estimator π_n is maintained as follows.*
If $a \in \left(0, \frac{\alpha}{d+4}\right)$, then

$$MWISE[\pi_n] = \frac{1}{4} \frac{I_1}{(1-2a\xi)^2} h_n^4 + \mathbf{o}(h_n^4).$$

If $a = \frac{\alpha}{d+4}$, then

$$MWISE[\pi_n] = \frac{I_2}{2-(\alpha-ad)\xi} R(\mathbf{K}) \gamma_n h_n^{-d} + \frac{1}{4} \frac{I_1}{(1-2a\xi)^2} h_n^4 + \mathbf{o}(h_n^4).$$

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

$$MWISE[\pi_n] = \frac{I_2}{2-(\alpha-ad)\xi} R(\mathbf{K}) \gamma_n h_n^{-d} + \mathbf{o}\left(\gamma_n h_n^{-d}\right).$$

The following corollary ensures that the bandwidth which minimizes the $MWISE$ of π_n depends on the choice of the stepsize (γ_n). As a matter of fact, the corresponding $MWISE$ depends also on (γ_n).

Corollary 2.6. *Let assumptions (A_1), (A_2) and (A_4) hold. To minimize the $MWISE$ of π_n, the bandwidth (h_n) must be equal to*

$$\left(\left(\frac{d\,(1-2a\xi)^2}{2-(\alpha-ad)\xi}\frac{I_2}{I_1}R(\mathbf{K})\right)^{\frac{1}{d+4}}\gamma_n^{\frac{1}{d+4}}\right).$$

Hence, the corresponding $MWISE$ is determined by

$$MWISE[\pi_n]=\frac{d+4}{4d^{\frac{d}{d+4}}}\left(\frac{I_1}{(1-2a\xi)^2}\right)^{\frac{d}{d+4}}\left(\frac{I_2}{2-(\alpha-ad)\xi}\right)^{\frac{4}{d+4}}R(\mathbf{K})^{\frac{4}{d+4}}\gamma_n^{\frac{4}{d+4}}+\mathbf{o}\left(\gamma_n^{\frac{4}{d+4}}\right).$$

The following corollary holds in the special case where (γ_n) is chosen as $(\gamma_n)=(\gamma_0 n^{-1})$ in order to minimize the $MWISE[\pi_n]$.

Corollary 2.7. *Let assumptions (A_1), (A_2) and (A_4) hold. To minimize the $MWISE$ of π_n, we need to opt for the stepsize (γ_n) in $\mathcal{GS}(-1)$ such that $\lim_{n\to+\infty} n\gamma_n=\gamma_0$. Then the bandwidth (h_n) must be equal to*

$$\left(\left(\frac{d\,(\gamma_0(d+4)-2)}{2(d+4)}\frac{I_2}{I_1}R(\mathbf{K})\right)^{\frac{1}{d+4}}n^{-\frac{1}{d+4}}\right).$$

Consequently, the corresponding $MWISE$ is identified by

$$MWISE[\pi_n]=\frac{(d+4)^{\frac{3d+8}{d+4}}}{4^{\frac{d+6}{d+4}}d^{\frac{d}{d+4}}}\gamma_0^2\,(\gamma_0(d+4)-2)^{-\frac{(2d+4)}{d+4}}\,I_1^{\frac{d}{d+4}}I_2^{\frac{4}{d+4}}R(\mathbf{K})^{\frac{4}{d+4}}n^{-\frac{4}{d+4}}+\mathbf{o}\left(n^{-\frac{4}{d+4}}\right).$$

In order to get the optimal choice of (γ_n), we deduce that the minimum of $MWISE[\pi_n]$ is achieved at $\gamma_0=1$. Hence, we introduce the following corollary.

Corollary 2.8. *Let assumptions (A_1), (A_2) and (A_4) hold. To minimize the $MWISE$ of π_n, we must select the stepsize (γ_n) in $\mathcal{GS}(-1)$ such that $\lim_{n\to+\infty} n\gamma_n=1$. Therefore, the optimal bandwidth (h_n) must equal*

$$\left(\left(\frac{d(d+2)}{2(d+4)}\frac{I_2}{I_1}R(\mathbf{K})\right)^{\frac{1}{d+4}}n^{-\frac{1}{d+4}}\right). \tag{2.21}$$

As a result, the corresponding $MWISE$ is expressed by

$$MWISE[\pi_n]=\frac{(d+4)^{\frac{3d+8}{d+4}}}{4^{\frac{d+6}{d+4}}d^{\frac{d}{d+4}}(d+2)^{\frac{2d+4}{d+4}}}I_1^{\frac{d}{d+4}}I_2^{\frac{4}{d+4}}R(\mathbf{K})^{\frac{4}{d+4}}n^{-\frac{4}{d+4}}+\mathbf{o}\left(n^{-\frac{4}{d+4}}\right).$$

Remark 2.9. *Note that, for the particular case where the stepsize (γ_n) is in $\mathcal{GS}(-1)$ such that $\lim_{n\to+\infty} n\gamma_n=1$ and the bandwidth (h_n) is chosen such that $\lim_{n\to+\infty} nh_n^{d+4}=0$ (which corresponds to undersmoothing), the asymptotic normality of our proposed estimator is indicated as follows*

$$\sqrt{nh_n^d}\,(\pi_n(y|x)-\pi(y|x))\xrightarrow[n\to+\infty]{\mathcal{D}}\mathcal{N}\left(0\,,\,\frac{d+4}{2(d+2)}R(\mathbf{K})\frac{\pi(y|x)(1-\pi(y|x))}{f_X(x)}\right). \tag{2.22}$$

The statistical inference of the CCDF multivariate non-recursive estimator $\widetilde{\pi}_n$ is addressed in our next section.
The following results can be handled in nearly the same way as π_n. The unique difference lies in the fact that it pertains to a non-recursive case. (See Hall *et al.* (1999) for more details of the univariate case.)

2.4 Asymptotic properties of $\widetilde{\pi}_n$

In order to tackle the asymptotic properties of our estimator π_n, we need first to introduce the following proposition which provides the bias and the variance of $\widetilde{\pi}_n$.

2.4.1 Bias and variance of $\widetilde{\pi}_n$

Proposition 2.10. *Let assumptions (A_1) and (A_4) hold. Then the bias and variance of Nadaraya-Watson's estimator are displayed as follows.*

1. *The bias of $\widetilde{\pi}_n$:*

$$\mathbb{E}[\widetilde{\pi}_n(y|x)]-\pi(y|x) = \frac{1}{2f_X(x)}\left(\sum_{j=1}^{d}\mu_j(\mathbf{K})\left[\pi_{jj}^{(2)}(y|x)f_X(x)+2\pi_j^{(1)}(y|x)f_X{}_j^{(1)}(x)\right]\right)h_n^2+\mathbf{o}\left(h_n^2\right).$$

2. *The variance of $\widetilde{\pi}_n$:*

$$Var[\widetilde{\pi}_n(x)] = R(\mathbf{K})\,\frac{\pi(y|x)(1-\pi(y|x))}{f_X(x)}\frac{1}{nh_n^d}+\mathbf{o}\left(\frac{1}{nh_n^d}\right).$$

The following proposition yields the distribution convergence rate of the non-recursive estimator.

2.4.2 Asymptotic normality of $\widetilde{\pi}_n$

Theorem 2.11. *Let assumptions (A_1) and (A_4) hold and suppose that $nh_n^{d+4} \underset{n\to+\infty}{\longrightarrow} 0$. Then,*

$$\sqrt{nh_n^d}\left(\widetilde{\pi}_n(y|x)-\pi(y|x)\right) \underset{n\to+\infty}{\overset{\mathcal{D}}{\longrightarrow}} \mathcal{N}\left(0\,,\; R(\mathbf{K})\,\frac{\pi(y|x)(1-\pi(y|x))}{f_X(x)}\right). \tag{2.23}$$

Remark 2.12. *It is obvious to infer from the expressions (2.22) and (2.23) that our CCDF proposed estimator is better than non-recursive one in terms of variance.*

In the next subsection, we exhibit the expression of the Mean Weighted Integrated Squared Error of Nadaraya-Watson's estimator.

2.4.3 Asymptotic expression of $MWISE[\widetilde{\pi}_n]$

Corollary 2.13. *The $MWISE$ expression of the non-recursive CCDF estimator is given by*

$$MWISE[\widetilde{\pi}_n] = \frac{1}{4}I_1h_n^4+I_2R(\mathbf{K})\frac{1}{nh_n^d}+\mathbf{o}\left(h_n^4+\frac{1}{nh_n^d}\right).$$

Proposition 2.14. *Let assumptions (A_1) and (A_4) hold. To minimize the $MWISE$ of $\widetilde{\pi}_n$, the bandwidth (h_n) must be equal to*

$$\left(\left(d\,\frac{I_2}{I_1}R(\mathbf{K})\right)^{\frac{1}{d+4}}n^{-\frac{1}{d+4}}\right). \tag{2.24}$$

Hence, the corresponding MWISE is determined by

$$MWISE[\widetilde{\pi}_n] = \frac{d+4}{4d^{\frac{d}{d+4}}} I_2^{\frac{4}{d+4}} I_1^{\frac{d}{d+4}} R(\mathbf{K})^{\frac{4}{d+4}} n^{-\frac{4}{d+4}} + \mathbf{o}\left(n^{-\frac{4}{d+4}}\right).$$

2.5 Bandwidth selection

Although theoretical asymptotic study yields the optimal bandwidth, the fact that we do not know the density function makes it hard to interpret it in practice. Hence, kernel smoothing in non-parametric statistics requires the choice of a bandwidth parameter. This choice is crucial to obtain a good rate of consistency of the kernel estimators. It has a significant influence on the size of the bias. One has to find an appropriate bandwidth that produces an estimator which has a good balance between the bias and the variance of the estimator of the function $a(\cdot,\cdot)$ as well as $f(\cdot)$. It is worth noticing that the bandwidth selection methods reported in the literature can be divided into three broad classes: the cross-validation techniques, the plug-in ideas and the bootstrap procedure. In this investigation, we are basically interested in the plug-in method. Altman and Leger (1995) developed an efficient method of bandwidth selection, which minimizes an estimate of the mean weighted integrated squared error, using the density function as a weight function. For this reason, we followed the work of Slaoui (2014a).

2.5.1 Plug-in bandwidth selection:

As a result to the plug-in procedure, based on the expression of the $MWISE$, we estimate the unknown quantities I_1 and I_2 by elaborating asymptotic unbiased estimators. Basically, we recall $(b_n) \in \mathcal{GS}(-\delta)$, as introduced in 1.15. In the following and for the sake of simplicity, the kernel $\mathbf{K}$ we shall use is considered as a product of univariate kernels given in (1.14).
Moreover, we assume that K_b stands for a kernel with bandwidth b_n such that $\delta = \frac{2}{5}$, and $K^{(2)}_{b'}$ corresponds to the second derivative of a kernel $K_{b'}$ with the associated bandwidth b'_n such that $\delta = \frac{3}{14}$. Note that our choice of the parameter δ is based on the work of Slaoui (2014a).
In addition, we note:

$$I_1 = \mu^2(K)\left(J_1 - 2J_2 + J_3\right),$$

where

$$J_1 = \int_{\mathbb{R}^{d+q}} \left(\sum_{j=1}^{d} a^{(2)}_{jj}(x,y)\right)^2 f_{X,Y}(x,y)dxdy, \qquad J_3 = \int_{\mathbb{R}^{d+q}} \left(\sum_{j=1}^{d} f^{(2)}_{X\,jj}(x)\right)^2 \pi^2(y|x) f_{X,Y}(x,y)dxdy,$$

$$J_2 = \int_{\mathbb{R}^{d+q}} \left(\sum_{j=1}^{d} a^{(2)}_{jj}(x,y)\right)\left(\sum_{j=1}^{d} f^{(2)}_{X\,jj}(x)\right) \pi(y|x) f_{X,Y}(x,y)dxdy, \qquad \mu(K) = \int_{\mathbb{R}} z^2 K(z)dz.$$

Recursive estimator π_n:

To estimate the optimal bandwidth (2.25), we need to estimate I_1 and I_2. Here we can write

$$a_n(x,y) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k\, h_k^{-d} \chi_k(y)\, \mathbf{K}\left(\frac{x - X_k}{h_k}\right) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k h_k^{-d} \prod_{i=1}^{d} K\left(\frac{x - X_{ki}}{h_k}\right) \chi_{ki}(y)$$

and

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

Estimation of I_1:

$$\widehat{J}_1 = \frac{\Pi_n^2}{n} \sum_{\substack{i,j,k=1 \\ i\neq j\neq k}}^{n} \Pi_j^{-1}\Pi_k^{-1}\gamma_j\gamma_k {b'}_j^{-(d+2)} {b'}_k^{-(d+2)} \left[\sum_{v=1}^{d} K_{b'}^{(2)}\left(\frac{X_{iv}-X_{jv}}{b'_j}\right)\prod_{\substack{l=1 \\ l\neq v}}^{d} K_b\left(\frac{X_{il}-X_{jl}}{b_j}\right)\right] \times \left[\sum_{v=1}^{d} K_{b'}^{(2)}\left(\frac{X_{iv}-X_{kv}}{b'_k}\right)\prod_{\substack{l=1 \\ l\neq v}}^{d} K_b\left(\frac{X_{il}-X_{kl}}{b_k}\right)\right]\prod_{s=1}^{q}\chi_{js}(Y_{is})\chi_{ks}(Y_{is}),$$

$$\widehat{J}_2 = \frac{\Pi_n^2}{n} \sum_{\substack{i,j,k,u=1 \\ i\neq j\neq k\neq u}}^{n} \Pi_j^{-1}\Pi_k^{-1}\gamma_j\gamma_k {b'}_j^{-(d+2)} {b'}_k^{-(d+2)} \left[\sum_{v=1}^{d} K_{b'}^{(2)}\left(\frac{X_{iv}-X_{jv}}{b'_j}\right)\prod_{\substack{l=1 \\ l\neq v}}^{d} K_b\left(\frac{X_{il}-X_{jl}}{b_j}\right)\right] \times \left[\sum_{v=1}^{d} K_{b'}^{(2)}\left(\frac{X_{iv}-X_{kv}}{b'_k}\right)\prod_{\substack{l=1 \\ l\neq v}}^{d} K_b\left(\frac{X_{il}-X_{kl}}{b_k}\right)\right]\prod_{s=1}^{q}\chi_{us}(Y_{is})\chi_{js}(Y_{is}),$$

$$\widehat{J}_3 = \frac{\Pi_n^2}{n} \sum_{\substack{i,j,k,u,m=1 \\ i\neq j\neq k\neq u\neq m}}^{n} \Pi_j^{-1}\Pi_k^{-1}\gamma_j\gamma_k {b'}_j^{-(d+2)} {b'}_k^{-(d+2)} \left[\sum_{v=1}^{d} K_{b'}^{(2)}\left(\frac{X_{iv}-X_{jv}}{b'_j}\right)\prod_{\substack{l=1 \\ l\neq v}}^{d} K_b\left(\frac{X_{il}-X_{jl}}{b_j}\right)\right] \times \left[\sum_{v=1}^{d} K_{b'}^{(2)}\left(\frac{X_{iv}-X_{kv}}{b'_k}\right)\prod_{\substack{l=1 \\ l\neq v}}^{d} K_b\left(\frac{X_{il}-X_{kl}}{b_k}\right)\right]\prod_{s=1}^{q}\chi_{us}(Y_{is})\chi_{ms}(Y_{is}),$$

At this stage, we obtain

$$\widehat{I}_1 = \mu^2(K)\left(\widehat{J}_1 - 2\widehat{J}_2 + \widehat{J}_3\right).$$

Estimation of I_2:

$$\widehat{I}_2 = \frac{\Pi_n}{n} \sum_{\substack{i,k,u=1 \\ i\neq k\neq u}}^{n} \Pi_k^{-1}\gamma_k b_k^{-1} \prod_{l=1}^{d} K_b\left(\frac{X_{il}-X_{kl}}{b_k}\right)\prod_{s=1}^{q}\chi_{us}(Y_{is})\left(1-\chi_{ks}(Y_{is})\right),$$

As a result, the plug-in estimator of (2.25) is determined by

$$(h_n) = \left(\left(\frac{d(d+2)}{2(d+4)}\right)^{\frac{1}{d+4}}\left(\frac{\widehat{I}_2}{\widehat{I}_1}\right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{-\frac{1}{d+4}}\right), \tag{2.25}$$

Eventually, an estimator of $MWISE[\pi_n]$ is specified by

$$\widehat{MWISE}[\pi_n] = \frac{(d+4)^{\frac{3d+8}{d+4}}}{4^{\frac{d+6}{d+4}} d^{\frac{d}{d+4}}(d+2)^{\frac{d+6}{d+4}}}\left(\widehat{I}_1\right)^{\frac{d}{d+4}}\left(\widehat{I}_2\right)^{\frac{4}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{-\frac{4}{d+4}} + \mathbf{o}\left(n^{-\frac{4}{d+4}}\right).$$

Non-Recursive estimator $\widetilde{\pi}_n$:

To estimate the optimal bandwidth (2.24), we need to estimate I_1 and I_2.Therefore, we can state

$$\widetilde{a}_n(x,y) = \frac{1}{nh_n^d}\sum_{k=1}^{n}\chi_k(y)\,\mathbf{K}\left(\frac{x-X_k}{h_n}\right) = \frac{1}{nh_n^d}\sum_{k=1}^{n}\prod_{i=1}^{d}K\left(\frac{x-X_{ki}}{h_k}\right)\chi_{ki}(y)$$

and

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

Estimation of I_1:

$$\widetilde{J}_1 = \frac{1}{n^3b_n^6}\sum_{\substack{i,j,k=1\\ i\neq j\neq k}}^{n}\left[\sum_{v=1}^{d}K_{b'}^{(2)}\left(\frac{X_{iv}-X_{jv}}{b'_n}\right)\prod_{\substack{l=1\\ l\neq v}}^{d}K_b\left(\frac{X_{il}-X_{jl}}{b_n}\right)\right]$$
$$\times\left[\sum_{v=1}^{d}K_{b'}^{(2)}\left(\frac{X_{iv}-X_{kv}}{b'_n}\right)\prod_{\substack{l=1\\ l\neq v}}^{d}K_b\left(\frac{X_{il}-X_{kl}}{b_n}\right)\right]\prod_{s=1}^{q}\chi_{js}(Y_{is})\chi_{ks}(Y_{is}),$$

$$\widetilde{J}_2 = \frac{1}{n^3b_n^6}\sum_{\substack{i,j,k,u=1\\ i\neq j\neq k\neq u}}^{n}\left[\sum_{v=1}^{d}K_{b'}^{(2)}\left(\frac{X_{iv}-X_{jv}}{b'_n}\right)\prod_{\substack{l=1\\ l\neq v}}^{d}K_b\left(\frac{X_{il}-X_{jl}}{b_n}\right)\right]$$
$$\times\left[\sum_{v=1}^{d}K_{b'}^{(2)}\left(\frac{X_{iv}-X_{kv}}{b'_n}\right)\prod_{\substack{l=1\\ l\neq v}}^{d}K_b\left(\frac{X_{il}-X_{kl}}{b_n}\right)\right]\prod_{s=1}^{q}\chi_{us}(Y_{is})\chi_{js}(Y_{is}),$$

$$\widetilde{J}_3 = \frac{1}{n^4b_n^6}\sum_{\substack{i,j,k,u,m=1\\ i\neq j\neq k\neq u\neq m}}^{n}\left[\sum_{v=1}^{d}K_{b'}^{(2)}\left(\frac{X_{iv}-X_{jv}}{b'_n}\right)\prod_{\substack{l=1\\ l\neq v}}^{d}K_b\left(\frac{X_{il}-X_{jl}}{b_n}\right)\right]$$
$$\times\left[\sum_{v=1}^{d}K_{b'}^{(2)}\left(\frac{X_{iv}-X_{kv}}{b'_n}\right)\prod_{\substack{l=1\\ l\neq v}}^{d}K_b\left(\frac{X_{il}-X_{kl}}{b_n}\right)\right]\prod_{s=1}^{q}\chi_{us}(Y_{is})\chi_{ms}(Y_{is}),$$

Then, we obtain

$$\widetilde{I}_1 = \mu^2(K)\left(\widetilde{J}_1 - 2\widetilde{J}_2 + \widetilde{J}_3\right).$$

Estimation of I_2:

$$\widetilde{I}_2 = \frac{1}{n^2b_n}\sum_{\substack{i,k,u=1\\ i\neq k\neq u}}^{n}\prod_{l=1}^{d}K_b\left(\frac{X_{il}-X_{kl}}{b_n}\right)\prod_{s=1}^{q}\chi_{us}(Y_{is})\left(1-\chi_{ks}(Y_{is})\right),$$

As a result, the plug-in estimator of (2.24) is denoted by

$$(h_n) = \left(\left(\frac{\widetilde{I}_2}{\widetilde{I}_1}\right)^{\frac{1}{d+4}}R(\mathbf{K})^{\frac{1}{d+4}}n^{-\frac{1}{d+4}}\right), \tag{2.26}$$

Finally, a non-recursive estimator of $MWISE[\pi_n]$ is provided by

$$\widetilde{MWISE}[\widetilde{\pi}_n] = \frac{5}{4}\left(\widetilde{I}_2\right)^{\frac{4}{d+4}}\left(\widetilde{I}_1\right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{-\frac{4}{d+4}} + \mathbf{o}\left(n^{-\frac{4}{d+4}}\right).$$

The major aim of our next section lies in comparing the performance of our recursive estimator (2.4) with that of non-recursive Nadaraya-Watson one (2.6).

2.6 Numerical applications

Let's start our numerical studies with some simulations with different dimensions Models.

2.6.1 Simulation studies

In order to compare the proposed recursive estimator with the Nadaraya-Watson non-recursive one, we consider three sample sizes: n=100, 200 and 500, a fixed number of simulations: N=500 and four distribution models:

- Model 1: $(X, Y) \in \mathbb{R} \times \mathbb{R}$:
  $Y = 2\sin(\pi X) + \epsilon$, where X follows the binomial distribution $\mathcal{B}(2, 1/3)$ and ϵ follows the normal distribution $\mathcal{N}(0, 1)$.

- Model 2: $(X, Y) \in \mathbb{R}^2 \times \mathbb{R}$:
  $Y = \exp(-X/2) + \epsilon$, where X follows the poisson distribution $\mathbb{P}(1/2)$ and ϵ follows the normal distribution $\mathcal{N}(0, 1/2)$.

- Model 3: $(X, Y) \in \mathbb{R}^3 \times \mathbb{R}^2$:
  $Y = AX + \epsilon$ with $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \end{pmatrix}$, $X = 0 \times \mathbb{1}_{Z<=0.5} + \mathbb{1}_{Z>0.5}$ where Z follows the uniform distribution $\mathcal{U}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right)$ and ϵ follows the normal distribution $\mathcal{N}(0, 1/2)$.

- Model 4: $(X, Y) \in \mathbb{R}^{10} \times \mathbb{R}^{10}$:
  $Y = \exp(X) + \epsilon$, with $X = \lfloor Z \rfloor$ where Z follows the 10-dimensional normal distribution $\mathcal{N}(0_{10}, I_{10})$ and ϵ follows the normal distribution $\mathcal{N}(0, 1)$.

We denote by π_i^* the reference CCDF and by π_i the test CCDF. Then, we calculate the following two measures:

- Mean squared error: $MSE = \dfrac{1}{n}\sum_{i=1}^{n}(\pi_i - \pi_i^*)^2$.

- The linear correlation: $Cor = \dfrac{Cov(\pi_i, \pi_i^*)}{\sigma(\pi_i)\sigma(\pi_i^*)}$.

In what follows, we portray the different steps of the simulation algorithm in the multivariate case.

2.6.2 Simulation Algorithm

Algorithm 1 K is the Gaussian kernel, d the dimension size, n the simple size, Np the number of observations and N the number of iterations.

Input: K, d, n, Np and N.

1: A random initialization of $\widehat{\Pi}^{(0)}$ (resp. $\widetilde{\Pi}^{(0)}$.)

2: **for** $l = 1, \ldots, N$ **do**

3: A random sample vectors $X_1, \ldots, X_d$ and Y of length n.

4: A choice value for the recursive bandwidth vectors $h_1, \ldots, h_n$. (resp. the non-recursive bandwidth values h_n) using the plug-in approach given in (2.25) (resp. (2.26)).

5: The choice of the stepsize $(\gamma_n) = \left(n^{-1}\right)$.

6: We fix $x_1, \ldots, x_d$ and consider an arbitrary sampling vector T of y of length N_p.

7: $\widehat{\pi}_l(y|x) = \dfrac{\sum_{k=1}^{n} k\gamma_k \mathbb{1}_{\{Y_k \leqslant y\}} h_k^{-d} \prod_{i=1}^{d} K\left(\dfrac{x_i - X_{k_i}}{h_k}\right)}{\sum_{k=1}^{n} k\gamma_k h_k^{-d} \prod_{i=1}^{d} K\left(\dfrac{x_i - X_{k_i}}{h_k}\right)}$ for the multivariate recursive CCDF estimator. (resp. $\widetilde{\pi}_l(y|x) = \dfrac{\sum_{k=1}^{n} \mathbb{1}_{\{Y_k \leqslant y\}} \prod_{i=1}^{d} K\left(\dfrac{x_i - X_{k_i}}{h_k}\right)}{\sum_{k=1}^{n} \prod_{i=1}^{d} K\left(\dfrac{x_i - X_{k_i}}{h_k}\right)}$ for the multivariate non-recursive CCDF estimator).

$\widehat{\Pi}^{(l)} = \widehat{\pi}_l(T|x)$. (resp. $\widetilde{\Pi}^{(l)} = \widetilde{\pi}_l(T|x)$.)

8: **end for**

9: $\widehat{\pi} = N^{-1} \sum_{l=1}^{N} \widehat{\Pi}^{(l)}$ (resp. $\widetilde{\pi} = N^{-1} \sum_{l=1}^{N} \widetilde{\Pi}^{(l)}$.)

output: The vectors $\widehat{\pi}$ and $\widetilde{\pi}$.

			$x = 0$		$x = 2$	
Model	MSE/Cor	n	Nadaraya's estimator	Recursive estimator	Nadaraya's estimator	Recursive estimator
Model 1	MSE	100	3.671766e-07	**1.333452e-07**	1.342202e-05	**8.228683e-06**
		200	3.377698e-07	**1.185438e-07**	3.434738e-06	**2.050298e-06**
		500	1.748553e-07	**4.915888e-08**	1.407363e-06	**6.641233e-07**
	Cor	100	9.999992e-01	**9.999997e-01**	9.999773e-01	**9.999849e-01**
		200	9.999992e-01	**9.999997e-01**	9.999914e-01	**9.999947e-01**
		500	9.999997e-01	**9.999999e-01**	9.999962e-01	**9.999981e-01**

Table 2.1: Quantitative comparison between Nadaraya-Watson estimator and the proposed estimator with stepsizes $(\gamma_n) = (n^{-1})$ through a plug-in method for Model 1.

			$x = (0,0)$		$x = (1,1)$	
Model	MSE/Cor	n	Nadaraya's estimator	Recursive estimator	Nadaraya's estimator	Recursive estimator
Model 2	MSE	100	0.005299303	**0.002345599**	0.011464770	**0.005503603**
		200	0.004444445	**0.001976603**	0.009766589	**0.004850061**
		500	0.003609582	**0.001795659**	0.006988903	**0.004007790**
	Cor	100	0.993278442	**0.997042336**	0.985998000	**0.993297472**
		200	0.993932848	**0.997313718**	0.987516900	**0.993740343**
		500	0.994773451	**0.997402840**	0.990016792	**0.994209320**

Table 2.2: Quantitative comparison between the recursive estimator and the non-recursive one with stepsizes $(\gamma_n) = (n^{-1})$ through a plug-in method for Model 2.

			$x = (0,0,0)$		$x = (1,1,1)$	
Model	MSE/Cor	n	Nadaraya's estimator	Recursive estimator	Nadaraya's estimator	Recursive estimator
Model 3	MSE	100	0.005170397	**0.003118756**	0.007481887	**0.002756530**
		200	0.004968340	**0.002747727**	0.007215537	**0.002449973**
		500	0.004639918	**0.002444590**	0.007200206	**0.002260629**
	Cor	100	0.989849852	**0.993964835**	0.989210840	**0.996122260**
		200	0.990247670	**0.994651835**	0.989322443	**0.996398260**
		500	0.990928518	**0.995271000**	0.989458730	**0.996602105**

Table 2.3: Quantitative comparison between the recursive estimator and the non-recursive one with stepsizes $(\gamma_n) = (n^{-1})$ through a plug-in method for Model 3.

			$x = (0,0,0,0,0,0,0,0,0,0)$	
Model	MSE/Cor	n	Nadaraya's estimator	Recursive estimator
Model 4	MSE	100	0.012132600	**0.008651937**
		200	0.012042460	**0.007971624**
		500	0.009422018	**0.006973041**
	Cor	100	0.985115900	**0.989107240**
		200	0.982886350	**0.987784007**
		500	0.984891065	**0.987888565**

Table 2.4: Quantitative comparison between the recursive estimator and the non-recursive one with stepsizes $(\gamma_n) = (n^{-1})$ through a plug-in method for Model 4.

Departing from Tables 4.1, 2.2 and 2.4, we conclude that:

1. The MSE of the proposed recursive estimator with stepsize $(\gamma_n) = (n^{-1})$ through a plug-in method is smaller than that of Nadaraya-Watson's non-recursive estimator.

2. The estimators get closer to the true CCDF function as sample size increases, i.e., the MSE decreases as the simple size increases and therefore the Cor increases as the sample size increases.

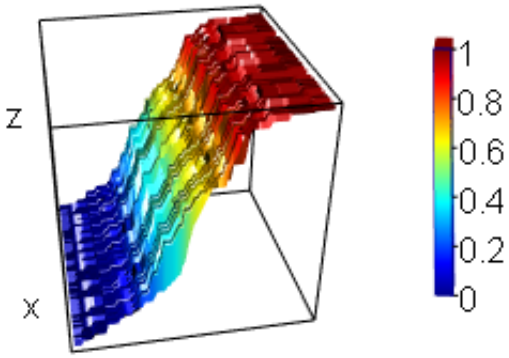


Figure 2.1: The reference CCDF for Model 1 for one simple simulation.

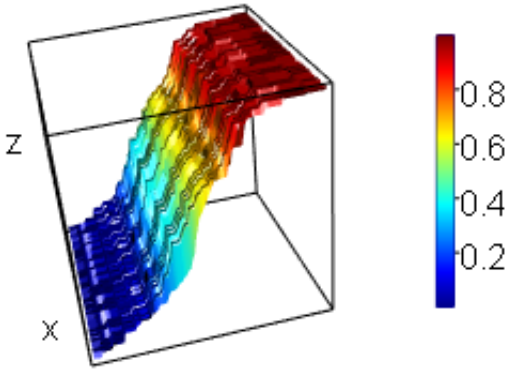


Figure 2.2: The recursive CCDF estimator for Model 1 for one simple simulation.

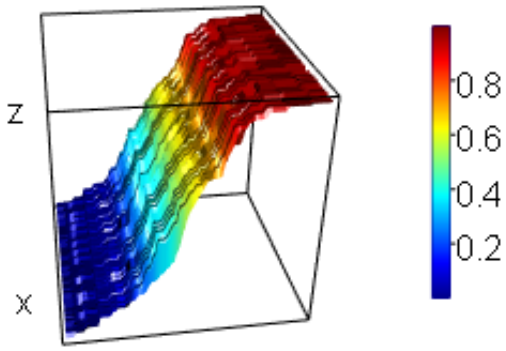


Figure 2.3: The non-recursive CCDF estimator for Model 1 for one simple simulation.

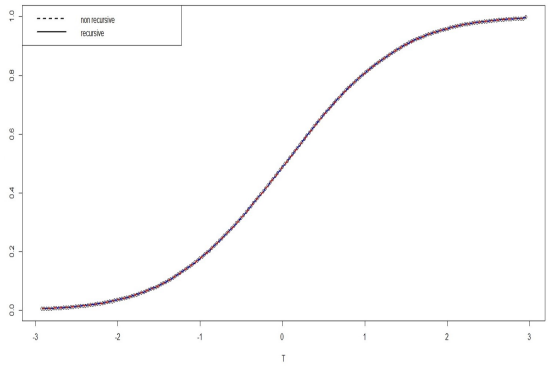


Figure 2.4: Qualitative comparison between the recursive estimator and the non-recursive one for Model 1 with $n = 200$, $N = 500$ and $x = 0$.

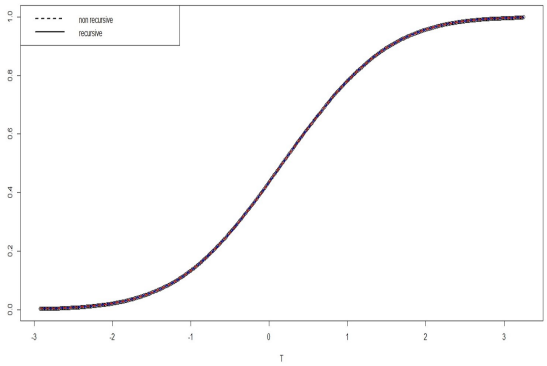


Figure 2.5: Qualitative comparison between the recursive estimator and the non-recursive one for Model 1 with $n = 500$, $N = 500$ and $x = 0$.

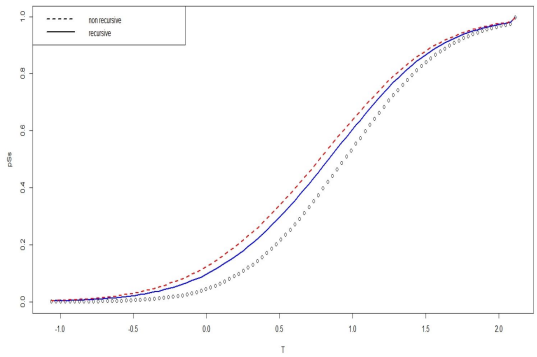


Figure 2.6: Qualitative comparison between the recursive estimator and the non-recursive one for Model 2 with $n = 100$, $N = 500$ and $x = (0, 0)$.

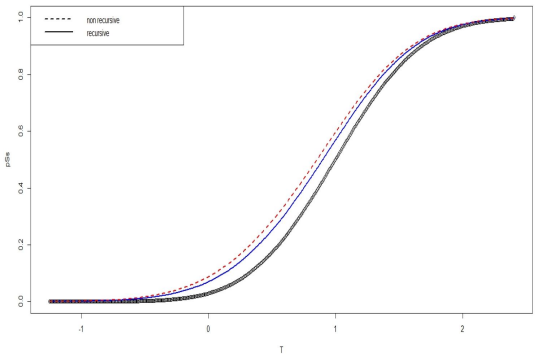


Figure 2.7: Qualitative comparison between the recursive estimator and the non-recursive one for Model 2 with $n = 500$, $N = 500$ and $x = (0, 0)$.

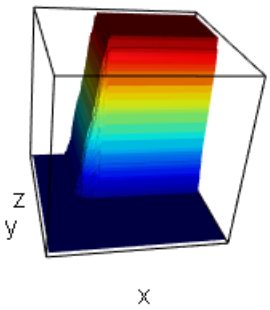


Figure 2.8: The reference CCDF for Model 3 for one simple simulation with $n = 500$ and $x = (1, 1, 1)$.

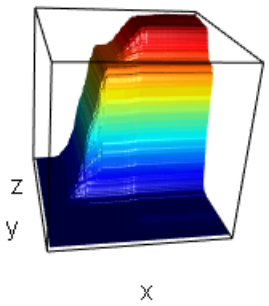


Figure 2.9: The recursive CCDF estimator for Model 3 for one simple simulation with $n = 500$ and $x = (1, 1, 1)$.

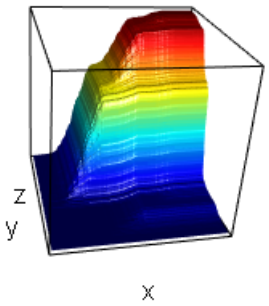


Figure 2.10: The non-recursive CCDF estimator for Model 3 for one simple simulation with $n = 500$ and $x = (1, 1, 1)$.

2.6.3 Real Datasets:

In this section, our focal point is to examine two real datasets Models, namely the Insurance Company Benchmark (COIL 2000) dataset as well as the French Hospital Data of COVID-19.

Application 1: The Insurance Company Benchmark (COIL 2000) dataset

The (COIL 2000) dataset is found in data.world website
https://data.world/uci/insurance-company-benchmark-coil-2000).
Information about customers consists of 86 variables and includes product usage data and sociodemographic data derived from zip area codes. The data are supplied by the Dutch data mining company Sentient Machine Research and rest on a real world business problem. The training set involves over 5000 descriptions of customers, including the information of whether or not they have a caravan insurance policy. A test set includes 4000 customers whom only the organizers know if they have a caravan insurance policy. This corresponds to a Dataset to train and validate prediction models and build up a description (5822 customer records). Each record consists of 86 attributes, incorporating sociodemographic data (attribute 1-43) and product ownership (attributes 44-86).
The sociodemographic data are derived from zip codes. All customers living in areas with the same zip code have the same sociodemographic attributes.
As far as our application is concerned, we shall consider the following two models:

- Model 1: X corresponds to the sociodemographic data attribute number 16 and Y stands for the whole 5822 observations of customer records.
- Model 2: X corresponds to the sociodemographic 5-dimensional data attributes number 6,8,11,12 and 13 and Y stands for the whole 5822 observations of customer records.

		x=0		x=1	
		Nadaraya's estimator	Recursive estimator	Nadaraya's estimator	Recursive estimator
Model 1	MSE	0.006979541	**0.004748803**	0.001999743	**0.001390595**
	Cor	0.987112335	**0.990653245**	0.995770041	**0.996574025**

		x=(0,0,0,0,0)		x=(1,1,1,2,2)	
		Nadaraya's estimator	Recursive estimator	Nadaraya's estimator	Recursive estimator
Model 2	MSE	0.000708417	**0.0005969373**	0.00414241	**0.003705582**
	Cor	0.996664739	**0.9979940979**	0.97540614	**0.976614363**

Table 2.5: Quantitative comparison between Nadaraya-Watson estimator and the proposed estimator with stepsizes $(\gamma_n) = (n^{-1})$ through a plug-in method for the Insurance Company Benchmark (COIL 2000) dataset case.

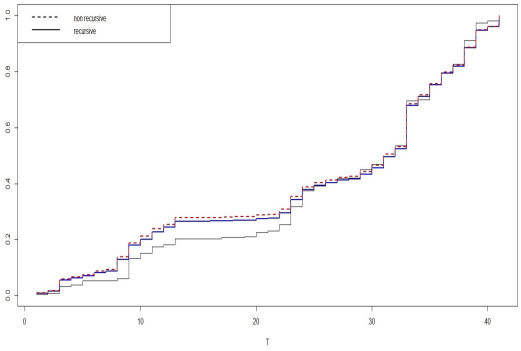


Figure 2.11: Qualitative comparison between the recursive estimator and the non-recursive one for the dataset Model 1 with $x = 1$.

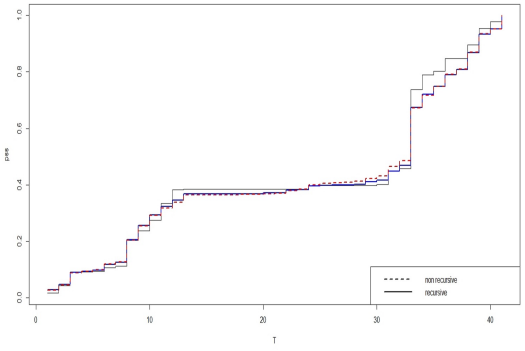


Figure 2.12: Qualitative comparison between the recursive estimator and the non-recursive one for the dataset Model 2 with $x = (0, 0, 0, 0, 0)$.

Application 2: French Hospital Data of COVID19

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

Description of the dataset

This dataset provides information on the hospital situation regarding the COVID-19 epidemic. We have opted for the first proposed file:
Hospital data related to the COVID-19 epidemic by department (dep), sex of the patient (sex), number of hospitalized patients (hosp), number of persons currently in intensive care or resuscitation (rea), number of persons currently in follow-up and rehabilitation care (SSR) or long-term care units (USLD), number of persons currently in conventional hospitalization (HospConv), number of persons currently hospitalized in another type of service (autres), cumulative number of persons returning home (rad) or cumulative number of dead persons (dc).
The data have been daily updated. For the current application, we considered the data of 28/07/2021, with a total of 150894 observations. For simplicity reason, we have chosen to just study the department of 'Vienne' database.
Therefore, for our application, we served of a dataframe of 1494 observations and 6 variables. Hence, we shall consider the following three models:

- Model 1: $X = dc$ and $Y = hosp$.
- Model 2: $X1 = sex$, $X2 = rea$, $X3 = dc$ and $Y = hosp$.
- Model 3: $X1 = sex$, $X2 = rea$, $X3 = dc$, $Y1 = hosp$ and $Y2 = rad$.

		x=119		x=17	
		Nadaraya's estimator	Recursive estimator	Nadaraya's estimator	Recursive estimator
Model 1	MSE	0.02538461	**0.01900265**	0.008079359	**0.007509241**
	Cor	0.86514927	**0.88894736**	0.781752320	**0.789593904**

		$x = (2, 0, 17)$		$x = (1, 0, 0)$	
		Nadaraya's estimator	Recursive estimator	Nadaraya's estimator	Recursive estimator
Model 2	MSE	0.02176777	**0.01416798**	0.01026319	**0.005681196**
	Cor	0.55644221	**0.63197054**	0.74621610	**0.819493783**
Model 3	MSE	0.02107916	**0.009991914**	0.04130757	**0.03504694**
	Cor	0.85362436	**0.920385555**	0.61896604	**0.64659332**

Table 2.6: Quantitative comparison between Nadaraya-Watson estimator and the proposed estimator with stepsizes $(\gamma_n) = (n^{-1})$ through a plug-in method for the COVID-19 epidemic dataset case.

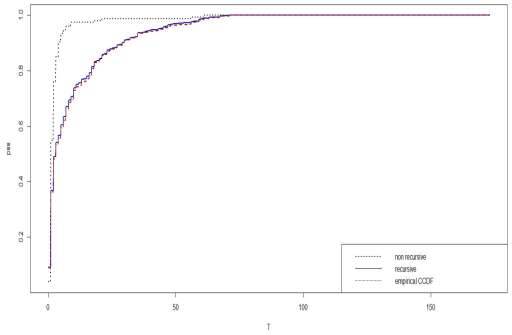


Figure 2.13: Qualitative comparison between the recursive estimator and the non-recursive one for the COVID-19 epidemic dataset Model 1 with $x = 17$.

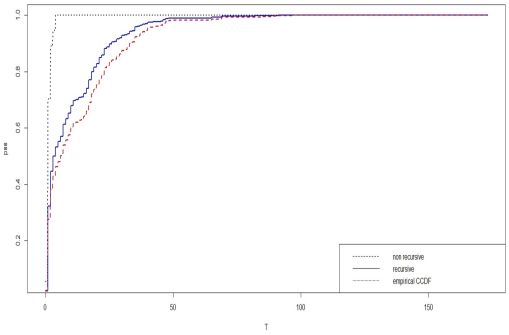


Figure 2.14: Qualitative comparison between the recursive estimator and the non-recursive one for the COVID-19 dataset Model 2 with $x = (2, 0, 17)$.

Data interpretation:
Referring to Tables 2.5, 2.6 and Figures 2.11, 2.12, 2.13 and 2.14 , we conclude that:

1. For all considered Models, the proposed recursive estimator with stepsize $(\gamma_n) = (n^{-1})$ through a plug-in method outperformed the non-recursive one in terms of estimation error MSE and Cor.

2. The proposed recursive estimator is closer to the true CCDF function, compared with Nadaraya-Watson's non-recursive estimator.

Concerning the COVID-19 epidemic Model 1, we can infer that, for a fixed number of deaths $x = 17$, the proportion of hospitalized cases less then 20 is 50% and the proportion of hospitalized cases less then 50 is 85%. Moreover, 99% of the population have a number of hospitalized cases less than 100.
Likewise, the COVID-19 epidemic Model 2 yielded the same results as model 1. Indeed, we recorded for the women gender a fixed value of sex $x_1 = 2$, a fixed number of REA persons $x_2 = 0$ and deaths $x_3 = 17$.

2.7 Conclusion

In this work, we elaborated a multivariate recursive CCDF estimator. We tackled 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 revealed that the use of our recursive estimator with an appropriate choice of the bandwidth and the stepsize enables us to get closer to the true conditional cumulative distribution function rather than non-recursive one.

2.8 Proofs

Throughout this section, devoted to the proofs of our main results, it is noteworthy that

· For all $x \in \mathbb{R}^d, y \in \mathbb{R}^q$,

$$\mathcal{Z}_n(x,y) := h_n^{-d}\chi_n(y)\,\mathbf{K}\left(\frac{x-X_n}{h_n}\right) \qquad \text{and} \qquad \mathcal{W}_n(x) := h_n^{-d}\,\mathbf{K}\left(\frac{x-X_n}{h_n}\right).$$

Proof of Proposition 2.2.
Our proof starts with the observation that, based on the expression (2.1) without assuming $a_0(x,y) = 0$ and with recurrence on n, we get

$$\begin{aligned} a_n(x,y) - a(x,y) &= (1-\gamma_n)a_{n-1}(x,y) + \gamma_n\mathcal{Z}_n(x,y) - a(x,y) \\ &= (1-\gamma_n)[a_{n-1}(x,y) - a(x,y)] + \gamma_n[\mathcal{Z}_n(x,y) - a(x,y)] \\ &= \prod_{i=1}^{n}(1-\gamma_i)[a_0(x,y) - a(x,y)] + \sum_{k=1}^{n-1}\prod_{i=k+1}^{n}(1-\gamma_i)\gamma_k(\mathcal{Z}_k(x,y) - a(x,y)) \\ &+ \gamma_n(\mathcal{Z}_n(x,y) - a(x,y)). \end{aligned}$$

Grounded on the fact that, $\Pi_n = \prod_{i=1}^{n}(1-\gamma_i)$ and $\prod_{i=k+1}^{n}(1-\gamma_i) = \Pi_n\Pi_k^{-1}$, we deduce

$$\forall\, x \in \mathbb{R}^d, y \in \mathbb{R}^q, \quad a_n(x,y) - a(x,y) = \Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\gamma_k(\mathcal{Z}_k(x,y) - a(x,y)) + \Pi_n[a_0(x,y) - a(x,y)].$$

Therefore,

$$\mathbb{E}[a_n(x,y)] - a(x,y) = \Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\gamma_k(\mathbb{E}[\mathcal{Z}_k(x,y)] - a(x,y)) + \Pi_n[a_0(x,y) - a(x,y)].$$

In order to determine the bias of a_n, we need to simply focus on the quantity $\mathbb{E}[\mathcal{Z}_k(x,y)] - a(x,y)$. Relying upon the assumption (A_1) and (A_4), it follows that

$$\begin{aligned} \mathbb{E}[\mathcal{Z}_k(x,y)] &= \int_{\mathbb{R}^{d+q}} h_k^{-d}\mathbf{K}\left(\frac{x-t}{h_k}\right)\chi_k(y)f_{(X,Y)}(t,u)dtdu \\ &= \int_{\mathbb{R}^d} h_k^{-d}\mathbf{K}\left(\frac{x-t}{h_k}\right)\mathbb{E}[\chi_k(y)|X=t]f_X(t)dt \\ &= \int_{\mathbb{R}^d} h_k^{-d}\,\mathbf{K}\left(\frac{x-t}{h_k}\right)\pi(y|t)f_X(t)dt \\ &= \int_{\mathbb{R}^d} h_k^{-d}\,\mathbf{K}\left(\frac{x-t}{h_k}\right)a(t,y)dt \\ &= \int_{\mathbb{R}^d} \mathbf{K}\,(z)\,a(x-zh_k,y)dz. \end{aligned}$$

Moreover, Taylor's expansion with integral remainder ensures that

$$\begin{aligned} \mathbb{E}[\mathcal{Z}_k(x,y)] - a(x,y) &= \int_{\mathbb{R}^d}\mathbf{K}(z)\,[a(x-zh_k,y) - a(x,y)]\,dz \\ &= \int_{\mathbb{R}^d}\mathbf{K}(z)\left[\sum_{i=1}^{d}\frac{\partial a}{\partial x_i}(x,y)z_ih_k + \int_0^1(1-t)\sum_{i,j=1}^{d}\frac{\partial^2 a}{\partial x_i\partial x_j}(x-tzh_k,y)z_iz_jh_k^2dt\right]dz \\ &= h_k\sum_{i=1}^{d}a_i^{(1)}(x,y)\int_{\mathbb{R}^d}\mathbf{K}(z)z_idz + h_k^2\sum_{i,j=1}^{d}\int_{\mathbb{R}^d}\int_0^1(1-t)\,a_{ij}^{(2)}(x-tzh_k,y)z_iz_j\,\mathbf{K}(z)dtdz. \end{aligned}$$

Referring to the assumption (A_1), $\int_{\mathbb{R}^d} \mathbf{K}(z) z_i dz = 0$ and $\mu_i(\mathbf{K}) = \int_{\mathbb{R}^d} z_i^2 \mathbf{K}(z) dz$, it can be inferred that

$$\mathbb{E}[\mathcal{Z}_k(x,y)] - a(x,y) = \frac{h_k^2}{2} \sum_{j=1}^{d} \mu_j(\mathbf{K}) a_{jj}^{(2)}(x,y) + h_k^2 \eta_k(x),$$

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

Since we have $a_{ij}^{(2)}$ which is bounded and continuous at x for all $i,j \in \{1,\ldots,d\}$, we obtain $\lim_{k \to +\infty} \eta_k(x) = 0$, which ensures that, $\eta_k(x) = \mathbf{o}(1)$.

Hence,

$$\begin{aligned}
\mathbb{E}[a_n(x,y)] - a(x,y) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k (\mathbb{E}[\mathcal{Z}_k(x,y)] - a(x,y)) + \Pi_n [a_0(x,y) - a(x,y)] \\
&= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \left(\frac{h_k^2}{2} \sum_{j=1}^{d} \mu_j(\mathbf{K}) a_{jj}^{(2)}(x,y) + h_k^2 \eta_k(x) \right) + \Pi_n [a_0(x,y) - a(x,y)] \\
&= \frac{1}{2} \sum_{j=1}^{d} \mu_j(\mathbf{K}) a_{jj}^{(2)}(x,y) \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k h_k^2 + \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k h_k^2 \eta_k(x) \\
&+ \Pi_n [a_0(x,y) - a(x,y)].
\end{aligned}$$

• For the case $a \leqslant \alpha/(d+4)$, we have $\lim_{n \to +\infty} (n\gamma_n) > 2a$ and then $1 - 2a\xi > 0$. The application of lemma 1.2 enables us to write

$$\begin{aligned}
\mathbb{E}[a_n(x,y)] - a(x,y) &= \frac{1}{2} \sum_{j=1}^{d} \mu_j(\mathbf{K}) a_{jj}^{(2)}(x,y) \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k h_k^2 + \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \mathbf{o}\left(h_k^2\right) + \mathbf{o}\left(\Pi_n\right) \\
&= \frac{1}{2(1-2a\xi)} \left(\sum_{j=1}^{d} \mu_j(\mathbf{K}) a_{jj}^{(2)}(x,y) \right) h_n^2 + \mathbf{o}\left(h_n^2\right) + \mathbf{o}(1) + \mathbf{o}\left(\Pi_n\right).
\end{aligned}$$

We thus obtain the following desired result

$$\mathbb{E}[a_n(x,y)] - a(x,y) = \frac{1}{2(1-2a\xi)} \left(\sum_{j=1}^{d} \mu_j(\mathbf{K}) a_{jj}^{(2)}(x,y) \right) h_n^2 + \mathbf{o}\left(h_n^2\right).$$

• For the case $a > \alpha/(d+4)$, we have $\lim_{n \to +\infty} (n\gamma_n) > \frac{\alpha - ad}{2}$ which yields that $h_n^2 = \mathbf{o}\left(\sqrt{\gamma_n h_n^{-d}}\right)$. Then, the use of lemma 1.2 leads to

$$\begin{aligned}
\mathbb{E}[a_n(x,y)] - a(x,y) =& \frac{1}{2} \sum_{j=1}^{d} \mu_j(\mathbf{K}) a_{jj}^{(2)}(x,y) \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \mathbf{o}\left(\sqrt{\gamma_k h_k^{-d}}\right) + \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \mathbf{o}\left(\sqrt{\gamma_k h_k^{-d}}\right) \\
&+ \mathbf{o}\left(\Pi_n\right) \\
=& \mathbf{o}\left(\sqrt{\gamma_n h_n^{-d}}\right).
\end{aligned}$$

Therefore, the claimed result (2.12) is established.

For the variance, and owing to the independence of X_i, for $i = 1, \ldots, n$, it's obvious that

$$\begin{aligned} Var[a_n(x,y)] &= Var[\Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \mathcal{Z}_k(x,y)] \\ &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2} \gamma_k^2 Var[\mathcal{Z}_k(x,y)] \\ &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2} \gamma_k^2 \left(\mathbb{E}[\mathcal{Z}_k^2(x,y)] - \mathbb{E}[\mathcal{Z}_k(x,y)]^2 \right). \end{aligned}$$

$$\begin{aligned} \mathbb{E}[\mathcal{Z}_k^2(x,y)] - \mathbb{E}[\mathcal{Z}_k(x,y)]^2 &= \int_{\mathbb{R}^d} h_k^{-2d} \mathbf{K}^2 \left(\frac{x-t}{h_k} \right) \mathbb{E}[\chi_k(y)^2 | X = t] f_X(t) dy \\ &- \left(\int_{\mathbb{R}^d} h_k^{-d} \mathbf{K} \left(\frac{x-t}{h_k} \right) \pi(y|t) f_X(t) dt \right)^2 \\ &= \int_{\mathbb{R}^d} h_k^{-d} \mathbf{K}^2(z)\, a(x - z h_k, y) dz - \left(\int_{\mathbb{R}^d} \mathbf{K}(z)\, a(x - z h_k, y) dz \right)^2. \end{aligned}$$

As matter of fact, the Taylor's expansions theorem ensures that

$$Var[a_n(x,y)] = \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2} \gamma_k^2 h_k^{-d} \left[a(x,y) \int_{\mathbb{R}^d} \mathbf{K}^2(z)\, dz + \nu_k(x) - h_k^d \eta_k(x) \right],$$

where,

$$\nu_k(x) = \int_{\mathbb{R}^d} \mathbf{K}^2(z) \left[a(x - z h_k, y) - a(x,y) \right] dz \text{ and } \eta_k(x) = \left(\int_{\mathbb{R}^d} \mathbf{K}(z)\, a(x - z h_k, y) dz \right)^2.$$

• For the case $a < \alpha/(d+4)$, we have $\lim_{n \to +\infty} (n\gamma_n) > 2a$ which provides that $\gamma_n h_n^{-d} = \mathbf{o}\left(h_n^4\right)$. By applying lemma 1.2, we infer that

$$\begin{aligned} Var[a_n(x,y)] &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2} \gamma_k^2 h_k^{-d} \left[a(x,y) R(\mathbf{K}) + \mathbf{o}(1) \right] \\ &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2} \gamma_k \mathbf{o}\left(h_k^4\right). \end{aligned}$$

Hence, we obtain the result

$$Var[a_n(x,y)] = \mathbf{o}\left(h_n^4\right).$$

• For the case $a \geqslant \alpha/(d+4)$, we have $\lim_{n \to +\infty} (n\gamma_n) > \frac{\alpha - ad}{2}$ and then $2 - (\alpha - ad)\xi > 0$. 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 1.2 ensures that

$$\begin{aligned} Var[a_n(x,y)] &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2} \gamma_k^2 h_k^{-d} \left[a(x,y) R(\mathbf{K}) + \nu_k(x) - h_k^d \eta_k(x) \right] \\ &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2} \gamma_k^2 h_k^{-d} \left[a(x,y) R(\mathbf{K}) + \mathbf{o}(1) \right] \\ &= \frac{1}{2 - (\alpha - ad)\xi} \frac{\gamma_n}{h_n^d} [a(x,y) R(\mathbf{K}) + \mathbf{o}(1)]. \end{aligned}$$

Thus, this leads to the result displayed in (2.14). □

Proof of Theorem 2.3.
Our proof rests upon the following decomposition, for $f_n \neq 0,\ n \geq 0$

$$\pi_n(y|x) - \pi(y|x) = A_n(x,y)\frac{f_X(x)}{f_n(x)}, \tag{2.27}$$

with

$$A_n(x,y) = \frac{1}{f_X(x)}\left(a_n(x,y) - a(x,y)\right) - \frac{\pi(y|x)}{f_X(x)}\left(f_n(x) - f_X(x)\right).$$

It follows from (2.27) that the asymptotic behavior of $\pi_n(y|x) - \pi(y|x)$ can be deduced from the one of $A_n(x,y)$.
1. **Bias of π_n:** Here, we can state

$$\mathbb{E}[A_n(x,y)] = \frac{1}{f_X(x)}\left(\mathbb{E}[a_n(x,y)] - a(x,y)\right) - \frac{\pi(y|x)}{f_X(x)}\left(\mathbb{E}[f_n(x)] - f_X(x)\right).$$

Now, using the first bias part of proposition 2.1 and proposition 2.2 and considering the fact that $a(x,y) = \pi(y|x)f_X(x)$, then by combining the assertions (2.7), (2.8),(2.11) and (2.12); we obtain the relations (2.15) and (2.16).
2. **Variance of π_n:** In order to confirm this statement, we have

$$\begin{aligned} Var[A_n(x,y)] = &\frac{1}{(f_X(x))^2}\,Var[a_n(x,y)] + \frac{(\pi(y|x))^2}{(f_X(x))^2}\,Var[f_n(x)] \\ &- 2\frac{\pi(y|x)}{(f_X(x))^2}\,Cov(a_n(x,y), f_n(x)). \end{aligned} \tag{2.28}$$

Given that the X_k's are independent, then for all $i = 1,\dots,n$, $k = 1,\dots,n$ and $i \neq k$, we have $Cov(\mathcal{Z}_k(x,y), \mathcal{W}_i(x)) = 0$. Using lemma 1.2, classical computations entail

$$Cov(a_n(x,y), f_n(x)) = \frac{R(\mathbf{K})}{2-(\alpha-ad)\xi}\,\pi(y|x)f_X(x)\,\frac{\gamma_n}{h_n^d} + \mathbf{o}\left(\frac{\gamma_n}{h_n^d}\right). \tag{2.29}$$

In fact, we have

$$\begin{aligned} Cov(a_n(x,y), f_n(x)) &= Cov\left(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k\,\mathcal{Z}_k(x,y), \Pi_n \sum_{i=1}^{n} \Pi_i^{-1}\gamma_i\,\mathcal{W}_i(x)\right) \\ &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2\,Cov\left(\mathcal{Z}_k(x,y), \mathcal{W}_k(x)\right) \\ &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2\,\left(\mathbb{E}\left[\mathcal{Z}_k(x,y)\,\mathcal{W}_k(x)\right] - \mathbb{E}\left[\mathcal{Z}_k(x,y)\right]\mathbb{E}\left[\mathcal{W}_k(x)\right]\right). \end{aligned}$$

Hence, the use of Taylor's expansion with integral remainder ensures that

$$\mathbb{E}\left[\mathcal{Z}_k(x,y)\,\mathcal{W}_k(x)\right] = \mathbb{E}\left[\chi_k(y)\,h_k^{-2d}\,\mathbf{K}^2\left(\frac{x-X_k}{h_k}\right)\right] = R(\mathbf{K})\pi(y|x)f_X(x)h_k^{-d} + \mathbf{o}\left(h_k^{-d}\right),$$

$$\mathbb{E}\left[\mathcal{Z}_k(x,y)\right] = \mathbb{E}\left[\chi_k(y)\,h_k^{-d}\,\mathbf{K}\left(\frac{x-X_k}{h_k}\right)\right] = \pi(y|x)f_X(x) + \mathbf{o}(1),$$

and $$\mathbb{E}\left[\mathcal{W}_k(x)\right] = \mathbb{E}\left[h_k^{-d}\,\mathbf{K}\left(\frac{x-X_k}{h_k}\right)\right] = f_X(x) + \mathbf{o}(1).$$

Thus, by applying lemma 1.2, we can assert

$$\begin{aligned} Cov(a_n(x,y), f_n(x)) &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 \Big(R(\mathbf{K})\pi(y|x) f_X(x) h_k^{-d} - \pi(y|x) f_X^2(x) + \mathbf{o}\left(h_k^{-d}\right) + \mathbf{o}(1) \Big) \\ &= \Pi_n^2 \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 h_k^{-d} \Big(R(\mathbf{K})\pi(y|x) f_X(x) + \mathbf{o}\left(1\right) \Big) \\ &= \frac{R(\mathbf{K})}{2-(\alpha-ad)\xi} \pi(y|x) f_X(x) \, \frac{\gamma_n}{h_n^d} + \mathbf{o}\left(\frac{\gamma_n}{h_n^d}\right). \end{aligned}$$

Consequently, relations (2.17) and (2.18) follow from the combination of assertions (2.9), (2.10), (2.13), (2.14) and (2.29).
For the case $a \geq \alpha/(d+4)$, we deduce with (2.28) that

$$Var[\pi_n(y|x)] = \frac{R(\mathbf{K})}{2-(\alpha-ad)\xi} \frac{\pi(y|x)(1-\pi(y|x))}{f_X(x)} \, \frac{\gamma_n}{h_n^d} + \mathbf{o}\left(\frac{\gamma_n}{h_n^d}\right).$$

Proceeding with the same reasoning applied for the case $a < \alpha/(d+4)$, we obtain the desired result (2.17). □

Proof of Theorem 2.4.
We can write, for all $n \geq 0$, $x \in \mathbb{R}^d, y \in \mathbb{R}^q$,

$$\begin{aligned} A_n(x,y) - \mathbb{E}[A_n(x,y)] =& \frac{1}{f_X(x)}[a_n(x,y) - \mathbb{E}[a_n(x,y)]] - \frac{\pi(y|x)}{f_X(x)}[f_n(x) - \mathbb{E}[f_n(x)]] \\ =& \frac{1}{f_X(x)} \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k \left(T_k(x,y) - \mathbb{E}[T_k(x,y)]\right), \end{aligned}$$

where

$$T_k(x,y) = \mathcal{Z}_k(x,y) - \pi(y|x)\mathcal{W}_k(x).$$

Here and subsequently, it is worth noting

$$S_k(x,y) := \Pi_k^{-1}\gamma_k \left(T_k(x,y) - \mathbb{E}[T_k(x,y)]\right).$$

Hence, we state

$$A_n(x,y) - \mathbb{E}[A_n(x,y)] = \frac{1}{f_X(x)} \Pi_n \sum_{k=1}^{n} S_k(x,y). \tag{2.30}$$

This proof falls naturally into the application of Lyapunov's theorem 1.14 for $S_k(x,y)$.
On the one hand, we can write

$$\begin{aligned} u_n^2 := \sum_{k=1}^{n} Var[S_k(x,y)] &= \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 Var[T_k(x,y)] \\ &= \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 Var\left[\mathcal{Z}_k(x,y) - \pi(y|x)\mathcal{W}_k(x)\right] \\ &= \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 \left(Var\left[\mathcal{Z}_k(x,y)\right] + \pi(y|x)^2 Var\left[\mathcal{W}_k(x)\right] \right. \\ &\quad \left. - 2\pi(y|x)\, Cov\left(\mathcal{Z}_k(x,y), \mathcal{W}_k(x)\right). \right. \end{aligned}$$

[

Moreover, we have

$$\begin{aligned} &Var\left[\mathcal{Z}_k(x,y)\right] = h_k^{-d}\Big(R(\mathbf{K})\,\pi(y|x)f_X(x) + \mathbf{o}(1)\Big),\\ &Var\left[\mathcal{W}_k(x)\right] = h_k^{-d}\Big(R(\mathbf{K})\,f_X(x) + \mathbf{o}(1)\Big),\\ \text{and}\qquad &Cov\left(\mathcal{Z}_k(x,y),\mathcal{W}_k(x)\right) = h_k^{-d}\Big(R(\mathbf{K})\,\pi(y|x)f_X(x) + \mathbf{o}(1)\Big). \end{aligned}$$

Therefore, by applying lemma 1.2, it can be deduced that

$$u_n^2 = \sum_{k=1}^{n} \Pi_k^{-2}\gamma_k^2 h_k^{-d}\Big(R(\mathbf{K})f_X(x)\pi(y|x)\left(1-\pi(y|x)\right) + \mathbf{o}(1)\Big) = \frac{f_X(x)^2}{\Pi_n^2}\frac{\gamma_n}{h_n^d}[\sigma^2(x,y) + \mathbf{o}\left(1\right)]. \tag{2.31}$$

On the other hand, we have

$$\begin{aligned} &\mathbb{E}[\mathcal{Z}_k^{2+p}(x,y)] = h_k^{-d(1+p)}\left[\pi(y|x)f_X(x)\int_{\mathbb{R}^d}\mathbf{K}^{2+p}\left(z\right)dz + \mathbf{o}(1)\right]\\ \text{and}\quad &\mathbb{E}[\mathcal{W}_k^{2+p}(x,y)] = h_k^{-d(1+p)}\left[f_X(x)\int_{\mathbb{R}^d}\mathbf{K}^{2+p}\left(z\right)dz + \mathbf{o}(1)\right]. \end{aligned}$$

Then, we can write

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

Therefore,

$$\begin{aligned} \mathbb{E}\left[|S_k(x,y)|^{2+p}\right] &= \Pi_k^{-(2+p)}\gamma_k^{2+p}\mathbb{E}\big[|T_k(x,y) - \mathbb{E}[T_k(x,y)]|^{2+p}\big]\\ &= \Pi_k^{-(2+p)}\gamma_k^{2+p}\mathbb{E}\left[|T_k(x,y) + \mathbf{o}(1)|^{2+p}\right]. \end{aligned}$$

Hence,

$$\mathbb{E}\left[|S_k(x,y)|^{2+p}\right] = \mathbf{O}\left(\Pi_k^{-(2+p)}\gamma_k^{2+p}\mathbb{E}\left[|T_k(x,y)|^{2+p}\right]\right).$$

This yields,

$$\begin{aligned} \sum_{k=1}^{n}\mathbb{E}[|S_k(x,y)|^{2+p}] &= \mathbf{O}\left(\sum_{k=1}^{n}\Pi_k^{-(2+p)}\gamma_k^{2+p}\mathbb{E}\left[|T_k(x,y)|^{2+p}\right]\right)\\ &= \mathbf{O}\left(\sum_{k=1}^{n}\Pi_k^{-(2+p)}\gamma_k^{2+p}\frac{1}{h_k^{d(1+p)}}\right). \end{aligned}$$

For the application of lemma 1.2, let us assume that there exists a positive real p such that

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

Then, we obtain

$$\sum_{k=1}^{n}\mathbb{E}[|S_k(x,y)|^{2+p}] = \mathbf{O}\left(\frac{\gamma_n^{1+p}}{\Pi_n^{2+p}h_n^{d(1+p)}}\right).$$

It follows that

$$\frac{1}{u_n^{2+p}}\sum_{k=1}^{n}\mathbb{E}[|S_k(x,y)|^{2+p}] = \mathbf{O}\left(\frac{\gamma_n^{1+p}}{u_n^{2+p}\Pi_n^{2+p}h_n^{d(1+p)}}\right).$$

As a sequel, with the assertion (2.31), we can write

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

Moreover, since we have

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

then, by applying Lyapunov theorem 1.14, we get

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

This implies

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

Additionally, the relations (2.27) and (2.30) ensure that

$$\frac{1}{u_n\Pi_n}f_X(x)\left(\pi_n(y|x)-\mathbb{E}[\pi_n(y|x)]\right) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0\,,1\right). \tag{2.32}$$

Given that

$$u_n^2 = \frac{f_X(x)^2}{\Pi_n^2}\frac{\gamma_n}{h_n^d}[\sigma^2(x,y)+\mathbf{o}\left(1\right)],$$

and by replacing u_n with its value in relation (2.32), we deduce that

$$\sqrt{\gamma_n^{-1}h_n^d}\left(\pi_n(y|x)-\mathbb{E}[\pi_n(y|x)]\right) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0\,,\sigma^2(x,y)\right). \tag{2.33}$$

Since we have $\sqrt{\gamma_n^{-1}h_n^{d+4}} \xrightarrow[n\to+\infty]{} \sqrt{c}$, then the convergence (2.19) follows from the combination of relations (2.15), (2.16) and convergence (2.33). □

Proof of Proposition 2.5.
Based on the relation (2.20) and by distinguishing the different possible cases according to the expressions of the Bias ((2.15) and (2.16)) as well as the Variance ((2.17) and (2.18)), one can prove this proposition and find the required result.
In fact, we first note

$$C_1 = \frac{1}{4}\frac{I_1}{(1-2a\xi)^2}, \quad C_2 = \frac{I_2}{2-(\alpha-ad)\xi}R(\mathbf{K}).$$

$$MWISE[\pi_n] = \begin{cases} C_1h_n^4+o(h_n^4) & \text{if} \quad a\in\left(0,\frac{\alpha}{d+4}\right) \\ C_2\gamma_nh_n^{-d}+C_1h_n^4+o(h_n^4) & \text{if} \quad a=\frac{\alpha}{d+4} \\ C_2\gamma_nh_n^{-d}+o\left(\gamma_nh_n^{-d}\right) & \text{if} \quad a\in\left(\frac{\alpha}{d+4},1\right). \end{cases}$$

Set $\alpha\in]1/2,1]$.

- If $a<\frac{\alpha}{d+4}$, $C_1h_n^4\in GS(-4a)$ with $-\frac{4\alpha}{d+4}<-4a$.
- If $a=\frac{\alpha}{d+4}$, $C_2\gamma_nh_n^{-d}+C_1h_n^4\in GS(-\frac{4\alpha}{d+4})$ with $-\frac{4\alpha}{d+4}=-4a$.
- If $a>\frac{\alpha}{d+4}$, $C_2\gamma_nh_n^{-d}\in GS(-\alpha+ad)$ with $-\frac{4\alpha}{d+4}<-\alpha+ad$.

It follows that, for a given α, to minimize the $MWISE[\pi_n]$, we select the smallest quantity which is $-\frac{4\alpha}{d+4}=-4a$. Therefore, the parameter a must be chosen equal to $\frac{\alpha}{d+4}$.

□

Chapter 3

The stochastic approximation method for semi-recursive multivariate kernel-type regression estimation

3.1 Introduction

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

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

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

An estimator of the r_φ regression was first developed by Roussas (1990) and improved by Einmahl and Mason (2000) to determine exact rates of uniform strong consistency of kernel-type function estimators. Later, Deheuvels and Mason (2004) established uniform and non-uniform asymptotic simultaneous confidence bands for functionals of the distribution based on kernel-type estimators.
The classical recursive regression estimator was addressed in Mokaddem *et al.* (2009b) for univariate framework and a multivariate extension of this estimator was carried out by Mokaddem and Pelletier(2016). Subsequently, Slaoui (2016) established the semi-recursive case and introduced a new estimator which is the fraction of a recursive regression by a recursive density function. As far as this research is concerned, our basic objective is to extend this estimator for kernel-type estimation with large choice of parameters and properties in a multivariate case. Note that, recently, Bouzebda and Slaoui (2020) explored general kernel type estimators for censored data defined by the stochastic approximation algorithm.

3.1.1 Presentation of the method

Let us start with the presentation of our stochastic approximation method. The stochastic approximation algorithm, which estimates recursively the regression function

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

at a given vector x, can be expressed as follows :

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

where (β_n) is a positive sequences of real numbers decreasing towards zero verifying (18).
Here, we consider that $a_0(x) = 0$, then by a recurrence, we get

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

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

$$r_{\varphi_n}(x) = \begin{cases} \dfrac{a_{\varphi_n}(x)}{f_n(x)} & \text{if} \quad f_n(x) \neq 0 \\ 0 & \text{otherwise} \end{cases}, \tag{3.1}$$

with f_n stands for the recursive density estimator given in (2.3).
Our first aim is to examine the asymptotic properties of our proposed semi-recursive estimator of a multivariate regression function. Then, we prove its performance.
By introducing the non-recursive estimator of a_φ given by

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

we shall compare our estimator to the generalized non-recursive kernel regression estimator of Nadaraya-Watson Nadaraya (1964) and Watson (1964) $\widetilde{r}_{\varphi_n}$ indicated by

$$\widetilde{r}_{\varphi_n}(x) = \begin{cases} \dfrac{\widetilde{a}_{\varphi_n}(x)}{\widetilde{f}_n(x)} & \text{if} \quad \widetilde{f}_n(x) \neq 0 \\ 0 & \text{otherwise} \end{cases}, \tag{3.2}$$

where $\widetilde{f}_n$ was given in (2.5).

Particular cases:

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

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

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

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

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

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

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

   A recursive estimator of r_{χ_t} was identified in Slama *et al.* (2021).

3.1.2 Notations and assumptions

For this section and under (A_1) and (A_2) given in 1.1.1, we provide the following notations and assumptions which will be intensively used for our theoretical main results.

(A_5) (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_6) (i) The functions f and a_φ are bounded and twice differentiable.

(ii) For all $i,j \in \{1,\dots,d\}$, $f_{ij}^{(2)} := \dfrac{\partial^2 f}{\partial x_i \partial x_j}$ and $a_{\varphi_{ij}}^{(2)} := \dfrac{\partial^2 a_\varphi}{\partial x_i \partial x_j}$ are bounded and continuous at x.

(iii) For all $p>0$, $s \longmapsto \displaystyle\int_{\mathbb{R}} |\varphi(t)|^{2+p} g(s,t)dt$ is a bounded function.

(iv) The function $s \longmapsto \displaystyle\int_{\mathbb{R}} \varphi(t)^2 g(s,t)dt$ is bounded and continuous at $s=x$.

Discussion of the assumptions:
It is to be noted that the assumption $(A_5)(iii)$ with regard to the limit of $(n\beta_n)$ as n goes to infinity is quite common within the context of stochastic approximation algorithms. More specifically, the limit $\xi_\beta := \lim\limits_{n\to+\infty} (n\beta_n)^{-1}$ is implied to be finite. Moreover, throughout the rest of this thesis manuscript, we shall use an other notation for $\lim\limits_{n\to+\infty} (n\gamma_n)^{-1}$ denoted ξ_α.
Furthermore, it is noteworthy that the assumptions given in (A_6) are fulfilled under the following conditions:

- For every $t \in \mathbb{R}$, the function $g(.,t)$ is twice continuously differentiable in $\mathbb{R}^d$.
- The functions $s \longmapsto \displaystyle\int_{\mathbb{R}} g(s,t)dt$, $s \longmapsto \displaystyle\int_{\mathbb{R}} \varphi(t) g(s,t)dt$ and $s \longmapsto \displaystyle\int_{\mathbb{R}} \varphi(t)^2 g(s,t)dt$ are bounded and continuous at $s=x$.
- For $p>0$, $s \longmapsto \displaystyle\int_{\mathbb{R}} |\varphi(t)|^{2+p} g(s,t)dt$ is a bounded function.
- For all $i,j \in \{1,\dots,d\}$, $\displaystyle\int_{\mathbb{R}} \left|\frac{\partial g}{\partial s_i}(x,t)\right| dt < +\infty$ and $\displaystyle\int_{\mathbb{R}} |\varphi(t)| \left|\frac{\partial g}{\partial s_i}(x,t)\right| dt < +\infty$.
- For all $i,j \in \{1,\dots,d\}$, $s \longmapsto \displaystyle\int_{\mathbb{R}} \frac{\partial^2 g}{\partial s_i \partial s_j}(s,t)dt$ and $s \longmapsto \displaystyle\int_{\mathbb{R}} \varphi(t) \frac{\partial^2 g}{\partial s_i \partial s_j}(s,t)dt$ are bounded functions continuous at $s=x$.

Those conditions on the density of the couple (X,Y) were applied in the non-recursive framework for the estimation of the regression function Nadaraya (1964), Watson (1964) and in the recursive framework Mokaddem *et al.* (2009b), Slaoui (2015).
Throughout this chapter, the following notations are used :

$$\xi_{\alpha,\beta} = \xi_\beta \xi_\alpha^{-1} := \lim_{n\to+\infty} (\gamma_n \beta_n^{-1}) \quad \text{and} \quad \xi_{\beta,\alpha} = (\xi_{\alpha,\beta})^{-1} := \lim_{n\to+\infty} (\beta_n \gamma_n^{-1}).$$

3.2 Main results

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

3.2.1 Bias and variance of a_{φ_n}

Proposition 3.1. *Under the assumptions* (A_1), (A_2), (A_5) *and* (A_6), *we obtain*

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

$$\mathbb{E}[a_{\varphi_n}(x)] - a_{\varphi}(x) = \frac{h_n^2}{2(1-2a\xi_\beta)} \sum_{j=1}^{d} \mu_j(\mathbf{K}) a^{(2)}_{\varphi_{jj}}(x) + o\left(h_n^2\right). \tag{3.3}$$

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

$$\mathbb{E}[a_{\varphi_n}(x)] - a_{\varphi}(x) = o\left(\sqrt{\beta_n h_n^{-d}}\right). \tag{3.4}$$

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

$$Var[a_{\varphi_n}(x)] = o\left(h_n^4\right). \tag{3.5}$$

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

$$Var[a_{\varphi_n}(x)] = \frac{\beta_n}{h_n^d} \frac{\mathbb{E}[\varphi(Y)^2 | X = x]}{2-(\beta - ad)\xi_\beta} f(x) R(\mathbf{K}) + o\left(\beta_n h_n^{-d}\right). \tag{3.6}$$

3.2.2 Bias and variance of r_{φ_n}

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

Theorem 3.2. *Under the assumptions* (A_1), (A_2), (A_5) *and* (A_6), *we obtain*

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

$$\mathbb{E}[r_{\varphi_n}(x)] - r_{\varphi}(x) = \frac{h_n^2}{f(x)} \left(\frac{\sum_{j=1}^{d} \mu_j(\mathbf{K}) a^{(2)}_{\varphi_{jj}}(x)}{2(1-2a\xi_\beta)} - \frac{r_\varphi(x) \sum_{j=1}^{d} \mu_j(\mathbf{K}) f^{(2)}_{jj}(x)}{2(1-2a\xi_\alpha)} \right) + o\left(h_n^2\right). \tag{3.7}$$

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

$$\mathbb{E}[r_{\varphi_n}(x)] - r_{\varphi}(x) = o\left(\sqrt{\beta_n h_n^{-d}}\right) \mathbf{1}_{\{\beta \leq \alpha\}} + o\left(\sqrt{\gamma_n h_n^{-d}}\right) \mathbf{1}_{\{\alpha < \beta\}}. \tag{3.8}$$

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

$$Var[r_{\varphi_n}(x)] = o\left(h_n^4\right). \tag{3.9}$$

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

$$\begin{aligned} Var[r_{\varphi_n}(x)] = \frac{\beta_n}{h_n^d} \frac{R(\mathbf{K})}{f(x)} \Bigg[& \frac{\mathbb{E}[\varphi(Y)^2 | X = x]}{2-(\beta - ad)\xi_\beta} \\ & - r_\varphi^2(x) \left(\frac{2}{1-(\beta - ad - \xi_\beta^{-1})\xi_\alpha} - \frac{\xi_{\alpha,\beta}}{2-(\alpha-ad)\xi_\alpha} \right) + o\left(\frac{\beta_n}{h_n^d}\right) \Bigg] \mathbf{1}_{\{\beta \leqslant \alpha\}} \\ & + \frac{\gamma_n}{h_n^d} \frac{R(\mathbf{K})}{f(x)} \Bigg[\frac{\mathbb{E}[\varphi(Y)^2 | X = x] \xi_{\beta,\alpha}}{2-(\beta-ad)\xi_\beta} \\ & - r_\varphi^2(x) \left(\frac{2}{1-(\alpha - ad - \xi_\alpha^{-1})\xi_\beta} - \frac{1}{2-(\alpha-ad)\xi_\alpha} \right) + o\left(\frac{\gamma_n}{h_n^d}\right) \Bigg] \mathbf{1}_{\{\alpha < \beta\}}. \end{aligned} \tag{3.10}$$

Therefore, the bias and the variance of the estimator r_{φ_n} defined by the stochastic approximation algorithm (3.1) depend heavily on the choice of the stepsizes (β_n) and (γ_n).

Remark 3.3. *Notice that, for the case where $(\gamma_n) = (\beta_n)$ and then $\alpha = \beta$, the expression (3.10) will be written as follows*

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

The asymptotic normality of the generalized semi-recursive estimator r_{φ_n} is indicated by the following theorem.

3.2.3 Weak pointwise convergence rate of r_{φ_n}

Theorem 3.4. *Under the assumptions (A_1), (A_2), (A_5) and (A_6), we obtain:*

1. ***For the case** $\beta \leq \alpha$:*

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

   $$\sqrt{\beta_n^{-1}h_n^d}\left(r_{\varphi_n}(x) - r_\varphi(x)\right) \underset{n\to+\infty}{\overset{\mathcal{D}}{\longrightarrow}} \mathcal{N}\Big(\sqrt{c}\mathbf{M}_\beta(x)\ ,\ \mathbf{\Sigma}_\beta(x)\Big), \tag{3.11}$$

   with

   $$\mathbf{M}_\beta(x) = \frac{1}{2f(x)}\left(\frac{\sum_{j=1}^{d}\mu_j(\mathbf{K})a^{(2)}_{\varphi_{jj}}(x)}{(1-2a\xi_\beta)} - \frac{r_\varphi(x)\sum_{j=1}^{d}\mu_j(\mathbf{K})f^{(2)}_{jj}(x)}{(1-2a\xi_\alpha)}\right) \tag{3.12}$$

   and

   $$\mathbf{\Sigma}_\beta(x) = \frac{R(\mathbf{K})}{f(x)}\left[\frac{\mathbb{E}[\varphi(Y)^2|X=x]}{2-(\beta-ad)\xi_\beta} - r_\varphi^2(x)\left(\frac{2}{1-(\beta-ad-\xi_\beta^{-1})\xi_\alpha} - \frac{\xi_{\alpha,\beta}}{2-(\alpha-ad)\xi_\alpha}\right)\right]. \tag{3.13}$$

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

   $$\frac{1}{h_n^2}\left(r_{\varphi_n}(x) - r_\varphi(x)\right) \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} \mathbf{M}_\beta(x).$$

2. ***For the case** $\beta > \alpha$:*

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

   $$\sqrt{\gamma_n^{-1}h_n^d}\left(r_{\varphi_n}(x) - r_\varphi(x)\right) \underset{n\to+\infty}{\overset{\mathcal{D}}{\longrightarrow}} \mathcal{N}\Big(\sqrt{c}\mathbf{M}_\gamma(x)\ ,\ \mathbf{\Sigma}_\gamma(x)\Big), \tag{3.14}$$

   with

   $$\mathbf{M}_\gamma(x) = \frac{1}{2f(x)}\left(\frac{\sum_{j=1}^{d}\mu_j(\mathbf{K})a^{(2)}_{\varphi_{jj}}(x)}{(1-2a\xi_\beta)} - \frac{r_\varphi(x)\sum_{j=1}^{d}\mu_j(\mathbf{K})f^{(2)}_{jj}(x)}{(1-2a\xi_\alpha)}\right) \tag{3.15}$$

and

$$\mathbf{\Sigma}_\gamma(x) = \frac{R(\mathbf{K})}{f(x)}\left[\frac{\mathbb{E}[\varphi(Y)^2|X=x]\xi_{\beta,\alpha}}{2-(\alpha-ad)\xi_\alpha} - r_\varphi^2(x)\left(\frac{2}{1-(\alpha-ad-\xi_\alpha^{-1})\xi_\beta} - \frac{1}{2-(\alpha-ad)\xi_\alpha}\right)\right]. \tag{3.16}$$

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

$$\frac{1}{h_n^2}\left(r_{\varphi_n}(x) - r_\varphi(x)\right) \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} \mathbf{M}_\gamma(x).$$

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

3.2.4 Strong pointwise convergence rate of r_{φ_n}

Theorem 3.5. *Under the assumptions* (A_1), (A_2), (A_5) *and* (A_6), *we get:*

1. ***For the case*** $\beta \leq \alpha$***:***

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

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

   is relatively compact and its limit set is the interval

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

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

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

2. ***For the case*** $\beta > \alpha$***:***

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

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

   is relatively compact and its limit set is the interval

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

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

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

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

3.3 Optimal choice of the stepsizes

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

$$I_1=\int_{\mathbb{R}^d}\left(\sum_{j=1}^{d}\mu_j(\mathbf{K})a^{(2)}_{\varphi_{jj}}(x)\right)^2 f(x)dx,\quad I_4=\int_{\mathbb{R}^d}\mathbb{E}[\varphi(Y)^2|X=x]f^2(x)dx,$$
$$I_2=\int_{\mathbb{R}^d}\left(\sum_{j=1}^{d}\mu_j(\mathbf{K})a^{(2)}_{\varphi_{jj}}(x)\right)\left(\sum_{j=1}^{d}\mu_j(\mathbf{K})f^{(2)}_{jj}(x)\right)r_{\varphi}(x)f(x)dx,$$
$$I_3=\int_{\mathbb{R}^d}\left(\sum_{j=1}^{d}\mu_j(\mathbf{K})f^{(2)}_{jj}(x)\right)^2 r^2_{\varphi}(x)f(x)dx,\quad I_5=\int_{\mathbb{R}^d}r^2_{\varphi}(x)f^2(x)dx.$$

3.3.1 Asymptotic expressions of $MWISE[r_{\varphi_n}]$

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

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

Proposition 3.6. *We first note,*

$$C_1=\frac{I_1}{(1-2a\xi_\beta)^2}-\frac{2I_2}{(1-2a\xi_\beta)(1-2a\xi_\alpha)}+\frac{I_3}{(1-2a\xi_\alpha)^2},\quad C_2=\frac{I_4}{2-(\beta-ad)\xi_\beta},$$
$$C_3=\frac{I_4\xi_{\beta,\alpha}}{2-(\beta-ad)\xi_\beta},\qquad C_4=I_5\left(\frac{2}{1-(\beta-ad-\xi_\beta^{-1})\xi_\alpha}-\frac{\xi_{\alpha,\beta}}{2-(\alpha-ad)\xi_\alpha}\right),$$
$$C_5=I_5\left(\frac{2}{1-(\alpha-ad-\xi_\alpha^{-1})\xi_\beta}-\frac{1}{2-(\alpha-ad)\xi_\alpha}\right).$$

1. ***For the case*** $\beta\leq\alpha$***:***

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

2. ***For the case*** $\beta > \alpha$***:***

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

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

Corollary 3.7. *Let assumptions* (A_1), (A_2), (A_5) *and* (A_6) *hold. To minimize the* $MWISE$ *of* r_{φ_n}*, the bandwidth* (h_n) *needs to be equal to the following expressions.*

1. ***For the case*** $\beta \leq \alpha$***:***

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

Hence, the corresponding $MWISE$ *is specified by*

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

2. ***For the case*** $\alpha < \beta$ **:**

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

Thus, the corresponding $MWISE$ *is expressed by*

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

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

Proposition 3.8. *Let assumptions* (A_1), (A_2), (A_5) *and* (A_6) *hold. It is worth noting,*

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

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

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

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

   Consequently, the corresponding $MWISE$ *is determined by*

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

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

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

   Therefore, the corresponding $MWISE$ *is specified by*

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

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

$$\left(\left(\frac{d(d+2)}{2(d+4)}\right)^{\frac{1}{d+4}} \left(\frac{I_4 - I_5}{I_1 - 2I_2 + I_3}\right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{\frac{-1}{d+4}}\right). \tag{3.17}$$

Thus, the corresponding $MWISE$ is indicated by

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

Remark 3.9. *Note that for the particular case where the stepsize* (β_n) *is in* $\mathcal{GS}(-1)$ *such that* $\lim_{n\to+\infty} n\beta_n = 1$, (γ_n) *is in* $\mathcal{GS}(-1)$ *such that* $\lim_{n\to+\infty} n\gamma_n = 1$ *and the bandwidth* (h_n) *is chosen such that* $\lim_{n\to+\infty} nh_n^{d+4} = 0$ *(which corresponds to undersmoothing), the asymptotic normality of the proposed estimator is represented as follows*

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

3.4 Asymptotic properties of $\widetilde{r}_{\varphi_n}$

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

Proposition 3.10. *Let assumptions* (A_1) *and* (A_6) *hold. Then, the asymptotic properties of Nadaraya-Watson's estimator are denoted as follows.*

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

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

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

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

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

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

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

$$\left(d^{\frac{1}{d+4}}\left(\frac{I_4 - I_5}{I_1 - 2I_2 + I_3}\right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{-\frac{1}{d+4}}\right). \tag{3.18}$$

Therefore, the corresponding MWISE is expressed by

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

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

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

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

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

3.5 Bandwidth selection

Kernel smoothing in non-parametric statistics requires the choice of a bandwidth parameter. There are numerous methods for bandwidth selection, namely the cross-validation method, the bootstrap procedure and the second generation plug-in approach.
First of all, we adopt (1.14) and the kernel $\mathbf{K}$ we shall use is considered as a product of univariate kernels. Hence, we have

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

and

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

Let us start by introducing our bandwidth selection methods.

3.5.1 Plug-in method

In statistics, Altman and Leger (1995) set forward an efficient method of bandwidth selection, a plug-in estimate which minimizes an estimate of the mean weighted integrated squared error, using the density function as a weight function. Since the $MWISE$ depends on the unknown quantities I_j, $j = 1, \ldots, 5$, we attempt to construct an asymptotic unbiased estimator of those quantities.

For this purpose, we let $\mu(K) = \int_{\mathbb{R}} z^2 K(z)dz$ and

$$I_i = \mu^2(K) I_i', \; i = 1, 2, 3,$$

where

$$I_1' = \int_{\mathbb{R}^d} \left(\sum_{j=1}^{d} a_{\varphi_{jj}}^{(2)}(x) \right)^2 f(x)dx.$$

$$I_2' = \int_{\mathbb{R}^d} \left(\sum_{j=1}^{d} a_{\varphi_{jj}}^{(2)}(x) \right) \left(\sum_{j=1}^{d} f_{jj}^{(2)}(x) \right) r_\varphi(x) f(x)dx.$$

$$I_3' = \int_{\mathbb{R}^d} \left(\sum_{j=1}^{d} f_{jj}^{(2)}(x) \right)^2 r_\varphi(x) f(x)dx.$$

At this stage of analysis, in order to estimate the optimal bandwidth (3.17), we need to estimate I_j, $j = 1, \ldots, 5$. For this purpose, we recall that K_b is a kernel and b_n is the associated bandwidth, such that $\delta = 2/5$, and $K_{b'}^{(2)}$ is the second derivative of a kernel $K_{b'}$ with the associated bandwidth $b_n^{'}$ such that $\delta = 3/14$.

Semi-Recursive estimator r_{φ_n}:

To estimate the optimal bandwidth (3.17), we need to estimate I_j, $j = 1, \ldots, 5$.

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

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

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

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

Therefore, we obtain

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

Estimation of I_4 and I_5:

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

and

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

As a result, the plug-in estimator of (3.17) is denoted in terms of :

$$h_n = \left(\left(\frac{d(d+2)}{2(d+4)} \right)^{\frac{1}{d+4}} \left(\frac{\widehat{I}_4 - \widehat{I}_5}{\widehat{I}_1 - 2\widehat{I}_2 + \widehat{I}_3} \right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{\frac{-1}{d+4}} \right), \tag{3.19}$$

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

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

Non-Recursive estimator $\widetilde{r}_{\varphi_n}$:

To estimate the optimal bandwidth (3.18), we need to estimate I_j, $j = 1, \dots, 5$.
Estimation of I_1, I_2 and I_3: For the non-recursive case, the plug-in estimate yields

$$\widetilde{I'}_1 = \frac{1}{n^3 b'^{2(d+2)}_n} \sum_{\substack{i,j,k=1 \\ i \neq j \neq k}}^{n} \left[\sum_{t=1}^{d} K^{(2)}_{b'} \left(\frac{X_{it} - X_{jt}}{b'_n} \right) \prod_{\substack{l=1 \\ l \neq t}}^{d} K_b \left(\frac{X_{il} - X_{jl}}{b_n} \right) \right]$$
$$\times \left[\sum_{t=1}^{d} K^{(2)}_{b'} \left(\frac{X_{it} - X_{kt}}{b'_n} \right) \prod_{\substack{l=1 \\ l \neq t}}^{d} K_b \left(\frac{X_{il} - X_{kl}}{b_n} \right) \right] \varphi(Y_j)\varphi(Y_k),$$

$$\widetilde{I'}_2 = \frac{1}{n^3 b'^{2(d+2)}_n} \sum_{\substack{i,j,k=1 \\ i \neq j \neq k}}^{n} \left[\sum_{t=1}^{d} K^{(2)}_{b'} \left(\frac{X_{it} - X_{jt}}{b'_n} \right) \prod_{\substack{l=1 \\ l \neq t}}^{d} K_b \left(\frac{X_{il} - X_{jl}}{b_n} \right) \right]$$
$$\times \left[\sum_{t=1}^{d} K^{(2)}_{b'} \left(\frac{X_{it} - X_{kt}}{b'_n} \right) \prod_{\substack{l=1 \\ l \neq t}}^{d} K_b \left(\frac{X_{il} - X_{kl}}{b_n} \right) \right] \varphi(Y_j)\varphi(Y_i),$$

$$\widetilde{I'}_3 = \frac{1}{n^4 b'^{2(d+2)}_n} \sum_{\substack{i,j,k,m=1 \\ i \neq j \neq k \neq m}}^{n} \left[\sum_{t=1}^{d} K^{(2)}_{b'} \left(\frac{X_{it} - X_{jt}}{b'_n} \right) \prod_{\substack{l=1 \\ l \neq t}}^{d} K_b \left(\frac{X_{il} - X_{jl}}{b_n} \right) \right]$$
$$\times \left[\sum_{t=1}^{d} K^{(2)}_{b'} \left(\frac{X_{it} - X_{kt}}{b'_n} \right) \prod_{\substack{l=1 \\ l \neq t}}^{d} K_b \left(\frac{X_{il} - X_{kl}}{b_n} \right) \right] \varphi(Y_i)\varphi(Y_m),$$

Therefore, we obtain

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

Estimation of I_4 and I_5:

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

and

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

As a consequence, the plug-in estimator of (3.18) is indicated by

$$h_n = \left(\left(\frac{\widetilde{I}_4 - \widetilde{I}_5}{\widetilde{I}_1 - 2\widetilde{I}_2 + \widetilde{I}_3} \right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{-\frac{1}{d+4}} \right), \tag{3.20}$$

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

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

3.5.2 Wild Bootstrap approach

The basic idea of the wild bootstrap introduced in Hardle and Marron (1991) lies in resampling from the estimated residuals

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

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

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

Such distribution is expressed by

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

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

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

The bootstrapped bandwidth h^* is then indicated by:

$$h^* = \underset{h \in H}{argmin} \left(\frac{1}{NB} \sum_{i=1}^{NB} (r_n^*(X_i, h) - r_n(X_i, g))^2 \right), \tag{3.21}$$

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

3.6 Confidence intervals

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

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

with

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

and

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

In fact, since we have (3.28), the Confidence Intervals for means with unknown standard deviation approach ensure

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

with

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

is an estimator of (3.16).
Therefore, a confidence interval for the coverage error is given by

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

3.7 Numerical applications

The main target of this section is to perform a simulation study comparing the performance of our semi-recursive estimator (3.1) to that of Nadaraya-Watson (3.2) from confidence interval point of view. Throughout this section, we consider the regression model defined as

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

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

3.7.1 Simulation studies

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

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

Moreover, we shall consider the case where $\lambda = 0.05$ which yield $t_{\frac{\lambda}{2}} = 1.96$.

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

- The stepsizes $(\beta_n, \gamma_n) = \left(\beta_0 n^{-1}, \gamma_0 n^{-1}\right)$, where $\beta_0 = 1$ and $\gamma_0 = 1$.
- The bandwidth (h_n) which is chosen to be equal to (3.19) for plug-in recursive estimator (resp. (3.21) for bootstrapped recursive one).

For this special case, we set

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

When applying our estimator $\widetilde{r}_{\varphi_n}$, we have to opt for the following quantity:

- The bandwidth (h_n) which is chosen to be equal to (3.20) for plug-in non-recursive estimator (resp. (3.21) for bootstrapped non-recursive one).

For this special case, we set

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

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

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

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

- Model 1: X follows the normal distribution $\mathcal{N}(0,5)$ and $r_\varphi(x) = \dfrac{1}{1+\exp(-x)}$.
- Model 2: X follows the standard normal distribution $\mathcal{N}(0,1)$ and $r_\varphi(x) = \cos(x)$.
- Model 3: X follows the standard bivariate normal distribution $\mathcal{N}\left(\binom{0}{0}, \left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)\right)$ and $r_\varphi(x_1, x_2) = \exp(-x_1^2) + \sin(x_2)$.

				Plug-in		Bootstrap	
Model	σ_ε		n	Nadaraya's estimator	Recursive estimator	Nadaraya's estimator	Recursive estimator
Model 1	0.01	MSE	50	0.00316476	0.00282563	0.00307640	**0.00273586**
			100	0.00176918	0.00165111	0.00171129	**0.00159965**
			200	0.00111536	0.00103373	0.00108654	**0.00099738**
			500	0.00090141	0.00085514	0.00088022	**0.00083644**
		Cor	50	0.99499632	0.99573655	0.99512281	**0.99585851**
			100	0.99690637	0.99723254	0.99699979	**0.99731362**
			200	0.99797460	0.99819921	0.99802550	**0.99825906**
			500	0.99828980	0.99841832	0.99832927	**0.99845168**
		$MAIC$	50	0.05291563	0.05051224	0.04785009	**0.04556449**
			100	0.03647338	0.03533060	0.02995457	**0.02911132**
			200	0.02599057	0.02479223	0.02181338	**0.02103482**
			500	0.02089086	0.02039699	0.01663979	**0.01614454**
Model 2	0.01	MSE	50	0.04793966	0.04395630	0.04683956	**0.04317905**
			100	0.04366443	0.03475473	0.04086334	**0.03142012**
			200	0.01696163	0.00593385	0.01698463	**0.00592159**
			500	0.01596187	0.00438506	0.01534741	**0.00436068**
		Cor	50	0.98491608	0.98493530	0.98295729	**0.98600537**
			100	0.98738420	0.98756152	0.98746333	**0.98772430**
			200	0.98387705	0.99434872	0.98379161	**0.99423778**
			500	0.99639055	0.99739599	0.99640605	**0.99739693**
		$MAIC$	50	0.53178680	0.50550970	0.52328180	**0.47993540**
			100	0.43686240	0.41610057	0.42394440	**0.37903800**
			200	0.45451390	0.36258430	0.42906110	**0.35669820**
			500	0.36321757	0.32708432	0.34213073	**0.30261850**

Table 3.1: Quantitative comparison between Nadaraya-Watson estimator and the proposed estimator with stepsizes $(\beta_n, \gamma_n) = (n^{-1}, n^{-1})$ through a plug-in method and a bootstrap one in the unidimensional case.

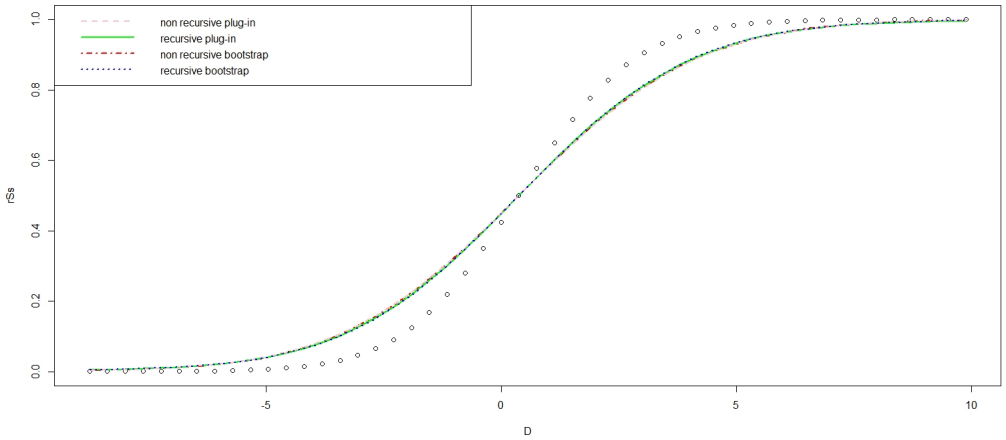


Figure 3.1: Qualitative comparison between the Nadaraya-Watson estimator and the recursive estimator for Model 1 with n=50 and $\sigma_\varepsilon = 0.01$.

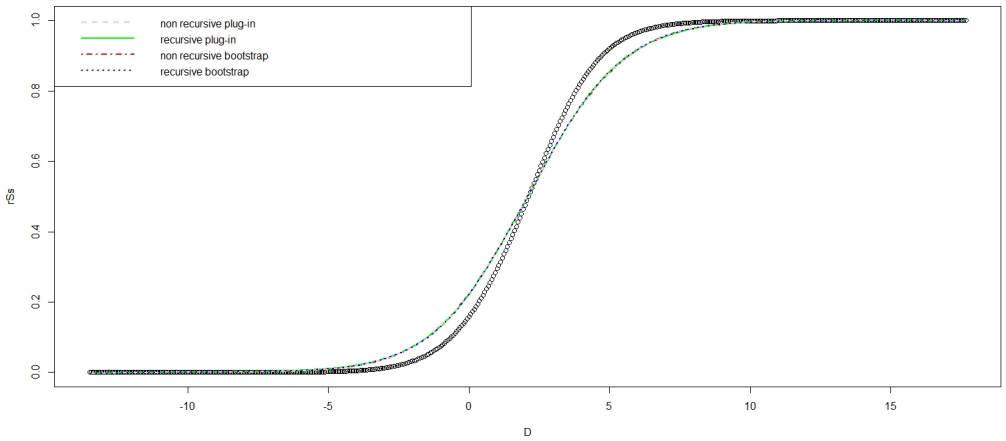


Figure 3.2: Qualitative comparison between the Nadaraya-Watson estimator and the recursive estimator for Model 1 with n=500 and $\sigma_\varepsilon = 0.01$.

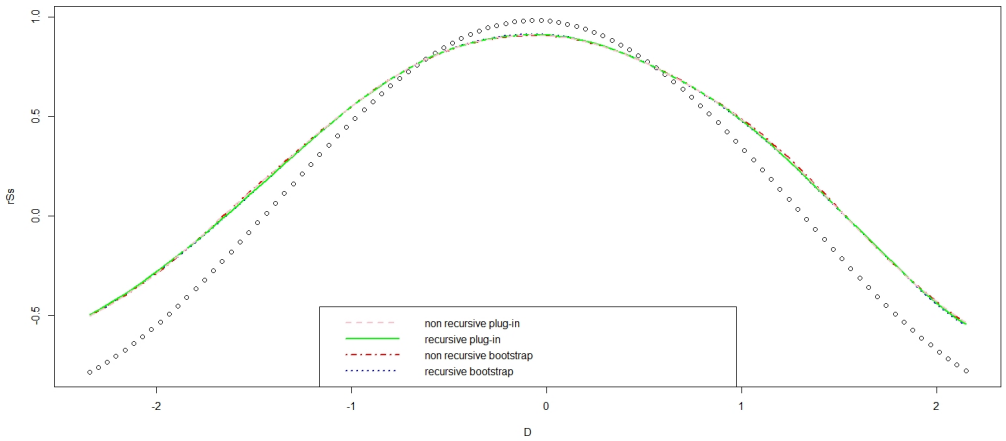


Figure 3.3: Qualitative comparison between the Nadaraya-Watson estimator and the recursive estimator for Model 2 with n=100 and $\sigma_\varepsilon = 0.1$.

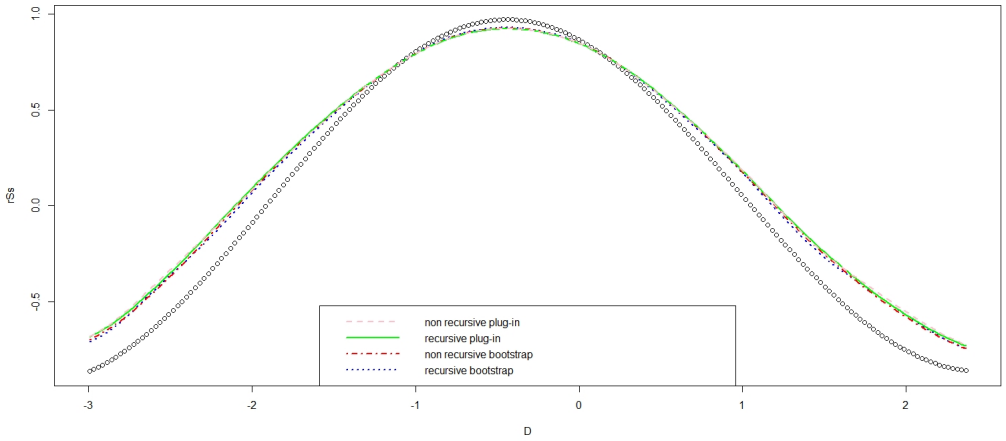


Figure 3.4: Qualitative comparison between the Nadaraya-Watson estimator and the recursive estimator for Model 2 with n=200 and $\sigma_\varepsilon = 0.1$.

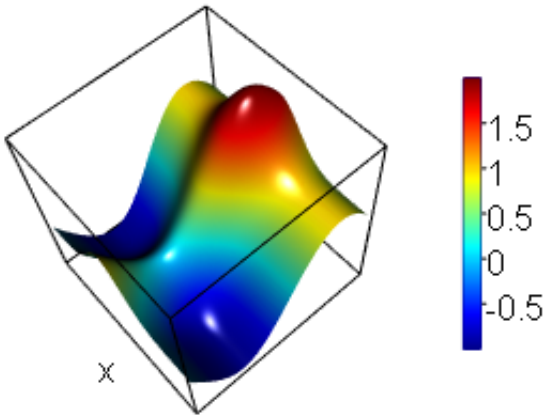


Figure 3.5: The reference regression function for Model 3 for one simple simulation with $n = 500$.

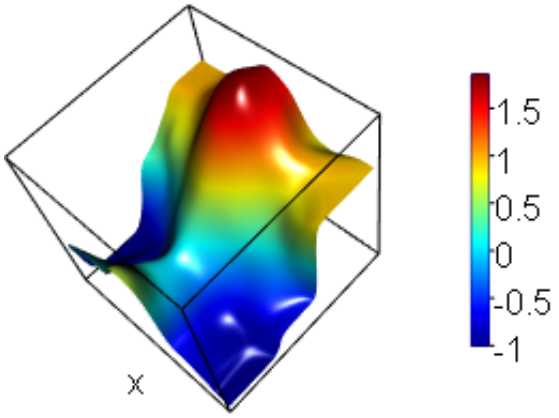


Figure 3.6: The recursive regresion estimator for Model 3 for one simple simulation with $n = 500$.

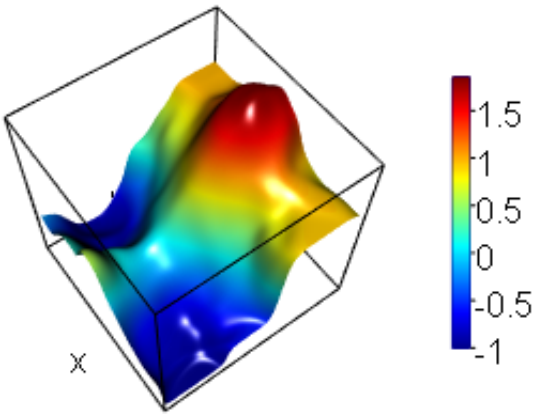


Figure 3.7: The non-recursive regression estimator for Model 3 for one simple simulation with $n = 500$.

3.7.2 Real Datasets

Application 1: French Hospital Data of COVID19

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

Description of the dataset
This dataset provides information on the hospital situation regarding the COVID-19 epidemic. We have chosen the first proposed file:
Hospital data related to the COVID-19 epidemic by department (dep) and sex (sex) of the patient: number of hospitalized patients (hosp), number of persons currently in intensive care or resuscitation (rea), number of persons currently in follow-up and rehabilitation care (SSR) or long-term care units (USLD), number of persons currently in conventional hospitalization (HospConv), number of persons currently hospitalized in another type of service (autres) or cumulative number of persons having returned home (rad), cumulative number of persons who died (dc).
The data are daily updated. For the current application, we have selected the data of 28/07/2021, with a total of 150894 observations. For simplicity reasons, we opted for focusing just on the department of 'Paris' database.
As a matter of fact, our application rests upon a dataframe of 1494 observations and 6 variables. The following two models are considered :

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

	Plug-in		Bootstrap	
Model	Nadaraya's estimator	Recursive estimator	Nadaraya's estimator	Recursive estimator
Model 1	3.648179	3.647376	3.648197	**3.647362**
Model 2	2.852123	2.514630	2.852662	**2.514177**

Table 3.2: Quantitative comparison between Nadaraya-Watson estimator and the proposed one with stepsizes $(\beta_n, \gamma_n) = (n^{-1}, n^{-1})$ through plug-in method and the bootstrap one.

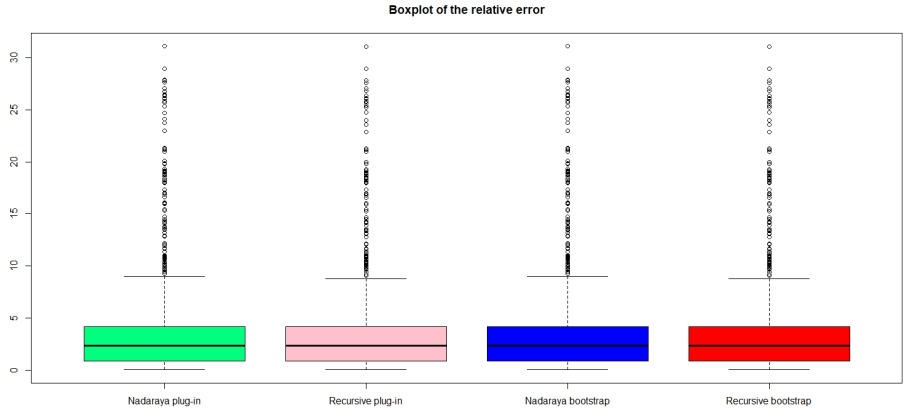


Figure 3.8: Box-plot of the relative error estimation of the four considered estimators for the bivariate COVID-19 application Model 1.

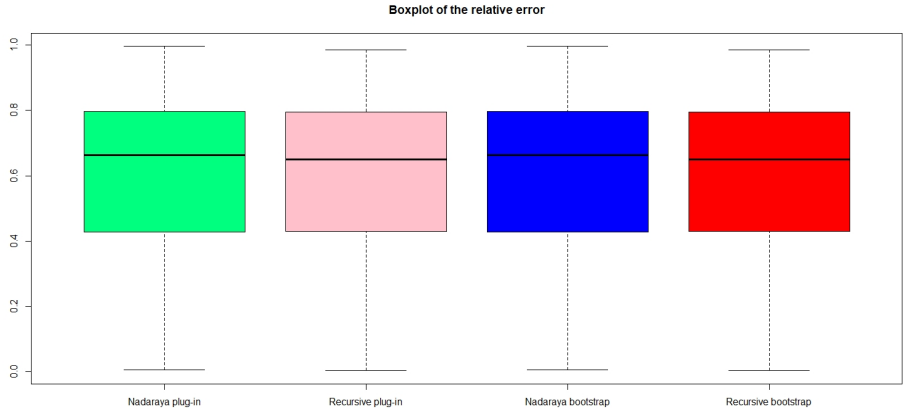


Figure 3.9: Box-plot of the relative error estimation of the four considered estimators for the bivariate COVID-19 application Model 1.

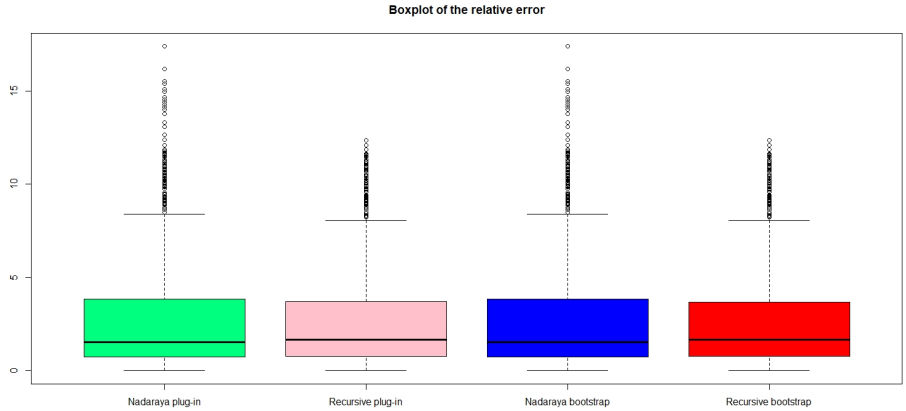


Figure 3.10: Box-plot of the relative error estimation of the four considered estimators for the bivariate COVID-19 application Model 2.

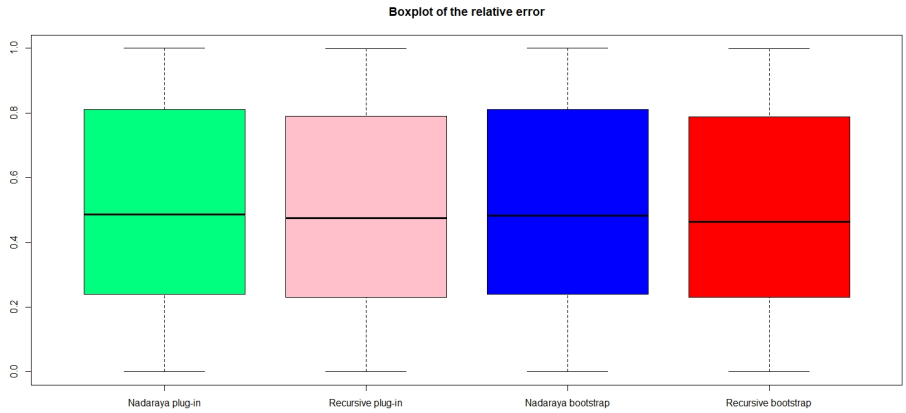


Figure 3.11: Box-plot of the relative error estimation of the four considered estimators for the bivariate COVID-19 application Model 2.

Application 2: Plasmodium falciparum Parasite Load

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

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

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

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

	Plug-in		Bootstrap	
Model	Nadaraya's estimator	Recursive estimator	Nadaraya's estimator	Recursive estimator
Model 3	0.8057840	**0.8015399**	0.8058742	0.8019406
Model 4	0.8019984	0.7962954	0.8025314	**0.7954731**

Table 3.3: Quantitative comparison between Nadaraya-Watson estimator and the proposed one with stepsizes $(\beta_n, \gamma_n) = (n^{-1}, n^{-1})$ through plug-in method and the bootstrap one.

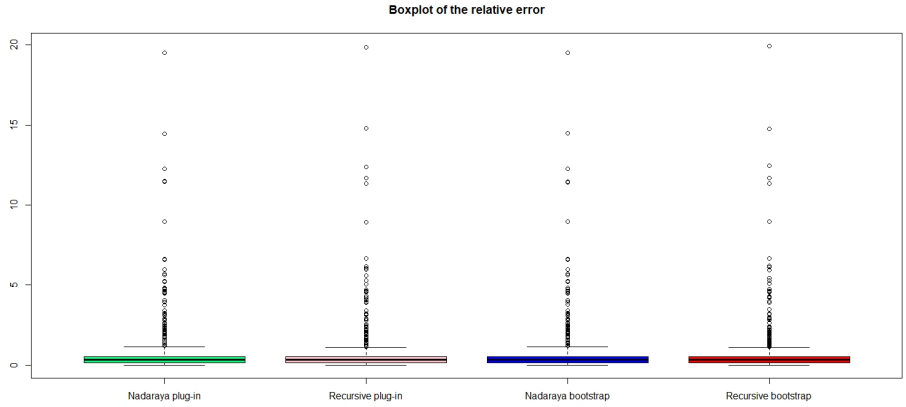


Figure 3.12: Box-plot of the relative error estimation of the four considered estimators for the multivariate PL application Model 3.

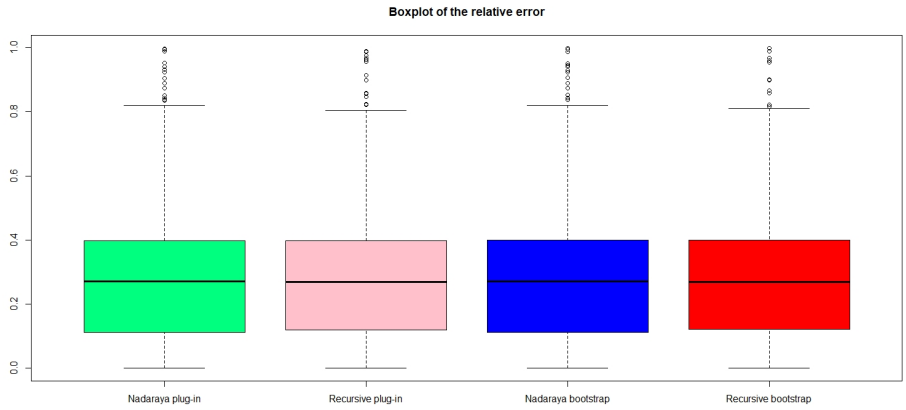


Figure 3.13: Box-plot of the relative error estimation of the four considered estimators for the multivariate PL application Model 3.

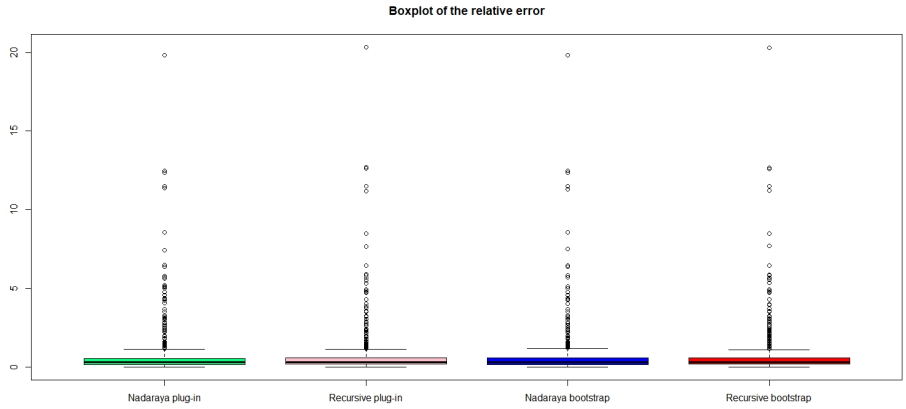


Figure 3.14: Box-plot of the relative error estimation of the four considered estimators for the multivariate PL application Model 4.

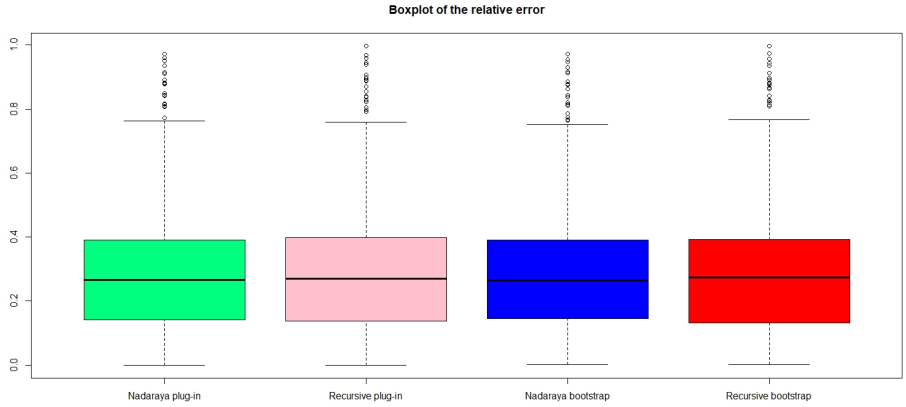


Figure 3.15: Box-plot of the relative error estimation of the four considered estimators for the multivariate PL application Model 4.

3.8 Conclusion

This chapter reports an extension of the semi-recursive regression function estimator. Initially, we tackled the asymptotic properties of the proposed estimator in order to demonstrate that our estimator asymptotically follows a normal distribution. The proposed estimator was compared to the non-recursive multivariate Nadaraya Watson regression estimator. Basically, we revealed that using a specific bandwidth selection, the plug-in approach as well as the bootstrap procedure, and particular stepsizes couple $(\gamma_n, \beta_n) = \left(n^{-1}, n^{-1}\right)$; the proposed estimator (3.1) often provides better results compared to the non-recursive Nadaraya Watson's one in terms of estimation error. The simulation studies and real datasets illustrate our findings. Even if the bootstrap approach outperforms the plug-in method, it's not quite accurate to assert that one method is better than the other. They are indistinguishable and it has been widely proven that they behave similarly. We recommend the reader to consult Delaigle and Gijbels (2004) for a detailed comparison of practical bandwidth selection procedures. In conclusion, the use of our recursive estimator, with an appropriate choice of the bandwidth, enables us to get closer to the true regression function rather than non-recursive one.

3.9 Proofs

Throughout this section, we will need the following notations:

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

Proof of Proposition 3.1.
This proof is mainly based on the same concept as the second chapter, by assuming $a := a_\varphi$. To this extent, we just briefly outline the proof. We have

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

Hence,

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

Bias of a_{φ_n} : Resting upon the assumptions (A_1) and (A_6) and by applying Taylor's development formula for a_φ, we deduce that

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

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

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

Since $a_{\varphi_{ij}}^{(2)}$ is bounded and continuous at x, we deduce that $\lim\limits_{k\to+\infty}\eta_k(x) = 0$.
For the case $a \leqslant \beta/(d+4)$, we have $\lim\limits_{+\infty}(n\beta_n) > 2a$ and then $1-2a\xi_\beta > 0$. Hence, the application of lemma 1.2 provides

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

Thus the result can be written as

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

For the case $a > \beta/(d+4)$, we have $\lim\limits_{+\infty}(n\beta_n) > \frac{\beta-a}{2}$, which ensures that $h_n^2 = o\left(\sqrt{\beta_n h_n^{-d}}\right)$.
Therefore, the application of lemma 1.2 entails

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

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

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

We have

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

with

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

Thus,

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

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

Resting upon (A_6), we have that the function $s \longmapsto \mathbb{E}[\varphi(Y)^2|X=s]f(s) = \int_{\mathbb{R}}\varphi(y)^2 g(s,y)dy$ is bounded and continuous at $s=x$ which ensures that $\lim_{k\to+\infty}\nu_k(x) = 0$.

For the case $a \geqslant \beta/(d+4)$, we have $\lim_{+\infty}(n\beta_n) > \frac{\beta-ad}{2}$ and therefore $2-(\beta-ad)\xi_\beta > 0$. Since we have $h_k\kappa_k(x) = o(1)$ and $\nu_k(x) = o(1)$, then the application of lemma 1.2 yields

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

Thus, the result is indicated in terms of

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

For the case $a < \beta/(d+4)$, we have $\lim_{+\infty}(n\beta_n) > 2a$, which ensures that $\beta_n h_n^{-d} = o\left(h_n^4\right)$. By applying lemma 1.2, we obtain

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

□

Proof of Theorem 3.2.
This proof is based on the following observation

$$r_{\varphi_n}(x) - r_\varphi(x) = D_n(x)\frac{f(x)}{f_n(x)},\ f_n \neq 0 \tag{3.22}$$

with

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

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

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

Combining the bias of $a_{\varphi_n}(x)$ ((3.3) and (3.4)) as well as that of $f_n(x)$ ((2.7) and (2.8)) yields the desired results (3.7) and (3.8).
For the variance, we get

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

1. **For the case $\beta \leq \alpha$:**
   Since X_k's are independent, then for all $i \neq k, Cov(\mathcal{Z}_k(x), \mathcal{W}_i(x)) = 0$ and by applying lemma 1.2, classical computations entail

$$Cov(a_{\varphi_n}(x), f_n(x)) = r_\varphi(x)f(x)R(\mathbf{K})\beta_n h_n^{-d}\left(\frac{1}{1-(\beta - ad - \xi_\beta^{-1})\xi_\alpha} + o(1)\right). \tag{3.23}$$

   In fact, we have

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

Consequently, (3.9) and (3.10) follow from the combination of the variance of $a_{\varphi_n}(x)$ ((3.5) and (3.6)), as well as from that of $f_n(x)$ ((2.9) and (2.10)) and the covariance expression (3.23).
It is noteworthy that, for the case $a \geq \beta/(d+4)$, we deduce

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

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

□

Proof of Theorem 3.4.
We have

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

with

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

We note

$$S_k(x) = Q_k^{-1}\beta_k\left(T_k(x) - \mathbb{E}[T_k(x)]\right). \tag{3.24}$$

Hence, we can write

$$D_n(x) - \mathbb{E}[D_n(x)] = \frac{1}{f(x)}Q_n\sum_{k=1}^{n}S_k(x). \tag{3.25}$$

Now, we are trying to apply Lyapunov's theorem 1.14 for $S_k(x)$. For this reason, we assume

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

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

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

then, the application of lemma 1.2 ensures that

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

On the other side, we have

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

Therefore,

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

Hence,

$$\mathbb{E}\left[|S_k(x)|^{2+p}\right] = O\left(Q_k^{-2-p}\beta_k^{2+p}\frac{1}{h_k^{d(1+p)}}\right). \tag{3.26}$$

As a consequence,

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

In what follows, let us suppose that $a \geq \beta/(d+4)$ and assume that there is $p > 0$, such that

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

The application of lemma 1.2 yields

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

Hence,

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

Thus, we deduce

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

In addition, since we have

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

therefore, by applying the Lyapunov theorem, we get

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

which implies

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

Moreover, (3.22) and (3.25) ensure that

$$f(x)Q_n^{-1}\upsilon_n^{-1}\left(r_{\varphi_n}(x)-\mathbb{E}[r_{\varphi_n}(x)]\right) \underset{n\to+\infty}{\overset{\mathcal{D}}{\longrightarrow}} \mathcal{N}\left(0\,,1\right). \tag{3.27}$$

Given that $\upsilon_n^2 = \dfrac{\beta_n}{h_n^d}\dfrac{f(x)^2}{Q_n^2}[\mathbf{\Sigma}_\beta(x)+o\left(1\right)]$, where $\mathbf{\Sigma}_\beta(x)$ is defined in (3.12) and by replacing υ_n with its value in (3.27), we conclude that

$$\sqrt{\beta_n^{-1}h_n^d}\left(r_{\varphi_n}(x)-\mathbb{E}[r_{\varphi_n}(x)]\right) \underset{n\to+\infty}{\overset{\mathcal{D}}{\longrightarrow}} \mathcal{N}\left(0\,,\mathbf{\Sigma}_\beta(x)\right). \tag{3.28}$$

The convergence in (3.11) then follows from the application of Lyapounov's Theorem and the combination between (3.7), (3.8) and (3.28).
The case $a < \beta/(d+4)$ is fulfilled in the convergence in probability. By applying the Bienaymé–Chebyshev inequality, we get

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

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

$$\frac{1}{h_n^2}\left(r_{\varphi_n}(x)-r_\varphi(x)\right) \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} \mathbf{M}_\beta(x),$$

with $\mathbf{M}_\beta(x)$ is provided in (3.12). □

For the next proof of the strong convergence rate, we first need to introduce a real LIL version of the Mokaddem and Pelletier (2007b) theorem 1.
Let $(X_n)_{n\geq1}$ be a sequence of independent random real variables with $\mathbb{E}[X_n]=0$. Set

$$\zeta_n = \sum_{k=1}^{n}X_k \qquad \text{and} \qquad B_n = \sum_{k=1}^{n}\mathbb{E}[X_k^2].$$

We suppose that we have the following assumptions

(AS_1) (i) $\mathbb{E}[|X_n|^3] < +\infty$.

(ii) $\dfrac{1}{n\sqrt{(n)}} \sum_{k=1}^{n} \mathbb{E}[|B_n^{-1/2} X_k|^3] \leq \dfrac{1}{\ln(B_n)}$.

(AS_2) (i) There exists a positive quantity Γ such that $\lim_{n\to+\infty} H_n^2 B_n = \Gamma$.

(ii) $\lim_{n\to+\infty} H_n^{-2} = +\infty$ and $\lim_{n\to+\infty} \dfrac{H_n^2}{H_{n-1}^2} = 1$.

Note that the assumption $(AS_1)(ii)$ is the condition which (together with the Lyapunov condition) ensures that ζ_n satisfies the central limit theorem

$$H_n \zeta_n \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}(0, \Gamma).$$

Theorem 3.11. *Laws of the iterated logarithm (LIL) for ζ_n, under assumptions (AS_1) and (AS_2), ensure that, with probability one, the sequence*

$$\left(\frac{H_n \zeta_n(x)}{\sqrt{2 \ln \ln(H_n^{-2})}} \right)$$

is relatively compact and its limit set is the interval

$$[-\Gamma, \Gamma].$$

Proof of Theorem 3.5.
For this proof, we state

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

where S_k is given in (3.24).
Here, we consider the case $\beta \leq \alpha$. We suppose that $a \geq \frac{\beta}{d+4}$ and set $\beta_0 = h_0 = 1$ and $H_n^2 = Q_n^2 \beta_n^{-1} h_n^d$, then we get

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

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

$$\ln(H_n^{-2}) = (2 - (\beta - ad)\xi) s_n + o(s_n). \tag{3.29}$$

Since $2-(\beta-ad)\xi > 0$ and s_∞ diverges, then we deduce that $\lim_{n\to+\infty} H_n^{-2} = +\infty$ and $\lim_{n\to+\infty} \frac{H_{n-1}^{-2}}{H_n^{-2}} = 1$. Moreover, we have $\sum_{k=1}^{n} Var[S_k(x)] = \frac{\beta_n}{h_n^d}\frac{f(x)^2}{Q_n^2}[\mathbf{\Sigma}_\beta(x)+o(1)]$, where $\mathbf{\Sigma}_\beta(x)$ is defined in (3.16). From this perspective, it's obvious that

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

Considering the particular case of $p=1$ in (3.26), we have $\mathbb{E}\left[|S_k(x)|^3\right] = O\left(Q_k^{-3}\beta_k^3 h_k^{-2d}\right)$ and then we deduce that

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

The application of Theorem (3.11) then ensures that, with probability one, the sequence

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

is relatively compact and its limit set is the interval

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

On account of (3.29), we have $\lim_{n\to+\infty}\frac{\ln\ln(H_n^{-2})}{\ln(sn)} = 1$, and referring to (3.22) and (3.25), we deduce that

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

is relatively compact and its limit set is the interval

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

The combination between (3.7) and (3.8) then entails

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

is relatively compact and its limit set is the interval

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

where $\mathbf{M}_\beta(x)$ is defined in (3.12) and $\mathbf{\Sigma}_\beta(x)$ is provided in (3.16).

Now we suppose that $a < \frac{\beta}{d+4}$. Set $H_n^{-2} = Q_n^{-2}h_n^4(\ln\ln(Q_n^{-2}h_n^4)^{-1})$, then we get

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

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

$$\ln(H_n^{-2}) = (2-4a\xi)\,s_n + o\left(s_n\right). \tag{3.30}$$

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

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

Considering the particular case of $p=1$ in (3.26), we have $\mathbb{E}\left[|S_k(x)|^3\right] = O\left(Q_k^{-3}\beta_k^3h_k^{-2d}\right)$ and then we deduce that

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

The application of Theorem (3.11) then ensures that, with probability one, the sequence

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

is relatively compact and its limit set is 0.

On account of (3.29), we have $\lim_{n\to+\infty}\frac{\ln\ln(H_n^{-2})}{\ln\ln(Q_n^{-2}h_n^4)} = 1$, and referring to (3.22) and (3.25), we deduce that

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

The combination between (3.7) and (3.8) then entails

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

where $\mathbf{M}_{\beta}(x)$ is defined in (3.12). □

Chapter 4

Non-parametric multivariate kernel regression estimation to describe cognitive processes and mental representations

4.1 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 Perret and Olive(2019)). 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., Bonin *et al.* (2002), Perret and Laganaro (2013)). 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 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 McGill(1963), McCormack and Wright (1964) and Luce(1986)), we resorted to non-parametric approach. In this task, 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 rely on the recursive approach of estimation using stochastic approximation method. 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 Cousineau and Chartier (2010)). Afterword, Slaoui (2017) 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.

4.1.1 Presentation of the method

Let (X,T) be a random vector with values in $\mathbb{R}^d \times \mathbb{R}$ with a joint density function $h(x,t)$ and let f denote the probability density of X. Moreover, let $(X_1,T_1),\ldots,(X_n,T_n)$ be independent random vectors identically distributed as (X,T). Assuming that $T_1,\ldots,T_n$ are subjects to missing data, the observed random variables are then Y_i and δ_i, where

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

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

$$\psi_i := \mathbb{P}[\delta_i = 1|T_i], \text{ for all } i \in \{1,\ldots,n\} \quad \text{and} \quad \psi = \lim_{n\to+\infty} \psi_n.$$

In the remainder, Y is considered as the response variable of interest and X its associated regressor vector variable.
Our basic purpose in this chapter 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

$$m : x \longmapsto \mathbb{E}[T|X=x]f(x) = \int_{\mathbb{R}} t\, h(x,t)dt$$

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 $U_n(x)$ is an observation of the function ϕ at the point $m_{n-1}(x)$. By choosing

$$U_n(x) = Y_n \psi_n^{-1}\, h_n^{-d} \mathbf{K}\left(\frac{x - X_n}{h_n}\right) - m_{n-1}(x),$$

the stochastic approximation algorithm that we consider to estimate recursively the regression function m at a vector x can be expressed by :

$$m_n(x) = (1-\beta_n)m_{n-1}(x) + \beta_n\, Y_n \psi_n^{-1}\, h_n^{-d} \mathbf{K}\left(\frac{x - X_n}{h_n}\right). \tag{4.1}$$

Throughout this section, we consider that $m_0(x) = 0$. It follows that

$$m_n(x) = Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k\, Y_k \psi_k^{-1} h_k^{-d} \mathbf{K}\left(\frac{x - X_k}{h_k}\right). \tag{4.2}$$

In this chapter, we consider the following recursive estimator of the regression function p at the vector x

$$p_n(x) = \begin{cases} \dfrac{m_n(x)}{f_n(x)} & \text{if} \quad f_n(x) \neq 0 \\ 0 & \text{otherwise} \end{cases}, \tag{4.3}$$

with f_n stands for the recursive density estimator given in (2.3).

We explore the asymptotic properties of our proposed multivariate recursive kernel regression estimator. Afterwards, by introducing the non-recursive estimator of m given by

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

we compare our proposed estimator to the multivariate non-recursive generalized Nadaraya-Watson's regression estimator indicated by

$$\widetilde{p}_n(x) = \begin{cases} \dfrac{\widetilde{m}_n(x)}{\widetilde{f}_n(x)} & \text{if} \quad \widetilde{f}_n(x) \neq 0 \\ 0 & \text{otherwise} \end{cases}, \tag{4.4}$$

with $\widetilde{f}_n$ stands for the non-recursive density estimator given in (2.5).

4.1.2 Notations and assumptions

For this section and under (A_1) given in 1.1.1 and (A_5) provided in 3.1.2, the assumptions upon which we shall rely in this chapter are the following.

Assumptions:

(A_7) (i) The functions f and m are bounded and twice differentiable.

(ii) For all $i, j \in \{1, \ldots, d\}$, $f_{ij}^{(2)} := \dfrac{\partial^2 f}{\partial x_i \partial x_j}$ and $m_{ij}^{(2)} := \dfrac{\partial^2 m}{\partial x_i \partial x_j}$ are bounded and continuous at x.

(iii) For all $p > 0$, $s \longmapsto \displaystyle\int_{\mathbb{R}} |t|^{2+p} h(s,t)dt$ is a bounded function.

(iv) The function $s \longmapsto \displaystyle\int_{\mathbb{R}} t^2 h(s,t)dt$ is bounded and continuous at $s = x$.

4.2 Main results

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

4.2.1 Bias and variance of m_n

Proposition 4.1. *Let assumptions* $(A1)$, (A_5) *and* (A_7) *hold. Hence, we obtain*

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

$$\mathbb{E}[m_n(x)] - m(x) = \frac{h_n^2}{2(1 - 2a\xi_\beta)} \sum_{j=1}^{d} \mu_j(\mathbf{K}) m_{jj}^{(2)}(x) + o\left(h_n^2\right). \tag{4.5}$$

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

$$\mathbb{E}[m_n(x)] - m(x) = o\left(\sqrt{\beta_n h_n^{-d}}\right). \tag{4.6}$$

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

$$Var[m_n(x)] = o\left(h_n^4\right). \tag{4.7}$$

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

$$Var[m_n(x)] = \frac{\beta_n}{h_n^d} \psi_n^{-1} \frac{\mathbb{E}[T^2 | X = x]}{2 - (\beta - ad)\xi_\beta} f(x) R(\mathbf{K}) + o\left(\beta_n h_n^{-d}\right). \tag{4.8}$$

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

4.2.2 Bias and variance of p_n

Theorem 4.2. *Let assumptions* $(A1)$, (A_5) *and* (A_7) *hold. Hence we obtain*

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

$$\mathbb{E}[p_n(x)] - p(x) = \frac{1}{2(1-2a\xi_\beta)} \frac{h_n^2}{f(x)} \sum_{j=1}^{d} \mu_j(\mathbf{K}) \left(m_{jj}^{(2)}(x) - p(x) f_{jj}^{(2)}(x) \right) + o\left(h_n^2\right). \quad (4.9)$$

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

$$\mathbb{E}[p_n(x)] - p(x) = o\left(\sqrt{\beta_n h_n^{-d}}\right). \quad (4.10)$$

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

$$Var[p_n(x)] = o\left(h_n^4\right). \quad (4.11)$$

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

$$Var[p_n(x)] = \frac{\beta_n}{h_n^d} \frac{\psi_n^{-1}}{2-(\beta-ad)\xi_\beta} \frac{R(\mathbf{K})}{f(x)} \Big(\mathbb{E}[T^2|X=x] - \psi p^2(x) \Big) + o\left(\frac{\beta_n}{h_n^d}\right). \quad (4.12)$$

The bias and the variance of the estimator p_n defined by the stochastic approximation algorithm (4.3) then heavily depend on the choice of the stepsizes (β_n).

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).

4.2.3 Asymptotic normality of p_n

Theorem 4.3. *Let assumptions* $(A1)$, (A_5) *and* (A_7) *hold. 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*

$$\sqrt{\beta_n^{-1} h_n^d \psi_n} \left(p_n(x) - p(x) \right) \underset{n\to+\infty}{\overset{\mathcal{D}}{\longrightarrow}} \mathcal{N}\Big(\sqrt{c} M(x) \, , \, \Sigma(x) \Big), \quad (4.13)$$

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

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

4.3 Optimal choice of the stepsizes

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 $f^3(x)$. 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 Slaoui (2016), and which shall be detailed later.

4.3.1 Asymptotic expressions of $MWISE$ of p_n

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

$$MWISE[p_n] = \int_{\mathbb{R}^d} \left(\mathbb{E}[p_n(x)] - p(x)\right)^2 f^3(x)dx + \int_{\mathbb{R}^d} Var[p_n(x)]f^3(x)dx. \tag{4.14}$$

In the sequel, we will need the folowing notations

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

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

Proposition 4.4. *For simplicity, we first set*

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

It follows that

$$MWISE[p_n] = \begin{cases} \frac{1}{4}C_1 h_n^4 + o(h_n^4) & \text{if} \quad a \in \left(0, \frac{\beta}{d+4}\right) \\ C_2 R(\mathbf{K})\beta_n h_n^{-d}\psi_n^{-1} + \frac{1}{4}C_1 h_n^4 + o(h_n^4) & \text{if} \quad a = \frac{\beta}{d+4} \\ C_2 R(\mathbf{K})\beta_n h_n^{-d}\psi_n^{-1} + o\left(\beta_n h_n^{-d}\right) & \text{if} \quad a \in \left(\frac{\beta}{d+4}, 1\right) \end{cases}.$$

The corollary below 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 4.5. *Let assumptions $(A1)$, (A_5) and (A_7) 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 $(\beta_n) = (\beta_0 n^{-1})$. We can check easily that the optimal choice of β_0 is obtained by getting β_0 equal to 1.

Corollary 4.6. *Let assumptions $(A1)$, (A_5) and (A_7) 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*

$$\left(\left(\frac{d(d+2)}{2(d+4)}\right)^{\frac{1}{d+4}} \left(\frac{I_4 - \psi I_5}{I_1 - 2I_2 + I_3}\right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{\frac{-1}{d+4}} \psi_n^{\frac{-1}{d+4}}\right). \tag{4.15}$$

Thus, the corresponding $MWISE$ is provided by

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

4.4 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 4.7. *Let assumptions* (A_1) *and* (A_7) *hold. 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

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

Corollary 4.8. *Let assumptions* (A_1) *and* (A_7) *hold. To minimize the MWISE of* $\widetilde{p}_n$*, the bandwidth* (h_n) *must be equal to*

$$\left(d^{\frac{1}{d+4}} \left(\frac{I_4 - \psi I_5}{I_1 - 2I_2 + I_3} \right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{-\frac{1}{d+4}} \psi_n^{-\frac{1}{d+4}} \right). \tag{4.16}$$

Then, the corresponding MWISE is specified by

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

It is obvious that, the use of such bandwidth (4.16), 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.

4.5 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 Delaigle and Gijbels (2004). In this chapter, based on the previous work conducted by Slaoui (2014a), Slaoui (2014b), Slaoui (2016) for unidimensional data, we propose a second generation Plug-in bandwidth data-driven procedures in the multivariate data for regression estimation.

4.5.1 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. Altman and Leger (1995) proposed an efficient method of bandwidth selection, a plug-in estimate. Since the $MWISE$ depends on unknown quantities I_j, $j = 1 \ldots 5$, we suggest elaborating an asymptotic unbiased estimator of those quantities. For this purpose, we adopt the relation (1.14) and let $\mu(K) = \int_{\mathbb{R}} z^2 K(z) dz$ and

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

where

$$I_1' = \int_{\mathbb{R}^d} \left(\sum_{j=1}^{d} m_{jj}^{(2)}(x) \right)^2 f(x)dx,$$

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

$$I_3' = \int_{\mathbb{R}^d} \left(\sum_{j=1}^{d} f_{jj}^{(2)}(x) \right)^2 p^2(x)f(x)dx.$$

At this stage of analysis, in order to estimate the optimal bandwidth (4.15), we need to estimate I_j, $j = 1, \ldots, 5$. For this purpose, we assume that K_b is a kernel and b_n is the associated bandwidth, such that $\delta = 2/5$, and $K_{b'}^{(2)}$ is the second derivative of a kernel $K_{b'}$ with the associated bandwidth $b_n^{'}$ such that $\delta = 3/14$.

Multivariate recursive kernel regression estimator under missing data p_n:

Here, we can state

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

and

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

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:

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

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

$$\widehat{I'}_3 = \frac{Q_n^2}{n} \sum_{\substack{i,j,k,m=1 \\ i\neq j\neq k\neq m}}^{n} Q_j^{-1} Q_k^{-1} \beta_j \beta_k b'^{-(d+2)}_j b'^{-(d+2)}_k \left[\sum_{t=1}^{d} K^{(2)}_{b'} \left(\frac{X_{it} - X_{jt}}{b'_j} \right) \prod_{\substack{l=1 \\ l\neq t}}^{d} K_b \left(\frac{X_{il} - X_{jl}}{b_j} \right) \right]$$
$$\times \left[\sum_{t=1}^{d} K^{(2)}_{b'} \left(\frac{X_{it} - X_{kt}}{b'_k} \right) \prod_{\substack{l=1 \\ l\neq t}}^{d} K_b \left(\frac{X_{il} - X_{kl}}{b_k} \right) \right] Y_i \psi_i^{-1} Y_m \psi_m^{-1},$$

Therefore, we obtain

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

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

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

and

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

It follows that, the plug-in bandwidth selection estimator of (4.15) is expressed by

$$(h_n) = \left(\left(\frac{d(d+2)}{2(d+4)} \right)^{\frac{1}{d+4}} \left(\frac{\widehat{I}_4 - \psi \widehat{I}_5}{\widehat{I}_1 - 2\widehat{I}_2 + \widehat{I}_3} \right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{\frac{-1}{d+4}} \psi_n^{-\frac{1}{d+4}} \right), \tag{4.17}$$

Then, the plug-in estimator of $MWISE[p_n]$ is equal to

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

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

Here, we can state

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

and

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

In order to estimate the optimal bandwidth (4.16), 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:

$$\widetilde{I'}_1 = \frac{1}{n^3 b_n'^{2(d+2)}} \sum_{\substack{i,j,k=1 \\ i\neq j\neq k}}^{n} \left[\sum_{t=1}^{d} K_{b'}^{(2)}\left(\frac{X_{it}-X_{jt}}{b_n'}\right)\prod_{\substack{l=1\\ l\neq t}}^{d} K_b\left(\frac{X_{il}-X_{jl}}{b_n}\right)\right] \times \left[\sum_{t=1}^{d} K_{b'}^{(2)}\left(\frac{X_{it}-X_{kt}}{b_n'}\right)\prod_{\substack{l=1\\ l\neq t}}^{d} K_b\left(\frac{X_{il}-X_{kl}}{b_n}\right)\right] Y_j\psi_j^{-1}Y_k\psi_k^{-1},$$

$$\widetilde{I'}_2 = \frac{1}{n^3 b_n'^{2(d+2)}} \sum_{\substack{i,j,k=1 \\ i\neq j\neq k}}^{n} \left[\sum_{t=1}^{d} K_{b'}^{(2)}\left(\frac{X_{it}-X_{jt}}{b_n'}\right)\prod_{\substack{l=1\\ l\neq t}}^{d} K_b\left(\frac{X_{il}-X_{jl}}{b_n}\right)\right] \times \left[\sum_{t=1}^{d} K_{b'}^{(2)}\left(\frac{X_{it}-X_{kt}}{b_n'}\right)\prod_{\substack{l=1\\ l\neq t}}^{d} K_b\left(\frac{X_{il}-X_{kl}}{b_n}\right)\right] Y_j\psi_j^{-1}Y_i\psi_i^{-1},$$

$$\widetilde{I'}_3 = \frac{1}{n^4 b_n'^{2(d+2)}} \sum_{\substack{i,j,k,m=1 \\ i\neq j\neq k\neq m}}^{n} \left[\sum_{t=1}^{d} K_{b'}^{(2)}\left(\frac{X_{it}-X_{jt}}{b_n'}\right)\prod_{\substack{l=1\\ l\neq t}}^{d} K_b\left(\frac{X_{il}-X_{jl}}{b_n}\right)\right] \times \left[\sum_{t=1}^{d} K_{b'}^{(2)}\left(\frac{X_{it}-X_{kt}}{b_n'}\right)\prod_{\substack{l=1\\ l\neq t}}^{d} K_b\left(\frac{X_{il}-X_{kl}}{b_n}\right)\right] Y_i\psi_i^{-1}Y_m\psi_m^{-1},$$

Therefore, we obtain

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

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

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

and

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

Hence, the plug-in bandwidth selection estimator of (4.16) is indicated by

$$(h_n) = \left(d^{\frac{1}{d+4}}\left(\frac{\widetilde{I}_4 - \psi\widetilde{I}_5}{\widetilde{I}_1 - 2\widetilde{I}_2 + \widetilde{I}_3}\right)^{\frac{1}{d+4}} R(\mathbf{K})^{\frac{1}{d+4}} n^{-\frac{1}{d+4}} \psi_n^{-\frac{1}{d+4}}\right), \tag{4.18}$$

It follows that, the plug-in non-recursive estimator of $MWISE[p_n]$ is equal to

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

4.6 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 Perret and Olive(2019). 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 Perret and Bonin (2019). 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., Bonin *et al.* (2002); Perret and Laganaro (2013)).
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 Bonin *et al.* (2002); Perret and Laganaro (2013)).

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 Bonin *et al.* (2002); Perret and Laganaro (2013)).

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 Bonin *et al.* (2002); Perret and Laganaro (2013)). 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 Bonin *et al.* (2002); Perret and Laganaro (2013)). 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 Bonin *et al.* (2002); Perret and Laganaro (2013)). Experimental work Perret *et al.* (2014) suggests that this variable influences the orthographic wordform encoding processes.

Problematic:

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 followed:

$$Y_i = T_i \times \mathbb{1}_{\{T_i \text{ is observed}\}}.$$

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:

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

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: $MSE = \dfrac{1}{n}\sum_{i=1}^{n}(p_i - p_i^*)^2$.
- The Mean Relative Error: $MRE = \dfrac{1}{n}\sum_{i=1}^{n}\left|\frac{p_i - p_i^*}{p_i^*}\right|$.

Algorithm 2 $X_1, \ldots, X_4$ 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, $X_1, \ldots, X_4$, $\mathbf{K}$, n and Np.

1: A choice value for the recursive bandwidth vectors $h_1, \ldots, h_n$. (resp. the non-recursive bandwidth values h_n) using the plug-in approach provided in (4.15) (resp. (4.18)).
2: The choice of the stepsize $(\beta_n) = \left(n^{-1}\right)$ (then, $(Q_n) = \left(n^{-1}\right)$).
3: An arbitrary sampling vectors $x_1, \ldots, x_4$.
4: The estimation of ψ is carried out according to the algorithm proposed in Slaoui (2017)
5: **for** $l = 1, \ldots, Np$ **do**
6: $\widehat{Y}_l = \dfrac{\sum_{k=1+(l-1)n}^{ln} k\beta_k Y_k \psi_k^{-1} \prod_{i=1}^{4} h_{ki}^{-1} \prod_{i=1}^{4} K\left(\frac{x_i - X_{ki}}{h_{ki}}\right)}{\sum_{k=1+(l-1)n}^{ln} k\beta_k \prod_{i=1}^{4} h_{ki}^{-1} \prod_{i=1}^{4} K\left(\frac{x_i - X_{ki}}{h_{ki}}\right)}$ for the multivariate recursive kernel regression estimator (resp. $\widetilde{Y}_l = \dfrac{\sum_{k=1+(l-1)n}^{ln} Y_k \psi_k^{-1} \prod_{i=1}^{4} K\left(\frac{x_i - X_{ki}}{h_{ni}}\right)}{\sum_{k=1+(l-1)n}^{ln} \prod_{i=1}^{4} K\left(\frac{x_i - X_{ki}}{h_{ni}}\right)}$ for the multivariate non-recursive kernel regression estimator).
7: **end for**

output: $\widehat{Y}_1, \ldots, \widehat{Y}_{Np}$ and $\widetilde{Y}_1, \ldots, \widetilde{Y}_{Np}$.

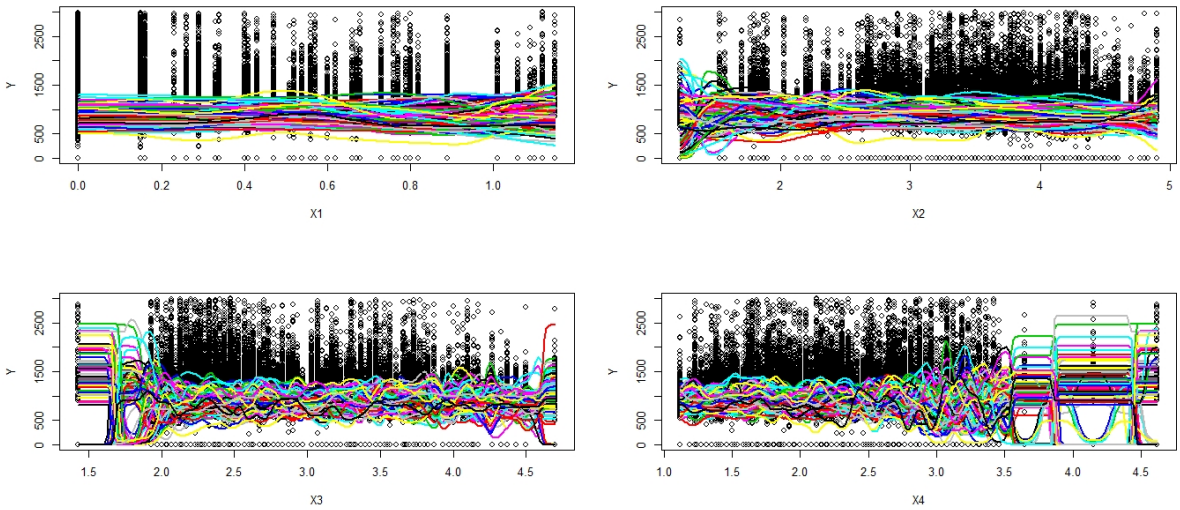


Figure 4.1: Participants' behavior representations, the regression between the reactions time variable $Y = CRTs$ and each covariate ($X_1 = H$, $X_2 = IA$, $X_3 = Ivar$ and $X_4 = AoA$) with the entire database (a total of 137 participants).

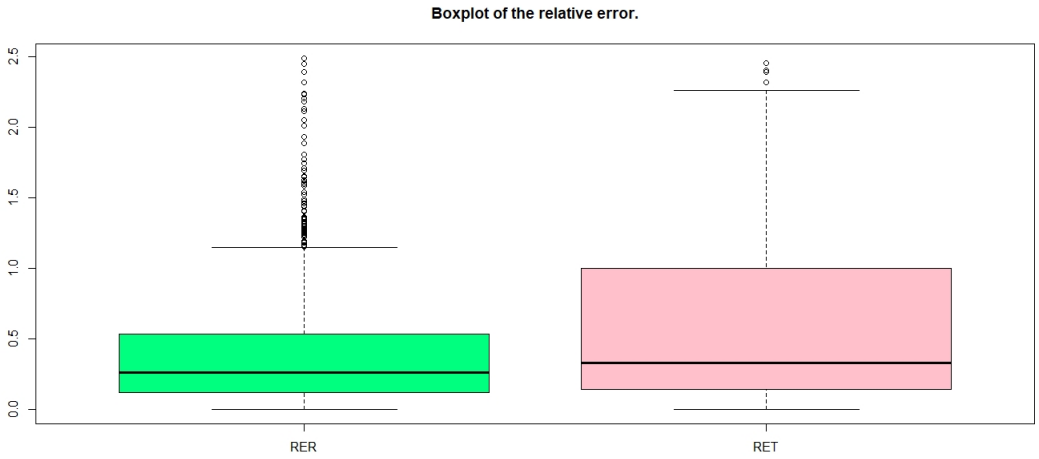


Figure 4.2: Box-plot of the relative error estimation of both considered estimators, the recursive one on the left and the non-recursive one on the right.

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 4.1: 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 shall use the MSE as a reference vector. Thus, we resort to the k-means method to specify the maximum number of needed clusters.

Algorithm 3 Participants classification algorithm:
Y is the response variable with $Y =$ `CRTs`, $\widehat{Y}_1, \ldots, \widehat{Y}_{Np}$ are the predicted multivariate recursive kernel regression estimators and $\widetilde{Y}_1, \ldots, \widetilde{Y}_{Np}$ are the predicted multivariate non-recursive kernel regression estimators.

Input: Y, $\widehat{Y}_1, \ldots, \widehat{Y}_{Np}$ and $\widetilde{Y}_1, \ldots, \widetilde{Y}_{Np}$.

1: Start with writing Y in a matrix form participant per participant.
2: **for** $l = 1, \ldots, Np$ **do**
3: $MSER_l = \frac{1}{n}\sum_{i=1}^{n}(\widehat{Y}_i - Y_i)^2$. 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).
4: **end for**
5: A classification of the remote distance through `kmeans` package in `R`.

output: The classification list using both considered estimators.

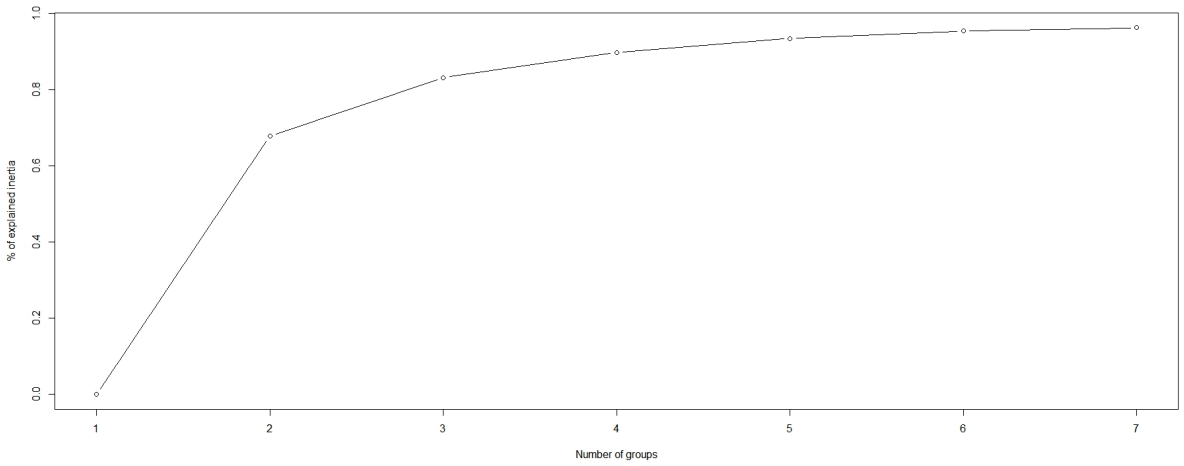


Figure 4.3: The `elbow` method of selecting the optimal number of clusters ($k = 3$) for K-means clustering on the MSE vector.

Result Analysis:

Departing form figure 4.2 and table 4.1, we deduce that the proposed recursive estimator outperforms the non-recursive one in terms of mean relative error estimation. Meanwhile, figures 4.1 and 4.3 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., Schmiedek et al.(2007); Unsworth *et al.* (2010)). 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.

4.7 Conclusion

In this research work, 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 (4.3) 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.

4.8 Proofs

Throughout this section, we use the following notations:

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

Proof of Proposition 4.1.
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,

$$\mathbb{E}[m_n(x)] - m(x) = Q_n \sum_{k=1}^{n} Q_k^{-1} \beta_k (\mathbb{E}[\mathcal{Z}_k(x)] - m(x)) + Q_n [m_0(x) - m(x)]. \tag{4.19}$$

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

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

A Taylor expansion of m around x ensures that

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

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 $i, j \in \{1, \ldots, d\}$, 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 1.2 ensures that

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

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 $\lim\limits_{n\to+\infty}(n\beta_n) > \frac{\beta - a}{2}$, which yields $h_n^2 = o\left(\sqrt{\beta_n h_n^{-d}}\right)$. Hence,

the application of lemma 1.2 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

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

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

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

Thus,

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

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

$$\text{and } \eta_k(x) = \left(\int_{\mathbb{R}^d}\mathbf{K}(z)\,m(x-zh_k)\,dz\right)^2$$

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 1.2 ensures that

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

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 1.2 ensures that

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

□

Proof of Theorem 4.2.
For this proof, let us note that for $f_n(x) \neq 0$, we have

$$p_n(x) - p(x) = A_n(x)\frac{f(x)}{f_n(x)}, \tag{4.20}$$

with

$$A_n(x) = \frac{1}{f(x)}\left(m_n(x) - m(x)\right) - \frac{p(x)}{f(x)}\left(f_n(x) - f(x)\right). \tag{4.21}$$

It follows from (4.20) 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 (4.5), (4.6), (2.7) and (2.8) in order to obtain (4.9) and (4.10). 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

$$\begin{aligned}
Cov(m_n(x), f_n(x)) &= Cov\left(Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k Y_k \psi_k^{-1} h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right), Q_n \sum_{i=1}^{n} Q_i^{-1}\beta_i h_i^{-d}\mathbf{K}\left(\frac{x-X_i}{h_i}\right)\right) \\
&= Q_n \sum_{k=1}^{n} Q_k^{-1}\beta_k Q_n \sum_{i=1}^{n} Q_i^{-1}\beta_i \, Cov\left(Y_k\psi_k^{-1} h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right), h_i^{-d}\mathbf{K}\left(\frac{x-X_i}{h_i}\right)\right) \\
&= Q_n^2 \sum_{k=1}^{n} Q_k^{-2}\beta_k^2 Cov\left(Y_k\psi_k^{-1} h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right), h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right)\right) \\
&= Q_n^2 \sum_{k=1}^{n} Q_k^{-2}\beta_k^2 \left(\mathbb{E}\left[Y_k\psi_k^{-1} h_k^{-2d}\mathbf{K}^2\left(\frac{x-X_k}{h_k}\right)\right]\right. \\
&\quad \left. - \mathbb{E}\left[Y_k\psi_k^{-1} h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right)\right]\mathbb{E}\left[h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right)\right]\right) \\
&= Q_n^2 \sum_{k=1}^{n} Q_k^{-2}\beta_k^2 \left(\mathbb{E}[T|X=x]f(x)R(\mathbf{K})h_k^{-d} - \mathbb{E}[T|X=x]f^2(x)\right) + o\left(h_k^{-d}\right) \\
&= Q_n^2 \sum_{k=1}^{n} Q_k^{-2}\beta_k^2 h_k^{-d}\left(p(x)f(x)R(\mathbf{K}) + o(1)\right) \\
&= \frac{\beta_n h_n^{-d}}{2-(\beta-ad)\xi_\beta} p(x)f(x)R(\mathbf{K}) + o\left(\beta_n h_n^{-d}\right). \qquad (4.22)
\end{aligned}$$

Consequently, (4.11) and (4.12) follow from the combination of (4.7), (4.8), (2.9), (2.10) and (4.22). For the case $a \geq \beta/(d+4)$, we can deduce

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

□

Proof of Theorem 4.3.
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

$$p_n(x) - \mathbb{E}[p_n(x)] = \frac{1}{f(x)} Q_n \sum_{k=1}^{n} S_k(x). \qquad (4.23)$$

On the other hand, we attempt to apply Lyapunov's theorem 1.14 for $S_k(x)$. For this reason, we consider

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

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 1.2 , it can be inferred that

$$\begin{aligned}\upsilon_n^2 &= \sum_{k=1}^{n} Q_k^{-2}\beta_k^2 h_k^{-d}\psi_k^{-1}\Big(\mathbb{E}[T^2|X=x]f(x)R(\mathbf{K}) + o(1)\Big) \\ &\quad + p(x)^2\sum_{k=1}^{n} Q_k^{-2}\beta_k^2 h_k^{-d}\Big(f(x)R(\mathbf{K}) + o(1)\Big) - 2p(x)\sum_{k=1}^{n} Q_k^{-2}\beta_k^2 h_k^{-d}\Big(p(x)f(x)R(\mathbf{K}) + o(1)\Big) \\ &= \frac{\beta_n}{h_n^d}\psi_n^{-1}\frac{f(x)^2}{Q_n^2}[\Sigma + o(1)]. \qquad (4.24)\end{aligned}$$

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,

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

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 1.2 , 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

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

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

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

which implies

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

Moreover, (4.23) ensures that

$$f(x) Q_n^{-1} \upsilon_n^{-1} (p_n(x) - \mathbb{E}[p_n(x)]) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}(0,1). \tag{4.25}$$

Then, the combination of (4.24) and (4.25) ensures that

$$\sqrt{\beta_n^{-1} h_n^d \psi_n} (p_n(x) - \mathbb{E}[p_n(x)]) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}(0,\Sigma). \tag{4.26}$$

Hence, the application of Lyapounov's Theorem coupled with the combination of (4.9), (4.10) and (4.26) ensures the convergence in (4.13). □

Conclusion and perspectives

In this research work, we elaborated a multivariate recursive functional estimators. The basic merit of recursive estimators resides in the fact that one can update the estimation with each additional new observation. Therefore, instead of re-running the data each time, it is possible to rewrite our considered estimator as a combination of two (or more) estimators, where each estimator is based on separate dataset, which can be very interesting to keep the computational cost reasonably low. For this purpose, we investigated the stochastic approximation method in order to set forward four basic recursive estimators, namely the multivariate distribution function estimator, the multivariate conditional cumulative distribution function estimator, the multivariate kernel-type regression estimator and the propensity score regression estimator. We first tackled the asymptotic properties of each 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 revealed that, using a specific bandwidth selection, namely the cross-validation method, the plug-in procedure as well as the bootstrap technique, and a particular choice of the stepsize; for all the cases, each proposed recursive estimator yielded better results compared to the corresponding non-recursive one in terms of estimation error. The simulation studies and real datasets illustrate our findings. Following our numerical applications, we can demonstrate that our multivariate functional recursive estimators are closer to the true functional than the multivariate non-recursive ones.

At this stage of reflection, it would be appropriate to assert that the present thesis would be worthwhile in terms of opening up further promising directions of investigation and for providing valuable perspectives for future researches. First of all, we shall state some ongoing research works in progress following the same spirit of multivariate functional estimation in order to construct some new classes of recursive estimators.

The conditional hazard function

The conditional hazard function $H(y|x)$ is defined by, for all real y and x,

$$H(y|x) := \frac{f(y|x)}{S(y|x)}$$

where

- $f(y|x)$ denotes the conditional density of Y given $X = x$.
- $S(y|x)$ denotes the conditional survival function defined by, for all real y and x, $S(y|x) := 1 - \pi(y|x)$.

Then, an estimator of the hazard function is expressed by

$$H_n(y|x) = \frac{\widehat{f}_n(y|x)}{S_n(y|x)},$$

with

$$\widehat{f}_n(y|x) = \frac{\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k\, h_k^{-2d}\mathbf{K_1}\left(\frac{y-X_k}{h_k}\right)\mathbf{K}\left(\frac{x-X_k}{h_k}\right)}{\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k\, h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right)},$$

and

$$S_n(y|x) = 1 - \pi_n(y|x) = \frac{\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k\ (1-\chi_k(y))\, h_k^{-d}\,\mathbf{K}\left(\frac{x-X_k}{h_k}\right)}{\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k\, h_k^{-d}\,\mathbf{K}\left(\frac{x-X_k}{h_k}\right)}.$$

The conditional variance

An estimator of the conditional variance $Var\,(Y|X)$ is determined by

$$\sigma_n^2(x) = \sigma_{\varphi_n}^2(x) := \frac{\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k\ \left(\varphi(Y_k)\ - r_{\varphi_n}(x)\right)^2 h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right)}{\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k\, h_k^{-d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right)}.$$

It is noteworthy that the empirical estimator of the conditional variance is specified by

$$\sigma_n^2(x) = \frac{1}{n}\sum_{i=1}^{n}\left(\varphi(Y_i) - r_n(X_i)\right)^2.$$

An other perspective approach concerning the estimation of stochastic process is given in the following.

Distribution function under Markov Renewal Process

A Markov Renewal Process (MRP) $(J_n, S_n)_{n\geq 0}$ is a random process that generalizes the notion of Markov jump processes, where $(J_n)_{n\geq 0}$ is a Markov chain and the process $(S_n)_{n\geq 0}$ are the jump times. Here, we may consider the distribution function associated with the sojourn time in state i before going to state j,

$$F_{ij}(x) = \mathbb{P}(X_{n+1} \leq x | J_n = i, J_{n+1} = j),$$

where, for $n \geq 0$, $J_0, J_1, \ldots, J_n$ are the consecutive states to be visited by the (MRP) and $(X_n)_{n\geq 0}$ defined by $X_0 = S_0 = 0$ and $X_n = S_n - S_{n-1}$ for $n \geq 1$, are the sojourn times in these states.
Then, an estimator of Jumps process with two conditions is identified by

$$\widehat{F}_n(x,y,t) = \frac{a_n(x,y,t)}{f_n(x,y)}$$

with

$$a_n(x,y,t) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k\, \chi_k(y) h_k^{-2d}\mathbf{K}\left(\frac{y-X_k}{h_k}\right)\mathbf{K}\left(\frac{x-X_k}{h_k}\right).$$

and

$$f_n(x,y) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k\, h_k^{-2d}\mathbf{K}\left(\frac{x-X_k}{h_k}\right)\mathbf{K}\left(\frac{y-X_k}{h_k}\right).$$

It's noteworthy that, even if our proposed recursive estimators outperformed the non-recursive ones, we still have the possibility to ameliorate their performances in order to minimize the risk and guaranty a much faster convergence. For this purpose, we will attempt to propose other efficient estimators using new techniques of risk optimization.

Further future research directions would be to extend our findings to the following topics.

1. Extend our findings to the setting of serially dependent observations, α-mixing framework like in Khardani and Slaoui (2019).

2. Extend our findings to the case of functional data like in Slaoui (2019) and Slaoui (2020) as well as spatial data like in Bouzebda and Slaoui (2020).

3. Consider the k-nearest neighbours smoothing with functional regressor, see Almanjahie *et al.* (2020) in finite dimensional data and Kara et al. (2017) in the case of functional data.

4. Explore the idea developed in the recent work of Boukabour and Masmoudi (2020) through considering some semi-parametric Bayesian networks approaches based on the idea developed in our last chapter.

Bibliography

Abdelkader, B. (2014). Asymptotic normality of the recursive kernel estimate of conditional cumulative distribution function.

Altman, N. and Leger, C. (1995). Bandwidth Selection for Kernel Distribution Function Estimation. *J. Statist. Plann. Inference*, **46**, 195–214.

Al-Awadhi, F., Kaid, Z., Laksaci, A., Ouassou, I. and Rachdi, M. (2019). Functional data analysis: local linear estimation of the L_1-conditional quantiles. *Stat. Methods Appl.*, **28** , 217–240.

Almanjahie, I M., Chikr Elmezouar, Z., Laksaci, A. and Rachdi, M. (2018). kNN local linear estimation of the conditional cumulative distribution function: Dependent functional data case. *C. R. Math. Acad. Sci. Paris*, **356**, 1036–1039.

Almanjahie, I. M, Aissiri, K. A., Laksaci, A. and Chikr Elmezouar, Z. (2020). The k nearest neighbors smoothing of the relative-error regression with functional regressor *Comm. Statist. Theory Methods.*

Amiri A. (2010). Estimateurs fonctionnels récursifs et leurs applications à la prévision. *Thèse en Mathématiques générales [math.GM]. Université d'Avignon.*

Benziadi, F., Laksaci, A. and Tebboune, F. (2016). Recursive kernel estimate of the conditional quantile for functional ergodic data. *Comm. Statist. Theory Methods*, **45**, 3097–3113.

Berlinet, A., Gannoun, A. and Matzner-Lø ber, E. (1998). Normalité asymptotique d'estimateurs convergents du mode conditionnel. *Canad. J. Statist.*, **26**, 365–380.

Berlinet, A., Gannoun, A. and Matzner-Lø ber, E. (1998). Propriétés asymptotiques d'estimateurs convergents des quantiles conditionnels. *C. R. Acad. Sci. Paris Sér. I Math.*, **326**, 611–614.

Blum, J.R. (1954) Multidimensional stochastic approximation methods. *Ann. Math. Statist.*, **25**, 737–744.

Bonin, P., Chalard, M., Meot, A. and Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology*, **93**, 89–114. *JPSS J. Probab. Stat. Sci.*, **12**, 117–126.

Bouanani, O., Rahmani, S. and Ait-Hennani, L. (2020). Local linear conditional cumulative distribution function with mixing data. *Arab. J. Math. (Springer)*, **9**, 289–307.

Boukabour, S. and Masmoudi, A. (2020). Semiparametric Bayesian networks for continuous data. *Comm. Statist. Theory Methods* doi:10.1080/03610926.2020.1738486.

Bouzebda, S. and Slaoui, Y. (2020). *Nonparametric Recursive Method for Kernel-Type Function Estimators for Censored Data*, *J. Stoch. Anal.* **3**, 19.

Brunel, E., Comte, F. and Lacour, C. (2010). Minimax estimation of the conditional cumulative distribution function. *Sankhya A*, **72**, 293–330.

Chikr-Elmezouar, Z., Almanjahie, I M., Laksaci, A. and Rachdi, M. (2019). FDA: strong consistency of the kNN local linear estimation of the functional conditional density and mode. *J. Nonparametr. Stat.*, **31**, 175–195.

Chilinski, P. and Silva, R. (2020). Neural likelihoods via cumulative distribution function. *arXiv:1811.00974v2[stat.ML].*

Chowdhury, M., Wu, C. and Modarres, R. (2018). Nonparametric estimation of conditional distribution functions with longitudinal data and time-varying parametric models. *Metrika*, **81**, 61–83.

Cousineau, D. and Chartier, S. (2010). Outliers detection and treatment: a review. *International Journal of Psychological Research* **3**, 58–67.

Deheuvels, P. and Mason, D. M. (2004). *General asymptotic confidence bands based on kernel-type function estimators*, *Stat. Inference Stoch. Process.* **7**, 225–277.

Deheuvels, P. (1973). *Sur l'estimation séquentielle de la densité. C. R. Acad. Sci. Paris Sér. A-B* **276**, A1119–A1121.

Delaigle, A and Gijbels, I. (2004). Practical bandwidth selection in deconvolution kernel density estimation, *Comput. Statist. Data Anal.* **45**, 249–267.

Dippon, J. and Renz, J. (1997). Weighted means in stochastic approximation of minima. *SIAM J. Control Optim.*, **35**, 1811–1827.

Dippon, J. (2003). Accelerated randomized stochastic optimization. *Ann. Statist.*, **31**, 1260–1281.

Duflo, M. (1997). *Random iterative models*, volume 34 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin. Translated from the 1990 French original by Stephen S. Wilson and revised by the author.

Einmahl, U. and Mason, D. M. (2000). *An empirical process approach to the uniform consistency of kernel-type function estimators*, *J. Theoret. Probab.* **13**, 1–37.

Fan, J., Yao, Q. and Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, **83**, 189–206.

Ferraty, F. and Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Comput. Statist.*, **17**, 545–564.

Ferrigno, S. Foliguet, B. Myriam, M. and Muller-Gueudin, A. (2014). Certainty bands for the conditional cumulative distribution function and applications

Galambos, J. and Seneta E. (1973). Regularly Varying Sequences. *Proc. Amer. Math. Soc.* **41**, 110–116.

Gannoun, A., Girard, S., Guinot, C. and Saracco, J. (2002). Reference curves based on non-parametric quantile regression. *Statistics in Medicine*, **21**, 3119–3135.

Hall, P., Wolff, R C. L. and Yao, Q. (1999). Methods for estimating a conditional distribution function. *J. Amer. Statist. Assoc.*, **94**, 154–163.

Hardle, W. and Marron, J.S.(1985). Optimal Bandwidth Selection in Nonparametric Regression Function Estimation. *The Ann. Statist.* **13**, 1465–1481.

Hardle, W. and Marron, J. S. (1991). *Bootstrap Simultaneous Error Bars for Nonparametric Regression*, *The Ann. Statist.* **19**, 778–796.

Hill, P.D. (1985). Kernel Estimation of a Distribution Function. *Comm. Statist. Theory Methods*, **14**, 605–620.

Honda, T. (2000). Nonparametric estimation of a conditional quantile for α-mixing processes. *Ann. Inst. Statist. Math.*, **52**, 459–470.

Honda, T. (2000). Nonparametric estimation of the conditional median function for long-range dependent processes. *J. Japan Statist. Soc.*, **30**, 129–142.

Jmaei, A. Slaoui Y. and Dellagi, W. (2017). Recursive distribution estimators defined by stochastic approximation method using Bernstein polynomials *J. Nonparametr. Stat.*, **29**, 792–805.

Kara, L.Z., Laksaci, A., Rachdi, M. and Vieu, P. (2017) Data-driven kNN estimation in nonparametric functional data analysis, *J. Multivariate Anal.* **153**, 176–188.

S. Khardani and Y. Slaoui. Recursive kernel density estimation and optimal bandwidth selection under alpha-mixing data, *J.* Stat. Theory Pract., 13(36), 2019.

Kiefer, J. and Wolfowitz, J. (1952). Stochastic Estimation of the Maximum of a Regression Function. *Ann. Math. Statist.*, **23**, 462–466.

Kiwitt, S. and Neumeyer, N. (2012). Estimating the conditional error distribution in nonparametric regression. *Scand. J. Stat.*, **39**, 259–281.

Kushner, H.J. and Clark, D.S. (1978). Stochastic approximation methods for constrained and unconstrained systems. *Springer*, New York.

Kara, L.Z., Laksaci, A., Rachdi, M. and Vieu, P. (2017) Data-driven kNN estimation in nonparametric functional data analysis, *J. Multivariate Anal.* **153**, 176–188.

Laksaci, A. and Hachemi, N. (2012). Note on the functional linear estimate of conditional cumulative distribution function. *JPSS J. Probab. Stat. Sci.*, **10**, 153–160.

Laksaci, A. and Maref, F. (2009). Conditional cumulative distribution estimation and its applications. *JPSS J. Probab. Stat. Sci.*, **7**, 57–69.

Li, Q., Lin, J. and Racine, J S. (2013). Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *J. Bus. Econom. Statist.*, **31**, 57–65.

Luce, R.D. (1986). Response times: their role in inferring elementary mental organization. *Handbook of Mathematical Psychology*New York: Oxford.

McCormack, P.F. and Wright, N.M. (1964). The positive skew observed in reaction time distributions. *Canadian Journal of Psychology* **18**, 43–51.

McGill, W.J. (1963). Stochastic latency mechanisms. *Handbook of Mathematical Psychology* **1**, 309–360. New York: John Wiley and Sons, Inc.

Milet, J. AND Nuel, G. AND Watier, L. AND Courtin, D. AND Slaoui, Y. AND Senghor, P. AND Migot-Nabias, F. AND Gaye, O. AND Garcia, A. *Genome Wide Linkage Study, Using a 250K SNP Map, of Plasmodium falciparum Infection and Mild Malaria Attack in a Senegalese Population*, *PLOS ONE* **5** (2010), 1–11.

Mokkadem, A. and Pelletier, M. (2007). *A companion for the Kiefer–Wolfowitz–Blum stochastic approximation algorithm*, *Ann.Statist.* **35**,1749–1772.

Mokkadem, A. and Pelletier, M. (2007). *Compact law of the iterated logarithm for matrix-normalized sums of random vectors*, *Teor. Veroyatn. Primen.* **52**, 752–767.

Mokkadem, A. Pelletier, M. and Slaoui, Y. (2009a). The stochastic approximation method for the estimation of a multivariate probability density. *J. Statist. Plann. Inference.* **139**, 2459–2478.

Mokkadem, A. Pelletier, M. and Slaoui, Y. (2009b). Revisiting Revesz's Stochastic Approximation Method for the Estimation of a Regression Function. *ALEA Lat. Amer.J. Probab.Math. Statist.* **6**, 63–114.

Mokkadem, A. and Pelletier, M. (2016). The Multivariate Revesz's Online Estimator of a Regression Function and Its Averaging.. *M. Math. Meth. Stat.* **25**, 151–167.

Nadaraya, E.A. (1964). Some New Estimates for Distribution Functions. *Theory Probab. Appl.*, **9**, 497–500.

Ould-Saïd, E. (2006). A strong uniform convergence rate of kernel conditional quantile estimator under random censorship. *Statist. Probab. Lett.*, **76**, 579–586.

Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *Ann. Math. Statist.* **33**, 1065–1076.

Perret, C. and Bonin, P. (2019). Which variables should be controlled for to investigate picture naming in adults? A Bayesian meta-analysis. *Behavior Research Methods* **51**, 2533–2545.

Perret, C. and Laganaro, M. (2013). Why are written naming latencies (not) longer than spoken naming? Reading and Writing *An Interdisciplinary Journal* **26**, 225–239.

Perret, C. and Olive, T. (2019). Spelling and Writing Words: Theoretical and Methodological Advances. *Brills Edition.*

Perret, C., Bonin, P., and Laganaro, M. (2014). Exploring the multiple-level hypothesis of AoA effects in spoken and written picture naming using a topographic ERP analysis. *Brain and Language* **135**, 20–31.

Sandra Plancade. (2013) Adaptive estimation of the conditional cumulative distribution function from current status data. *J. Statist. Plann. Inference*, **143**, 1466–1485.

Polyak, B.T. and Tsybakov, A.B. (1990). Optimal orders of accuracy for search algorithms of stochastic optimization. *Problems Inform. Transmission.*, **26**, 126–133.

Reiss, R.D. (1981). Non parametric Estimation of Smooth Distribution Function. *Scand. J. Statist.*, **8**, 116–119.

Révész, P. (1973). Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes. *I. Studia Sci. Math. Hung.*, **8**, 391–398.

Révész, P. (1977). How to Apply the Method of Stochastic Approximation in the Non parametric Estimation of a Regression Function. *Math. Operationsforsch. Statist. Ser. Statistics.*, **8**, 119–126.

Robbins, H. and Monro, S. (1951). A Stochastic Approximation Method. *Ann. Statist.*, **22**, 400–407.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.

Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *Ann. Math. Statist.*, **27**, 832–837.

Roussas, G. (1990). *Nonparametric regression estimation under mixing conditions*, *Stochastic Process. Appl.* **36**, 107–116.

Ruppert, D. (1982). Almost Sure Approximations to the Robbins–Monro and Kiefer–Wolfowitz Processes with Dependent Noise. *Ann. Probab.*, **10**, 178–187.

Sarda, P. (1993). Smoothing parameter selection for smooth distribution functions. *J. Statist Plann. Inference.*, **35**, 65–75.

Schmiedek, F., Oberauer, K., Wilhelm, O., Süβ, H.-M. and Wittmann, W.W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, **136**, 414–429.

Slama, S. and Slaoui, Y. (2021). Multivariate distribution function estimation using stochastic approximation method, *International Journal of Mathematics and Statistics*, vol. 22, no. 2, pp. 31–59.

Slama, S. Slaoui, Y. and Fathallah, H. (2022). *Statistical inference for multivariate conditional cumulative distribution function estimation by stochastic approximation method. Statistics, Optimization & Information Computing.* 10(3), 789-814, https://doi.org/10.19139/soic-2310-5070-1416.

Slama, S., Slaoui, Y. Le Du, G. and Perret, C. (2022). Non parametric multivariate kernel regression estimation to describe cognitive processes and mental representations. *Statistics, Optimization & Information Computing.* 10(4), 1021-1043. https://doi.org/10.19139/soic-2310-5070-1507.

Slama, S. Slaoui, Y. and Fathallah, H. (2022). *The stochastic approximation method for semi-recursive multivariate kernel-type regression estimation. Theory of Stochastic Processes.*

Slaoui, Y. (2006). Application des méthodes d'approximations stochastiques à l'estimation de la densité et de la régression. Mathématiques [math]. *Université de Versailles-Saint Quentin en Yvelines*, HAL Id : tel-00131964, version 1.

Slaoui, Y. (2013). Large and Moderate Principles for Recursive Kernel Density Estimators Defined by Stochastic Approximation Method. *Serdica Math. J.*, **39**, 53–82.

Slaoui, Y. (2014a). Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method. *Journal of Probability and Statistics*, ID 739640, doi:10.1155/2014/739640.

Slaoui, Y. (2014b). The stochastic approximation method for the estimation of a distribution function. *Math. Methods Statist.* **23**, 306–325.

Slaoui, Y. (2015). Plug-in bandwidth selector for recursive kernel regression estimators defined by stochastic approximation method. *Statistica Neerlandica.* **69**, 483–509.

Slaoui, Y. (2016). Optimal bandwidth selection for semi-recursive kernel regression estimators. *Statistics and Its Interface.* **9**, 375–388.

Slaoui, Y. (2017). Recursive kernel density estimators under missing data. *Comm. Statist. Theory Methods.* **18**, 9101–9125.

Slaoui, Y. (2018). Bias reduction in kernel density estimation. *J. Nonparametr. Stat.* **30**, 505–522.

Slaoui, Y. (2019). Wild Bootstrap Bandwidth Selection of Recursive Nonparametric Relative Regression for Independent Functional Data, *J. Multivariate Anal.*, **173**, 494–511.

Slaoui, Y. (2019). Large and moderate deviation principles for nonparametric recursive kernel distribution estimators defined by stochastic approximation method. *Opuscula Mathematica*, **39**, 733–746.

Slaoui, Y. and Jmaei, A. (2019). Recursive density estimators based on Robbins-Monro's scheme and using Bernstein polynomials. *Stat. Interface* **12**, 439–455.

Slaoui, Y. and Khardani, S. (2020). Adaptive recursive kernel conditional density estimators under censoring data. *ALEA Lat. Am. J. Probab. Math. Stat.***17**,389–417.

Slaoui, Y. (2020). Recursive non-parametric regression estimation for independent functional data, *Statist. Sinica*, **30**, 417–437.

Stute, W. (1986). Conditional Empirical Processes *Ann. Statist.*, **14**, 638–647.

Tsybakov, A.B. (1990). Recurrent Estimation of the Mode of a Multidimensional Distribution. *Probl. Inf. Transm.*, **8**, 119–126.

Unsworth, N., Redick, T.S., Lakey, C.E., and Young, D.L. (2010). Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation. *Intelligence*, **38**, 111–122.

Veraverbeke, N., Gijbels, I. and Omelka, M. (2014) Preadjusted non-parametric estimation of a conditional distribution function. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, **76**, 399–438.

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

Wegman, E.J. and Davies, H.I. (1979). Remarks on some recursive estimators of a probability density. *Ann.Statist.* **7**, 316–327.

Wolverton, C.T. and Wagner, T.J. (1969). Asymptotically optimal discriminant functions for pattern classification. *IEEE Trans. Inform. Theory.* **15**, 258–265.

Yu, K. and Jones, M. C. (1998) Local linear quantile regression. *J. Amer. Statist. Assoc.*, **93**, 228–237.