N° d'ordre :

REPUBLIQUE ALGERIENNE DEMOCRATIQUE & POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR & DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DJILLALI LIABES
FACULTE DES SCIENCES EXACTES
SIDI BEL ABBÈS

THESE DE DOCTORAT

Présentée par Mokhtaria Mebsout

Domaine : *Mathématique Informatique*
Filière: *Mathématique*
Intitulé de la formation: *Statistique, Mathématique Appliquées à l'économie et la Finance.*

Intitulée

Estimation non paramétrique fonctionnelle de la régression robuste avec paramètre d'échelle inconnu.

Soutenue le 15 /02/2022

Devant le jury composé de :

Président :

Mr BENAISSA Samir (Professeur à l'Université Djillali Liabès)

Examinateurs :

Mr MECHAB Boubaker (Professeur à l'Université Djillali Liabès)

Mr AZZOUZIBadr Eddine (MCA à l'ESM,Tlemcen)

Directeur de thèse :

Mr ATTOUCH Mohamed Kadi (Professeur à l'Université Djillali Liabès)

Co-Directeur de thèse :

Mr LAKSACI ALI (Professeur à l'Université Djillali Liabès)

Année universitaire : 2021/2022

Dédicaces

Remercimerciments et louanges à ***Dieu Tout Puissant*** *pour m'avoir donné la foi et la force d'accomplir ce travail .Prière et Salut soient sur Notre Chére* ***Maître et Prophète "Mohamed"*** *et sur sa famille et ses fidèles compagnons.*

A mes très chers parents " Mama et Papa"

J'adresse toute ma tendresse et ma reconnaissance à ma très ***chére Mère****, mon très* ***chére Père****. Affables, aimables, honorables : vous présentez pour moi le symbole de la bonté par excellence, en témoignage et en gratitude de leur dévouement, de leur soutien permanent durant toutes mes années d'études, leur sacrifies illimité, leur réconfort moral, eux qui ont consentitant d'effort pour mon éducation, mon instruction et pour me voir atteindre ce but, pour tout cela et pour ce qui ne peut être dit, mes affectations sans limite*

Je tiens à dédier ce mémoire aussi :

À mon très cher mari "Ben Attou" et sa famille.

Quand je t'ai connu, j'ai trouvé l'homme de ma vie, mon âme soeur et la lumière de mon chemin.

Tes sacrifices, ton soutien moral et matériel, ta gentillesse sans égal, ton profond attachement m'ont permis de réussir mes études.

Que dieu réunisse nos chemins pour un long commun serein et que ce travail soit témoignage de ma reconnaissance et de mon amour sincère et fidèle.

À ma très cher sœur "Hayet" et à mes très chères frères "Kada" , "Abd El-Hak" et "Younes"

Ma très cher sœur et mes très chères frères présents dans tous mes moments d'examens par leur soutien moral et matériel, sont la source de mon inspiration et mon courage, à qui je dois de l'amour et de la reconnaissance, et qui ont générés leur confiances en moi

Je vous exprime à travers ce travail mes sentiments de fraternité et d'amour.

À *mes Chére Amis pour leur soutien tout au long de mon séjour a l'université, et pour tous les moments de joies et de peines qu'on a passé ensemble.*

A tous ceux qui ont une relation de prés ou de loin avec la réalisation du présent rapport.

Remerciements

Nous tenons à exprimer nos remerciement tout d'abord au grand dieu qui nous donné la volonté et la patience pour réaliser se Modeste travail.

Je tiens à exprimer toute ma gratitude à mon directeur de thèse Monsieur ***le Professeur Mohammed Kadi Attouch*** *qui nous a permis de bénéficier de son encadrement.pour sa rigueur et sa enthousaisme dans la supervision de ce travail et pour son temps précieux qu'il m'a consacré, et a engagé à suivre les différentes parties de mon travail, et bien conseillé enrichissant ce mémoire, pour ses conseils, pour son aimable attention, pour son savoir qu'il ma communiqué, pour disponibilité, pour ses encouragements, et sa grande patience*

Mes sincères remerciements et ma gratitude vont aussi au ***Professeur Laksaci Ali****. Votre aide et vos remarques pertinentes m'ont beaucoup aidé à améliorer la qualité de ce travail.*

Je voudrais aussi remercier chaleureusement chacun des membres du jury qui me fond le grand honneur d'y participer.

Je tiens aussi à exprimer ma reconnaissance à tous les membres du laboratoire de Statistique et Processus Stochastiques (LSPS) ainsi les enseignants du département de Probabilité et Statistiquesqui, tout au long de mon cursus universitaire, m'ont appris du mieux qu'ils pouvaient, tout ce que je sais aujourd'hui, ainsi que mes collègues de la faculté des sciences Exactes de l'unversité de Sidi Bel-Abbés.

Enfin, je tiens à remercier vivement tout ma famille pour leur soutien constant : mes parents, mon mari, ma chère soeur, ma belle mère, mes deux frères. Je les remercie tous de m'avoir supportée et encouragée pendant les moments de doute. Je n'aurai jamais pu faire cette thèse sans eux.

Table des matières

Résumé

Au cours de la dernière décennie, la branche de la statistique qui traite l'analyse des données fonctionnelles a connu un grand essor en termes des développements théoriques et de la diversité de ses domaines d'application. Dans cette thèse, nous nous intéressons à l'estimation robuste qui présente une approche alternative aux méthodes de régression classiques, par exemple lorsque les observations sont affectées par la présence de données aberrantes. Récemment, ces estimateurs robustes ont été considérés pour des modèles avec données fonctionnelles. A priori l'estimation robuste, et en particulier l'estimation robuste de la fonction de régression est un thème ayant eu un grand intérêt en statistique non paramétrique. Dans ce cadre, on se propose dans ce travail d'étudier l'estimation robuste de la fonction de régression, dans le cas où les observations sont de nature fonctionnelle.

Dans un premier temps, on considère une suite d'observations indépendantes identiquement distribuées. Dans ce contexte, nous établissons la convergence d'une famille d'estimateurs robustes basée sur la méthode du noyau, nous montrons, sous des hypothèses générales, sont convergence presque complète (a.com.) avec taux de convergence. A titre illustratif, notre résultat est appliqué à la discrimination des courbes, aux problèmes de la prévision, et à la construction des intervalles de confiance.

Dans un second temps, nous étudions un estimateur de noyau récursif de la fonction de hasard conditionnelle, sous des certaines hypothèses on montre que l'estimation récursive du noyau des trois paramètres (densité conditionnelle, distribution conditionnelle, hasard conditionnelle) sont normalement distribuée asymptotiquement avec manquant au hasard.

Abstract

During the last decade, the branch of statistics which deals with the analysis of functional data has experienced a great boom in terms of theoretical developments and the diversity of its fields of application. In this thesis, we are interested in robust estimation which presents an alternative approach to classical regression methods, for example when the observations are affected by the presence of outliers. Recently, these robust estimators have been considered for models with functional data. A priori the robust estimation, and in particular the robust estimation of the regression function, is a topic which has had a great interest in nonparametric statistics. In this context, we propose in this work to study the robust estimation of the regression function, in the case where the observations are of a functional nature.

First, we consider a series of observations independent identically distributed. In this context, we establish the convergence of a family of robust estimators based on the kernel method, we show, under general assumptions, its almost complete convergence (a.com.) With rate of convergence. By way of illustration, our result is applied to the discrimination of curves, to the problems of forecasting, and to the construction of confidence intervals.

Secondly, we study a recursive kernel estimator of the conditional hazard function, under certain assumptions we show that the recursive estimate of the kernel of the three parameters (conditional density, conditional distribution, conditional hazard) are normally distributed asymptotically with missing at radom.

Chapitre 1

Introduction

1.1 Contexte bibliographique

1.1.1 Données fonctionnelles

La statistique nous aide à mieux comprendre le fonctionnement du monde dans lequel nous vivons. Pour d'écrire, mesurer et observer les phénomènes, elle utilise des variables. Dans ce sens, la variable est la base de la méthode statistique. Nous distinguons deux types de statistiques, statistique paramétrique et statistique non paramétrique. La statistique paramétrique est le cadre classique de la statistique. Le modèles statistique y est d'écrit par un nombre fini de paramètres. Par opposition, en statistique non paramétrique, le modèle n'est pas décrit par un nombre fini de paramètres. Divers cas de figures peuvent se présenter, comme par exemple : On s'autorise toutes les distributions possibles, i.e. on ne fait aucune hypothèse sur la forme/nature/type de la distribution des variables aléatoires. On travaille sur des espaces fonctionnels, de dimension infinie. Exemple : les densités continues sur $[0;1]$, ou les densités monotones sur $\mathbb{R}$. Le nombre de paramètres du modèle n'est pas fixé et varie (augmente) avec le nombre d'observations. Le support de la distribution est discret et varie (augmente) avec le nombre d'observations. La statistique non paramétrique grandissent chez de de nombreux auteurs et dans différents domaines. En effet, celle-ci possède un champ d'application très large permettant d'expliquer certains phénomènes mal modélisés jusqu'à présent, telles que les séries chronologiques et prédire les réalisations futures. Les progrès des procédés de recueil de données ont permis d'offrir la possibilité aux statisticiens de disposer de plus en plus souvent d'observation de variables fonctionnelles. Ces données sont modélisées comme étant des réalisations d'une variable aléatoire prenant ses valeurs dans un espace abstrait de dimension éventuellement infinie.

La branche de l'analyse statistique pour des variables fonctionnelles ont pris une ampleur considérable dans ces derniers temps, ces domaines en statistique constitue un champs de recherches connait actuellement un grand essor chez de nombreux auteurs plus en plus un grand succès auprès de la commuauté des statisticiens et dans des différents domaines. La preuve de ce succès les plusieurs publications scientifiques données sur ce sujet, et les nombreuses applications pratiques. C'est le cas notamment lorsque l'on s'intéresse aux techniques d'estimation lorsque les données sont fonctionnelles. Les premiers travaux dans ce domaine datent des années 50, par Rao et Tucker (1958) qui ont étudié l'estimation de maximum de vraisemblance pour la distribution multinomiale.

La modélisation statistique de données fonctionnelles provenant de diverses branches des sciences (économétrie, biologie, en-environnement), où cette dernière est liée à l'étude des expérience d'ensembles de données dans lesquels les observations peuvent être généralement collectées sous forme de courbes ou à des surfaces. sous cette supposition, l'analyse statistique se concentre sur un cadre de dimension infinie pour les données étudiées. Ce type de statistique modernes a reçu beaucoup d'attention au cours des 20 dernières années, et il a été popularisé dans le livres de Ramsay et Silverman.Ce type de données apparait dans de nombreux domaines de la statistique appliqués :l'environnement, la chimiométrie,les sciences météorologiques, ect. A titre théorique, un échantillon de données fonctionnelle peut être impliqué dans de nombreuses statistiques différentes problèmes, comme par exemple : classification et analyse en composantes principales (ACP) ou études longitudinales, régression et prédiction. La récente monographie de Ferraty et Vieu résume nombre de leurs contributions au non-estimation paramétrique avec des données fonctionnelles ; entre autres propriétés, la cohérence de la densité conditionnelle, la distribution conditionnelle et les estimations de régression sont établis dans le cas iid ainsi que dans des conditions de dépendance (fort mélange) et les techniques différentes sont appliquées à plusieurs exemples d'échantillons de données fonctionnelles. L'importance de ce sujet s'explique par le perfectionnement des outils de mesure et par le développement des outils informatiques et leurs capacités de stockage, ainsi, de part la richesse en applications et les problèmes théoriques soulevés, cette branche de la statistique est devenue le centre d'intérêt de plusieurs études, tant sur le plan pratique que sur le plan théorique. Cette importance devient deux principales raisons à l'engouement suscité par le traitement des variables fonctionnelles : (1) cela permet d'utiliser et de développer des outils théoriques performants, (2) cela offre un énorme potentiel en terme d'applications notamment en imagerie, agro-alimentaire, reconnaissance de formes, géophysique, économétrie, en-environnement. De plus, cette thématique couvre tous les domaines concernés par la comunauté de statisticiens : des plus appliqués au plus théoriques sans prédominance de l'une sur l'autre.

En raison du développement et de la rapidité qu'a connu le domaine informatique dans la façon de récolter et manipuler les données, d'autre alternatives sont devenues obligatoires afin de soulever cette difficulté et d'étudier les données dans leur propre dimension.

Historiquement, les premiers résultats considérant des observations sous forme de trajectoires ont été obtenus par Obukhov(1960), Holmstrom (1963) en climatologie, Deville (1974) en démographie, Molenaar et Boomsma (1987) puis Kirkpatrick et Heckman (1989) en génétique.

1.1.2 Régression pour variables fonctionnelles

L'étude du modèles de régression adaptés à des données fonctionnelles est un domaine important de la statistique fonctionnelle. On y retrouve des situations très différentes suivant que la variable explicative, la variable réponse ou les deux variables sont de nature fonctionnelle. Nous verrons à travers cette étude bibliographique que c'est un domaine en plein essor, que certaines situations ont été plus étudiées que d'autres et qu'il reste encore de nombreuses questions ouvertes. Enfin nous verrons que l'étude des séries temporelles ou la discrimination de courbes peuvent être directement reliées à des problèmes de régression faisant intervenir des variables fonctionnelles. Les références seront regroupées et présentées suivant la nature de la variable explicative X et de la variable réponse Y .

On peut commencer par le cas ou l'on s'intéresse de la régression d'une variable réelle sur une variable fonctionnelle. C'est la forme de régression qui a été le plus étudiée dans la littérature. Comme dans le cas multivarié, et les différentes méthodes ont été proposées selon la nature de l'opérateur de régression. Le premier modèle de régression considéré est le modèle linéaire fonctionnel, introduit et étudié par Ramsay et Dalzell (1991) puis Hastie et Mallows (1993), dans lequel on suppose que l'opérateur de régression est linéaire. Ainsi Preda (1999), puis Preda et Saporta (2002, 2004, 2005a, 2005b) proposent des méthodes basées sur une généralisation de la méthode "partial least square" au cas fonctionnel. Cardot et al. (1999), dans le cas d'une variable explicative non bruit´ee, puis Crambes (2007), dans le cas d'une variable explicative bruitée, adoptent une méthode de régression sur composantes principales fonctionnelles (introduite par Bosq, 1991).

Toutes les méthodes que nous venons d'évoquer concernent l'estimation de l'espérance conditionnelle. Cependant, la régression sur variable fonctionnelle peut être réalisée à partir de l'estimation d'autres quantités liées à la distribution conditionnelle de Y sachant X. Un des premiers exemples auxquels on peut penser est celui de la régression par quantiles conditionnels pour lesquels des méthodes d'estimation linéaire ont été proposées et étudiées par Cardot et al. (2004a, 2004b,2005). Une méthode d'estimation des quantiles conditionnels à partir de l'estimation à noyau de la fonction de répartition conditionnelle a également été proposée et étudiée par Ferraty, Rabhi et al. (2005), Ferraty et al. (2006), Ferraty et Vieu (2006a) et Ezzahrioui (2007). En ce qui concerne la régression par le mode conditionnel, la méthode proposée et étudiée par Ferraty, Laksaci et al. (2005), Ferraty et Vieu (2006a), Ferraty et al. (2006), Dabo-Niang et Laksaci (2006) et Ezzahrioui (2007) est basée sur l'estimation de la densité conditionnelle par des estimateurs à noyau. Les méthodes à noyaux développées pour estimer la fonction de répartition et la fonction de densité conditionnelles (voir Laksaci, 2007) permettent de considérer le problème de l'estimation de la fonction de hasard conditionnelle (voir Ezzahrioui, 2007, et Ferraty et al., 2008). On pourra regarder également les récents travaux de Attouch et al. (2007) sur les estimateurs robustes en régression sur une variable fonctionnelle.

Dans ces derniers temps la modélisation par la régression (paramétrique ou non-paramétrique) a fait l'objet d'une littérature très dense dans ce domaine de la statistique fonctionnelle. Par ailleurs, dans le contexte paramétrique, Ramsay et Silverman (2002,2005) ont apporté une grande contribution, ce qui est considéré comme une nouvelle réalisation pour la statistique, et ils représentent un large éventail de méthode pour des variables fonctionnelles, en particulier pour des variables indépendantes. Tandis que le livre de Bosq (2000) traite des modèles fonctionnels linéaires dépendants (modèle d'auto-régression). Cardot et al. (1999) ont construit un estimateur pour le modèle de régression linéaire Hilbertien en utilisant l'analyse en composantes principales fonctionnelles. Cet estimateur est construit à l'aide des propriétés spectrales de la version empirique de l'opérateur de variance-covariance de la variable explicative fonctionnelle. La convergence presque complète, ainsi que la convergence en norme L^2 pour une version régularisée ont été établi pour cet estimateur par les mêmes auteurs (2000). Cardot et al. (2004), ont introduit un estimateur pour les quantiles conditionnels vu comme forme linéaire continue définie sur un espace de Hilbert. Notons que le modèle le plus fréquemment rencontré en statistique non paramétrique est le modèle de régression qui décrit la relation entre deux ou plusieurs variables aléatoires. Ce modèle a été déjà pris en considération par plusieurs auteurs. Les propriétés et les résultats obtenus de l'estimateur de la fonction de régression dans le cas des variables indépendantes et identiquement distribuées (i.i.d) par Nadaraya (1964). Watson (1964) a établi la convergence uniforme de cet estimateur, la convergence uniforme presque sure est obtenue par Devroye (1978) et la normalité asymptotique du même estimateur a été établie par Roussas (1989). Dans la même année Gyorfie 12 Introduction générale a obtenu les résultats asymptotiques de l'estimateur de la fonction de régression sur des processus α mélangeants, Vieu (1991) a donné les termes asymptotiques exactes de l'erreur quadratique de l'estimateur à noyau de la fonction de régression.

1.1.3 Analyse non paramétrique robuste

La "robustesse" implique une insensibilité ax écarts dus à une non conformité aux hypothèses sous-jacentes à un modèle probabiliste. Le plus souvent, nous voulons obtenir des méthodes dont les hypothèses d'application ne soient pas trop restrictives. Les méthodes classiques d'Analyse Statistique nécessitent très souvent une distribution Gaussienne des données. Les méthodes robustes garantiront que le résultat est bon pour une très grande collection de distributions sans pour autant être les meilleures pour une en particulier. L'Analyse de Variance classique, par exemple, nécessite que les résidus suivent une loi Gaussienne avec un même écart-type pour chacun des groupes étudiés ; ce n'est pas une méthode robuste.

La robustesse d'une procédure statistique usuelle (estimation, test) est une question très importante en statistique. Elle permet de contrôler la stabilité de cette procédure relativement à la déviation du modèle et/ou des observations. Notons que ce problème a fait l'objet d'un long débat à la fin du XIX siècle, plusieurs scientifiques avaient déjà une idée relativement claire de cette notion de robustesse. En faite, le premier travail mathématique sur l'estimation robuste semble remonté en 1818 avec le travail de Laplace P. S. dans son deuxième supplément de la théorie analytique des probabilités. Plus exactement, le terme "robuste" a été introduit en 1954 par G. E. P. Box. Mais cette notion ne fut reconnu comme un champ de recherche qu'au milieu des années soixante. C'est surtout avec les travaux de Huber P.J. (1964), Hampel F.R. (1971) q'une théorie cohérente de la statistique robuste a été développé en se basant sur des critères de type min-max et utilise essentiellement des arguments de convexité. D'un autre point de vu d'autres auteurs Huber (1973 et 1981), Andrews (1974), Krasker et Welsh (1982), ont développé des méthodes automatiques d'ajustement robuste, qui ont l'avantage d'être aussi efficaces que la méthode des moindres carrés lorsqu'il n'y a pas de points aberrants, mais plus efficaces en présence d'observations atypiques ou encore lorsque la distribution de l'erreur dans le modèle suit une distribution à queues lourdes.

L'estimation robuste présente une approche alternative aux méthodes de régression classiques, par exemple lorsque les observations sont affectées par la présence de données aberrantes. Récemment, ces estimateurs robustes ont été considérés pour des modèles avec données fonctionnelles. A priori L'estimation robuste, et en particulier l'estimation robuste de la fonction de régression est un thème ayant eu un grand intérêt en statistique non paramétrique. Il s'agit d'un domaine dans lequel les premiers résultats furent établis au début des années soixante par Huber (1964), dont il a obtenu la consistance et la normalité asymptotique d'une classe d'estimateurs pour cette fonction. Robinson (1984), Hardle (1984) et Hardle et Tsybakov (1989) ont établi sous des conditions de mélange la normalité asymptotique d'une famille d'estimateurs à pondération issue de la méthode à noyau pour la fonction de régression. Parallèlement, Boente et Fraiman (1989, 1990) ont utilisé l'estimateur de Robinson (1984) pour étudier simultanément les deux paramètres de position et d'échelle. La consistance des estimateurs construits a été obtenue sous des conditions générales et dans les deux cas indépendants et dépendants. La convergence uniforme de l'estimateur robuste de la fonction de régression a été obtenu par Collomb et Hardle (1986), en considérant des observations ϕ-mélangeantes. Une méthode alternative d'estimation robuste de la fonction de régression a été proposé par Fan et al. (1994). Cette méthode permet d'englober plusieurs modèles non paramétrique et de robustifier la régression classique. Laïb et Ould-Saïd (2000) ont adapté l'estimateur de Collomb et Hardle (1986) pour le modèle d'autorégression d'un processus stationnaire ergodique. Ils ont obtenu la convergence uniforme de cet estimateur même lorsque la fonction objective est non bornée. Cai et Ould-Saïd (2003) ont utilisé une version robuste de l'estimation par la méthode des polynômes locaux pour la fonction de régression Ils ont établi, sous des conditions standards et lorsque les observations sont α mélangeantes, la normalité asymptotique et la convergence presque sûre de ces estimateurs.

1.1.4 Fonction de hasard

La fonction de hasard joue un rôle très important en statistique, elle se pose dans différent domaines y compris économétries, épidemiology, science de l'environement et beaucoup d'autre.
Pour le taux de hasard inconditionnel, plusieurs méthodes sont proposées et étudiées à partir d'un échantillon $X_1, X_2, ..., X_n$, dans le cas i.i.d. et le cas univarié (voir Singpurvwalla and Wong (1983), Hassani et al. (1986)) et par Shanbhag et Kotz (1987). Pour les cas multivariés et dépendants, nous renvoyons le lecteur à Roussas (1989), Sarda et Vieu (1989) et Lecoutre et Ould-Saïd (1995). En effet, d'un point de vue pratique, il semble plus réaliste d'abandonner l'indépendance et de la remplacer par un mode de dépendance. La forte cohérence de l'estimateur à noyau de la densité conditionnelle et du taux de risque conditionnel, lorsque la variable aléatoire d'intérêt est soumise à la censure, a été obtenu par Lecoutre et Ould-Saïd (1992). Spierdijk (2005) a proposé un nouvel estimateur non paramétrique pour le taux de risque conditionnel, basé sur l'estimation par des méthodes locale linéaire. Il a montré que l'estimateur du taux de risque par la méthode locale linéaire est cohérent et asymptotiquement normale.
Dans la statistique fonctionnelle, les premiers résultats de l'estimation non paramétrique de ce modèle sont obtenus par Ferraty et al. (2008), où ils ont définie l'estimateur à noyau de la fonction de hasard conditionnelle, ces auteurs ont établi la convergence presque complète de cet estimateur lorsque les observations sont indépendantes et identiquement distribuées, ils ont traité aussi le cas censurés.

Dans le même contexte, Ezzahrioui (2007) a étudié la normalité asymptotique. Le cas α-melangeant a ete traite par Quintela-Del-Rio (2010). Ce dernier a établi la convergence presque complète et la normalité asymptotique de l'estimateur proposé par Ferraty et al. (2008). L'auteur a illustré ces resultats asymptotiques par une application sur des données sismiques. On pourra regarder également le récent article de Laksaci et Mechab (2010) sur l'estimation de la fonction de hasard conditionnelle pour des données fonctionnelles spatialement dépendantes.

1.1.5 Estimation récursive

L'estimation récursive est devenu banal, prenant sa place en tant que composante essentielle dans la plupart des cours portant sur la théorie du contrôle et des systèmes, le traitement du signal et de l'image, l'estimation statistique et l'économétrie, et il devient de plus en plus important dans d'autres cours de sciences appliquées, tels que les sciences de la terre et de l'atmosphère (hydrologie, océanographie, science de l'atmosphère) ainsi que certains cours en sciences sociales (psychologie, sociologie). packages informatiques existants tels que CAPTAIN Toolbox, la boite à outils d'identification des systèmes (SID) dans Matlab et la boite à outils STAMP pour les économétries, comprennent des outils d'estimation récursive et sont devenue populaires dans divers domaines de la science et des sciences sociales.
Il est bien connu que la méthode d'estimation récursive permet de mettre à jour l'estimation pour chaque information supplémentaire cette fonctionnalité est très utile dans ce domaine de l'analyse des données fonctionnelles, parce que les observations sont disponibles réele de l'information.
L'estimation récursive non paramétrique a été étudiée par de nombreux auteurs, Wolverton et Wagner ont introduit la méthode d'estimation récursive, ils ont étudié les fonctions discriminantes asymptotiquement optimales pour la classification des modèles. Masry (1986) a étudié l'estimation récursive de la densité de probabilité pour un processus faiblement stationnaire dépendent. Amiri (2012) a établi les propriétés asymptotique de l'estimateur de régression récursive avec application dans la prédiction non paramétrique, aussi Amiri et al. (2014) ont établi les propriétés asymptotiques de l'estimateur de régression récursive avec covariable fonctionnelle. Laksaci et al. (2012) ont étudié la convergence presque complète du noyau récursif fonctionnel du quantile conditionnel pour les données ergodiques.

1.2 Brève présentation de cadre général de la thèse

considérons que $Z_i = (X_i, Y_i)_{i=1\cdots n}$ soit n copies du vecteur aléatoire, réparti de manière identique comme (X, Y) et valorisé dans $\mathcal{F} \times \mathbb{R}$, ou $\mathcal{F}$ est un espace semi-métrique , d désignant le semi-métrique. Pour tout $x \in \mathcal{F}$, nous considérons une fonction de Borel à valeur réelle ψ, et énonçons le modèle de la covariation entre X_i and Y_i. Notre fonction de régression non paramétrique, notée θ_x est implicitement définie comme un zéro par rapport à (w.r.t.) t de ce qui suit équation. Définissons (X, Y) un élément aléatoire dans $\mathcal{F} \times \mathbb{R}$ et soit

$$\Psi(x, t, \sigma) = \mathbb{E}\left(\psi\left(\frac{Y-t}{\sigma}\right)/X = x\right), \tag{1.2.1}$$

Ou $\psi : \mathbb{R} \times \mathbb{R}$ est une fonction impaire, bornée et continue satisfaisant certaines conditions de régularité à indiquer ci-dessous. La mesure d'échelle conditionnelle peut être prise comme la médiane conditionnelle de l'écart absolu par rapport à la la médiane conditionnelle, c'est -à-dire

$$s(x) = MED(|Y - m(x)|/X = x) = MAD_c(F_Y^x(.)), \tag{1.2.2}$$

ou $m(x) = MED(Y/X = x)$ est la médiane de la distribution conditionnelle. Notre estimation non paramétrique robuste de $\theta(x)$ est donné par la solution $\widehat{\theta}(x)$ de $\widehat{\Psi}(x, t, \widehat{s}(x)) = 0$, ou

$$\widehat{\Psi}(x, t, \widehat{\sigma}) = \int \psi\left(\frac{y-t}{\widehat{\sigma}}\right) d\widehat{F}(y/X = x) = \sum_{i=1}^{n} w_i(x)\psi\left(\frac{Y_i - t}{\widehat{\sigma}}\right), \tag{1.2.3}$$

ou $w_i(x) = \dfrac{K\left(\frac{d(X_i,x)}{h_k}\right)}{\displaystyle\sum_{i=1}^{n} K\left(\frac{d(X_i,x)}{h_k}\right)}$.

1.3 Brève présentation des résultats

Résultats : Cas i.i.d.

Dans cette partie les observations sont considérées indépendantes. Cas i.i.d concerne les résultats obtenu par Azzedine et al. (2008) sur la convergence presque complète de ces estimateurs. Si la fonction objective ψ vérifie certaines conditions de régularité, sous des propriété de concentration de la mesure de probabilité de la variable fonctionnelle sur des petites boules.

Le paramètre fonctionnel étudié noté θ_x , est solution de l' équation suivante :

$$\Psi(x, t, \sigma) = \mathbb{E}\left(\psi\left(\frac{Y-t}{\sigma}\right)/X = x\right) = 0, \tag{1.3.1}$$

Nous supposons que, pour tout $x \in \mathcal{F}$, θ_x existe et est unique, est un zéro par rapport à t de 3.3.9 (voir, par exemple, Boente et Fraiman (1989) ou Koul et Stute (1998) pour l'existence et l'unicité de θ_x Pour tout $(x, t) \in \mathcal{F} \times \mathbb{R}$, nous proposons un estimateur non para- métrique de $\Psi(x, t, \sigma)$ définie par

$$\widehat{\Psi}(x,t,\widehat{\sigma}) = \int \psi\left(\frac{y-t}{\widehat{\sigma}}\right) d\widehat{F}(y/X=x) = \sum_{i=1}^{n} w_i(x)\psi\left(\frac{Y_i - t}{\widehat{\sigma}}\right), \qquad (1.3.2)$$

where $w_i(x) = \dfrac{K\left(\frac{d(X_i,x)}{h_k}\right)}{\sum_{i=1}^{n} K\left(\dfrac{d(X_i,x)}{h_k}\right)}$. où K est un noyau et $h = h_n$ (est une suite de nombres réels positifs qui tend vers zéro quand n tend vers l'infini) Un estimateur naturel du paramètre θ_x, noté $\widehat{\theta}$, est un zéro par rapport à t de

$$\widehat{\Psi}(x,t,\widehat{\sigma}) = 0$$

. L'objectif principal est d'obtenir le taux de convergence presque complète pour x. Certains simulation a été donné pour montrer comment mettre en ouvre notre méthodologie pour les données fonctionnelles et le comportement de notre estimateur. Hypothèses Dans ce qui suit, x est un point fixe en $\mathcal{F}$, $\mathcal{N}_x$ désigne un voisinage fixe de x, et nous présentons les hypothèses suivantes :

(**H1**) Soit dénote par $\phi_x(h) = \mathbb{P}(X \in B(x,h)) = \mathbb{P}[X \in \{x' \in \varepsilon; d(x,x') < h\}]$, et nous suppose que $\phi_x(h)$ est continue, strictement croissant dans un voisinage de 0 et $\phi_x(0) = 0$.

(**H2**) La fonction Ψ et tel que

- **i)** La fonction $\Psi(x,t,\sigma)$ is of class $\mathcal{C}^2$ on $[\theta_x - \tau, \theta_x + \tau]$, $\tau > 0$.
- **ii)** $\forall (a_1,a_2) \in [\theta_x - \tau, \theta_x + \tau] \times [\theta_x - \tau, \theta_x + \tau]$, $\forall (x_1,x_2) \in \mathcal{N}_x \times \mathcal{N}_x$,
  $|\Psi(x_1,a_1,\sigma) - \Psi(x_2,a_2,\sigma)| \leq Cd^{\gamma_1}(x_1,x_2) + |a_1 - a_2|^{\gamma_1}, \gamma_1,\ \gamma_2 > 0.$
- **iii)** Pour chaque fixe $t \in [\theta_x - \tau, \theta_x + \tau]$, la fonction $\Psi(x,t,\sigma)$ is continue au point x.

(**H3**) **i)** ψ est une fonction monotone par rapport au deuxième composant (w.r.t).

- **ii)** $\psi : \mathbb{R} \to \mathbb{R}$ est une fonction impaire, strictement convexe et croissante, continuellement différentiable avec dérivé borné ψ, tel que $\xi(u) = u\psi'(u)$.
- **iii)** pour chaque fixe $t \in [\theta_x - \tau, \theta_x + \tau]$, $\mathbb{E}\left[\psi\left(\frac{Y-t}{\sigma}\right)/X\right] < C < \infty, a.s.$

(**H4**) Le noyau K est une fonction bornée non négative avec support [0,1), tel que $|\int_0^1 K(u)du| < \infty$, et satisfait la condition de Lipschitz d'ordre $\gamma_1 > 0$

$$\exists L_K < \infty, |K(u) - K(v)| \leq L_K|u-v|^{\gamma_1},$$

and

i) Si $K(1) = 0$, K est différentiable avec la dérivée K' and

$$-\infty < \inf_{u\in[0,1]} K'(u) \leq \sup_{u\in[0,1]} K'(u) = \|K'\|_\infty < 0.$$

ii) If $K(1) > 0$, il existe deux constantes réelles $C_1, C_2, 0 < C_1 < C_2 < \infty$ such that

$$0 < C_1 \mathbb{I}_{[0,1]} < K < C_2 \mathbb{I}_{[0,1]} < \infty.$$

(**H5**) Soit $S_{\mathcal{F}}$ un compact de $\mathcal{F}$ tel que

i) $F(y/X = x)$ est symétrique autour de θ_x.

ii) $F(y/X = x)$ a une médiane unique $m(x)$.

iii) Pour chaque y fixe $F(y/X = x)$ est une fonction uniformément continue de x au voisinage de $S_{\mathcal{F}}$.

iv) La condition d'équicontinuité suivante donne,

$$\forall \epsilon > 0, \exists \delta > 0; |u - v| < \delta \Rightarrow \sup_{x\in S_{\mathcal{F}}} |F(u/X = x) - F(v/X = x)| < \epsilon.$$

(**H6**) Les fonctions ϕ_x et $\Gamma_{S_{\mathcal{F}}}$ sont telles que

i) $\exists\ C > 0$ et $\exists\ \eta_0 > 0$ tels que pour tout $\eta < \eta_0$, $\phi'_x(\eta) < C$. If $K(1) = 0$, la fonction ϕ_x satisfait la condition supplémentaire :

$$\exists\ C > 0,\ \text{et}\ \ \exists\ \eta_0 > 0, \text{tels que} \ \forall\ 0 < \eta < \eta_0 \ \int_0^\eta \phi_x(u)du > C\eta\phi_x(\eta).$$

ii) Pour n assez grand,

$$\frac{(\log(n))^2}{n\phi_x(h)} < \Gamma_{S_{\mathcal{F}}}\left(\frac{\log(n)}{n}\right) < \frac{n\phi_x(h)}{\log(n)}.$$

(**H7**) La séquence $h = h_n$ et telle $h_n \to 0$,

$$n\phi_x(h_n) \to \infty \ \text{and} \ \frac{n\phi_x(h_n)}{\log(n)} \ \text{as} \ n \to \infty.$$

(**H8**) **i)** $\forall (y_1, y_2) \in S_{\mathbb{R}} \times S_{\mathbb{R}}, \forall (x_1, x_2) \in S_{\mathcal{F}} \times S_{\mathcal{F}}$,

$$|F(y_1/X = x_1) - F(y_2/X = x_2)| \leq C\left(d(x_1, x_2)^{\gamma_1} + C|y_1 - y_2|^{\gamma_2}\right).$$

ii) La fonction $F(y/X = x)$ et uniformément Lipschitz dans un voisinage $S_{\mathcal{F}}$ de $\mathcal{F}$, et il existe deux constantes $C > 0$ et $\gamma_1 > 0$, telles que, pour $x_1, x_2 \in S_{\mathcal{F}}$,

$$\sup_{y\in\mathbb{R}} |F(y/X = x_1) - F(y/X = x_2)| \leq Cd^{\gamma_1}(x_1, x_2).$$

(**H9**) ϵ-entropie de Kolmogorov's de $S_{\mathcal{F}}$, satisfait

$$\sum_{i=1}^{\infty} n \exp\left\{(1-\beta)\Gamma_{S_{\mathcal{F}}}\left(\frac{\log(n)}{n}\right)\right\} < \infty \qquad \text{for some } \beta > 1. \quad (1.3.3)$$

(**H10**) La suite $(X_i, Y_i)_{i\in\mathbb{N}}$ satisfait :

i) $\exists a > 0, \exists c > 0 : \forall n \in \mathbb{N}, \alpha(n) \leq cn^{-a}$.

ii) $\begin{cases} \forall i \neq j, \forall t \in [\theta_x - \tau, \theta_x + \tau], \\ \quad \mathbb{E}\left[\psi\left(\frac{Y_i - t}{\sigma}\right)\psi\left(\frac{Y_j - t}{\sigma}\right) \mid X_i, X_j\right] \leq C < \infty, \\ \mathbb{P}\left((X_i, X_j) \in B(x,r) \times B(x,r)\right) = \varphi_x(r) > 0. \end{cases}$

iii) Il existe $\eta > 0$, tel que

$$Cn^{2(2-a)/(a+1)+\eta} \leq \max\left(\phi_x^2(r), \varphi_x(r)\right) \leq n^{2a/(2-a)p},$$

ou $a > \max\left(5/2, \frac{4p+1+\sqrt{24p+1}}{p-1}\right)$.

1.3.1 Resultats de convergence uniforme forte

Dans cette section, nous obtiendrons des résultats de convergence uniformes sur des ensembles compacts de l'estimateur de régression $\widehat{\theta}$ défini comme une solution de $\widehat{\Psi}(x, t, \widehat{s}(x)) = 0$. Théorème 1.3.1 généraliser le résultat obtenu dans le lemme (6.5) en [13].notons que c'est la version fonctionnelle du théorème(3.1) in [8].

Theorem 1.3.1 *Soit $S_{\mathcal{F}} \subset \mathcal{F}$ un ensemble compact. Supposons que $(H1)$, $(H4)$, $(H5)iii)iv)$, $(H6)$ et $(H7)$ tient, alors nous avons que*

$$\sup_{x\in S_{\mathcal{F}}} |\widehat{F}(y/X = x) - F(y/X = x)| \overset{a.s}{\to} 0 \text{ as } n \to 0. \quad (1.3.4)$$

lemma 1.3.1 *Soit, $S_{\mathcal{F}} \subset \mathcal{F}$, un ensemble compact d'interieur non vide. Supposons que $(H1)$, $(H4)$ et $(H6)$, puis, pour $j = 0, 1$,nous avons*

a) pour tout $\epsilon > 0$,

$$\sup_{y\in\mathbb{R}} \sup_{x\in S_{\mathcal{F}}} \mathbb{P}\{|\upsilon_j(x) - \mathbb{E}[\upsilon_j(x)]| > \epsilon\}$$

$$\leq C \exp\left(-\frac{\epsilon^2 n\phi_x(h)}{2C'\|K\|_\infty^2\left(1 + \frac{2\epsilon}{C'\|K\|_\infty}\right)}\right) + Cnr^{-1}\left(\frac{r}{\epsilon}\right).$$

b) pour tout $\epsilon > 0$,

$$\mathbb{P}\{\sup_{x\in S_{\mathcal{F}}} \mathbb{P}|\upsilon_j(x) - \mathbb{E}[\upsilon_j(x)]| > \epsilon\} \leq N_\rho(S_{\mathcal{F}}) \exp\left(-\frac{\epsilon^2 n\phi_x(h)}{a_1(1 + \epsilon a_2)}\right)$$

$$+Cnr^{-1}\left(\frac{r}{\epsilon}\right).$$

c) il existe, $c > 2$, tel que, pour tout $\epsilon_0 > c$ et $n \geq n_0$, nous avons

$$\sup_{y\in\mathbb{R}} \mathbb{P}\left\{\theta_n^{-1} \sup_{x\in S_{\mathcal{F}}} \mathbb{P}|\upsilon_j(x) - \mathbb{E}[\upsilon_j(x)]| > \epsilon_0\right\}$$
$$\leq \exp\left(-\frac{\epsilon_0^2}{8(1+\epsilon_0)}\Gamma_{S_{\mathcal{F}}}\left(\frac{\log(n)}{n}\right)\right) + n(\log n)^{-2}\left(\frac{4C(\log n)^2}{n\epsilon}\right)^{a+1}.$$

1.3.2 Taux de convergence uniformes et forts

Théorème 1.3.2 étant donné les taux de convergence presque complets pour les estimateurs de la distribution conditionnelle empirique et pour les M-estimateurs locaux de la fonction de régression.

Theorem 1.3.2 *Soit, $S_{\mathcal{F}} \subset \mathcal{F}$ un ensemble compact. Supposons que $(H1)$, $(H4)$, $(H5iii)iv)$, $(H6)$, $(H7)$, $(H8)$, et $(H9)$ tient. Si, γ_1 defini dans $(H8)$, est tel que, $\gamma_1 < \frac{1}{2}$, de plus, supposons que, $\left(\frac{n}{\log(n)}\right)^{1-\gamma_1} \phi_x(h) \leq C$ tient. Soit $\widehat{\theta}(x)$ un estimateur robust avec la probabilité 1, ils existes des constantes réelles, $0 < T_1 < T_2$, telles que, $T_1 < \widehat{\theta}(x) < T_2$, pour tous$x \in S_{\mathcal{F}}$ et $n > n_0$.*

puis,

$$\sup_{x\in S_{\mathcal{F}}} \sup_{y\in\mathbb{R}} |\widehat{\theta}(x) - \theta(x)| = O_{a.co}\left(h^{\eta} + \sqrt{\frac{\Gamma_{S_{\mathcal{F}}}(\log(n)/n)}{n\phi_x(h)}}\right).$$

Theorem 1.3.3 *Soit $S_{\mathcal{F}} \cup \mathcal{F}$ un ensemble compact. Supposons que $(H1), (H4), (H5)$, $(H6), (H7)$ and $(H9)$ tient. En outre, supposons que σ et θ sont Lipschitz fonction de l'ordre γ_1 et γ_2 respectivement. Ensuite, si $\gamma = \min(\gamma_1, \gamma_2)$, tel que $\gamma > 1$, et supposer que*

$$h\left(\frac{n}{\log(n)}\right)^{1-\gamma} \leq C_{\gamma} \quad for\ all\ n \geq 1 \tag{1.3.5}$$

ou

$$\phi_x(h)\left(\frac{n}{\log(n)}\right)^{1-\gamma} \leq C_{\gamma} \quad for\ all\ n \geq 1 \quad for\ all\ n \geq 1 \tag{1.3.6}$$

est complet.

Soit, $\widehat{s}(x)$ un estimateur a l'échelle avec la probabilité 1, et $\theta(x)$ be la solution unique de $\Psi(x, a, s(x)) = 0$. Supposons qu'il existe des constantes réelle, $0 < T_1 < T_2$, telle que $T_1 < \widehat{s}(x) < T_2$ pour tous $x \in S_{\mathcal{F}}$, et $n > n_0$.

Alors, si $\widehat{\theta}(x)$ est une solution de $(2.2.3)$, tel que $\sup_{x\in S_{\mathcal{F}}} |\widehat{\theta}(x) - \theta(x)| \overset{a.so}{\rightarrow} 0$, nous avons

$$sup_{x\in S_{\mathcal{F}}} |\widehat{\theta}(x) - \theta(x)| = O_{a.co}\left(h^{\gamma} + \sqrt{\frac{\Gamma_{S_{\mathcal{F}}}(\log(n)/n)}{n\phi_x(h)}}\right).$$

Résultats : Normalité asymptotique

1.4 L'estimation récursive de la fonction d'hasard conditionnelle avec missing at random

Soit $(X_i, Y_i)_{i=1,\dots,n}$ une suite de processus ergodiques strictement stationnaires. Où X_i soit des valeurs dans l'espace semi-métrique $(\mathcal{F}, d)$ et Y_i sont des variables aléatoires à valeurs réelles. N_x désignera un voisinage fixe de x. Nous supposons que la version régulière de la probabilité conditionnelle de Y étant donné X existe. De plus,nous supposons que, pour tout $x \in N_x$ la fonction de distribution conditionnelle de Y étant donné $X = x$, l'estimateur récursif du noyau de la fonction de hasard conditionnelle $h^x(y)$ tel que

$$h^x(y) = \frac{f^x(y)}{1 - F^x(y)}, \quad \text{for } y \in \mathbb{R} \ \text{ and } \ F^x(y) < 1$$

est

$$\widehat{h}^x(y) = \frac{\widehat{f}^x(y)}{1 - \widehat{F}^x(y)}, \quad \forall y \in \mathbb{R}.$$

On définit l'estimateur a noyau récursif de la fonction de distribution conditionnelle par

$$\widehat{F}^x(y) = \frac{\sum_{i=1}^{n} K(a_i^{-1} d(x, X_i)) H(b_i^{-1}(y - Y_i))}{\sum_{i=1}^{n} K(a_i^{-1} d(x, X_i))}$$

L'estimateur $\widehat{f}^x(.)$ de noyau récursif de $f^x(.)$ est donné par

$$\widehat{f}^x(y) = \frac{\sum_{i=1}^{n} b_i^{-1} K(a_i^{-1} d(x, X_i)) H'(b_i^{-1}(y - Y_i))}{\sum_{i=1}^{n} K(a_i^{-1} d(x, X_i))}.$$

Où K est le noyau, H est une fonction strictement croissante, H' est la dérivée de H et a_i, b_i sont des suites de nombre réel positif tel que $\lim_{n \to +\infty} a_n = \lim_{n \to +\infty} b_n = 0$.

Cependant, dans le cas d'un aléatoire manquant pour la variable de réponse, un échantillon incomplet disponible de taille n à partir de (X, Y, δ) est $\{(X_i, Yi, \delta i), 1 \leq i \leq n\}$; où X_i est observé complètement, $\delta i = 1$ if Yi est observé, et $\delta i = 0$ sinon. pendant ce temps la variable aléatoire de Bernoulli δ est stisfaite de

$$\mathbb{P}(\delta = 1|X = x, Y = y) = \mathbb{P}(\delta = 1|X = x) = p(x)$$

Où $p(x)$ est un fonction opérateur appelé probabilité conditionnelle de la réponse d'observation étant donné le prédicteur et qui est souvent inconnu. ce mécanisme montre que δ et Y are conditionnellement indépendants étant donnée X.
Par conséquent, l'éstimateur de $F^x(.)$ et $f^x(.)$ est donné par :

$$\widehat{F}^x(y) = \frac{\displaystyle\sum_{i=1}^{n} \delta i K(a_i^{-1} d(x, X_i)) H(b_i^{-1}(y - Y_i))}{\displaystyle\sum_{i=1}^{n} \delta_i K(a_i^{-1} d(x, X_i))} \tag{1.4.1}$$

and

$$\widehat{f}^x(y) = \frac{\displaystyle\sum_{i=1}^{n} \delta i b_i^{-1} K(a_i^{-1} d(x, X_i)) H'(b_i^{-1}(y - Y_i))}{\displaystyle\sum_{i=1}^{n} \delta_i K(a_i^{-1} d(x, X_i))}. \tag{1.4.2}$$

Theorem 1.4.1 *Sous hypothèses (H1)-(H5), nous avons pour tous $x \in \mathcal{A}$*

$$\left(\frac{n\varphi_n(x)}{p(x)^2 \sigma_h^2(x, y)}\right)^{1/2} (\widehat{h}^x(y) - h^x(y)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad as\ n \to \infty \tag{1.4.3}$$

ou$\mathcal{A} = \{x, \quad \sigma_h^2(x, y) \neq 0\}$ and $\sigma_h^2(x, y) = \dfrac{\alpha_2 h^x(y)}{\alpha_1^2(1 - F^x(y))}$
avec $\alpha_1 = K(1) - \int_0^1 K'(s)\beta_x(s)ds$ and $\alpha_2 = K^2(1) - \int_0^1 (K^2(s))'\beta_x(s)ds.$

lemma 1.4.1 *Sous hypothèses de la théorème3.3.1, nous avons*

$$\left(\frac{n\varphi_n(x)}{p(x)^2 \sigma_F^2(x, y)}\right)^{1/2} (\widehat{F}^x(y) - F^x(y)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad as \quad n \to \infty \tag{1.4.4}$$

ou $\sigma_F^2(x, y) = \dfrac{p(x)\alpha_2 F^x(y)\big(1 - F^x(y)\big)}{\alpha_1^2}.$

lemma 1.4.2 *Sous hypothèses de la théorème 3.3.1, nous avons*

$$\left(\frac{n\varphi_n(x)}{p(x)^2 \sigma_f^2(x, y)}\right)^{1/2} (\widehat{f}^x(y) - f^x(y)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad as\ n \to \infty \tag{1.4.5}$$

ou $\sigma_f^2(x, y) = \dfrac{p(x)\alpha_2 f^x(y)}{\alpha_1^2}.$

1.5 Outils et propriétés

definition 1.5.1 *On dit que la suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ converge presque complètement (p.co.) vers la variable aléatoire X, si et seulement si*

$$\forall \epsilon > 0; \sum_{n=0}^{\infty} P(|X_n - X| > \epsilon) < \infty.$$

On l'écrit

$$X_n \xrightarrow{p.co.} X.$$

definition 1.5.2 *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires, $(u_n)_{n\in\mathbb{N}*}$ une suite de nombres réels positifs. On dit que $(X_n)_{n\in\mathbb{N}}$ converge presque complètement (p.co.) vers la variable aléatoire X avec la vitesse de convergence un si et seulement si*

$$\forall \epsilon > 0, \ \sum_{n=1}^{\infty} P(|X_n - X| > \epsilon u_n) < \infty.$$

On note

$$X_n - X = O(u_n). \quad p.co.$$

remarque 1.5.1 *cette convergence implique à la fois la convergence presque sûre et la convergence en probabilité.*

definition 1.5.3 *On dit qu'une suite de variables aléatoires $(X_n)_{n\geq 1}$ est une suite de différences de martingales par rapport à la tribu $\mathcal{F}_n$ s'il satisfait aux conditions suivantes :*

1. $\mathbb{E}(|X_n|) < \infty$.
2. $\mathbb{E}(X_n|\mathcal{F}_{n-1}) = 0$. *p.co.*

lemma 1.5.1 *"Inégalité de* **Hoeffding***"*
Soit $X_1, ..., X_n$ des variables aléatoires réelles centrées, indépendantes et de même loi (i, i, d) définies sur l'espace de probabilité, tels qu'il existe deux réels positifs θ_1 et θ_2 vérifiant $X_1 < \theta_1$ et $\mathbb{E}X_1^2 < \theta_2$ alors, pour tout $\varepsilon \in]0, \frac{\theta_1}{\theta_2}[$ on a

$$\mathbb{P}\left(n^{-1}\left|\sum_{i=1}^{n} X_i\right| \geq \varepsilon\right) \leq 2\exp\left(\frac{-n\varepsilon^2}{4\theta_2}\right).$$

lemma 1.5.2 *"Inégalité de type* **Fuk-Nagaev** *sous mélange algébrique"*
Soit $\{\Delta_i, i \in \mathbb{N}\}$ une famille de variables aléatoires à valeur dans $\mathbb{R}$ fortement mélangeantes, de coefficient de mélange algébriquement décroissant. On pose

$$S_n^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} |Cov(\Delta_i, \Delta_j)|$$

Si $\forall i,\ ||\Delta_i||_\infty < \infty$, *alors, pour tout* $\varepsilon > 0$ *et pour tout* $r > 1$, *on a :*

$$\mathbb{P}\left[\left|\sum_{k=1}^{n} \Delta_k\right| > 4\varepsilon\right] \leq 4\left(1 + \frac{\varepsilon^2}{r s_n^2}\right)^{-\frac{r}{2}} + 2ncr^{-1}\left(\frac{2r}{\varepsilon}\right)^{a+1}. \qquad (1.5.1)$$

Cette inégalité est en fait une extension au cadre de variables fortement mélangeantes de l'inégalité de Hoeffding.

Plan de la thèse

Ce travail est présenté essentiellement en quatre chapitres et ils organisées comme suit :

Le premier chapitre est consacré à le contexte bibliographique et représentative, où nous avons apporté une contribution au thème de l'estimation non paramétrique fonctionnelle et à l'estimation par la fonction de régression, et la fonction récursive. nous avons donc parler sur les données fonctionnelles, régression pour variables fonctionnelles, analyse non paramétrique robuste, fonction de hasard, et on a terminée par donnée quelque notion sur l'estimation récursive, Puis nous avons ontamées une présentation de cadre générale de la these, ou nous avons mentionné les résultats obtenus.

Le second chapitre, on a étudié la régression robuste dans lequel l'objectif principal de cet travail est de prouver la convergence presque complète (avec taux) pour l'estimateur proposé.

Le troisième chapitre traite de la normalité asymptotique de l'estimateur récursive fonctionnelle de la fonction de risque dans le cas des données ergodiques.

A la fin de cette thése on se propose une court conclusion générale.

Bibliographie

[1] Attouch, M., Laksaci, A., and Ould-Said, E.(2010). Asymptotic normality of a robust estimator of the regression function for functional time series data, J. of the Korean Statist. Soc., Vol. 39, No. 4, pp. 489-500.

[2] Attouch, M., Laksaci, A., and Ould Said, E. (2012). Strong uniform convergence rate of a robust estimator of the regression function for functional and dependent processes, Journal of Japan Statistical Society, Vol. 42, No. 3, 125-143.

[3] Attouch, M.K., Kaid, Z., and Louhab, H.(2019) : Asymptotic normality of a robust kernel estimator of the regression function for functional ergodic data : case of an unknown scale parameter, C. R. Math. Acad. Sci. Paris, Vol. 357, No. 5, pp. 478-481.

[4] Azzedine, N., Laksaci, A. and Ould-Said, E. (2008). On robust nonparametric regression estimation for a functional regressor, Statistics & Probability Letters, Vol. 78, No. 18, 3216-3221.

[5] Benhenni, K., Ferraty, F., Rachdi, M. and Vieu, P. (2007). Local smoothing regression with functional data, Comput. Statist., Vol.22, No. 3, 353-369.

[6] Boente, G. and Fraiman, R. (1989). Robust nonparametric regression estimation for dependent observations, Annals of Statistics, Vol. 17, No. 3, 1242-1256.

[7] Boente, G. and Fraiman, R. (1989). Nonparametric regression estimation, J. Multivariate Anal. Vol. 29, No. 2, 180-198.

[8] Boente, G. and Fraiman, R. (1991). Strong Uniform Convergence Rates for Some robust equivariant nonparametric regression estimates for mixing processes, International Statistical Institute, Vol. 59, No. 3, 355-372.

[9] Boente, G. and Vahnovanb, A. (2015). Strong convergence of robust equivariant nonparametric functional regression estimators, Statistic & Probability Letters, Vol. 100, 1-11.

[10] Chikobvu, D., and Sigauke, C. (2012). Regression-SARIMA modelling of daily peak electricity demand in South Africa, Journal of Energy in Southern Africa, Vol. 23, No. 3, 23-30.

[11] Collomb,G. (1982). Prédiction non paramétrique : Etude de l'erreur quadratique du préedictogramme, C. R.Acad. Sci. Paris Sér. I Math, Vol. 294, No. 4, 59-62.

[12] Ferraty, F., Laksaci, A. and Vieu, Ph. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models, Statistical Inference for Stochastic Processes, Vol. 9, No. 3, 47-76.

[13] Ferraty, F., Laksaci, A., Tadj, A. and Vieu, Ph. (2010). Rate of uniform consistency for nonparametric estimates with functional variables, Journal of statistical planning and inference, Vol. 140, No. 2, 335-352.

[14] Ferraty, F., Vieu, Ph., (2006). *Nonparametric functional data analysis.* Springer-Verlag Publisher, United States of America.

[15] Gheriballah, A. Laksaci, A. and Sekkal, S. (2013). Nonparametric M-regression for functional ergodic data, Journal Statistics et Probability Letters, Vol. 83, No.3, 902-908.

[16] Goia, A. May, C. and Fusai, G. (2010). Functional clustering and linear regression for peak load forecasting, International Journal of Forecasting, Vol. 26, No. 4, 700-711. Statistics, Vol. 35, No. 1, 73-101.

[17] Hyndman, Rob J., and Fan, S. (2010). Density Forecasting for Long-Term Peak Electricity Demand, IEEE Transaction on Power Systems, Vol. 25, No. 2, 1142-1153.

[18] Kolmogorov, A. N. (1956). Asymptotic characteristics of some completely bounded metric spaces, Doklady Akademii Nauk SSSR, Vol. 1, No. 8, 585-589.

[19] Kolmogorov, A. N. and Tikhomirov, V .M. (1959). ϵ-entropy and ϵ-capacity, Uspekhi Mat. Nauk, Vol. 14, 3-86. (English version in, Transl. Amer. Math. Soc. Transl. Ser, Vol. 2, No.1, 277-364 (1961).

[20] Masry, E. (2005). Nonparametric regression estimation for dependent functional data : asymptotic normality, Stoch. Proc. and their Appl., Vol.

[21] Rachdi M, Laksaci A, Demongeot J, Abdali A, Madani F. (2014). Theoretical and practical aspects of the quadratic error in the local linear estimation of the conditional density for functional data. Comput. Statist Data Anal., Vo.73, No. 2, 53-68.

[22] Ramsay, J. and Silverman, B. (2002). *Applied functional data analysis : methods and case studies.* Springer.

[23] Ramsay, J. and Silverman, B. (2005). *Functional data analysis.* Springer Series in Statistics.

[24] C. J. Stone. (2005). Nonparametric M-regression with free knot splines, J. Stat. Plan. Inf., Vol. 130, No. 1-2, 183-206.

[25] Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblements dépendants.* (In french). Mathématiques & Applications, Vol. 31, Springer Verlag, New-York.

Chapitre 2

Nonparametric M-regression with scale parameter for functional dependent data

[1]**Mebsout Mokhtaria,** [2,⋆]**Attouch Mohammed Kadi and** [3]**Fetitah Omar**

[1,2,3] Laboratoire de Statistique Processus Stochastiques
Djillali Liabes University
Sidi Bel Abbes, 22000
Algeria
[1]mokhtariamebsout@yahoo.com ; [3]fetitah-omar@hotmail.com

[2] Mathematic Department, King Khalid University
Abha, 61421
Saudi Arabia.
[2]attou_kadi@yahoo.fr ;

⋆Corresponding Author

Abstract

In this paper, we study the equivariant nonparametric robust regression estimation relationship between a functional dependent random covariable and a scalar response. We consider a new robust regression estimator when the scale parameter is unknown. The consistency result of the proposed estimator is studied, namely the uniform almost complete convergence (with rate). Thus, suitable topological considerations are needed, implying changes in the convergence rates, which are quantified by entropy considerations. The benefits of considering robust estimators are illustrated on two real data sets where the robust fit reveals the presence of influential outliers.

Keywords : Scale parameter ; Robust function ; Functional dependent data ; Nonparametric estimation

MSC 2010 No. : 62G35, 62G08, 62G05

2.1 Introduction

Studying the relationship between a random variable Y and a set of covariates X in comparison with usual regression methods is a very relevant topic and at the same time it considered a common problem in nonparametric statistics, and there are several ways to explain this relationship. In many applications, the covariates X can be seen as functions recorded over a period of time instead of finite-dimensional vectors, (see, [13]) for an extensive discussion on nonparametric statistics for functional data.

In this general framework, statistical models adapted to infinite dimensional data have been recently studied. We refer to [11], [23] and [13] for a description of different procedures for functional data. Linear nonparametric regression estimators in the functional setting, that is, estimators based on a weighted average of the response variables, have been considered by several authors such as [5] and [12] who also considered estimators of the conditional quantiles. The literature on robust proposals for nonparametric regression estimation is sparse. Motivated by its flexibility when data are affected by outliers, the robust regression was widely studied in nonparametric functional statistics. Indeed, It was firstly introduced by [4] who proved the almost-complete convergence of this model in the independent and identically distributed (i.i.d.) case. Since this work, several results on the nonparametric robust functional regression were realized (see, for instance, [1], [2], [3], [15], [9] and references therein for some key references on this topic). Notice that all these results are obtained when the scale parameter is supposed to be known.

In this sense, we extend some of the previous works in two directions. On the one hand, we generalize the proposal given in the Euclidean case by [7] to provide robust equivariant estimators for the regression function in the functional case, that is, in the case where the covariates are in an infinite-dimensional space. On the other side, we extend the proposal given in [4] to allow for an unknown scale, and heteroscedastic models are provided in [9]. The main goal of this paper is to study the uniform convergence of this nonparametric estimator with an unknown scale parameter when the explanatory variable X is valued in infinite dimension space and the observations $(X_n, Y_n)_{n\geq 1}$ are strongly mixing.

The paper is organized as follows. In Section 2, we state our notation and introduce the robust equivariant estimators. Section 3 contains the main results of this paper, namely, the uniform convergence consistency and uniform convergence rates over compact sets of the equivariant local M-estimators. In Section 4, we examine the performances of our proposed estimator with two real data sets applications. Proofs are relegated to the Appendix.

2.2 Basic definitions and notation

Consider $Z_i = (X_i, Y_i)_{i=1\cdots n}$ be n copies of random vector, identically distributed as (X, Y) and is valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a semi-metric space, d denoting the semi-metric. For any $x \in \mathcal{F}$, we consider a real-valued Borel function ψ, and stated the model of the covariation between X_i and Y_i. Our nonparametric regression function, denoted by θ_x is implicitly defined as a zero with respect to (w.r.t.) t of the following equation.

Let us define (X, Y) be a random element in $\mathcal{F} \times \mathbb{R}$ and let

$$\Psi(x, t, \sigma) = \mathbb{E}\left(\psi\left(\frac{Y-t}{\sigma}\right)/X = x\right), \tag{2.2.1}$$

where $\psi : \mathbb{R} \times \mathbb{R}$ is an odd, bounded and continuous function satisfying some regularity conditions to be stated below. In the following, we assume that the equation (2.2.1) allows θ as a unique solution(see for instance, [7]for sufficient conditions for existence and uniqueness of θ). In addition, our robustification method allows us to consider the functional nonparametric regression model with a scale of the error assumed to be unknown, where $\sigma(.)$ is a measure of spread for the conditional distribution of Y given $X = x$. We return to [24] for other examples of the function ψ.

The conditional scale measure can be taken as the conditional median of the absolute deviation from the conditional median, that is

$$s(x) = MED(|Y - m(x)|/X = x) = MAD_c(F_Y^x(.)), \tag{2.2.2}$$

where $m(x) = MED(Y/X = x)$ is the median of the conditional distribution.

Note that $s(x)$ which corresponds to a robust measure of the conditional scale, usually equals $\sigma(X)$ up to a multiplicative constant, when $\epsilon = \sigma(X)u$ with u independent of X. For instance, the median of the absolute deviation is usually calibrated so that $MAD(\phi) = 1$, where ϕ states for the distribution function of a standard normal random variable. In this case, when the errors u have a Gaussian distribution, we x have that $s(x) = MAD_c(F_Y^x(.)) = \sigma(x)$. To obtain estimators of $\theta(x)$ we plug into (2.2.1) an estimator of $F_Y^x(y)$, which will be taken as $\widehat{F}(y/X = x)$. Denote by $\widehat{s}(x)$ a robust estimator of the conditional scale, for instance, $\widehat{s}(x) = MAD_c\left(\widehat{F}(./X = x)\right)$, the scale measure defined in (2.2.2) evaluated in $\widehat{F}(y/X = x)$. With this notation, the robust nonparametric estimator of $\theta(x)$ is given by the solution $\widehat{\theta}(x)$ of $\widehat{\Psi}(x, t, \widehat{s}(x)) = 0$, where

$$\widehat{\Psi}(x, t, \widehat{\sigma}) = \int \psi\left(\frac{y-t}{\widehat{\sigma}}\right) d\widehat{F}(y/X = x) = \sum_{i=1}^{n} w_i(x)\psi\left(\frac{Y_i - t}{\widehat{\sigma}}\right), \tag{2.2.3}$$

where $w_i(x) = \dfrac{K\left(\frac{d(X_i,x)}{h_k}\right)}{\displaystyle\sum_{i=1}^{n} K\left(\frac{d(X_i,x)}{h_k}\right)}$.

2.3 Main results

Uniform convergence results and uniform convergence rates over compacts for the local M-estimators are derived under some general assumptions that are described below. From now on, $\xrightarrow{a.co}$ and a.co. stand for almost complete convergence while $\xrightarrow{a.s}$ stands for almost sure convergence.

Throughout this paper, when no confusion will be possible, we will denote by C and C' some strictly positive generic constants. x is a fixed point in $\mathcal{F}$ and $\mathcal{N}_x$ denotes a fixed neighborhood of x. We considerer $S_{\mathbb{R}} \subset \mathbb{R}$ and $S_{\mathcal{F}} \subset \mathcal{F}$ a compact subsets of non empty interior. For $r > 0$, let $B(x,r) =: \{x' \in \mathcal{F}/d(x',x) < r\}$.

In order to define the strong mixing property, introduce the following notations. Denote by $\mathcal{F}_1^k$ the σ-algebra generated by $(X_1,Y_1),\ldots,(X_k,Y_k)$ and $\mathcal{F}_{k+n}^{\infty}$ that generated by $(X_{k+n},Y_{k+n}),\ldots$.
Let define, for any $n \geq 1$,

$$\alpha(n) = \sup_{A\in\mathcal{F}_1^k} \sup_{B\in\mathcal{F}_{k+n}^{\infty}} \{|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|\}. \tag{2.3.1}$$

The process $(X_n,Y_n)_{n\geq 1}$ is said to be strongly mixing if

$$\lim_{n\to\infty} \alpha(n) = 0. \tag{2.3.2}$$

There exists many processes fulfilling the strong mixing property. We quote, here, the usual ARMA processes which are geometricaly strongly mixing, *i.e.*, there exists $\rho \in (0,1)$ and $a > 0$ such that, for any $n \geq 1$, $\alpha(n) \leq a\rho^n$ (see, *e.g.*, [25] for more detail).

Before giving the main asymptotic result, we need some assumptions.

(**H1**) Let denote by $\phi_x(h) = \mathbb{P}(X \in B(x,h)) = \mathbb{P}[X \in \{x' \in \varepsilon; d(x,x') < h\}]$, and we suppose that $\phi_x(h)$ is continuous, strictly increasing in a neighborhood of 0 and $\phi_x(0) = 0$.

(**H2**) The function Ψ is such that

i) The function $\Psi(x,t,\sigma)$ is of class $\mathcal{C}^2$ on $[\theta_x-\tau,\theta_x+\tau]$, $\tau > 0$.

ii) $\forall(a_1,a_2) \in [\theta_x-\tau,\theta_x+\tau]\times[\theta_x-\tau,\theta_x+\tau]$, $\forall(x_1,x_2) \in \mathcal{N}_x\times\mathcal{N}_x$,
$|\Psi(x_1,a_1,\sigma) - \Psi(x_2,a_2,\sigma)| \leq Cd^{\gamma_1}(x_1,x_2) + |a_1-a_2|^{\gamma_1}, \gamma_1,\ \gamma_2 > 0.$

iii) For each fixed $t \in [\theta_x-\tau,\theta_x+\tau]$, the function $\Psi(x,t,\sigma)$ is continuous at the point x.

(**H3**) **i)** ψ is a monotone function w.r.t. the second component.

ii) $\psi : \mathbb{R} \to \mathbb{R}$ is an odd, strictly convex and increasing function, continuously differentiable with bounded derivative ψ, such that $\xi(u) = u\psi'(u)$.

iii) For each fixed $t \in [\theta_x - \tau, \theta_x + \tau]$, $\mathbb{E}\left[\psi\left(\frac{Y-t}{\sigma}\right)/X\right] < C < \infty, a.s.$

(**H4**) The kernel K is a bounded nonnegative function with support [0,1), such that $|\int_0^1 K(u)du| < \infty$, and satisfies Lipschitz condition of order $\gamma_1 > 0$

$$\exists L_K < \infty, |K(u) - K(v)| \leq L_K|u - v|^{\gamma_1},$$

and

i) If $K(1) = 0$, K is a differentiable with derivative K' and

$$-\infty < \inf_{u\in[0,1]} K'(u) \leq \sup_{u\in[0,1]} K'(u) = \|K'\|_\infty < 0.$$

ii) If $K(1) > 0$, there exist two real constants $C_1, C_2, 0 < C_1 < C_2 < \infty$ such that

$$0 < C_1\mathbb{I}_{[0,1]} < K < C_2\mathbb{I}_{[0,1]} < \infty.$$

(**H5**) Let $S_\mathcal{F}$ be a compact of $\mathcal{F}$ such that :

i) $F(y/X = x)$ is symmetric around θ_x.

ii) $F(y/X = x)$ has a unique median $m(x)$.

iii) For each y fixed $F(y/X = x)$ is a uniformly continuous function of x in a neighborhood of $S_\mathcal{F}$.

iv) The following equicontinuity condition holds,

$$\forall\epsilon > 0, \exists\delta > 0; |u - v| < \delta \Rightarrow \sup_{x\in S_\mathcal{F}} |F(u/X = x) - F(v/X = x)| < \epsilon.$$

(**H6**) The functions ϕ_x and $\Gamma_{S_\mathcal{F}}$ are such that :

i) $\exists\ C > 0$ and $\exists\ \eta_0 > 0$ such that for all $\eta < \eta_0$, $\phi'_x(\eta) < C$. If $K(1) = 0$, the function ϕ_x satisfy the additional condition :

$\exists\, C > 0$, and $\exists\, \eta_0 > 0$, such that $\forall\, 0 < \eta < \eta_0$

$$\int_0^\eta \phi_x(u)du > C\eta\phi_x(\eta).$$

ii) For n large enough,

$$\frac{(\log(n))^2}{n\phi_x(h)} < \Gamma_{S_\mathcal{F}}\left(\frac{\log(n)}{n}\right) < \frac{n\phi_x(h)}{\log(n)}.$$

(**H7**) The sequence $h = h_n$ is such that $h_n \to 0$,

$$n\phi_x(h_n) \to \infty \text{ and } \frac{n\phi_x(h_n)}{\log(n)} \text{ as } n \to \infty.$$

(**H8**) **i)** $\forall (y_1, y_2) \in S_{\mathbb{R}} \times S_{\mathbb{R}}, \forall (x_1, x_2) \in S_{\mathcal{F}} \times S_{\mathcal{F}}$,

$$|F(y_1/X = x_1) - F(y_2/X = x_2)| \leq C\left(d(x_1, x_2)^{\gamma_1} + C|y_1 - y_2|^{\gamma_2}\right).$$

ii) The function $F(y/X = x)$ is uniformly Lipschitz in a neighborhood $S_{\mathcal{F}}$ of $\mathcal{F}$, and there exists a constant $C > 0$ and $\gamma_1 > 0$, such that, for $x_1, x_2 \in S_{\mathcal{F}}$,

$$\sup_{y\in\mathbb{R}} |F(y/X = x_1) - F(y/X = x_2)| \leq C d^{\gamma_1}(x_1, x_2).$$

(**H9**) The Kolmogov's ϵ-entropy of $S_{\mathcal{F}}$, satisfy

$$\sum_{i=1}^{\infty} n \exp\left\{(1-\beta)\Gamma_{S_{\mathcal{F}}}\left(\frac{\log(n)}{n}\right)\right\} < \infty \qquad \text{for some } \beta > 1. \quad (2.3.3)$$

(**H10**) The sequence $(X_i, Y_i)_{i\in\mathbb{N}}$ satisfies :

i) $\exists a > 0, \exists c > 0 : \forall n \in \mathbb{N}, \alpha(n) \leq cn^{-a}$.

ii) $\begin{cases} \forall i \neq j, \forall t \in [\theta_x - \tau, \theta_x + \tau], \\ \mathbb{E}\left[\psi\left(\frac{Y_i - t}{\sigma}\right)\psi\left(\frac{Y_j - t}{\sigma}\right) \mid X_i, X_j\right] \leq C < \infty, \\ \mathbb{P}\left((X_i, X_j) \in B(x, r) \times B(x, r)\right) = \varphi_x(r) > 0. \end{cases}$

iii) There exists $\eta > 0$, such that

$$Cn^{2(2-a)/(a+1)+\eta} \leq \max\left(\phi_x^2(r), \varphi_x(r)\right) \leq n^{2a/(2-a)p},$$

where $a > \max\left(5/2, \frac{4p+1+\sqrt{24p+1}}{p-1}\right)$.

The quantity $\Gamma_S(\varepsilon) = \log(N_\varepsilon(S))$, where $N_\varepsilon(S)$ denote the minimal number of open balls in $\mathcal{F}$ of radius ε which is necessary to cover $S_{\mathcal{F}}$. This concept was introduced by Kolmogorov in the mid-1950s (see, [19]) and it represents a measure of the complexity of a set, in sense that, high entropy means that much information is needed to describe an element with an accuracy ε.

2.3.1 Uniform strong convergence results

In this section, we will obtain uniform convergence results over compact sets of the regression estimator $\widehat{\theta}$ defined as a solution of $\widehat{\Psi}(x, t, \widehat{s}(x)) = 0$. Theorem 2.3.1 generalize the result obtained in Lemma (6.5) in [13]. Note that it is the functional version of Theorem (3.1) in [8].

Theorem 2.3.1 *Let $S_{\mathcal{F}} \subset \mathcal{F}$ be a compact set. Assume that $(H1), (H4)$, $(H5)iii)iv), (H6)$ and $(H7)$ holds, then we have that*

$$\sup_{x \in S_{\mathcal{F}}} |\widehat{F}(y/X = x) - F(y/X = x)| \stackrel{a.s}{\to} 0 \text{ as } n \to 0. \tag{2.3.4}$$

proof 2.3.1 *The proofs are analogous to the one given in [9]. Moreover, to prove Theorem 2.3.1 and 2.3.2, we need to fixing some notation.*

Given fixed $y \in \mathbb{R}$, we denote by $W_i^j = \mathbb{I}_{(-\infty,y]}(Y_i)$, and for $j = 0, 1$

$$\tilde{\upsilon}_j(x) = \frac{1}{n} \sum_{i=1}^{n} W_i^j \frac{K_i(x)}{\mathbb{E}K_1(x)}, \ \forall n \geq 1, \tag{2.3.5}$$

$$\upsilon_j(x) = \frac{1}{n} \sum_{i=1}^{n} W_i^j \frac{K_i(x)}{\phi_x(h)}, \tag{2.3.6}$$

with

$$K_i(x) = K\left(\frac{d(x, X_i)}{h}\right),$$

hence

$$\widehat{F}(y/X = x) = \frac{\tilde{\upsilon}_1(x)}{\tilde{\upsilon}_0(x)}.$$

As in [11], we have the following bounds

$$\sup_{x \in S_{\mathcal{F}}} |\widehat{F}(y|X = x) - F(y|X = x)| \leq \frac{1}{\inf_{x \in S_{\mathcal{F}}} \tilde{\upsilon}_0(t)} \sup_{x \in S_{\mathcal{F}}} |\tilde{\upsilon}_1(x) - \mathbb{E}(\tilde{\upsilon}_1(x))|$$

$$+ \frac{1}{\inf_{x \in S_{\mathcal{F}}} \tilde{\upsilon}_0(t)} \sup_{x \in S_{\mathcal{F}}} |\tilde{\upsilon}_0(x) - \mathbb{E}(\tilde{\upsilon}_0(x))|$$

$$+ \frac{1}{\inf_{x \in S_{\mathcal{F}}} \tilde{\upsilon}_0(x)} \sup_{x \in S_{\mathcal{F}}} |\mathbb{E}(\tilde{\upsilon}_1(x)) - F(y/X = x)\mathbb{E}(\tilde{\upsilon}_0(x))|.$$

To prove 2.3.4, we need to show that, $\forall y \in \mathbb{R}$,

$$\sup_{x \in S_{\mathcal{F}}} |\tilde{\upsilon}_j(x) - \mathbb{E}(\tilde{\upsilon}_j(x))| \stackrel{a.co}{\to} 0, \tag{2.3.7}$$

$$\sup_{x \in S_{\mathcal{F}}} |\mathbb{E}(\tilde{\upsilon}_1(x)) - F(y/X = x)\mathbb{E}(\tilde{\upsilon}_0(x))| \to 0, \tag{2.3.8}$$

and for some $a > 0$, we have

$$\sum_{n \geq 1} \mathbb{P}\left(\inf_{x \in S_{\mathcal{F}}} \tilde{\upsilon}_0(x) < a\right) < \infty. \tag{2.3.9}$$

Note that Lemmas (4.3) and (4.4) in [13], the assumptions $(H1), (H4)$ and $(H6)$, imply that, there exists a constants such that

$$\forall x \in S_{\mathcal{F}}, \qquad \exists\, 0 < C < C' < \infty, \quad C\phi_x(h) < \mathbb{E}\left[K_1(x)\right] < C'\phi_x(h). \tag{2.3.10}$$

Then, if $\tilde{C} = 1/C$, we have

$$|\tilde{\upsilon}_j(x) - \mathbb{E}(\tilde{\upsilon}_j(x))| \leq \tilde{C}|\upsilon_j(x) - \mathbb{E}(\upsilon_j(x))|.$$

Therefore, to obtain (2.3.7), it will be enough to show that

$$\sup_{x \in S_{\mathcal{F}}} |\upsilon_j(x) - \mathbb{E}(\upsilon_j(x))| \overset{a.co}{\to} 0. \tag{2.3.11}$$

Note that, $\mathbb{E}[\upsilon_0(x)] = 1$, and using the fact that

$$\inf_{x \in S_{\mathcal{F}}} \tilde{\upsilon}_0(x) \geq \inf_{x \in S_{\mathcal{F}}} \mathbb{E}[\upsilon_0(x)] - \sup_{x \in S_{\mathcal{F}}} |\tilde{\upsilon}_0(x) - \mathbb{E}[\tilde{\upsilon}_0(x)]|,$$

we can deduce (2.3.11) *and* (2.3.9) *by using the following Lemmas.*

lemma 2.3.1 *Let, $S_{\mathcal{F}} \subset \mathcal{F}$, be a compact set of non empty interior. Assume $(H1)$, $(H4)$ and $(H6)$, then, for $j = 0, 1$, we have*

a) For any $\epsilon > 0$,

$$\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{F}}} \mathbb{P}\{|\upsilon_j(x) - \mathbb{E}[\upsilon_j(x)]| > \epsilon\} \leq$$

$$C \exp\left(-\frac{\epsilon^2 n \phi_x(h)}{2C'\|K\|_\infty^2 \left(1 + \frac{2\epsilon}{C'\|K\|_\infty}\right)}\right) + Cnr^{-1}\left(\frac{r}{\epsilon}\right).$$

b) For any $\epsilon > 0$,

$$\mathbb{P}\{\sup_{x \in S_{\mathcal{F}}} \mathbb{P}|\upsilon_j(x) - \mathbb{E}[\upsilon_j(x)]| > \epsilon\} \leq N_\rho(S_{\mathcal{F}}) \exp\left(-\frac{\epsilon^2 n \phi_x(h)}{a_1(1 + \epsilon a_2)}\right)$$

$$+ Cnr^{-1}\left(\frac{r}{\epsilon}\right).$$

c) There exists, $c > 2$, such that, for any $\epsilon_0 > c$ and $n \geq n_0$, we have $\sup_{y \in \mathbb{R}} \mathbb{P}\left\{\theta_n^{-1} \sup_{x \in S_{\mathcal{F}}} \mathbb{P}|\upsilon_j(x) - \mathbb{E}[\upsilon_j(x)]| > \epsilon_0\right\} \leq$

$$\exp\left(-\frac{\epsilon_0^2}{8(1+\epsilon_0)} \Gamma_{S_{\mathcal{F}}}\left(\frac{\log(n)}{n}\right)\right) + n(\log n)^{-2}\left(\frac{4C(\log n)^2}{n\epsilon}\right)^{a+1}.$$

proof 2.3.2 *Proof of Part a).*
Let denote by, $Z_i = W_i^j \Pi_i - \mathbb{E}(W_i^j \Pi_i)$, where $|W_i^j| \leq 1$ and $\Pi_i = K_i(x)/\phi_x(h)$, the kernel K is bounded, and let $C_1 = \|K\|_\infty$, such that

$$D_n(x) = \upsilon_j(x) - \mathbb{E}[\upsilon_j(x)]. \tag{2.3.12}$$

Thus,

$$D_n(x) = \frac{1}{n\phi_x(h)} \sum_{i=1}^n \left[W_i^j K_i(x) - \mathbb{E}\left[W_i^j K_i(x)\right]\right], \quad \text{for,} \quad j = 0, 1.$$

We have

$$\frac{1}{n\phi_x(h)}\left[W_i^j K_i(x) - \mathbb{E}\left[W_i^j K_i(x)\right]\right] \leq \frac{1}{n\phi_x(h)}\left[K_i(x) - \mathbb{E}\left[K_i(x)\right]\right],$$

and denote by

$$\Delta_i(x) = \frac{1}{\phi_x(h)}\left[K_i(x) - \mathbb{E}\left[K_i(x)\right]\right].$$

We need to evaluate the variance term $S_n^2(x)$*,*

$$S_n^2(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} cov|\Delta_i(x), \Delta_j(x)| =: S_n^{2*}(x) + nVar[\Delta_1(x)],$$

where

$$S_n^{2*}(x) = \sum_{i=1}^{n}\sum_{j\neq i} cov|\Delta_i(x), \Delta_j(x)|.$$

Next, we evaluate the asymptotic behavior of $S_n^{2*}(x)$*.*

Following [20], we define the sets

$$L_1 = \{(i,j) \quad \text{such that} \quad 1 \leq |i-j| \leq m_n\},$$

and

$$L_2 = \{(i,j) \quad \text{such that} \quad m_n + 1 \leq |i-j| \leq n-1\},$$

where, $m_n \longrightarrow \infty$*, as* $n \to \infty$ *.*

Let, E_1 *and* E_2 *, be the sums of the covariances over* L_1 *and* L_2 *, respectively. Then*

$$S_n^{2*}(x) = \underbrace{\sum_{L_1} \left|cov(\Delta_i(x), \Delta_j(x))\right|}_{J_1} + \underbrace{\sum_{L_2} \left|cov(\Delta_i(x), \Delta_j(x))\right|}_{J_2}.$$

We stared by, J_1

$$J_1 = \sum_{|i-j|\leq m_n} |cov(\Delta_i(x), \Delta_j(x))|,$$

where

$$cov(\Delta_i(x), \Delta_j(x)) = \frac{1}{\phi_x^2(h)} cov(K_i(x) - \mathbb{E}[K_i(x)], K_j(x) - \mathbb{E}[K_j(x)]).$$

So

$$\begin{aligned} J_1 &\leq \frac{1}{\phi_x^2(h)} \sum_{L_2} |\mathbb{E}[K_i(x)K_j(x)] - \mathbb{E}^2[K_1(x)]| \\ &\leq \tilde{C} n m_n \phi_x(h)^{\frac{1-a}{a}} n^{\frac{-1}{a}}. \end{aligned}$$

For J_2*, we have,*

$$J_2 = \sum_{|i-j|\geq m_n} |cov(\Delta_i(x), \Delta_j(x))|.$$

We use Davydov-Rio's inequality ([25], p. 87) for mixing processes, to leads, for all $i \neq j$ *,*

$$cov|\Delta_i(x), \Delta_j(x)| \leq 4\alpha(|i-j|).$$

Finally,

$$J_2 \leq Cn^2 m_n^{-a}.$$

So

$$\sum_{i\neq j} |cov(\Delta_i(x), \Delta_j(x))| = Cnm_n \phi_x(h)^{\frac{1-a}{a}} n^{\frac{-1}{a}} + \tilde{C}n^2 m_n^{-a}.$$

choosing $m_n = \left(\frac{\phi_x(h)}{n}\right)^{-\frac{1}{a}}$*, then*

$$\begin{aligned} J_1 + J_2 &= Cn\phi_x(h)^{-\frac{1}{a}} n^{\frac{1}{a}} \phi_x(h)^{\frac{a+1}{a}} n^{\frac{-1}{a}} + \tilde{C}n^2 \left(\left(\frac{\phi_x(x)}{n}\right)^{-\frac{1}{a}}\right)^{-a} \\ &= Cn\phi_x^{-1}(h) + \frac{\tilde{C}n}{\phi_x(h)}. \end{aligned}$$

Putting, $C = 2C_2/n$ *and* $\tilde{C} = 4C_1\epsilon/n\phi_x^2(h)$*, we conclude that*

$$S_n^2(x) \leq \frac{2C_2}{\phi_x(h)} \left(1 + \frac{4\epsilon C_1/\phi_x(h)}{2C_2/\phi_x(h)}\right).$$

Then, Applying Fuk-Nagaev's inequality, (see, [13] p.237.Proposition A.11.), we can get

$$\begin{aligned} \mathbb{P}\{|\upsilon_j(x) - \mathbb{E}[\upsilon_j(x)]| > \epsilon\} &= \mathbb{P}(|\frac{1}{n}\sum_{i=1}^{n} Z_i| > \epsilon) \\ &= \leq C\left\{exp\left(-\frac{\epsilon^2 n\phi_x(h)}{2C_2\left(1 + \frac{2\epsilon}{C'C_1}\right)}\right) + nr^{-1}\left(\frac{r}{\epsilon}\right)\right\}. \end{aligned}$$

Proof of Part (b).
Let denote by, $\rho = \rho_n > 0$*, a numeric sequence such that,* $\rho_n \to 0$*, consider a covering of* $S_{\mathcal{F}}$ *by balls of radius* ρ *, i.e.,* $S_{\mathcal{F}} \subset \bigcup_{k=1}^{l} B(x_k, \rho)$*, where* $l = N_\rho(S_{\mathcal{F}})$*.*

Let, $D_n(x) = \upsilon_j(x) - \mathbb{E}\upsilon_j(x)$*, where* $\upsilon_j(x) = (1/n)\sum_{i=1}^{n} W_i^j K_i(x)/\phi_x(h)$*, and for all* $x \in x \in B(x_k, \rho)$*, consider*

$$k(x) = \arg \min_{k\in 1,2........N_\rho(S_{\mathcal{F}})} d(x - x_k),$$

then

$$D_n(x) = \tilde{D}_n(x) + D_n(x_k).$$

Therefore,

$$\sup_{x \in S_{\mathcal{F}}} |\upsilon_j(x) - \mathbb{E}\left[\upsilon_j(x)\right]| \leq \max_{1\leq k\leq l} |D_n(x_k)| + \max_{1\leq k\leq l} \sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)} |\tilde{D}_n(x)|,$$

it entail that

$$\mathbb{P}\left(\sup_{x \in S_{\mathcal{F}}} |\upsilon_j(x) - \mathbb{E}\left[\upsilon_j(x)\right]| > \epsilon\right) \leq \lambda_n + \delta_n,$$

where $\delta_n = \mathbb{P}\left(\max_{1\leq k\leq l} |D_n(x_k)| > \frac{\epsilon}{2}\right)$ and $\lambda_n = \mathbb{P}\left(\max_{1\leq k\leq l} \sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)} |\tilde{D}_n(x)| > \frac{\epsilon}{2}\right)$.

Using the obtained result in part a), leads to

$$\lambda_n \leq \sum_{k=1}^{l} \mathbb{P}\left(|D_n(x_k)| > \frac{\epsilon}{2}\right) \leq l \sup_{x\in S_{\mathcal{F}}} \mathbb{P}\left(|D_n(x)| > \frac{\epsilon}{2}\right)$$

$$\leq l\left\{\exp\left(-\frac{\epsilon^2 n\phi_x(h)}{8C_2\left(1+C_3\frac{\epsilon}{2}\right)}\right) + nr^{-1}\left(\frac{r}{\epsilon}\right)\right\},$$

where, $C_3 = 2/C'\|K\|_\infty$ *and* $C_2 = C'\|K\|_\infty^2$.

Then, the proof of the part b) can be obtained by using the following inequality

$$\delta_n \leq \sum_{k=1}^{l} \mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)} |\tilde{D}_n(x)| > \frac{\epsilon}{2}\right) \leq$$

$$l \max_{1\leq k\leq l} \mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)} |\tilde{D}_n(x)| > \frac{\epsilon}{2}\right).$$

We consider two cases $K(1) = 0$ *and* $K(1) > 0$.

We begin by $K(1) = 0$, *where* K *is Lipschitz of order one in* $[0,1]$, *we have that*

$$|\tilde{D}_n(x)| = \left|\frac{1}{n\phi_x(h)} \sum_{i=1}^{n} \left(W_i^j K_i(x) - W_i^j K_i(x_k)\right)\right.$$

$$\left.-\mathbb{E}\left[\frac{1}{n\phi_x(h)} \sum_{i=1}^{n} \left(W_i^j K_i(x) - W_i^j K_i(x_k)\right)\right]\right|$$

$$\leq \tfrac{1}{n\phi_x(h)} \textstyle\sum_{i=1}^{n} \left\{|K_i(x) - K_i(x_k)| + \mathbb{E}|K_i(x) - K_i(x_k)|\right\},$$

when

$$\frac{1}{\phi_x(h)} \sum_{i=1}^{n} |K_i(x_k) - K_i(x_k)| = \frac{1}{\phi_x(h)} \sum_{i=1}^{n} |K_i(x_k) - K_i(x_k)|\, \mathbb{I}_{B(x,h)\cup B(x_k,h)}(X_i).$$

Thus, we conclude,

$$\sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)}|\tilde{D}_n(x)| \leq \frac{C\rho}{h\phi_x(h)}\sup_{x\in S_{\mathcal{F}}}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{B(x_k,h)\cup B(x_k,h)}(X_i)\right.$$
$$\left.+\mathbb{E}(\mathbb{I}_{B(x_k,h)\cup B(x_k,h)}(X_1))\right\}$$
$$\leq \frac{C\rho}{h\phi_x(h)}\sup_{x\in S_{\mathcal{F}}}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{B(x_k,h+\rho)}(X_i)+\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\mathbb{I}_{B(x_k,h+\rho)}(X_i))\right\}$$
$$\leq \frac{C\rho}{h\phi_x(h)}\sup_{x\in S_{\mathcal{F}}}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{I}_{B(x_k,h+\rho)}(X_i)+\mathbb{E}\mathbb{I}_{B(x_k,h+\rho)}(X_1)\right)\right\}. \tag{2.3.13}$$

We denote by, $Z_i=\frac{\rho}{h\phi_x(h)}\mathbb{I}_{B(x_k,h+\rho)}(X_i)$, *and suppose that*

$$\phi_x(h+\rho)\leq\phi_x(h)+C\rho\leq 2\phi_x(h). \tag{2.3.14}$$

Therefore,

$$\frac{\rho}{h\phi_x(h)}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{B(x_k,h+\rho)}(X_i)+\mathbb{E}(\mathbb{I}_{B(x_k,h+\rho)}(X_i))\right\}\leq\frac{1}{n}\sum_{i=1}^{n}|Z_i-\mathbb{E}[Z_i]|,$$

we can deduce that

$$\mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)}|\tilde{D}_n(x)|>\frac{\epsilon}{2}\right)\leq\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}|Z_i-\mathbb{E}[Z_i]|>\frac{\epsilon}{4C}\right).$$

Moreover, we apply Fuck-Nagaev's exponential inequality, (Proposition A.11 in [13]), to get

$$\mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)}|\tilde{D}_n(x)|>\frac{\epsilon}{2}\right)\leq C(A_1(x)+A_2(x)), \tag{2.3.15}$$

where

$$A_1(x)=\left(1+\frac{\epsilon^2}{rS_n^2}\right)^{\frac{-r}{2}}\qquad\text{and}\qquad A_2(x)=nr^{-1}\left(\frac{r}{\epsilon}\right)^{a+1}. \tag{2.3.16}$$

Firstly, we must calculate the term

$$S_n^2(x)=\sum_{i=1}^{n}\sum_{j=1}^{n}|cov(\Lambda_i(x),\Lambda_j(x))|=S_n^{2*}(x)+nVar(\Lambda_1(x)),$$

where

$$S_n^{2*}=\sum_{i\neq j}|cov(\Lambda_i(x),\Lambda_j(x))|,$$

and $\Lambda_i = \mathbb{I}_{B(x_k,h+\rho)}(X_i) + \mathbb{E}\mathbb{I}_{B(x_k,h+\rho)}(X_1)$.

Next, we evaluate the asymptotic behavior of $S_n^{2*}(x)$.

Following the decomposition used in [20], we define the sets

$$F_1 = \{(i,j) \quad \text{such that} \quad 1 \leq |i-j| \leq \nu_n\},$$

and

$$F_2 = \{(i,j) \quad \text{such that} \quad \nu_n + 1 \leq |i-j| \leq n-1\},$$

where $\nu_n \longrightarrow \infty$, *as* $n \to \infty$. *Let,* Γ_1 *and* Γ_2, *be the sums of covariances over* F_1 *and* F_2, *respectively.*

Hense,

$$S_n^{2*}(x) = \underbrace{\sum_{F_1} |cov(\Lambda_i(x), \Lambda_j(x))|}_{\Gamma_1} + \underbrace{\sum_{F_2} |cov(\Lambda_i(x), \Lambda_j(x))|}_{\Gamma_2}.$$

We stared by evaluate the term Γ_1, *note that,*

$$\begin{aligned} cov(\Lambda_i(x), \Lambda_j(x)) &= cov(\mathbb{I}_{B(x_k,h+\rho)}(X_i) + \mathbb{E}\mathbb{I}_{B(x_k,h+\rho)}(X_j), \\ &\qquad \mathbb{I}_{B(x_k,h+\rho)}(X_j) + \mathbb{E}\mathbb{I}_{B(x_k,h+\rho)}(X_j)) \\ &= \mathbb{E}[\mathbb{I}_{B(x_k,h+\rho)}(X_i)\mathbb{I}_{B(x_k,h+\rho)}(X_j)] + \mathbb{E}[\mathbb{I}_{B(x_k,h+\rho)}(X_i)]\mathbb{E}[\mathbb{I}_{B(x_k,h+\rho)}(X_j)]. \end{aligned}$$

Thus,

$$\begin{aligned} |cov(\Lambda_i(x), \Lambda_j(x))| &\leq C_1\mathbb{E}\left(\mathbb{I}_{B(x_k,h+\rho)\times B(x_k,h+\rho)}(X_i, X_j)\right) \\ &+ C_1\mathbb{E}[\mathbb{I}_{B(x_k,h+\rho)}(X_i)]\mathbb{E}[\mathbb{I}_{B(x_k,h+\rho)}(X_j)] \\ &\leq C_1\mathbb{P}\left((X_i, X_j) \in B(x_k, h+\rho) \times B(x_k, h+\rho)\right) \\ &+ C_1\mathbb{P}\left(X_i \in B(x_k, h+\rho)\right)\mathbb{P}\left(X_j \in B(x_k, h+\rho)\right). \end{aligned} \tag{2.3.17}$$

Under the assumptions $(H2)$, $(H4)$ *and* $(H10-i)$, *we have*

$$\begin{aligned} \Gamma_1 &\leq Cn\nu_n\phi_x(h+\rho)\left(\left(\frac{\phi_x(h+\rho)}{n}\right)^{\frac{1}{a}} + \phi_x(h+\rho)\right) \\ &\leq Cn\nu_n\phi_x(h+\rho)^{\frac{a+1}{a}} n^{-\frac{1}{a}}. \end{aligned}$$

For F_2, *we can write*

$$\Gamma_2 = \sum_{|i-j|\geq\nu_n} |cov(\Lambda_i(x), \Lambda_j(x))|,$$

by Davydov-Rio's inequality ([25], p. 87) for mixing processes, we have, for all $i \neq j$,

$$cov|\Lambda_i(x), \Lambda_j(x))| \leq 4\alpha(|i-j|).$$

Then,

$$\Gamma_2 = C \sum_{|i-j| \geq \nu_n} \alpha(|i-j|),$$

and by α-mixing condition $(H10, i)$, we have

$$\sum_{\Gamma_2} cov|\Lambda_i(x), \Lambda_j(x))| \leq C_1 n^2 \alpha(\nu_n) \leq C_1 n^2 \nu_n^{-a}.$$

We put, $\nu_n = \left(\frac{\phi_x(h+\rho)}{n}\right)^{\frac{-1}{a}}$, we obtain

$$\Gamma_1 + \Gamma_2 \leq \bar{C} n \phi_x(h+\rho)^{\frac{a+1}{a}} n^{\frac{1}{a}} n^{\frac{-1}{a}} + \acute{C} n^2 \left(\left(\frac{\phi_x(h+\rho)}{n}\right)^{\frac{-1}{a}}\right)^{-a},$$

where, $\bar{C} = C \dfrac{\rho^2}{h^2 n \phi_x(h)^2}$ and $\acute{C} = \dfrac{\rho}{nh}\epsilon$.

When we replace ν_n, and we use (2.3.14), we can deduce that

$$\Gamma_1 + \Gamma_2 \leq \frac{\rho^2 \phi_x(h+\rho)}{h^2 \phi_x(h)^2} + \frac{C\epsilon\rho\phi_x(h+\rho)}{h\phi_x(h)^2} \leq C\left(2\frac{\rho^2}{h^2\phi_x(h)} + 2\epsilon\frac{\rho}{h\phi_x(h)}\right).$$

Finally,

$$S_n^2(x) = O\left(2C\left(\frac{\rho^2}{h^2\phi_x(h)} + \epsilon\frac{\rho}{h\phi_x(h)}\right)\right).$$

We use (2.3.16), to conclude that

$$\begin{aligned} A_1(x) &\leq \left(1 + \frac{\epsilon^2}{r2Cn\left(2\frac{\rho^2}{h^2\phi_x(h)} + 2\epsilon\frac{\rho}{h\phi_x(h)}\right)}\right)^{\frac{-r}{2}} \\ &\leq \exp\left(-\frac{\epsilon^2}{2C\left(2\frac{\rho^2}{h^2\phi_x(h)} + 2\epsilon\frac{\rho}{h\phi_x(h)}\right)}\right), \end{aligned}$$

we put $r = (\log n)^2$, thus,

$$A_2(x) = n(\log n)^{-2}\left(\frac{(\log n)^2}{\epsilon}\right)^{a+1}.$$

Moreover, we use $A_1(x)$ and $A_2(x)$, to get

$$\mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)} |\tilde{D}_n(x)| > \frac{\epsilon}{2}\right) \leq \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} |Z_i - \mathbb{E}[Z_i]| > \frac{\epsilon}{4C}\right)$$

$$\leq exp\left(-\epsilon^2 n\phi_x(h)\frac{1}{32C^3\left(2\frac{\rho^2}{h^2}+\frac{\epsilon}{2C}\frac{\rho}{h}\right)}\right)+n(\log n)^{-2}\left(\frac{4C(\log n)^2}{n\epsilon}\right)^{a+1}. \tag{2.3.18}$$

Then, using the fact that, $\dfrac{\rho}{h}\to 0$, *then, if* $n\geq n_0$, *we have*
$\dfrac{\rho}{h}<\left(\dfrac{1}{4}\right)\left(\dfrac{1}{2\parallel K\parallel_\infty}\right)$, *and* $\dfrac{2\rho^2}{h^2}<\dfrac{1}{4}$.

Thus,

$$32C^3\left(2\frac{\rho^2}{h^2}+\frac{\epsilon n}{2C}\frac{\rho}{h}\right)+n(\log n)^{-2}\left(\frac{4C(\log n)^2}{n\epsilon}\right)^{a+1}\leq$$

$$8\left(1+\epsilon\frac{1}{\parallel K\parallel_\infty}\right)+n(\log n)^{-2}\left(\frac{4C(\log n)^2}{n\epsilon}\right)^{a+1}.$$

Then, the case $K(1)=0$ *is achieved.*

The second case is, $K(1)>0$, *let define,*

$$\tilde{\tilde{D}}_n(x)=\frac{1}{n}\sum_{i=1}^{n}\left(W_j^i\frac{K_i(x)}{\phi_x(h)}-W_j^i\frac{K_i(x_k)}{\phi_x(h)}\right).$$

Then,
$|\tilde{D}_n(x)|\leq|\tilde{\tilde{D}}_n(x)|+\mathbb{E}|\tilde{\tilde{D}}_n(x)|$. *As in Lemma 8 of [13], if* $x\in S_{\mathcal{F}}\cap B(x_k,\rho)$, *we have*

$$\begin{aligned}
|\tilde{\tilde{D}}_n(x)| &\leq \frac{1}{n\phi_x(h)}\left\{\sum_{i=1}^{n}|K_i(x)-K_i(x_k)|\mathbb{I}_{B(x_k,\rho)\cap B(x,\rho)}(X_i)\right.\\
&+ \sum_{i=1}^{n}|K_i(x)|\mathbb{I}_{B(x_k,\rho)^c\cap B(x,\rho)}(X_i)\\
&+ \left.\sum_{i=1}^{n}|K_i(x)|\mathbb{I}_{B(x_k,\rho)\cap B(x,\rho)^c}(X_i)\right\}\\
&\leq C\frac{\rho}{h\phi_x(h)}\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{B(x_k,\rho)\cap B(x,\rho)}(X_i)
\end{aligned}$$

$$\begin{aligned}
&+ \quad \| K \|_\infty \frac{1}{\phi_x(h)} \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(x_k,\rho)^c \cap B(x,\rho)}(X_i) \\
&+ \quad \| K \|_\infty \frac{1}{\phi_x(h)} \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(x_k,\rho) \cap B(x,\rho)^c}(X_i) \\
&\leq \quad C \frac{\rho}{h\phi_x(h)} \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(x_k,\rho)}(X_i) \\
&+ \quad \| K \|_\infty \frac{1}{\phi_x(h)} \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(x_k,\rho)^c \cap B(x_k,h+\rho)}(X_i) \\
&+ \quad \| K \|_\infty \frac{1}{\phi_x(h)} \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(x_k,\rho) \cap B(x_{k,h-\rho})^c}(X_i) \\
&\leq \quad \frac{1}{n} \sum_{i=1}^n Z_i + Z_{n,1} + Z_{n,2},
\end{aligned}$$

where Z_i *is defined in the first case, and* $Z_{n,j} = \frac{1}{n}\sum_{i=1}^n W_{i,j}$ *where*

$$W_{i,1} = \| K \|_\infty \frac{2}{\phi_x(h)} \mathbb{I}_{B(x_k,\rho)^c \cap B(x_k,h+\rho)}(X_i)$$

and $W_{i,2} = \|K\|_\infty \dfrac{2}{\phi_x(h)} \mathbb{I}_{B(x_k,\rho) \cap B(x_k,h-\rho)^c}(X_i)$.

Therefore, $W_{i,j} \leq 2/|K\|_\infty/\phi_x(h)$, *and let consider*

$$\Lambda_{i,1} = \| K \|_\infty \frac{2}{\phi_x(h)} \left[\mathbb{I}_{B(x_k,\rho)^c \cap B(x_k,h+\rho)}(X_i) - \mathbb{E}[\mathbb{I}_{B(x_k,\rho)^c \cap B(x_k,h+\rho)}(X_i)] \right],$$

and

$$\Lambda_{i,2} = \| K \|_\infty \frac{2}{\phi_x(h)} \left[\mathbb{I}_{B(x_k,\rho) \cap B(x_k,h-\rho)^c}(X_i) - \mathbb{E}[\mathbb{I}_{B(x_k,\rho) \cap B(x_k,h-\rho)^c}(X_i)] \right].$$

Thus, for $k = 1, 2$, *we have*

$$S^2_{n_{i,k}}(x) = \sum_{i=1}^n \sum_{j=1}^n |cov(\Lambda_{i,k}(x), \Lambda_{j,k}(x))| = S^{2*}_{n_{i,k}}(x) + nVar(\Lambda_1(x)) \leq S^2_n(x),$$

with

$$S^{2*}_{n_{i,k}} = \sum_{i \neq j} |cov(\Lambda_{i,k}(x), \Lambda_{j,k}(x))|.$$

Then, we can get

$$\sup_{x \in S_{\mathcal{F}} \cap B(x_k,\rho)} \tilde{\tilde{D}}_n(x) \leq \frac{C}{n} \sum_{i=1}^n Z_i + Z_{n,1} + Z_{n,2}.$$

Notice that,

$$\mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)}|\tilde{D}_n(x)|>\frac{\epsilon}{2}\right)\leq\mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)}|\tilde{\tilde{D}}_n(x)|>\frac{\epsilon}{4}\right)$$
$$\leq\mathbb{P}\left(\frac{C}{n}\sum_{i=1}^{n}Z_i+Z_{n,1}+Z_{n,2}>\frac{\epsilon}{4}\right).$$

Using the result obtained in the case $K(1)=0$, we deduce that

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}Z_i>\frac{\epsilon}{8C}\right) &\leq exp\left(-\epsilon^2 n\phi_x(h)\frac{1}{128\left(1+\epsilon\frac{1}{16\|K\|_\infty}\right)}\right)\\ &+ n(\log n)^{-2}\left(\frac{8C(\log n)^2}{n\epsilon}\right)^{a+1}.\end{aligned}$$

Then, for concluding the proof we use a similar inequality for $\mathbb{P}(Z_{n,j}>\frac{\epsilon}{16})$, for $j=1,2$.

We can found the bound of $\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}|W_{i,j}-\mathbb{E}[W_{i,j}]|>\frac{\epsilon}{32}\right)$, for $j=1,2$, by using the Proposition A.11 in [13], if $C_1=4\parallel K\parallel_\infty^2$ and $C_2=4\parallel K\parallel_\infty$, its follows that

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}|W_{i,j}-\mathbb{E}[W_{i,j}]|>\frac{\epsilon}{32}\right)\leq 2exp\left(-\epsilon^2 n\phi_x(h)\frac{1}{2048\left(C_1\frac{\rho}{\phi_x(h)}+\epsilon C_2\right)}\right)$$
$$+n(\log n)^{-2}\left(\frac{4C(\log n)^2}{n\epsilon}\right)^{a+1}.$$

Proof of Part c).
The proof is similar to the one given in part b), and then omitted.

lemma 2.3.2 *Let, $\tilde{v}_j(x)$ be defined in (2.3.5) for $j=0,1$. Under $(H1)$ and $(H4)$, if $h_n\to 0$, we have that*

a)

$$\sup_{y\in\mathbb{R}}\sup_{x\in S_{\mathcal{F}}}|\mathbb{E}\left[\tilde{v}_1(x)\right]-F(y/X=x)\mathbb{E}\left[\tilde{v}_0(x)\right]|\to 0.$$

b) If the assumption $H8$ holds also, we deduce that

$$\sup_{y\in\mathbb{R}}\sup_{x\in S_{\mathcal{F}}}|\mathbb{E}\left[\tilde{v}_1(x)\right]-F(y/X=x)\mathbb{E}\left[\tilde{v}_0(x)\right]|=O(h^{\eta_1}).$$

proof 2.3.3 *Proof of Part b).*
Let, $Z(x)$ be a random variable where, $Z(x)=d(x,X)/h$ and $P_{Z(x)}$ his probability function. By the assumption $(H8-ii)$, and the boundness of K, we get

$$\delta_1(x,y)=C\int_{S_{\mathcal{F}}^{\epsilon}}d^{\gamma_1}(x,u)K\left(\frac{d(x,u)}{h}\right)dP_X(u)\leq Dd^{\gamma_1}\int v^{\gamma_1}K(v)dP_{Z(x)}(v)$$

$$\leq Dd^{\gamma_1}\int_0^1 K(v)dP_{Z(x)}(v) \leq D_1 d^{\gamma_1}\phi_x(h).$$

Then, for all $y \in \mathbb{R}$ and $x \in S_{\mathcal{F}}$

$$\frac{C_1}{\phi_x(h)}\delta_1(x,y) \leq D_1 d^{\gamma_1}\phi_x(h), \tag{2.3.19}$$

which conclude the proof.

The part a) is direct consequence of the result of part b).

Let, $F_n(y/X = x)$ be a sequence of conditional distribution function verifying

$$\sup_{y\in\mathbb{R}}\sup_{x\in S_{\mathcal{F}}}|F_n(y/X = x) - F(y/X = x)| \to 0. \tag{2.3.20}$$

Then, if F verifies the assumption $(H5)$, there exist a positive constants $T_1 \leq T_2$, such that, $s_n(x) = MAD_C(F_n(./X = x))$ satisfies

$$T_1 \leq s_n(x) \leq T_2, \forall x \in S_{\mathcal{F}},\ andn > n_0.$$

The Proof of this result is analogous to Lemma 3.1 in [8]. Using Lemma 2.3.1 part.b), when $\rho_n = \frac{\log n}{n}$, we obtain, for any $\epsilon > 0$,

$$\mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}}|\upsilon_j(x) - \mathbb{E}\left[\upsilon_j(x)\right]| > \epsilon\right) \leq \exp\left(\Gamma_{S_{\mathcal{F}}}(\rho_n) - \frac{\epsilon^2 n\phi_x(h)}{a_1(1+\epsilon a_2)}\right) + Cnr^{-1}\left(\frac{r}{\epsilon}\right)$$

$$\leq \exp\left\{-n\phi_x(h)\left(\frac{\epsilon^2}{a_1(1+\epsilon a_2)} - \frac{\Gamma_{S_{\mathcal{F}}}(\rho_n)}{n\phi_x(h)}\right)\right\} + Cnr^{-1}\left(\frac{r}{\epsilon}\right).$$

On the other hand, $H6$ ii) implies that, for $n > n_0$

$$\frac{\Gamma_{S_{\mathcal{F}}}(\rho_n)}{n\phi_x(h)} < \frac{1}{2}\frac{\epsilon^2}{a_1(1+\epsilon a_2)}.$$

If we take $c = 2a_1$, we obtain

$$\mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}}|\upsilon_j(x) - \mathbb{E}\left[\upsilon_j(x)\right]|\right) \leq \exp\left\{-n\phi_x(h)\frac{\epsilon^2}{c(1+\epsilon a_2)}\right\} + Cnr^{-1}\left(\frac{r}{\epsilon}\right).$$

Since, $|\tilde{\upsilon}_j(x) - \mathbb{E}\left[\tilde{\upsilon}_j(x)\right]| \leq \tilde{C}|\upsilon_j(x) - \mathbb{E}\left[\upsilon_j(x)\right]|$, thus

$$\mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}}|\tilde{\upsilon}_j(x) - \mathbb{E}\left[\tilde{\upsilon}_j(x)\right]| > \tilde{C}\epsilon\right) \leq \exp\left\{-n\phi_x(h)\frac{\epsilon^2}{c(1+\epsilon a_2)}\right\} + Cnr^{-1}\left(\frac{r}{\epsilon}\right).$$

Then, let $\mathcal{A} = \{\inf_{x\in S_{\mathcal{F}}}\tilde{\upsilon}_0(x) \leq \frac{1}{2}\}$ and remark that, $\mathbb{E}\left[\tilde{\upsilon}_0(x)\right] = 1$, we can deduce that

$$\mathbb{P}(\mathcal{A}) \leq \mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}}|\tilde{\upsilon}_j(x) - \mathbb{E}\left[\tilde{\upsilon}_j(x)\right]| > \frac{1}{2}\right) \leq \exp(-An\phi_x(h)),$$

where $A^{-1} = 2\tilde{C}c(2\tilde{C} + a_2)$.

Therefore, using that $\frac{n\phi_x(h)}{\log(n)} \to \infty$, *then* $\sup_{x\in S_{\mathcal{F}}} |\tilde{v}_j(x) - \mathbb{E}[\tilde{v}_j(x)]| \overset{a.co}{\to} 0$, *for* $j = 0, 1$ *and*

$$\mathbb{P}\left(\inf_{x\in S_{\mathcal{F}}} \tilde{v}_0(x) \leq \frac{1}{2}\right) < \exp(-An\phi_x(h)). \tag{2.3.21}$$

This results give the proof of (2.3.11), *and combined with 2.3.1.a) and (2.3.7), we can deduce that,* $\forall y \in \mathbb{R}$,

$$\sup_{x\in S_{\mathcal{F}}} |\hat{F}(y|X = x) - F(y|X = x)| \overset{a.s}{\to} 0. \tag{2.3.22}$$

For each, $q \in \mathbb{Q}$, *define* $\mathcal{N} = \{w \in \Omega : \sup_{x\in S_{\mathcal{F}}} |\widehat{F}(q/X = x) - F(q/X = x)| \nrightarrow 0\}$ *and* $\mathcal{N} = \cup_{q\in\mathbb{Q}}\mathcal{N}(q)$.

Then, (2.3.22) *implies that* $\mathbb{P}(\mathcal{N}) = 0$. *Let* $w \in \Omega,\ w \in \mathcal{N}$, *then,*

$$\sup_{x\in S_{\mathcal{F}}} |\widehat{F}(q/X = x) - F(q/X = x)| \to 0, \text{ for all}\, q \in \mathbb{Q}.$$

Given $\epsilon > 0$, *by the assumption* $H5$ *iii), there exist* $a, b \in \mathbb{Q}$, *such that,* $F(b/X = x) > 1 - \epsilon$ *and* $F(a/X = x) < \epsilon$, $\forall x \in S_{\mathcal{F}}$.

Moreover, the equicontinuity condition of F, *given in the hypothesis* $(H5-iv)$, *entails that there exist* $a = y_1 < y_2 < \cdots < y_l = b; y_i \in \mathbb{Q}$, *such that,* $|F(y/X = x) - F(y_i/X = x)| < \epsilon$, *for* $x \in S_{\mathcal{F}}$ *and for any* y.

Let, $n_0 \in \mathbb{N}$, *such that for,* $n > n_0$,

$$\max_{1\leq i\leq l} \sup_{x\in S_{\mathcal{F}}} |\widehat{F}(y_i/X = x) - F(y/X = x) < \epsilon.$$

Then, it is easy to see that

$$\max_{y\in\mathbb{R}} \sup_{x\in S_{\mathcal{F}}} |\widehat{F}(y/X = x) - F(y/X = x) < 2\epsilon,\ \forall n \geq n_0,$$

which is the claimed result.

2.3.2 Uniform strong convergence rates

Theorems 2.3.2 give almost complete convergence rates for the estimators of the empirical conditional distribution and for the local M-estimators of the regression function.

Theorem 2.3.2 *Let,* $S_{\mathcal{F}} \subset \mathcal{F}$ *be a compact set. Assume that* $(H1)$, $(H4)$, $(H5iii)iv)$, $(H6)$, $(H7)$, $(H8)$, *and* $(H9)$ *holds. If,* γ_1 *defined in* $(H8)$, *is such that,* $\gamma_1 < \frac{1}{2}$, *furthermore, assume that,* $\left(\frac{n}{\log(n)}\right)^{1-\gamma_1} \phi_x(h) \leq C$ *holds. Let* $\widehat{\theta}(x)$

be a robust estimator such that with probability 1, there exists real constants, $0 < T_1 < T_2$, such that, $T_1 < \widehat{\theta}(x) < T_2$, for all $x \in S_{\mathcal{F}}$ and $n > n_0$.

Then,

$$\sup_{x\in S_{\mathcal{F}}}\sup_{y\in\mathbb{R}}|\widehat{\theta}(x)-\theta(x)| = O_{a.co}\left(h^{\eta}+\sqrt{\frac{\Gamma_{S_{\mathcal{F}}}(\log(n)/n)}{n\phi_x(h)}}\right).$$

***proof 2.3.4** Suppose that, $T_1 \leq \widehat{\theta} \leq T_2$, thus, for any $\varepsilon \in \mathbb{R}$,*

$$\sup_{x\in S_{\mathcal{F}}}|\Psi(x,\theta(x)+\varepsilon,\widehat{s}(x))-\widehat{\Psi}(x,\theta(x)+\varepsilon,\widehat{s}(x))| \leq$$

$$\frac{1}{T_1}\|\psi\|_V \sup_{x\in S_{\mathcal{F}}}\sup_{y\in\mathbb{R}}|F(y/X=x)-\widehat{F}(y/X=x)|.$$

So, if $\tilde{\theta}_n = h^{\gamma_1}+\sqrt{\frac{\Gamma_{S_{\mathcal{F}}}(\log(n)/n)}{n\phi_x(h)}}$, and for each $\tau > 0$, we have

$$\sup_{x\in S_{\mathcal{F}}}\sup_{-\tau<\varepsilon<\tau}|\Psi(x,\theta(x)+\varepsilon,\widehat{s}(x))-\widehat{\Psi}(x,\theta(x)+\varepsilon,\widehat{s}(x))| = O_{a.s}(\tilde{\theta}_n).$$

Then, using the fact that, $\sup_{x\in S_{\mathcal{F}}}|\widehat{\theta}(x)-\theta(x)| \overset{a.s}{\to} 0$, we can easily get

$$\sup_{x\in S_{\mathcal{F}}}|\Psi(x,\widehat{\theta}(x),\widehat{s}(x))-\widehat{\Psi}(x,\widehat{\theta}(x),\widehat{s}(x))| = O_{a.s}(\tilde{\theta}_n). \tag{2.3.23}$$

Note that,
$\Psi(x,\theta(x),\widehat{s}(x))-\Psi(x,\widehat{\theta}(x),\widehat{s}(x))+\Psi(x,\widehat{\theta}(x),\widehat{s}(x))-\widehat{\Psi}(x,\widehat{\theta}(x),\widehat{s}(x))=0$
and using results of (2.3.23), we get

$$\sup_{x\in S_{\mathcal{F}}}|\widehat{\Psi}(x,\theta(x),\widehat{s}(x))-\widehat{\Psi}(x,\widehat{\theta}(x),\widehat{s}(x))| = O_{a.s}(\tilde{\theta}_n). \tag{2.3.24}$$

Denote by, $\Psi'(x,t,\sigma) = \partial\Psi(x,u,\sigma)/\partial u|_{u=t}$, the Mean Value Theorem leads to

$$\widehat{\Psi}(x,\theta(x),\widehat{s}(x))-\widehat{\Psi}(x,\widehat{\theta}(x),\widehat{s}(x)) = (\widehat{\theta}(x)-\theta(x))[\Psi(x,\xi_n,\widehat{s}(x))]), \tag{2.3.25}$$

where $\xi_n(x) \in [\theta(x),\widehat{\theta}(x)]$.

Thanks to assumption $(H3-ii)$, we get

$$\inf_{-\tau<\varepsilon<\tau}\inf_{x\in S_{\mathcal{F}}} -\frac{\partial}{\partial u}\Psi(x,\theta(x),\widehat{s}(x))|_{u=a} > c_0 > 0. \tag{2.3.26}$$

Let, $\mathcal{N}$ be the set where (2.3.24), (2.3.26) and $\sup_{x\in S_{\mathcal{F}}}|\widehat{\theta}(x)-\theta(x)| \longrightarrow 0$ hold, then, $\mathbb{P}(\mathcal{N}) = 0$. Since, for $w \in \mathcal{N}$ and $\tau > 0$, $|\xi_n - \theta(x)| \leq \tau$,
$\sup_{x\in S_{\mathcal{F}}}|\Psi(x,\theta(x),\widehat{s}(x))-\Psi(x,\widehat{\theta}(x),\widehat{s}(x))| = (\tilde{\theta}_n)$.

Therefore, we get that

$$\inf_{x\in S_{\mathcal{F}}} -\Psi'(x,\xi_n,\widehat{s}(x)) > c_0 > 0.$$

The combination of this result together with (2.3.24),(2.3.25) and (2.3.26) achieves the proof of Theorem 2.3.2.

Theorem 2.3.3 *Let $S_{\mathcal{F}} \cup \mathcal{F}$ be a compact set. Assume that $(H1), (H4), (H5), (H6), (H7)$ and $(H9)$ holds. Moreover, assume that σ and θ are Lipschitz function of order γ_1 and γ_2 respectively. Then, if $\gamma = \min(\gamma_1, \gamma_2)$, such that $\gamma > 1$, and assume that*

$$h\left(\frac{n}{\log(n)}\right)^{1-\gamma} \leq C_\gamma \quad \text{for all } n \geq 1 \tag{2.3.27}$$

or

$$\phi_x(h)\left(\frac{n}{\log(n)}\right)^{1-\gamma} \leq C_\gamma \quad \text{for all } n \geq 1 \quad \text{for all } n \geq 1 \tag{2.3.28}$$

is fulfilled.

Let, $\widehat{s}(x)$ be a robust scale estimator with probability 1, and $\theta(x)$ be the unique solution of $\Psi(x, a, s(x)) = 0$. Suppose that exists real constants, $0 < T_1 < T_2$, such that $T_1 < \widehat{s}(x) < T_2$ for all $x \in S_{\mathcal{F}}$, and $n > n_0$.

Then, if $\widehat{\theta}(x)$ is a solution of (2.2.3), such that $\sup_{x\in S_{\mathcal{F}}} |\widehat{\theta}(x) - \theta(x)| \overset{a.so}{\to} 0$, we have

$$sup_{x\in S_{\mathcal{F}}} |\widehat{\theta}(x) - \theta(x)| = O_{a.co}\left(h^\gamma + \sqrt{\frac{\Gamma_{S_{\mathcal{F}}}(\log(n)/n)}{n\phi_x(h)}}\right).$$

proof 2.3.5 (Theorem 2.3.3) *Our goal is to show,*

$$A_n = \sup_{x\in S_{\mathcal{F}}} \sup_{|\varepsilon|<\tau} \sup_{T_1<\sigma<T_2} |\widehat{\Psi}(x, \theta(x)+\varepsilon, \sigma) - \Psi(x, \theta(x)+\varepsilon, \sigma)| = O_{a.co}(\tilde{\theta}_n).$$

Let denote

$$W_{i,\varepsilon,\sigma}(x) = \psi\left(\frac{Y_i - \theta(x) - \varepsilon}{\sigma}\right), \qquad \tilde{\upsilon}_1(x, \varepsilon, \sigma) = \frac{1}{n}\sum_{i=1}^{n} W_{i,\varepsilon,\sigma}(x)\frac{K_i(x)}{\mathbb{E}K_1(x)},$$

and

$$\upsilon_1(x, \varepsilon, \sigma) = \frac{1}{n}\sum_{i=1}^{n} W_{i,\varepsilon,\sigma}(x)\frac{K_i(x)}{\phi_x(h)}.$$

Since, $\widehat{\Psi}(x, \theta(x)+\varepsilon, \sigma) = \tilde{\upsilon}_1(x, \varepsilon, \sigma)/\tilde{\upsilon}_0(x)$, where $\tilde{\upsilon}_0(x)$ is defined in (2.3.7), so

$$\begin{aligned}
|\widehat{\Psi}(x, \theta(x)+\varepsilon, \sigma) - \Psi(x, \theta(x)+\varepsilon, \sigma)| \leq \frac{1}{\inf_{x\in S_{\mathcal{F}}} \tilde{\upsilon}_0(x)} \Bigg[& \sup_{x\in S_{\mathcal{F}}} |\tilde{\upsilon}_0(x) - \mathbb{E}\tilde{\upsilon}_0(x)| \\
& + \sup_{x\in S_{\mathcal{F}}} \sup_{|\varepsilon|<\tau} \sup_{T_1<\sigma<T_2} |\tilde{\upsilon}_1(x, \varepsilon, \sigma) - \mathbb{E}\tilde{\upsilon}_1(x, \varepsilon, \sigma)| \\
+ \sup_{x\in S_{\mathcal{F}}} \sup_{|\varepsilon|<\tau} \sup_{T_1<\sigma<T_2} |\mathbb{E}\tilde{\upsilon}_1(x, \varepsilon, \sigma) - \Psi(x, \theta(x)+\varepsilon, \sigma)\mathbb{E}\tilde{\upsilon}_0(x)| & \Bigg].
\end{aligned} \tag{2.3.29}$$

By the assumption $(H2-ii)$*, and for some constant* $C > 0$*, if* $d(X_1, x) < h_n < 1$*, then*

$$\sup_{x\in S_{\mathcal{F}}} \sup_{|\varepsilon|<\tau} \sup_{T_1<\sigma<T_2} |\Psi(x_1, \theta(x)+\varepsilon, \sigma) - \Psi(x, \theta(x)+\varepsilon, \sigma)|$$

$$\leq C[d(X_1, x)^{\gamma_1} + d(X_1, x)^{\gamma_2}]$$
$$\leq Cd(X_1, x)^{\gamma},$$

where $\gamma = \min\{\gamma_1, \gamma_2\}$*.*

Since, $K_1(x) \leq \mathbb{I}_{B(x,h)}(X_1)$*, we have, for* n *large enough,*

$$\begin{aligned} &|\mathbb{E}\tilde{\upsilon}_1(x, \varepsilon, \sigma) - \Psi(x, \theta(x)+\varepsilon, \sigma)\mathbb{E}\tilde{\upsilon}_0(x)| \\ &\quad \leq \mathbb{E}|\Psi(X_1, \theta(x)+\varepsilon, \sigma) - \Psi(x, \theta(x)+\varepsilon, \sigma)| \frac{K_1(x)}{\mathbb{E}K_1(x)} \\ &\quad \leq Ch^{\gamma}. \end{aligned}$$

Therefore, if $\tilde{\theta}_n = \theta_n + h^{\gamma}$*, then,*

$$B_{2n} = \sup_{x\in S_{\mathcal{F}}} \sup_{|\varepsilon|<\tau} \sup_{T_1<\sigma<T_2} |\mathbb{E}\tilde{\upsilon}_1(x, \varepsilon, \sigma) - \Psi(x, \theta(x)+\varepsilon, \sigma)\mathbb{E}\tilde{\upsilon}_0(x)| \leq C\tilde{\theta}_n.$$

Let $\mathcal{A}_n = \inf_{x\in S_{\mathcal{F}}} \tilde{\upsilon}_0(x) < \frac{1}{2}$ *and* $\epsilon_0 > C$*. Using* (2.3.29)*, we conclude that*

$$\mathbb{P}(A_n > 4\epsilon_0) \leq \mathbb{P}(\mathcal{A}_n) + \mathbb{P}(\sup_{x\in S_{\mathcal{F}}} |\tilde{\upsilon}_0(x) - \mathbb{E}\tilde{\upsilon}_0(x)| > \epsilon_0\theta_n) + \mathbb{P}(\tilde{B}_{1n} > \epsilon_0\theta_n),$$

where

$$\tilde{B}_{1n} = \sup_{x\in S_{\mathcal{F}}} \sup_{|\varepsilon|<\tau} \sup_{T_1<\sigma<T_2} |\tilde{\upsilon}_1(x, \varepsilon, \sigma) - \mathbb{E}\tilde{\upsilon}_1(x, \varepsilon, \sigma)|.$$

By Lemma 2.3.1 part b), with $\rho_n = \log(n)/n$*, and using* (2.3.3)*, we get*

$$\sum_{n\geq 1} \mathbb{P}(\mathcal{A}_n < \infty).$$

When Lemma 2.3.1 part c), permit to obtain

$$\mathbb{P}(\sup_{x\in S_{\mathcal{F}}} |\tilde{\upsilon}_0(x) - \mathbb{E}\tilde{\upsilon}_0(x)| > \epsilon_0\theta_n)) < \infty, \text{ for } \epsilon_0 > c.$$

So, it suffices to show that

$$\sum_{n\geq 1} \mathbb{P}(\tilde{B}_{1n} > \epsilon_0\theta_n) < \infty.$$

Note that, $\mathbb{P}(\tilde{B}_{1n} > \epsilon_0\theta_n) \leq \mathbb{P}(B_{1n} > \epsilon_1\theta_n)$*, where* $\epsilon_1 = \epsilon_0 C$*, and*

$$B_{1n} = \sup_{x\in S_{\mathcal{F}}} \sup_{|\varepsilon|<\tau} \sup_{T_1<\sigma<T_2} |\upsilon_1(x, \varepsilon, \sigma) - \mathbb{E}\upsilon_1(x, \varepsilon, \sigma)|.$$

Therefore, we need to prove that, for some c_1 *and for some* $\epsilon_1 > c_1$, $\sum_{n\geq 1} \mathbb{P}(B_{1n} > \epsilon_1\theta_n) < \infty$.

Let, $\rho_n = \log(n)/n$ *and* $x_1, \cdots, x_l$*, such that,* $S_{\mathcal{F}} \cup \subset_{j=1}^{l} B(x_j, \rho_n)$*, where* $l = N_{\rho_n}(S_{\mathcal{F}})$.

Then, if $S_n(x,\varepsilon,\sigma) = \upsilon_1(x,\varepsilon,\sigma) - \mathbb{E}\upsilon_1(x,\varepsilon,\sigma)$ and $\tilde{S}_{n,k}(x,\varepsilon,\sigma) = S_n(x,\varepsilon,\sigma) - S_n(x_k,\varepsilon,\sigma)$*, we have that*

$$S_n(x,\varepsilon,\sigma) = S_n(x_k,\varepsilon,\sigma) + \tilde{S}_{n,k}(x,\varepsilon,\sigma),$$

and

$$\begin{aligned}
\mathbb{P}(B_{1n} > \epsilon_1\theta_n) &= \mathbb{P}(\sup_{x\in S_{\mathcal{F}}} \sup_{|\varepsilon|<\tau} \sup_{T_1<\sigma<T_2} S_n(x,\varepsilon,\sigma) > \epsilon_1\theta_n) \\
&\leq \mathbb{P}(\sup_{|\varepsilon|<\tau} \sup_{T_1<\sigma<T_2} \max_{1\leq k\leq l} |S_n = (x_k,\varepsilon,\sigma)| > \frac{\epsilon_1}{2}\theta_n) \\
&+ \mathbb{P}(\sup_{T_1<\sigma<T_2} \sup_{|\varepsilon|<\tau} \max_{1\leq k\leq l}) \sup_{x\in B(x_k,\rho_n)\cap S_{\mathcal{F}}} |\tilde{S}_{n,k}(x,\varepsilon,\sigma)| > \frac{\epsilon_1}{2}\theta_n).
\end{aligned}$$

Note that,

$$|\tilde{S}_{n,k}(x,\varepsilon,\sigma)| \leq |\upsilon_1(x,\varepsilon,\sigma) - \upsilon_1(x_k,\varepsilon,\sigma)| + |\mathbb{E}\upsilon_1(x,\varepsilon,\sigma) - \mathbb{E}\upsilon_1(x_k,\varepsilon,\sigma)|.$$

Then, by Lipschitz condition of ψ*, we have*

$$|W_{i,\varepsilon,\sigma}(x_1) - W_{i,\varepsilon,\sigma}(x_2)| \leq c_\theta \|\psi'\|_\infty d(x_1,x_2)^{\gamma_1}/A,$$

for all ε *and* $T_1 < \sigma < T_2$*, where* c_θ *stands for the Lipschitz constant of* θ.

Therefore, if $d(x_k, x) \leq \rho_n < 1$*, we have*

$$\begin{aligned}
|\upsilon_1(x,\varepsilon,\sigma) - \upsilon_1(x_k,\varepsilon,\sigma)| &\leq \|\psi\|_\infty \frac{1}{n\phi_x(h)} \sum_{i=1}^{n} |K_i(x) - K_i(x_k)| \\
&+ \frac{c_\theta}{T_1}\|\psi'\|_\infty \frac{\rho_n^n}{n\phi_x(h)} \sum_{i=1}^{n} |K_i(x_k)|.
\end{aligned}$$

Denote by,

$\tilde{T}_1 = c_\theta \|K\|_\infty \|\psi'\|_\infty / T_1$*, then*

$$\max_{1\leq k\leq l} \sup_{x\in B(x_k,\rho_n)\cap S_{\mathcal{F}}} \sup_{T_1<\sigma<T_2} \sup_{|\varepsilon|<\tau} |\upsilon_1(x,\varepsilon,\sigma) - \upsilon_1(x_k,\varepsilon,\sigma)|$$

$$\leq \tilde{T}\upsilon_1 \max_{1\leq k\leq l} \frac{\rho_n^n}{n\phi_x(h)} \sum_{i=1}^{n} \mathbb{I}_{B(x_k,\rho)}(X_i)$$

$$+\|\psi\|_\infty \max_{1\leq k\leq l} \sup_{x\in B(x_k,\rho_n)\cap S_{\mathcal{F}}} \frac{1}{n\phi_x(h)} \sum_{i=1}^{n} |K_i(x) - K_i(x_k)|. \tag{2.3.30}$$

We consider two situation which that $\gamma \geq 1$ or $\gamma \leq 1$.
i) *If $\gamma \geq 1$, we have that $\rho_n^n < \rho_n$, so the bound* (2.3.30)*, leads to*

$$\max_{1\leq k\leq l} \sup_{x\in B(x_k,\rho_n)\cap S_{\mathcal{F}}} \sup_{T_1<\sigma<T_2} \sup_{|\varepsilon|<\tau} |\upsilon_1(x,\varepsilon,\sigma) - \upsilon_1(x_k,\varepsilon,\sigma)|$$

$$\leq \tilde{T}_1 \max_{1\leq k\leq l} \frac{\rho_n}{n\phi_x(h)} \sum_{i=1}^{n} \mathbb{I}_{B(x_k,\rho)}(X_i)$$

$$+\|\psi\|_\infty \max_{1\leq k\leq l} \sup_{x\in B(x_k,\rho_n)\cap S_{\mathcal{F}}} \frac{1}{n\phi_x(h)} \sum_{i=1}^{n} |K_i(x) - K_i(x_k)|.$$

Since, we have that

$$\sum_{i\geq 1} \mathbb{P}(\sup_{T_1<\sigma<T_2} \sup_{|\varepsilon|<\tau} \max_{1\leq k\leq l} \sup_{x\in B(x_k,\rho_n)\cap S_{\mathcal{F}}} |\tilde{S}_{n,k}(x,\varepsilon,\sigma)| > \frac{\epsilon_1}{2}\theta_n) < \infty. \tag{2.3.31}$$

ii) *If $\gamma < 1$ and 2.3.27 holds, taking $C_\gamma \rho_n/h < 1$, we conclude that*

$$\max_{1\leq k\leq l} \sup_{x\in B(x_k,\rho_n)\cap S_{\mathcal{F}}} \sup_{T_1<\sigma<T_2} \sup_{|\varepsilon|<\tau} |\upsilon_1(x,\varepsilon,\sigma) - \upsilon_1(x_k,\varepsilon,\sigma)|$$

$$\leq \tilde{T}_1 C_\gamma \max_{1\leq k\leq l} \frac{\rho_n}{nh\phi_x(h)} \sum_{i=1}^{n} \mathbb{I}_{B(x_k,\rho)}(X_i)$$

$$+\|\psi\|_\infty \max_{1\leq k\leq l} \sup_{x\in B(x_k,\rho_n)\cap S_{\mathcal{F}}} \frac{1}{n\phi_x(h)} \sum_{i=1}^{n} |K_i(x) - K_i(x_k)|,$$

then (2.3.31) *can be deduced from the proof of the Lemma 2.3.1 part c).*
iii) *If $\gamma < 1$ and* (2.3.28) *holds, the sequence $\theta_n^{-1}\rho_n^n$ is bounded. Then, defining*

$$Z_i = \frac{\rho_n^n}{n\phi_x(h)} \mathbb{I}_{B(x_k,\rho)}(X_i).$$

There exist, $\tilde{c}$, such that for $\epsilon_1 > \tilde{c}$, we can write

$$\sum_{i\geq 1} \mathbb{P}\left(\max_{1\leq k\leq l} \frac{\rho_n^n}{n\phi_x(h)} \sum_{i=1}^{n} \mathbb{I}_{B(x_k,\rho)}(X_i) > \epsilon_1 \frac{\theta_n}{4} \right) < \infty,$$

which conclude the proof of (2.3.31)*.*

Now, we need to find a bound for χ_n, where

$$\chi_n = \mathbb{P}\left(\sup_{T_1<\sigma<T_2} \sup_{|\varepsilon|<\tau} \max_{1\leq k\leq l} |S_n(x_k,\varepsilon,\sigma)| > \epsilon_1 \frac{\theta_n}{2} \right).$$

Let, $|\varepsilon| \leq \tau$, a finite covering by intervals $I_j^{\varepsilon} = [\varepsilon_j, \varepsilon_{j+1}]$, when $|\varepsilon_{j+1} - \varepsilon_j| < \alpha_n$ with $\alpha_n = 1/\sqrt{n}$.

Then, we have at most $M_{1,\alpha_n} = 2\tau\alpha_n^{-1}$ intervals. On the other hand, considered a covering of $T_1 < \sigma < T_2$ by intervals $I_j^{\sigma} = [\sigma_j, \sigma_{j+1}]$, such that, $|\sigma_{j+1} - \sigma_j| < \alpha_n$, where we have $M_{2,\alpha_n} = 2(T_1 - T_2)\alpha_n^{-1}$ intervals. Thus,

$$\chi_n \leq \chi_{n,1} + \chi_{n,2},$$

with,

$$\chi_{n,1} = \mathbb{P}\left(\max_{1\leq s\leq M_{2,\alpha_n}} \max_{1\leq j\leq M_{1,\alpha_n}} \max_{1\leq k\leq l} |S_n(x_k, \varepsilon_j, \sigma_s)| > \epsilon_1 \frac{\theta_n}{2}\right), \tag{2.3.32}$$

and

$$\chi_{n,2} = \mathbb{P}\left(\max_{1\leq j\leq M_{1,\alpha_n}} \max_{1\leq k\leq l} \sup_{T_1<\sigma<T_2} \sup_{|\varepsilon|<\tau} |S_n(x_k, \varepsilon_j, \sigma_s) - S_n(x_k, \varepsilon, \sigma)| > \epsilon_1 \frac{\theta_n}{2}\right). \tag{2.3.33}$$

Since, $|W_{i,\varepsilon,\sigma}| \leq \|\psi\|_\infty$, similar arguments to those used in Lemma 2.3.1, we obtain

$$\sup_{T_1<\sigma<T_2} \sup_{|\varepsilon|<\tau} \max_{1\leq k\leq l} \mathbb{P}\left(|S_n(x_k, \varepsilon, \sigma)| > \theta_n \frac{\epsilon_1}{4}\right) \leq \exp\left\{(1 - C\epsilon_1^2)\Gamma_{S_{\mathcal{F}}}\left(\frac{\log(n)}{n}\right)\right\}.$$

Moreover, if $C_2 = 2\tau(T_1 - T_2)$, we have that

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\leq j\leq M_{1,\alpha_n}} \max_{1\leq s\leq M_{2,\alpha_n}} \max_{1\leq k\leq l} |S_n(x_k, \varepsilon_j, \sigma_s)| > \theta_n \frac{\epsilon_1}{4}\right) \\
&\leq 2M_{1,\alpha_n} M_{2,\alpha_n} M_{\rho_n}(S_{\mathcal{F}}) \exp\left\{-C\epsilon_1^2 \Gamma_{S_{\mathcal{F}}}\left(\frac{\log(n)}{n}\right)\right\} \\
&\leq 2C_2\alpha_n^{-2} \exp\left\{(1 - C\epsilon_1^2)\Gamma_{S_{\mathcal{F}}}\left(\frac{\log(n)}{n}\right)\right\}.
\end{aligned}$$

Taking, ϵ_1 such that $(1 - C\epsilon_1^2) > 1 - \beta$, and $\alpha_n = 1/\sqrt{n}$, then

$$\chi_{n,1} \leq 2C_2 n \exp\left\{(1 - \beta)\Gamma_{S_{\mathcal{F}}}\left(\frac{\log(n)}{n}\right)\right\}.$$

The result $\sum_{n\geq 1} \chi_{n,1} < \infty$ follows as a direct consequence of (2.3.3), for n a large enough.

Concerning

$$\sum_{n\geq 1} \chi_{n,2} < \infty, \tag{2.3.34}$$

note that, if $a \in I_j^{(a)}$*, for all* $T_1 < \sigma < T_2$*, we have*

$$|\upsilon_1(x_k, \varepsilon_j, \sigma_s) - \upsilon_1(x_k, \varepsilon, \sigma)| \leq \frac{\|\psi'\|_\infty}{T_1} \frac{\alpha_n}{n\phi_x(h)} \sum_{i=1}^{n} K_i(x_k).$$

Then, if $\sigma \in I_s^{(\sigma)}$*, using that* $\zeta(t) = t\psi'(t)$ *is bounded, we obtain*

$$|\upsilon_1(x_k, \varepsilon_j, \sigma_s) - \upsilon_1(x_k, \varepsilon, \sigma)| \leq \frac{\|\zeta\|_\infty}{T_1} \frac{\alpha_n}{n\phi_x(h)} \sum_{i=1}^{n} K_i(x_k).$$

Consequently,

$$|\upsilon_1(x_k, \varepsilon_j, \sigma_s) - \upsilon_1(x_k, \varepsilon, \sigma)| \leq C_3 \frac{\alpha_n}{n\phi_x(h)} \sum_{i=1}^{n} K_i(x_k),$$

where $C_3 = (\|\psi'\|_\infty + \|\zeta\|_\infty)/T_1$.

Noting that, $(1/n\phi_x(h)) \sum_{i=1}^{n} K_i(x_k) = \upsilon_0(x_k)$*, we get*

$$\begin{aligned}
|S_n(x_k, \varepsilon, \sigma) - S_n(x_k, \varepsilon, \sigma)| &\leq |\upsilon_1(x_k, \varepsilon_j, \sigma_s) - \upsilon_1(x_k, \varepsilon_j, \sigma_s)| \\
&+ \mathbb{E}|\upsilon_1(x_k, \varepsilon_j, \sigma_s) - \upsilon_1(x_k, \varepsilon_j, \sigma_s)| \\
&\leq C_3 \frac{\alpha_n}{n\phi_x(h)} \sum_{i=1}^{n} K_i(x_k) + C_3 \frac{\alpha_n}{\phi_x(h)} \mathbb{E}K_1(x_k) \\
&\leq C_3 \alpha_n |\upsilon_0(x_k) - \mathbb{E}\upsilon_0(x_k)| + 2C_3 \frac{\alpha_n}{\phi_x(h)} \mathbb{E}K_1(x_k).
\end{aligned}$$

Using the fact that, $\mathbb{E}K_1(x_k) \leq C'\phi_x(h), \alpha_n = 1/\sqrt{n}$ *and* $\theta_n^{-1}\alpha_n \to 0$*, then* $\exists n_0$*, such that, for* $n > n_0$*, we have* $\theta_n^{-1}\alpha_n \leq \min(1/C_3, \epsilon_1/(16C'C_3))$.

Therefore,

$$\begin{aligned}
\theta_n^{-1}|S_n(x_k, \varepsilon, \sigma) - S_n(x_k, \varepsilon, \sigma)| &\leq C_3\theta_n^{-1}\alpha_n|\upsilon_0(x_k) - \mathbb{E}\upsilon_0(x_k)| + 2C'C_3\theta_n^{-1}\alpha_n \\
&\leq |\upsilon_0(x_k) - \mathbb{E}\upsilon_0(x_k)| + \frac{\epsilon_1}{8}.
\end{aligned}$$

Then, the right hand side of the last inequality does not depend of σ *and* a*, so we can write*

$$\chi_{n,2} \leq \max_{1 \leq k \leq l} \mathbb{P}\left(|\upsilon_0(x_k) - \mathbb{E}\upsilon_0(x_k)| > \frac{\epsilon_1}{8}\right).$$

So, for all $\epsilon_0 > 0$,

$$\sum_{n \geq 1} \mathbb{P}\left(\sup_{x \in S_{\mathcal{F}}} |\upsilon_0(x_k) - \mathbb{E}\upsilon_0(x_k)| > \epsilon_0\right) < \infty,$$

this completes the proof of (2.3.34) *and therefore, the proof of Theorem 2.3.3.*

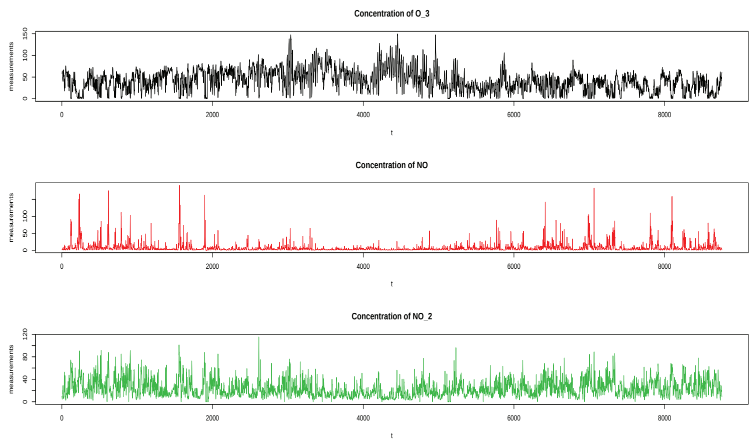

FIGURE 2.1 – The hourly measurements of O_3,NO and NO_2 concentration for the year 2018

2.4 Real data applications

The main objective of this part is to evaluate the good behavior of the proposed robust estimator for two real data applications and to show the efficiency of the robust estimator with unknown scale parameter compared to the one with fixed scale parameter.

2.4.1 Maximum Ozone Concentration

This section, we are interested in forecasting maximum values of ozone concentration. The data consist of hourly measurements of ozone (O_3) concentration together with additional chemical measurements such as NO and NO_2 concentrations (μg/m^3) (see, figure 2.1) in Leicester University monitoring site during the period from January 1^{st} to the 31^{st} December for the year 2018, (365 days). Data are available on the website (https ://uk-air.defra.gov.uk).

The original time series are :

$$O_{3,t}, NO_t \text{ and } NO_{2,t}, \quad t = 1, \ldots, 8760.$$

To fix the ideas, let's present the mathematical formulation of our prediction problem. Indeed, assume that we aim to predict the maximum air pollutant (Ozone O_3) concentration at day i, denoted by Y, using the curve of the daily emission of the gases observed the day before $i-1$. Formally, we assume that the output variable Y and the input variables are $Z = \left(X_{O_{3,i-1}}, X_{NO_{i-1}}, X_{NO_{2,i-1}}\right)$ is linked by the following regression formula,

$$Y_i = r(Z_i) + \epsilon_i, \qquad i = 1, \ldots, 364,$$

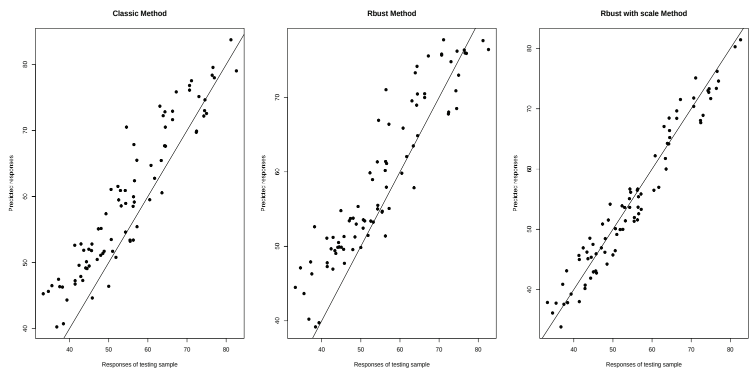

FIGURE 2.2 – Prediction of the maximum ozone of the last 64 days by classical and robust regression

is cut into 364 daily curves,

$$Z_i = \left\{(O_{3,24(i-1)+t}, NO_{24(i-1)+t}, NO_{2,24(i-1)+t}), t \in [\,0, 24\,[\right\}, \quad i = 1, \ldots, 364.$$

Second, we need to select a suitable semi-metric $d(.,.)$, kernel $K(.)$, smoothing parameter h_{opt} for our estimator. For that purpose, we choose the asymmetrical quadratic kernel defined as $K(u) = \frac{3}{4}\left(\frac{12}{11} - u^2\right)\mathbb{I}_{[0,1]}(u)$. According to the general guidelines provided in [13], we suggests to use standard PCA semi-metrics as following :

$$d^{PCA}(z_i, z_j) = d^{PCA}\left(x_{O_{3,i}}, x_{O_{3,j}}\right) + d^{PCA}\left(x_{NO_i}, x_{NO_j}\right) + d^{PCA}\left(x_{NO_{2,i}}, x_{NO_{2,j}}\right).$$

We choose $h_{opt} = \arg\min_h CV(h)$ for the estimator, where $CV(h) = \sum_{i=1}^n \left(Y_i - \widehat{\theta}_{(-i)}(Z_i)\right)^2$, and $\widehat{\theta}_{(-i)}(z)$ is the solution at t of :

$$\sum_{j=1, j\neq i}^{n} w_j(x)\psi\left(\frac{Y_j - t}{\widehat{\sigma}}\right) = 0.$$

Finally, we split our sample of 364 days into a learning sample containing the first 300 days and a testing sample with the last 64 days. Figure 2.2, shows the results obtained for the ozone prediction derived from the testing sample. The left panel represents the prediction under the classical method [13]. The central group represents the prediction by the robust method without scale parameter introduced by [4], while the right panel shows the forecast under our model (robust with scale parameter). The error used is the mean of squared error (MSE), expressed by

$$MSE = \frac{1}{64}\sum_{i=301}^{364}\left(Y_i - \widehat{\theta}(Z_i)\right)^2.$$

This is illustrated by the $MSE = 3.71488$ for the classic method, $MSE = 3.5238$ for the robust with fixed scale parameter, and $MSE = 3.025231$ for our proposed estimator.

Another important point for ensuring a good behavior of our method is to introduce the outliers in this learning sample. We multiply by 100 the response variable of some observations Y. We observe that, the robust method gives better results than the classical method in the presence of outliers (MSE for the classic is 67.97584, for the robust is 4.671861 and for the robust with scale parameter 4.15175).

As we can see, the robust with scale parameter method always gives good behavior in the tow cases, with and without outlier variables.

2.4.2 Peak electricity demand

The evolution of peak electricity demand can be considered as an essential system design metric for grid operators and planners. The peak demand forecasting of aggregated electricity demand has been widely studied in the statistical literature, and several approaches have been proposed to solve this issue, see, for instance, [10] and [16] for short-term peak forecasting and [17] for long-term density peak forecasting.

In this subsection, we are interested in the estimation of peak demand at the customer level. For a fixed day d let us denote by $(E_d(t_j))_{j=1,\ldots,24}$ the hourly measurements for the year 2016 (measured in MWh), retrieved from the smart metering device of a commercial center type of consumer (a large hypermarket). We have also acquired a dataset containing the historical hourly meteorological data regarding the temperature $(T_d(t_j))_{j=1,\ldots,24}$ (measured in Celsius degrees). The peak demand observed for the day d is defined as

$$P_d = \max_{j=1,\ldots,24} E_d(t_j).$$

Figure 2.3 provides a sample of 30 curves of hourly temperature measures and the associated electricity consumption curves. Therefore, our sample is formed as follows $(T_d, P_d)_{d=1,\ldots,366}$, where T_d is the predicted temperature curve for the day d and P_d the peak demand observed for the day d. We split our sample of 366 days into a learning sample containing the first 300 days and a testing sample with the last 66 days.

The choice of the kernel, the semi-metric and the bandwidth are similar to these used in the maximum ozone application.

Figure 2.4 shows the results obtained for the peak energy prediction derived from the testing sample for the three models (classical, robust and robust with scale parameter method). The associated MSE are 0.00102, 0.00098 and 0.00072 respectively for each methods.

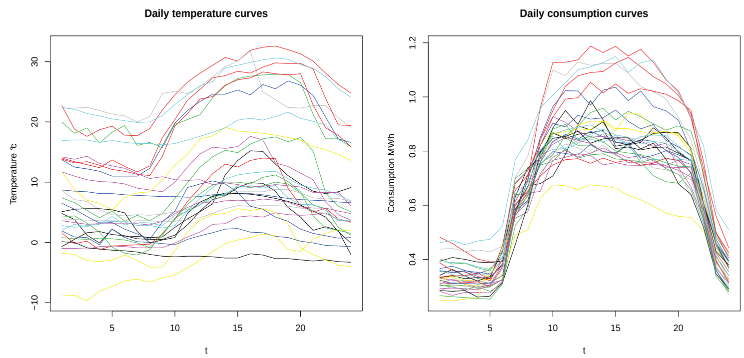

FIGURE 2.3 – Sample of 30 daily temperature curves and the associated energy consumption curves

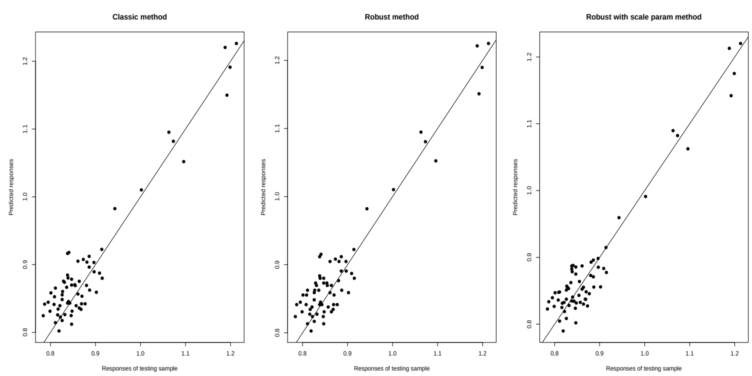

FIGURE 2.4 – Prediction of the last 66 days by three models

2.5 Conclusion

We provide in this paper the uniform almost complete consistency with rates of the robust regression function in case of unknown scale parameter. These results were obtained under sufficient standard conditions that allows one to explore different structural axes of the subject, such as the functional naturalness of the model and the data as well as the robustness of the regression function and the dependency of the observation. The real data applications (Maximum Ozone Concentration, Peak electricity demand) have also highlighted several attractive features of the robust regression with unknown scale parameter estimator. In terms of mean squared error (MSE) the proposed estimator performs competitively in comparison to existing estimators with know scale parameter.

Based on the experience of this paper on robust regression with unknown scale parameter, we guess that most of the techniques using nonparametric functional kernel smothers could be efficiently extended. For instance, challenging open questions in this sense could concern as well extensions to other forms of nonparametric predictors (like functional local linear ones, functional kNN ones, and many other ones $\cdots$). Extensions to other kinds of prediction models in which a preliminary kernel stage plays a crucial role (this would include many semiparametric regression models like functional single index models, and partial linear models, and many other ones $\cdots$).

Acknowledgment :

The authors would like to thank the Editor-in-Chief of AAM, and the anonymous reviewers for their valuable comments and suggestions which improved substantially the quality of this paper.

Bibliographie

[1] Attouch, M., Laksaci, A., and Ould-Said, E.(2010). Asymptotic normality of a robust estimator of the regression function for functional time series data, J. of the Korean Statist. Soc., Vol. 39, No. 4, pp. 489-500.

[2] Attouch, M., Laksaci, A., and Ould Said, E. (2012). Strong uniform convergence rate of a robust estimator of the regression function for functional and dependent processes, Journal of Japan Statistical Society, Vol. 42, No. 3, 125-143.

[3] Attouch, M.K., Kaid, Z., and Louhab, H.(2019) : Asymptotic normality of a robust kernel estimator of the regression function for functional ergodic data : case of an unknown scale parameter, C. R. Math. Acad. Sci. Paris, Vol. 357, No. 5, pp. 478-481.

[4] Azzedine, N., Laksaci, A. and Ould-Said, E. (2008). On robust nonparametric regression estimation for a functional regressor, Statistics & Probability Letters, Vol. 78, No. 18, 3216-3221.

[5] Benhenni, K., Ferraty, F., Rachdi, M. and Vieu, P. (2007). Local smoothing regression with functional data, Comput. Statist., Vol.22, No. 3, 353-369.

[6] Boente, G. and Fraiman, R. (1989). Robust nonparametric regression estimation for dependent observations, Annals of Statistics, Vol. 17, No. 3, 1242-1256.

[7] Boente, G. and Fraiman, R. (1989). Nonparametric regression estimation, J. Multivariate Anal. Vol. 29, No. 2, 180-198.

[8] Boente, G. and Fraiman, R. (1991). Strong Uniform Convergence Rates for Some robust equivariant nonparametric regression estimates for mixing processes, International Statistical Institute, Vol. 59, No. 3, 355-372.

[9] Boente, G. and Vahnovanb, A. (2015). Strong convergence of robust equivariant nonparametric functional regression estimators, Statistic & Probability Letters, Vol. 100, 1-11.

[10] Chikobvu, D., and Sigauke, C. (2012). Regression-SARIMA modelling of daily peak electricity demand in South Africa, Journal of Energy in Southern Africa, Vol. 23, No. 3, 23-30.

[11] Collomb,G. (1982). Prédiction non paramétrique : Etude de l'erreur quadratique du préedictogramme, C. R.Acad. Sci. Paris Sér. I Math, Vol. 294, No. 4, 59-62.

[12] Ferraty, F., Laksaci, A. and Vieu, Ph. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models, Statistical Inference for Stochastic Processes, Vol. 9, No. 3, 47-76.

[13] Ferraty, F., Laksaci, A., Tadj, A. and Vieu, Ph. (2010). Rate of uniform consistency for nonparametric estimates with functional variables, Journal of statistical planning and inference, Vol. 140, No. 2, 335-352.

[14] Ferraty, F., Vieu, Ph., (2006). *Nonparametric functional data analysis.* Springer-Verlag Publisher, United States of America.

[15] Gheriballah, A. Laksaci, A. and Sekkal, S. (2013). Nonparametric M-regression for functional ergodic data, Journal Statistics et Probability Letters, Vol. 83, No.3, 902-908.

[16] Goia, A. May, C. and Fusai, G. (2010). Functional clustering and linear regression for peak load forecasting, International Journal of Forecasting, Vol. 26, No. 4, 700-711.

Statistics, Vol. 35, No. 1, 73-101.

[17] Hyndman, Rob J., and Fan, S. (2010). Density Forecasting for Long-Term Peak Electricity Demand, IEEE Transaction on Power Systems, Vol. 25, No. 2, 1142-1153.

[18] Kolmogorov, A. N. (1956). Asymptotic characteristics of some completely bounded metric spaces, Doklady Akademii Nauk SSSR, Vol. 1, No. 8, 585-589.

[19] Kolmogorov, A. N. and Tikhomirov, V .M. (1959). ϵ-entropy and ϵ-capacity, Uspekhi Mat. Nauk, Vol. 14, 3-86. (English version in, Transl. Amer. Math. Soc. Transl. Ser, Vol. 2, No.1, 277-364 (1961).

[20] Masry, E. (2005). Nonparametric regression estimation for dependent functional data : asymptotic normality, Stoch. Proc. and their Appl., Vol. 115, No. 1, 155-177.

[21] Rachdi M, Laksaci A, Demongeot J, Abdali A, Madani F. (2014). Theoretical and practical aspects of the quadratic error in the local linear estimation of the conditional density for functional data. Comput. Statist Data Anal., Vo.73, No. 2, 53-68.

[22] Ramsay, J. and Silverman, B. (2002). *Applied functional data analysis : methods and case studies.* Springer.

[23] Ramsay, J. and Silverman, B. (2005). *Functional data analysis.* Springer Series in Statistics.

[24] C. J. Stone. (2005). Nonparametric M-regression with free knot splines, J. Stat. Plan. Inf., Vol. 130, No. 1-2, 183-206.

[25] Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblements dépendants*. (In french). Mathématiques & Applications, Vol. 31, Springer Verlag, New-York.

Chapitre 3

Asymptotic normality of a recursive estimation of the conditional hazard function with missing at random

[1]**Mebsout Mokhtaria,** [2,⋆]**Attouch Mohammed Kadi and** [3]**Mechab Boubakar**

[1,2,3] Laboratoire de Statistique Processus Stochastiques
Djillali Liabes University
Sidi Bel Abbes, 22000
Algeria
[1]mokhtariamebsout@yahoo.com ; [3]mechaboub@yahoo.fr

[2]attou_kadi@yahoo.fr ;

⋆Corresponding Author

Abstract

In this paper, we study the the recursive kernel estimator of the conditional hazard function with missing at random whenever functional stationary ergodic data are considered. Under the assumption of ergodicity, the novelty of our approach is that we do not require independence of the observations. It is shown that, under some wild conditions, the recursive kernel estimate of the three parameters (conditional density, conditional distribution and conditional hazard) are asymptotically normally distributed..

Keywords : Asymptotic normality, Conditional hazard function, Functional data, Recursive estimate, Stationary ergodic processes.

3.1 Introduction

Recently there has been an increasing interest in the study of functional data. For an overview of the present state on nonparametric functional data (FDA), we refer to the works of [13] and [11], and the references therein.
Conditional hazard estimation with a functional explanatory variable and a scalar response acquired considerable interest in the statistical literature. The first work was proposed by Ferraty et al. [1], where they introduce a kernel estimator and prove some asymptotic properties (with rates) in various situations including censored and/ or dependent variables. Quintela-del-Río [10] extended the results of Ferraty et al. [1] by calculating the bias and variance of these estimates, and establishing their asymptotic normality. In the case of completely observed data, another estimators have been proposed for the conditional hazard function by different approaches. In 2014, Attouch and Belabed [2] have studied the nonparametric estimator of the conditional hazard function using the k Nearest Neighbors (k-NN) estimation method and they

have shown its asymptotic properties in the case of independent data. Another approach has been proposed by Massim and Mechab [12] based on the local linear method, they have established the almost complete convergence of the proposed estimator.
In the case of the functional spatial data, the works on the conditional hazard function is limited and we can refer to [9] where they studied the almost complete convergence of the kernel type estimate. These authors studied in [8] the mean squared convergence rate and proved the asymptotic normality of the proposed estimator.
In recent years, the statistical modeling for functional ergodic data has been an increasing interest and a great importance in various fields. The general framework of ergodic functional data has been initiated by Laïb and Louani [6],[7] who stated consistencies with rates together with the asymptotic normality of the regression function estimate. In this topic of ergodic data, several works have been published, for example, Gheriballah et al. [4] showed the almost complete convergence (with rate) of a family of robust nonparametric estimators for regression function. More recently, Ardjoun et al. [5] treated the almost complete convergence and the asymptotic normality of the estimator of conditional mode, Benziadi et al. [3] studied the almost complete rate convergence of functional recursive kernel of the conditional quantiles.
This paper is organized as follows : Section 2 introduces the estimator of the conditional hazard function. In section 3 we will define some notations and hypothesis. The asymptotic normality of the proposed estimator of the conditional hazard function is given in Section 4. Finally, the proofs of our results are given in the Appendix.

3.2 The recursive estimation of the conditional hazard function with missing at random

Let $(X_i, Y_i)_{i=1,\dots,n}$ be a sequence of strictly stationary ergodic processes. Where X_i are values in semi-metric space $(\mathcal{F}, d)$ and Y_i are real-valued random variables. N_x will denote a fixed neighborhood of x. We assume that the regular version of the conditional probability of Y given X exists. Moreover, we suppose that, for all $x \in N_x$ the conditional distribution function of Y given $X = x$, the recursive kernel estimator of the conditional hazard function $h^x(y)$ such that

$$h^x(y) = \frac{f^x(y)}{1 - F^x(y)}, \quad \text{for } y \in \mathbb{R} \text{ and } F^x(y) < 1$$

is

$$\widehat{h}^x(y) = \frac{\widehat{f}^x(y)}{1 - \widehat{F}^x(y)}, \quad \forall y \in \mathbb{R}.$$

We define the recursive kernel estimator of the conditional distribution function by

$$\widehat{F}^x(y) = \frac{\sum_{i=1}^{n} K(a_i^{-1} d(x, X_i)) H(b_i^{-1}(y - Y_i))}{\sum_{i=1}^{n} K(a_i^{-1} d(x, X_i))}$$

The recursive kernel estimator $\widehat{f}^x(.)$ of $f^x(.)$ is given by

$$\widehat{f}^x(y) = \frac{\sum_{i=1}^{n} b_i^{-1} K(a_i^{-1} d(x, X_i)) H'(b_i^{-1}(y - Y_i))}{\sum_{i=1}^{n} K(a_i^{-1} d(x, X_i))}.$$

Where K is the kernel, H is a strictly increasing distribution function, H' is the derivative of H and a_i, b_i are a sequences of positive real numbers such that $\lim_{n \to +\infty} a_n = \lim_{n \to +\infty} b_n = 0$.
However, in the cas of missing random for the response variable, an available incomplete sample of size n from (X, Y, δ) is $\{(X_i, Yi, \delta i), 1 \le i \le n\}$; where X_i is observed completely, $\delta i = 1$ if Yi is observed, and $\delta i = 0$ otherwise. Meanwhile the Bernoulli random variable δ is stisfied with

$$\mathbb{P}(\delta = 1 | X = x, Y = y) = \mathbb{P}(\delta = 1 | X = x) = p(x)$$

Where $p(x)$ is a function operator, which is called the coditional probability of the observing response given the predictor and is often unknown. this mechanism shows that δ and Y are conditionally independent given X.

Therfore, the estimator of $F^x(.)$ and $f^x(.)$ are given by :

$$\widehat{F}^x(y) = \frac{\sum_{i=1}^{n} \delta i K(a_i^{-1} d(x, X_i)) H(b_i^{-1}(y - Y_i))}{\sum_{i=1}^{n} \delta_i K(a_i^{-1} d(x, X_i))} \tag{3.2.1}$$

and

$$\widehat{f}^x(y) = \frac{\sum_{i=1}^{n} \delta i b_i^{-1} K(a_i^{-1} d(x, X_i)) H'(b_i^{-1}(y - Y_i))}{\sum_{i=1}^{n} \delta_i K(a_i^{-1} d(x, X_i))}. \tag{3.2.2}$$

3.3 Notations and hypothesis

The assumptions that we will need in our study are the following :
The functional ergodic data is carried out by the following consideration : for $i = 1, ..., n$, we put $\mathcal{F}_k$ is the σ-algebra generated by $((X_1, Y_1), ..., (X_k, Y_k))$. We pose $\mathfrak{B}_k$ is the σ-algebra generated by $((X_1, Y_1), ..., (X_k, Y_k), X_{k+1})$.
We suppose that the strictly stationary ergodic process $(X_i, Y_i)_{i\in\mathbb{N}}$ satisfies

(H1) (i) The function $\phi(x, h) := \mathbb{P}(X \in B(x, h)) > 0, \forall h > 0,$ where $B(x, h) := \{x' \in \mathcal{F}/d(x', x) < h\}$.

ii) For all $i = 1, ..., n$ there exists a deterministic function $\phi_i(x, .)$ such that almost surely
$0 < \mathbb{E}^{\mathcal{F}_{i-1}}[K(X_i)] = \mathbb{P}(X_i \in B(x, h)|\mathcal{F}_{i-1}) \le \phi_i(x, h)$,
$\forall \ h > 0$ and $\phi_i(x, h) \to 0$ as $h \to 0$.

(iii) For all sequence $(h_i)_{i=1,...,n} > 0$,
$$\frac{1}{\sum_{i=1}^{n} \phi(x, h_i)} \sum_{i=1}^{n} \mathbb{P}(X_i \in B(x, h_i)|\mathcal{F}_{i-1}) \to 1.$$

(H2) (i) Let $\mathcal{S}$ be a compact set of $\mathbb{R}$, the conditional distribution function $F^x(.)$ is such that, $\forall y \in \mathcal{S}, \exists \beta > 0, \inf_{y\in\mathcal{S}}(1 - F(y|x)) > \beta, \forall (y_1, y_2) \in \mathcal{S} \times \mathcal{S}, \forall (x_1, x_2) \in N_x \times N_x$

$$|F(y_1|x_1) - F(y_2|x_2)| \le C_1(d(x_1, x_2)^{\beta_1} + |y_1 - y_2|^{\beta_2})$$

with $C_1 > 0, \beta_1 > 0, \beta_2 > 0$.

(ii) The density $f^x(.)$ is such that,$\forall y \in \mathcal{S}, \ \exists \alpha > 0 \ \ f^x(y) < \alpha, \forall (y_1,y_2)\in \mathcal{S} \times \mathcal{S}, \forall \ (x_1,x_2)\in N_x \times N_x$

$$|f^{x_1}(y_1) - f^{x_1}(y_2)| \le C_2(d(x_1, x_2)^{\beta_1} + |y_1 - y_2|^{\beta_2})$$

with $C_2 > 0, \beta_1 > 0, \beta_2 > 0$.

(H3) $\forall (y_1, y_2) \in \mathbb{R}^2 \quad |H^{(j)}(y_1) - H^{(j)}(y_2)| \leq C|y_1 - y_2|$, for $j = 0, 1$
$\displaystyle\int |t|^{\beta_2} H^{(1)}(t)dt < \infty$ and $\displaystyle\int H'^2(t)dt < \infty$.

(H4) K is a function with support (0,1) such that $0 < C_1 \mathbb{I}_{[0.1]} < K(t) < C_2 \mathbb{I}_{[0.1]} < \infty$.
where $\mathbb{I}_A$ is the indicator function

(H5) The bandwidths (a_i, b_i) satisfied : $\forall t \in [0, 1]$

(i) $$\lim_{n\to+\infty} \frac{\sum_{i=1}^n \phi(x, ta_i)}{\sum_{i=1}^n \phi(x, a_i)} = \tau_x(t)$$

where

$$K^2(1) - \int_0^1 (K^2(u))'\tau_x(u)du > 0 \text{ and } K(1) - \int_0^1 K'(u)\tau_x(u)du \neq 0$$

(ii) $$\lim_{n\to+\infty} \frac{\sqrt{n\phi_n(x)}}{\varphi_n(x)} \left(\sum_{i=1}^n a_i^{\beta_1} \phi(x, a_i) + \sum_{i=1}^n b_i^{\beta_2} \phi(x, a_i) \right) = 0$$

where $\varphi_n(x) = \mathbb{E}[K(X_1)] = n^{-1} \displaystyle\sum_{i=1}^n \phi(x, a_i)$.

and $\mathbb{E}[K^j(X_1)] = K^j(1)\phi(x, a_i) - \displaystyle\int_0^1 (K^j(u))'\phi(x, ua_i)du$

COMMENTS ON HYPOTHESES :

The condition (H1) involves the ergodic nature of the data and the small ball techniques used in this paper. The hypothesis (H1)(iii) is a direct consequence of Beck's theorem. The assumption (H2) presents the Lipschitz's condition to the conditional distribution function and conditional density function, it means that the both functions are continuous with respect to each variable and permits us to evaluate the bias term without using the differentiability. Hypothesis (H3) impose some regularity conditions upon the distribution function H used in our estimates. The condition (H4) is very standard in nonparametric function estimation. The assumptions (H5) is a technical condition.

Theorem 3.3.1 *Under hypotheses (H1)-(H5), we have for all $x \in \mathcal{A}$*

$$\left(\frac{n\varphi_n(x)}{p(x)^2 \sigma_h^2(x, y)} \right)^{1/2} (\widehat{h}^x(y) - h^x(y)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } n \to \infty \tag{3.3.1}$$

where $\mathcal{A} = \{x, \quad \sigma_h^2(x, y) \neq 0\}$ and $\sigma_h^2(x, y) = \dfrac{\alpha_2 h^x(y)}{\alpha_1^2(1 - F^x(y))}$
with $\alpha_1 = K(1) - \int_0^1 K'(s)\beta_x(s)ds$ and $\alpha_2 = K^2(1) - \int_0^1 (K^2(s))'\beta_x(s)ds$.

proof 3.3.1 *The proof of theorem is based on the following decomposition and lemmas bellow :*

$$\widehat{h}^x(y) - h^x(y) = \frac{1}{1 - \widehat{F}^x(y)} \left[\widehat{f}^x(y) - f^x(y) \right] + \frac{h^x(y)}{1 - \widehat{F}^x(y)} \left[\widehat{F}^x(y) - F^x(y) \right].$$

lemma 3.3.1 *Under hypothesis of the theorem 3.3.1, we have*

$$\left(\frac{n\varphi_n(x)}{p(x)^2\sigma_F^2(x,y)}\right)^{1/2}(\widehat{F}^x(y)-F^x(y)) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad as \quad n\to\infty \tag{3.3.2}$$

where $\sigma_F^2(x,y)=\dfrac{p(x)\alpha_2F^x(y)\big(1-F^x(y)\big)}{\alpha_1^2}$.

lemma 3.3.2 *Under hypothesis of the theorem 3.3.1, we have*

$$\left(\frac{n\varphi_n(x)}{p(x)^2\sigma_f^2(x,y)}\right)^{1/2}(\widehat{f}^x(y)-f^x(y)) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad as\ n\to\infty \tag{3.3.3}$$

where $\sigma_f^2(x,y)=\dfrac{p(x)\alpha_2f^x(y)}{\alpha_1^2}$.

proof 3.3.2 *3.3.1 :*
The proof of this lemma is based on the following decomposition in the following, we put, for any $x\in\mathcal{F}$, *and* $i=1,....,n$
$K_i=K(a_i^{-1}d(x,X_i))$ *and* $H_i=H(b_i^{-1}(Y_i-y))$.
We start by writing

$$\widehat{F}^x(y)-F^x(y)=\widehat{B}_{n,1}(x,y)+\frac{\widehat{R}_{n,1}(x,y)}{\widehat{F}_D(x)}+\frac{\widehat{Q}_{n,1}(x,y)}{\widehat{F}_D(x)}$$

where

$$\begin{aligned}
\widehat{Q}_{n,1}(x,y) &= (\widehat{F}_N^x(y)-\mathbb{E}[\widehat{F}_N^x(y)])-F^x(y)(\widehat{F}_D(x)-\mathbb{E}[\widehat{F}_D(x)]),\\
\widehat{B}_{n,1}(x,y) &= \frac{\mathbb{E}[\widehat{F}_N^x(y)]}{\mathbb{E}[\widehat{F}_D(x)]} \quad and \quad \widehat{R}_{n,1}(x,y)=-\widehat{B}_{n,1}(x,y)(\widehat{F}_D(x)-\mathbb{E}[\widehat{F}_D(x)])
\end{aligned}$$

with

$$\begin{aligned}
\widehat{F}_N^x(y) &= \frac{1}{n\varphi_n}\sum_{i=1}^n\delta_iK(a_i^{-1}d(x,X_i))H(b_i^{-1}(Y_i-y_i)),\\
\mathbb{E}[\widehat{F}_N^x(y)] &= \frac{1}{n\varphi_n}\sum_{i=1}^n\mathbb{E}^{\mathcal{F}_{i-1}}[\delta_iK(a_i^{-1}d(x,X_i))H(b_i^{-1}(Y_i-y_i))],\\
\widehat{F}_D(x) &= \frac{1}{n\varphi_n}\sum_{i=1}^n\delta_iK(a_i^{-1}d(x,X_i)),\\
\mathbb{E}[\widehat{F}_D(x)] &= \frac{1}{n\varphi_n}\sum_{i=1}^n\mathbb{E}^{\mathcal{F}_{i-1}}[\delta_iK(a_i^{-1}d(x,X_i))].
\end{aligned}$$

The proof of the lemma 3.3.1 is a consequence of the following lemmas, whose proofs are given in the Appendix

lemma 3.3.3 *Under the hypothesis of theorem 3.3.1*

$$\left(\frac{n\varphi_n(x)}{p(x)^2\sigma_F^2}\right)^{1/2}\widehat{Q}_{n,1}(x,y)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1)\quad as\ n\to\infty.$$

Corollary 3.3.1.1 *Under the hypothesis of theorem 3.3.1*

$$n\varphi_n(x)Var[\widehat{F}_N^x(y)]\longrightarrow\frac{p(x)\alpha_2F^x(y)\big(1-F^x(y)\big)}{\alpha_1^2}$$

lemma 3.3.4 *Under the hypothesis (H1) and (H5), we have*

$$\widehat{F}_D(y|x)-1=o_p(1).$$

lemma 3.3.5 *Under the hypothesis (H1), (H2) and (H3) we have*

$$\left(\frac{n\varphi_n(x)}{p(x)^2\sigma_F^2}\right)^{1/2}\widehat{B}_{n,1}(x,y)=o_p(1)\ \ as\ n\to\infty.$$

lemma 3.3.6 *Under the hypothesis (H1), (H2) and (H5) we have*

$$\left(\frac{n\varphi_n(x)}{p(x)^2\sigma_F^2}\right)^{1/2}\widehat{R}_{n,1}(x,y)=o_p(1)\ \ as\ n\to\infty.$$

proof 3.3.3 *3.3.2*
The proof is based on the following decomposition and lemmas bellow

$$\widehat{f}(y|x)-f(y|x)=\widehat{B}_{n,2}(x,y)+\frac{\widehat{R}_{n,2}(x,y)}{\widehat{F}_D(x)}+\frac{\widehat{Q}_{n,2}(x,y)}{\widehat{F}_D(x)}\tag{3.3.4}$$

where

$$\widehat{Q}_{n,2}(x,y)\ =\ (\widehat{f}_N^x(y)-\mathbb{E}[\widehat{f}_N^x(y)]-f_N^x(y)(\widehat{F}_D(x)-\mathbb{E}[\widehat{F}_D(x)]),$$

$$\widehat{B}_{n,2}(x,y)\ =\ \frac{\mathbb{E}[\widehat{f}_N^x(y)]}{\mathbb{E}[\widehat{F}_D(x)]}-f_N^x(y)\ \ and\ \ \widehat{R}_{n,2}(x,y)=-\widehat{B}_{n,2}(x,y)\left(\widehat{f}_N^x(y)-\mathbb{E}[\widehat{f}_N^x(y)]\right)$$

with

$$\widehat{f}_N^x(y)\ =\ \frac{1}{n\varphi_n}\sum_{i=1}^n\frac{\delta_ib_i^{-1}K(a_i^{-1}d(x,X_i))H'(b_i^{-1}(Y_i-y_i))}{\mathbb{E}\left(\delta_iK(a_i^{-1}d(x,X_i))\right)},$$

$$\mathbb{E}[\widehat{f}_N^x(y)]\ =\ \frac{1}{n\varphi_n}\sum_{i=1}^n\frac{\mathbb{E}^{\mathcal{F}_{i-1}}[\delta_ib_i^{-1}K(a_i^{-1}d(x,X_i))H'(b_i^{-1}(Y_i-y_i))]}{\mathbb{E}\left(\delta_iK(a_i^{-1}d(x,X_i))\right)}.$$

lemma 3.3.7 *Under the hypothesis of theorem 3.3.1*

$$\left(\frac{n\varphi_n(x)}{p(x)^2\sigma_f^2}\right)^{1/2}\widehat{Q}_{n,2}(x,y)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1)\quad as\ n\to\infty.$$

Corollary 3.3.1.2 *Under the hypothesis of theorem 3.3.1*

$$n\varphi_n(x)Var[\widehat{f}_N^x(y)]\longrightarrow\frac{p(x)\alpha_2 f^x(y)}{\alpha_1^2}$$

lemma 3.3.8 *Under the hypothesis (H1), (H2) and (H3) we have*

$$\left(\frac{n\varphi_n(x)}{p(x)^2\sigma_f^2}\right)^{1/2}\widehat{B}_{n,2}(x,y)=o_p(1)\ \ as\ n\to\infty.$$

lemma 3.3.9 *Under the hypothesis (H1), (H2) and (H5) we have*

$$\left(\frac{n\varphi_n(x)}{p(x)^2\sigma_f^2}\right)^{1/2}\widehat{R}_{n,2}(x,y)=o_p(1)\ \ as\ n\to\infty.$$

3.4 Appendix

proof 3.4.1 *3.3.1.1 :*
We bigin by the proof of the Corollary 3.3.1.1 For $i=1,\cdots,n$ *we consider the quantities* $K_i(x)=K(a_i^{-1}d(x,X_i)), H_i(y)=H(b_i^{-1}d(x,X_i))$ *let* $\widehat{F}_N^x(y)$ *(resp.)* $\widehat{F}_N^x(y)$ *be defined as*

$$\begin{aligned}\widehat{F}_N^x(y) &= \frac{1}{n\mathbb{E}K_1}\sum_{i=1}^n \delta_i K_i(x)H_i(y),\\ \mathbb{E}\left[\widehat{F}_N^x(y)\right] &= \frac{1}{n\mathbb{E}K_1}\sum_{i=1}^n \mathbb{E}^{\mathcal{F}_{i-1}}\left[\delta_i K_i(x)H_i(y)\right],\end{aligned}$$

By the definition of $\widehat{F}_N^x(y)$ *it follows that*

$$Var\left[\widehat{F}_N^x(y)\right]=\frac{1}{\left[n\mathbb{E}K_1\right]^2}Var\left[\sum_{i=1}^n\Gamma_i(x)\right]$$

where

$$\Gamma_i(x)=\delta_i K_i(x)H_i(y)-\mathbb{E}\left[\delta_i K_i(x)H_i(y)\right]$$

Now, we need to evaluate the variance of $\widehat{F}_N^x(y)$
we have that

$$\begin{aligned}
Var\left[\widehat{F}_N^x(y)\right] &= \frac{1}{\left[n\mathbb{E}K_1(x)\right]^2}\left[\sum_{i=1}^{n} Var\left[\Gamma_i(x)\right] + \sum_{i\neq j} Cov(\Gamma_i(x),\Gamma_j(x))\right] \\
&= \frac{1}{n\left[\mathbb{E}K_1(x)\right]^2} Var\left[\Gamma_1(x)\right] + \frac{1}{\left[n\mathbb{E}K_1(x)\right]^2}\sum_{i\neq j} Cov(\Gamma_i(x),\Gamma_j(x)) \\
&:= R_1 + R_2
\end{aligned}$$

We steady the term R_1
Let us calculate the quantity $Var\left[\Gamma_1(x)\right]$ *We have :*

$$\begin{aligned}
Var\left[\Gamma_1(x)\right] &= \mathbb{E}\left(\delta_1^2 K_1^2(x) H_1^2(y)\right) - \left[\mathbb{E}\left(\delta_1 K_1(x) H_1(y)\right)\right]^2 \\
&= p(x)\left[\mathbb{E}\left(K_1^2(x)H_1^2(y)\right) - \left[\mathbb{E}\left(K_1(x)H_1(y)\right)\right]^2\right] \\
&= p(x)\left[\mathbb{E}\left[K_1^2(x)\right]\frac{\mathbb{E}\left[K_1^2(x)H_1^2(y)\right]}{\mathbb{E}\left[K_1^2(x)\right]} - \left[\mathbb{E}K_1(x)\right]^2\left[\frac{\mathbb{E}\left[K_1(x)H_1(y)\right]}{\mathbb{E}K_1(x)}\right]^2\right]
\end{aligned}$$

So, by using the same arguments as those used by (Ferraty and Vieu, 2007), we get

$$\frac{1}{\phi(x,a_i)}\mathbb{E}\left[K_1^2(x)\right] \longrightarrow K^2(1) - \int_0^1 (K^2(u))'\tau_x(u)du + o(1)$$
$$\frac{\mathbb{E}\left[K_1^2(x)H_1^2(y)\right]}{\mathbb{E}\left[K_1^2(x)\right]} \longrightarrow F^x(y)(1-F^x(y))\int H'^2(t)dt$$

So we can write

$$\frac{\mathbb{E}\left[K_1^2(x)H_1^2(y)\right]}{\mathbb{E}\left[K_1^2(x)\right]} \longrightarrow F^x(y)(1-F^x(y))$$

which imply that

$$\frac{\phi(x,a_i)}{n\left[\mathbb{E}K_1(x)\right]^2} Var\left[\Gamma_1(x)\right] \longrightarrow \sigma_F^2(x,y)$$

So

$$n\varphi_n(x)Var[\widehat{F}_N^x(y)] \longrightarrow \sigma_F^2(x,y)$$

Now, we steady the term R_2*, we put two sets. Next, Following Masry (2005), we define the sets*

$$S_1 = \{(i,j) : 1 < |i-j| \le C_n\} \quad , \quad S_2 = \{(i,j) : C_n + 1 \le |i-j| \le n-1\}$$

Where C_n is a sequence of real numbers tends to ∞ as $n \to \infty$. We split the sum as follows :

$$\begin{aligned} R_2 &= \frac{1}{n^2(\varphi_n(x))^2} \sum_{(i,j)\in S_1} Cov(\delta_i K_i(x) H_i(y), \delta_j K_j(x) H_j(y)) \\ &+ \frac{1}{n^2(\varphi_n(x))^2} \sum_{(i,j)\in S_2} Cov(\delta_i K_i(x) H_i(y), \delta_j K_j(x) H_j(y)) \\ &:= J_1 + J_2 \end{aligned}$$

For J_1, Conditioning on .$(X_i; X_j)$, we have

$$\begin{aligned} Cov(\delta_i K_i(x) H_i(y), \delta_j K_j(x) H_j(y)) &\le \mathbb{E}\left[\mathbb{E}(H_i(y) H_j(y)/(X_i; X_j)) K_i(x) K_j(x)\right] \\ &\le \mathbb{E}(K_i(x) K_j(x)) \\ &\le C\mathbb{P}\left((X_i; X_j) \in B(x, a_i) \times (B(x, a_i))\right. \\ &\le C n C_n \varphi_n(a_i)^{\frac{a+1}{a}} n^{\frac{-1}{a}} \end{aligned}$$

Finaly, we have :

$$J_1 \le C C_n \varphi_n(a_i)^{\frac{1-a}{a}} n^{\frac{-a-1}{a}}$$

For J_2,we have,

$$J_2 = \sum_{(i,j)\in S_2} Cov(\delta_i K_i(x) H_i(y), \delta_j K_j(x) H_j(y))$$

we use Davydov-Rio's inequality (Rio, 2000, p. 87) for mixing processes. This leads, for all $i \ne j$, to

$$Cov(\delta_i K_i(x) H_i(y), \delta_j K_j(x) H_j(y)) \le 4\alpha(|i-j|)$$

finaly

$$J_2 \le C n^2 m_n^{-a}$$

So

$$\sum_{(i,j)\in S_2} Cov(\delta_i K_i(x) H_i(y), \delta_j K_j(x) H_j(y)) = C n m_n \phi(h)^{\frac{a+1}{a}} n^{\frac{-1}{a}} + \tilde{C} n^2 m_n^{-a}$$

choosing $\quad m_n = \left(\frac{\phi(h)}{n}\right)^{-\frac{1}{a}}$

So

$$\begin{aligned} J_1 + J_2 &= Cn\phi(h)^{-\frac{1}{a}} n^{\frac{1}{a}} \phi(h)^{\frac{a+1}{a}} n^{\frac{-1}{a}} + \tilde{C}n^2 \left(\left(\frac{\phi(x)}{n}\right)^{-\frac{1}{a}}\right)^{-a} \\ &= Cn\phi(h) + \frac{\tilde{C}n}{\phi(h)} \\ &= O\left(\frac{Cn}{\phi(h)}\right) \end{aligned}$$

donc the evidence above gave as the result of the corollary 3.3.1.1

proof 3.4.2 *3.3.3*

We use the same ideas in Laib and Louani (2011).
For all $i = 1, ..., n$, we define

$$\eta_{ni} = \left(\frac{n\varphi_n(x)}{p(x)^2\sigma_F^2}\right)^{\frac{1}{2}} (H_i(y) - F(y|x)) \frac{\delta_i K_i(x)}{n\varphi_n(x)}$$

and, we define $\xi_{ni} = \eta_{n_i} - \mathbb{E}^{\mathcal{F}_{i-1}}(\eta_{n_i})$.
It is easily seen that

$$\left(\frac{n\varphi_n(x)}{p(x)^2\sigma_F^2}\right)^{\frac{1}{2}} \widehat{Q}_{n,1}(x,y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{ni}$$

as ξ_{ni} is a triangular array of martingale differences according the σ-algebra $\mathcal{F}_{i-1}$, we are in position to apply the central limit theorem based on unconditional lindeberg condition to establish the asymptotic normality of $\widehat{Q}_{n,1}(x,y)$. This can be done if we establish the following statements :

a) $\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}^{\mathcal{F}_{i-1}}\left(\xi_{ni}^2\right) \longrightarrow 1$ *in probability*

b) $\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(\xi_{ni}^2 \mathbb{I}_{\xi_{ni}^2 > \epsilon n}\right) \longrightarrow 0$ *holds for any $\epsilon > 0$ (Lindeberg condition).*

proof 3.4.3 *of the part (a)*

$$\begin{aligned} \mathbb{E}^{\mathcal{F}_{i-1}}\left(\xi_{ni}^2\right) &= \mathbb{E}\left(\left(\eta_{ni} - \mathbb{E}^{\mathcal{F}_{i-1}}(\eta_{ni})\right)^2\right) \\ &= \mathbb{E}^{\mathcal{F}_{i-1}}\left(\eta_{ni}^2\right) - (\mathbb{E}^{\mathcal{F}_{i-1}}(\eta_{ni}))^2. \end{aligned}$$

The statement (a) follows then if we show that :

$$\text{(1)} \lim_{n\longrightarrow\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}^2\big(\eta_{ni}|\mathcal{F}_{i-1}\big)=0 \textit{ in probability,}$$

$$\text{(2)} \lim_{n\longrightarrow\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}^{\mathcal{F}_{i-1}}\left(\eta_{ni}^2\right)=1 \textit{ in probability.}$$

To prove (1), we put $A=\left(\frac{n\varphi_n(x)}{p(x)^2\sigma_F^2}\right)^{\frac{1}{2}}$*, we have*

$$\begin{aligned}
\mathbb{E}^{\mathcal{F}_{i-1}}(\eta_{ni}) &= \frac{A}{n\varphi_n(x)}\mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_iK_i(x)\big(H_i(y)-F(y|x)\big)\right)\\
\left|\mathbb{E}^{\mathcal{F}_{i-1}}(\eta_{ni})\right| &= \frac{A}{n\varphi_n(x)}\left|\mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_iK_i(x)\left[\big(\mathbb{E}^{\mathfrak{B}_{i-1}}\big(H_i(y)\big)-F(y|x)\big)\right]\right)\right|\\
\left|\mathbb{E}^{\mathcal{F}_{i-1}}(\eta_{ni})\right| &= \frac{A}{n\varphi_n(x)}\left|\mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_1K_1(x)\left[\big(\mathbb{E}^{X_i}\big(H_1(y)\big)-F(y|x)\big)\right]\right)\right|\\
&= \frac{A}{n\varphi_n(x)}p(x)\left|\mathbb{E}^{\mathcal{F}_{i-1}}\left(K_1(x)\left[\big(\mathbb{E}^{X_i}\big(H_1(b_i(y-Y_i))\big)-F(y|x)\big)\right]\right)\right|
\end{aligned}$$

under (H1) and (H4), we have

$$C\phi_i(x,a_i)\leq\mathbb{E}\big(K_i(x)|\mathcal{F}_{i-1}\big)\leq C'\phi_i(x,a_i)$$

Next, an integration by parts and a change of variable allow to get

$$\mathbb{E}^{X_i}(H_1(b_i(y-Y_i))=\int_{\mathbb{R}}H^{(1)}(t)F(y-b_it|X_i)dt.$$

Otherwise

$$\begin{aligned}
\left|\mathbb{E}^{X_i}(H_1(b_i(y-Y_i))-F(y|x)\right| &= \left|\int_{\mathbb{R}}H^{(1)}(t)F^{X_i}(y-b_it)-F(y|x)\right|dt.\\
&\leq \int_{\mathbb{R}}H^{(1)}(t)\left|F^{X_i}(y-b_it)-F(y|x)\right|dt
\end{aligned}$$

Thus,with (H2)(i) we obtain

$$\left|\mathbb{E}^{X_i}(H_1(b_i(y-Y_i))-F(y|x)\right|\leq C\int_{\mathbb{R}}H^{(1)}(b_i^{\beta_1}+|t|^{\beta_2}(b_i^{\beta_2})$$

So, we find

$$\left|\mathbb{E}^{X_i}\left(H_i(y)\right)-F(y|x)\right|\leq C'a_i^{\beta_1}$$

combining this results, we have

$$\begin{aligned}
\left|\mathbb{E}^{\mathcal{F}_{i-1}}(\eta_{ni})\right| &\leq AC''\frac{\phi_i(x,a_i)}{n\varphi_n(x,a_i)}a_i^{\beta_2}\\
\frac{1}{n}\left(\mathbb{E}^{\mathcal{F}_{i-1}}(\eta_{ni})\right)^2 &\leq a_i^{2\beta_1}\frac{1}{\sigma_F^2}\frac{\sum_{i=1}^n\left(a_i^{\beta_1}\phi_i(x,a_i)\right)^2}{n\varphi_n(x,a_i)}\longrightarrow 0 \textit{ (Under (H3)).}
\end{aligned}$$

finally, under hypothesis (H3) acheive the proof of lemma

Now, we have to prove (2), to do this, we observe that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}^{\mathcal{F}_{i-1}}\left((\eta_{ni})^2\right) &= \frac{A^2}{n\varphi_n^2(x,a_i)}\sum_{i=1}^{n}\mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_i^2K_i^2(x)\big(H_i(y)-F(y|x)\big)^2\right)\\
&= \frac{A^2}{n\varphi_n^2(x,a_i)}\sum_{i=1}^{n}\left[\mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_i^2K_i^2(x)H_i^2(y)\right)-2F(y|x)\mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_i^2K_i^2(x)H_i(y)\right)\right.\\
&+ \left.(F(y|x))^2\,\mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_i^2K_i^2(x)\right)\right].
\end{aligned}$$

We put

$$\begin{aligned}
I_1 &= \sum_{i=1}^{n}\mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_i^2K_i^2(x)H_i^2(y)\right),\\
I_2 &= \sum_{i=1}^{n}\mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_i^2K_i^2(x)H_i(y)\right),\\
I_3 &= \sum_{i=1}^{n}\mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_i^2K_i^2(x)\right).
\end{aligned}$$

We write

$$\begin{aligned}
I_1 &= F(y|x)\sum_{i=1}^{n}\mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_i^2K_i^2(x)\right)+\sum_{i=1}^{n}\mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_i^2K_i^2(x)H_i^2(y)\right)-F(y|x)\sum_{i=1}^{n}\mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_i^2K_i^2(x)\right)\\
&= F(y|x)\sum_{i=1}^{n}\mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_i^2K_i^2(x)\right)+\sum_{i=1}^{n}\mathbb{E}^{\mathcal{F}_{i-1}}\left[\mathbb{E}\left(H_i^2(y)|X_i\right)\delta_i^2K_i^2(x)\right]\\
&\quad -F(y|x)\sum_{i=1}^{n}\mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_i^2K_i^2(x)\right)\\
&\le p(x)^2\left(\sum_{i=1}^{n}\mathbb{E}^{\mathcal{F}_{i-1}}\left[\mathbb{E}\left(H_i^2(y)|X_i\right)K_i^2(x)\right]-F(y|x)\sum_{i=1}^{n}\mathbb{E}^{\mathcal{F}_{i-1}}\left(K_i^2(x)\right)\right).
\end{aligned}$$

Using the same argument as those used in proof of the part (1), we obtain

$$\frac{1}{n\varphi_n(x,a_i)}I_2 = o(1).$$

For I_3, we get

$$\mathbb{E}\left(\delta_i^2K_i^2(x)|\mathcal{F}_{i-1}\right) = p(x)^2\left(K^2(1)\phi_i(x,a_i)-\int_0^1\left(K^2(u)\right)'\phi_i(x,a_i)du\right)$$

so under (H1) we have

$$\begin{aligned}\frac{1}{n\varphi_n(x,a_i)}\sum_{i=1}^{n}\mathbb{E}\left(\delta_i^2K_i^2(x)|\mathcal{F}_{i-1}\right) &= p(x)^2\left(\frac{1}{n\varphi_n(x,a_i)}K^2(1)\sum_{i=1}^{n}\phi_i(x,a_i)\right.\\ &\quad\left.-\frac{1}{n\varphi_n(x,a_i)}\int_0^1\left(K^2(1)\right)'\sum_{i=1}^{n}\phi_i(x,ua_i)du\right)\\ &= p(x)^2\left(K^2(u)-\int_0^1\left(K^2(u)\right)'\frac{\sum_{i=1}^n\phi_i(x,a_i)}{n\varphi_n(x,a_i)}du\right)\\ &= o(1).\end{aligned}$$

By combining this results, we deduce that $\lim_{n\longrightarrow\infty}\sum_{i=1}^{n}\mathbb{E}\left(\eta_{ni}^2|\mathcal{F}_{i-1}\right) = 1$, *which complete the proof of part (a).*

proof 3.4.4 *of the part (b)*

The Lindeberg condition in which implies that

$$\mathbb{E}\left(\xi_{ni}^2\mathbb{I}_{\xi_{ni}>\epsilon n}\right)\le 4\mathbb{E}\left(\eta_{ni}^2\mathbb{I}_{\eta_{ni}>n\epsilon/2}\right).$$

Let $a>1$ *and* $b>1$ *such that* $\frac{1}{a}+\frac{1}{b}=1$ *making use the Holder's and Markov's inequalities one can write for all* $\epsilon>0$

$$\mathbb{E}\left(\eta_{ni}^2\mathbb{I}_{\eta_{ni}>n\epsilon/2}\right) \le \frac{\mathbb{E}\left(\eta_{ni}\right)^{2a}}{(n\epsilon/2)^{2a/b}}.$$

Taking $C_0\in\mathbb{R}_+^*$ *and* $2a=2+\gamma$, *from some* $\gamma>0$, *such that* $\mathbb{E}\left(|H_i(y)|^{2+\gamma}\right)<\infty$ *and* $\mathbb{E}\left(|H_i(y)-F(y|x)|^{2+\gamma}|X_i=u\right)=\overline{Z}_{2+\gamma}(u)$ *is a continuous function, we obtain*

$$\begin{aligned}4\mathbb{E}\left(\eta_{ni}^2\mathbb{I}_{\eta_{ni}>n\epsilon}\right) &\le C_0\left(\frac{n\varphi_n(x,a_i)}{p(x)^2\sigma_F^2}\right)^{2+\gamma}\frac{1}{(n\varphi_n(x,a_i))^{2+\gamma}}\left(|H_i(y)-F(y|x)|^{2+\gamma}\delta_i^{2+\gamma}K_i^{2+\gamma}(x)\right)\\ &\le C_0\left(\frac{n\varphi_n(x,a_i)}{p(x)^2n\sigma_F^2}\right)^{2+\gamma}\frac{\mathbb{E}\left(\mathbb{E}\big(|H_i(y)-F(y|x)|^{2+\gamma}\delta_i^{2+\gamma}K_i^{2+\gamma}(x)|X_i\big)\right)}{(n\varphi_n(x,a_i))^{2+\gamma}}\\ &\le C_0\left(\frac{n\varphi_n(x,a_i)}{p(x)^2n\sigma_F^2}\right)^{2+\gamma}\frac{p(x)^{2+\gamma}\mathbb{E}\left(K_i^{2+\gamma}(x)\overline{Z}_{2+\gamma}(x)\right)}{\left((\varphi_n(x,a_i))^{2+\gamma}\right)}\\ &\le C_0\left(\frac{n\varphi_n(x,a_i)}{p(x)^2n\sigma_F^2}\right)^{2+\gamma}(p(x)^2)^{2+\gamma}\\ &\quad\left(\frac{\mathbb{E}\left(K_i^{2+\gamma}(x)\left|\overline{Z}_{2+\gamma}(x)-\overline{Z}_{2+\gamma}(x)\right|+\left|\overline{Z}_{2+\gamma}(x)\right|\right)\mathbb{E}\left(K_i^{2+\gamma}(x)\right)}{(\varphi_n(x,a_i))^{2+\gamma}}\right)\\ &\le C_0\left(\frac{n}{\sigma_F^2}\right)^{2+\gamma}\mathbb{E}\left(K_i^{2+\gamma}(x)\right)\left(\left|\overline{Z}_{2+\gamma}(x)\right|+o(1)\right).\end{aligned}$$

Consequently $\frac{1}{n}\sum_{i=1}^{n} n\mathbb{E}\left(\xi_{ni}^2 \mathbb{I}_{\xi_{ni}>\epsilon n}\right) \longrightarrow 0$ *as* $n \longrightarrow \infty$ *which completes the proof of lemma.*

proof 3.4.5 *3.3.4*
Observe that

$$\begin{aligned}
1-\widehat{F}_D(x) &= 1-\frac{1}{n\varphi_n(x,a_i)}\sum_{i=1}^{n}\delta_i K_i \\
&= \frac{1}{p(x)n\varphi_n(x,a_i)}\sum_{i=1}^{n}\mathbb{E}K_1\mathbb{E}\delta_i K_i - \frac{1}{n\varphi_n(x,a_i)}\sum_{i=1}^{n}\delta_i K_i \\
&= \frac{1}{p(x)n\varphi_n(x,a_i)}\left(\sum_{i=1}^{n}\mathbb{E}K_1\mathbb{E}\delta_i K_i - p(x)\delta_i K_i\right) \\
&\leq \frac{1}{n\varphi_n(x,a_i)}\sum_{i=1}^{n}\Delta_i
\end{aligned}$$

where $\Delta_i = \mathbb{E}\delta_i K_i - \delta_i K_i$ *(*Δ_i *is a centered variable). Applying lamma of hoeffding on the variables* Δ_i*, we must bounded* $|\Delta_i|$ *and* $\mathbb{E}\Delta_i^2$*. Hence*

$$|\Delta_i| < \frac{C_1 p(x)}{n\phi_i(x,a_i)} = \theta_1$$

Under hypothesis (H4), we obtain

$$\mathbb{E}\Delta_i^2 \leq \mathbb{E}\delta_i^2 K_i^2 = \mathbb{E}\left(K_i^2\mathbb{E}^X(\delta_i^2)\right) \leq \frac{C_2 p(x)}{n\phi_i(x,a_i)} = \theta_2$$

Then for all $\varepsilon \in]0;\frac{\theta_1}{\theta_2}[$ *we have*

$$\begin{aligned}
\mathbb{P}\left(|\widehat{F}_D(x)-\mathbb{E}\widehat{F}_D(x)| > \sqrt{\frac{Logn}{n\phi_i(x,a_i)}}\right) &\leq 2exp\left(\frac{-\varepsilon^2 Logn}{4\phi_i(x,a_i)\theta_2}\right) \\
&= 2n^{\frac{-\varepsilon^2}{4\phi_i(x,a_i)\theta_2}} \\
&= 2n^{-C\varepsilon^2}
\end{aligned}$$

This finish the proof of this lemma

proof 3.4.6 *3.3.5*
We have

$$\widehat{B}_{n,1}(x,y) = \frac{\overline{F}_N(y|x)}{\overline{F}_D(x)}.$$

We write

$$\begin{aligned}
\left|\widehat{B}_{n,1}(x,y)\right| &= \frac{1}{\displaystyle\sum_{i=1}^{n}\mathbb{E}\big(\delta_i K_i(x)|\mathcal{F}_{i-1}\big)}\sum_{i=1}^{n}\big[\mathbb{E}\big[\delta_i K_i(x)\mathbb{E}\big[H_i(y)|\mathfrak{B}_{i-1}\big]|\mathcal{F}_{i-1}\big]\\
&- F(y|x)\mathbb{E}\big(\delta_i K_i(x)|\mathcal{F}_{i-1}\big)\big]\\
&= \frac{p(x)}{\displaystyle\sum_{i=1}^{n}\mathbb{E}\big(\delta_i K_i(x)|\mathcal{F}_{i-1}\big)}\sum_{i=1}^{n}\big[\mathbb{E}\big[K_i(x)\mathbb{E}\big[H_i(y)|X_i\big]|\mathcal{F}_{i-1}\big]\\
&- F(y|x)\mathbb{E}\big(\delta_i K_i(x)|\mathcal{F}_{i-1}\big)\big]\\
&\leq \frac{p(x)}{p(x)\displaystyle\sum_{i=1}^{n}\mathbb{E}\big(K_i(x)|\mathcal{F}_{i-1}\big)}\sum_{i=1}^{n}\big[\mathbb{E}\big[K_i(x)\mathbb{E}\big[H_i(y)|X_i\big]-F(y|x)|\mathcal{F}_{i-1}\big]\big].
\end{aligned}$$

Next, an integration by parts and change of variable allow to get

$$\mathbb{E}\big[H_i(y)|X_i\big] = \int_{\mathbb{R}} H^{(1)}(t)F(y-b_i t|X_i)dt$$

thus, we have

$$\left|\mathbb{E}\big[H_i(y)|X_i\big]-F(y|x)\right| \leq \int_{\mathbb{R}} H^{(1)}(t)\left|F(y-b_i t|X_i)-F(y|x)\right|dt \qquad (3.4.1)$$

under $(H2)$*, we obtain that*

$$\mathbb{I}_{B(x,a_i)}(X_i)\left|\mathbb{E}\big[H_i(y)|X_i\big]-F(y|x)\right| \leq C\int_{\mathbb{R}} H^{(1)}(t)(a_i^{\beta_1}+|t|\beta_2 b_i^{\beta_2})dt \quad (3.4.2)$$

and under (H4) we prove that

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\big(K_i(x)|\mathcal{F}_{i-1}\big) = o(1).$$

We achieve the proof of lemma 3.3.5.

proof 3.4.7 *3.3.6*
We write

$$\begin{aligned}
\widehat{R}_{n,1}(x,y) &= -\big(\widehat{B}_{n,1}(x,y)-F(y|x)\big)\big(\widehat{F}_N(x,y)-\overline{F}_N(x,y)\big)\\
&= -\left(\frac{\overline{F}_N(x,y)-F(y|x)\overline{F}_D(x,y)}{\overline{F}_D(x,y)}\right)\left(\widehat{F}_N(x,y)-\overline{F}_N(x,y)\right).
\end{aligned}$$

Clearly, it is suffices to show that

(a) $\left(\frac{\overline{F}_N(x,y)-F(y|x)\overline{F}_D(x,y)}{\overline{F}_D(x,y)}\right) = o(1),$

(b) $\left(\widehat{F}_N(x,y) - \overline{F}_N(x,y)\right) = o(1).$

The proof of the first hand uses arguments similar to those used in the proof of lemma 3.3.5 of the second part, will be established

(i) $\mathbb{E}\left(\widehat{F}_N(x,y) - \overline{F}_N(x,y)\right) = 0$, $Var\left(\widehat{F}_N(x,y) - \overline{F}_N(x,y)\right) \longrightarrow 0$ *as* $n \longrightarrow \infty$.

For all $i = 1, ..., n$, *we put*

$$\Lambda_i(x,y) = \frac{1}{n\varphi_n(x,a_i)}\left[\delta_i K_i(x) H_i(y) - \mathbb{E}\left(\delta_i K_i(x) H_i(y)|\mathcal{F}_{i-1}\right)\right]$$

where $\delta_i(x,y)$ *is a triangular array of martingale differences according to the* $\sigma-$*algebra* $\mathcal{F}_{i-1}$ *next by (H1)(ii) and (H4) we obtain*

$$\widehat{F}_N(x,y) - \overline{F}_N(x,y) = \sum_{i=1}^{n} \Lambda_i(x,y)$$

$\mathbb{E}\left(\Lambda_i(x,y)\right) = 0$ *by definition of* $\Lambda_i(x,y)$, *we write*

$$\sum_{i=1}^{n}\left(\Lambda_i^2(x,y)\right) \leq \sum_{i=1}\mathbb{E}\left(\Lambda_i^2(x,y)\right).$$

Furthermore, by Jensen's inequality we have

$$\begin{aligned}
\mathbb{E}\left(\Lambda_i^2(x,y)\right) &\leq \frac{1}{(n\varphi_n(x,a_i))^2}\mathbb{E}\left(\delta_i^2 K_i^2(x) H_i^2(y)|\mathcal{F}_{i-1}\right) \\
&\leq \frac{p(x)^2}{(n\varphi_n(x,a_i)^2}\mathbb{E}\left(K_i^2(x)|\mathcal{F}_{i-1}\right) \\
&\leq \frac{p(x)^2}{(n\varphi_n(x,a_i))^2}\mathbb{P}\left(X_i(x,a_i)|\mathcal{F}_{i-1}\right) \\
&\leq \frac{p(x)^2}{(n\varphi_n(x,a_i))^2}\phi_i(x,a_i).
\end{aligned}$$

So, we obtain

$$\sum_{i=1}\mathbb{E}\left(\delta_i^2(x,y)\right) \leq \frac{p(x)^2\sum_{i=1}^n \phi_i(x,a_i)}{n^2\varphi_n^2(x,a_i))}.$$

We deduce under (H1)(ii) that var $\left(\widehat{F}_N(x,y) - \overline{F}_N(x,y)\right) \longrightarrow 0$ *as* $n \longrightarrow \infty$.

proof 3.4.8 *of the lemma 3.3.7*

We use the ideas in Laib and Louani For all $i = 1, ..., n$, *we define*

$$\eta'_{ni} = \left(\frac{\varphi_n(x)}{np(x)^2\sigma_f^2}\right)^{\frac{1}{2}}\left(b_i^{-1}H'_i(y) - f(y|x)\right)\frac{\delta_i K_i(x)}{n\varphi_n(x)}$$

and, we define $\xi'_{ni} = \eta'_{n_i} - \mathbb{E}\big(\eta'_{n_i}|\mathcal{F}_{i-1}\big)$.
It is easily seen that

$$\left(\frac{n\varphi_n(x)}{p(x)^2\sigma_f^2}\right)^{\frac{1}{2}} \widehat{Q}_{n,2}(x,y) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi'_{ni}.$$

We follow the same idea in the proof of lemma 3.3.1 to prove this, we establish the following statements :

a) $\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\xi'^2_{ni}|\mathcal{F}_{i-1}\right) \longrightarrow 1$ *in probability,*

b) $\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\xi'^2_{ni}\mathbb{I}_{\xi'^2_{ni}>\epsilon n}\right) \longrightarrow 0$ *holds for any* $\epsilon > 0$ *(Lindeberg condition).*

proof 3.4.9 *of the part (a)*

$$\begin{aligned}\mathbb{E}\left(\xi'^2_{ni}|\mathcal{F}_{i-1}\right) &= \mathbb{E}\left(\big(\eta'_{ni} - \mathbb{E}(\eta'_{ni}|\mathcal{F}_{i-1})\big)^2\right)\\ &= \mathbb{E}\left(\eta'^2_{ni}|\mathcal{F}_{i-1}\right) - \mathbb{E}^2\big(\eta'_{ni}|\mathcal{F}_{i-1}\big).\end{aligned}$$

The statement (a) follows then if we show that

(1) $\lim_{n\longrightarrow\infty}\dfrac{1}{n}\sum_{i=1}^{n}\mathbb{E}^2\big(\eta'_{ni}|\mathcal{F}_{i-1}\big) = 0$ *in probability,*

(2) $\lim_{n\longrightarrow\infty}\dfrac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\eta'^2_{ni}|\mathcal{F}_{i-1}\right) = 1$ *in probability.*

To prove (1), we put $A = \left(\frac{n\varphi_n(x)}{p(x)^2\sigma_f^2}\right)^{\frac{1}{2}}$
we have

$$\begin{aligned}\mathbb{E}\left(\eta'_{ni}|\mathcal{F}_{i-1}\right) &= \frac{A}{n\varphi_n(x)}\mathbb{E}\left(\big(b_i^{-1}H'_i(y) - f(y|x)\big)\delta_i K_i(x)|\mathcal{F}_{i-1}\right)\\ \mathbb{E}\left(\eta'_{ni}|\mathcal{F}_{i-1}\right) &= \frac{p(x)A}{n\varphi_n(x)}\mathbb{E}\left(\big(b_i^{-1}H'_i(y) - f(y|x)\big)K_i(x)|\mathcal{F}_{i-1}\right)\\ \left|\mathbb{E}\left(\eta'_{ni}|\mathcal{F}_{i-1}\right)\right| &= \frac{p(x)A}{n\varphi_n(x)}\left|\mathbb{E}\left[\left[\big(\mathbb{E}\big(b_i^{-1}H'_i(y)|\mathfrak{B}_{i-1}\big) - f(y|x)\big)K_i(x)\right]|\mathcal{F}_{i-1}\right]\right|\\ &= \frac{p(x)A}{n\varphi_n(x)}\left|\mathbb{E}\left[\left[\mathbb{E}\big(b_i^{-1}H'_i(y)|X_i\big) - f(y|x)K_i(x)\right]|\mathcal{F}_{i-1}\right]\right|\end{aligned}$$

under (H1) and (H4), we have

$$C\phi_i(x,a_i) \le \mathbb{E}\big(K_i(x)|\mathcal{F}_{i-1}\big) \le C'\phi_i(x,a_i).$$

Next, an integration by parts and a change of variable allow to get

$$\mathbb{E}\big(H_i'(y)|X_i\big) = b_i \int_{\mathbb{R}} H'(t) f(y - b_i t|X_i) dt.$$

Thus, we have

$$\left|\mathbb{E}\left(H_i'(y)|X_i\right) - b_i f(y|x)\right| \le C' a_i^{B_1}$$

combining this results, we have

$$\begin{aligned}
\left|\mathbb{E}\left(\eta_{ni}'|\mathcal{F}_{i-1}\right)\right| &\le \frac{p(x)A}{\varphi_n(x,a_i)} C'' \phi_i(x,a_i) a_i^{\beta_1} \\
\left|\mathbb{E}\left(\eta_{ni}'|\mathcal{F}_{i-1}\right)\right| &\le \left(\frac{\varphi_n(x)}{\sigma_f^2}\right)^{1/2} \frac{\phi_i(x,a_i)}{\varphi_n(x,a_i)} a_i \beta_1 \\
\frac{1}{n}\left(\mathbb{E}\left(\eta_{ni}'|\mathcal{F}_{i-1}\right)\right)^2 &\le \frac{a_i^{2\beta_1}}{\sigma_f} \frac{\sum_{i=1}^n \left(a_i^{\beta_1}\phi_i(x,a_i)\right)^2}{n\varphi_n(x,a_i)} \longrightarrow 0 \text{ (Under (H3))}.
\end{aligned}$$

Now, we have to prove (2), to do this, we observe that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(\left(\eta_{ni}'\right)^2|\mathcal{F}_{i-1}\right) &= \frac{A^2}{n\varphi_n^2(x,a_i)} \sum_{i=1}^n \mathbb{E}\left(\left(b_i^{-1}H_i'(y) - f(y|x)\right)^2 \delta_i^2 K_i^2(x)|\mathcal{F}_{i-1}\right) \\
&= \frac{A^2}{n\varphi_n^2(x,a_i)} \sum_{i=1}^n \Big[\mathbb{E}\left(b_i^{-2}\delta_i^2 K_i^2(x) H_i'^2(y)|\mathcal{F}_{i-1}\right) \\
&\quad -2f(y|x)\mathbb{E}\left(b_i^{-1}\delta_i^2 K_i^2(x) H_i'(y)|\mathcal{F}_{i-1}\right) + \left(f(y|x)\right)^2 \mathbb{E}\left(\delta_i^2 K_i^2(x)|\mathcal{F}_{i-1}\right)\Big].
\end{aligned}$$

We put

$$\begin{aligned}
I_1 &= \sum_{i=1}^n \mathbb{E}\left(b_i^{-2}\delta_i^2 K_i^2(x) H_i'^2(y)|\mathcal{F}_{i-1}\right), \\
I_2 &= \sum_{i=1}^n \mathbb{E}\left(b_i^{-1}\delta_i^2 K_i^2(x) H_i'(y)|\mathcal{F}_{i-1}\right), \\
I_3 &= \sum_{i=1}^n \mathbb{E}\left(\delta_i^2 K_i^2(x)|\mathcal{F}_{i-1}\right).
\end{aligned}$$

We write

$$\begin{aligned}
I_1 &= f^x(y)\sum_{i=1}^n \mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_i^2 K_i^2(x)\right) + \sum_{i=1}^n \mathbb{E}^{\mathcal{F}_{i-1}}\left(b_i^{-2}\delta_i^2 K_i^2(x) H_i^{'2}(y)\right)\\
&\quad - f^x(y)\sum_{i=1}^n \mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_i^2 K_i^2(x)\right)\\
&= f^x(y)\sum_{i=1}^n \mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_i^2 K_i^2(x)\right) + \sum_{i=1}^n \mathbb{E}^{\mathcal{F}_{i-1}}\left[b_i^{-2}\mathbb{E}\left(H_i^{'2}(y)|X_i\right)\delta_i^2 K_i^2(x)\right]\\
&\quad - f^x(y)\sum_{i=1}^n \mathbb{E}^{\mathcal{F}_{i-1}}\left(\delta_i^2 K_i^2(x)\right)\\
&\leq p(x)^2\left(\sum_{i=1}^n \mathbb{E}^{\mathcal{F}_{i-1}}\left[b_i^{-2}\mathbb{E}\left(H_i^{'2}(y)|X_i\right)K_i^2(x)\right] - f^x(y)\sum_{i=1}^n \mathbb{E}^{\mathcal{F}_{i-1}}\left(K_i^2(x)\right)\right)
\end{aligned}$$

Using the same argument as those used in proof of the part (1), we obtain

$$\frac{1}{n\varphi_n(x,a_i)} I_2 = o(1).$$

For I_3, we get

$$\mathbb{E}\left(\delta_i^2 K_i^2(x)|\mathcal{F}_{i-1}\right) = p(x)^2\left(K^2(1)\phi_i(x,a_i) - \int_0^1 \left(K^2(u)\right)'\phi_i(x,a_i)du\right)$$

so under (H1) we have

$$\begin{aligned}
\frac{1}{n\varphi_n(x,a_i)}\sum_{i=1}^n \mathbb{E}\left(\delta_i^2 K_i^2(x)|\mathcal{F}_{i-1}\right) &= p(x)^2\left(\frac{1}{n\varphi_n(x,a_i)}K^2(1)\sum_{i=1}^n \phi_i(x,a_i)\right.\\
&\quad \left. - \frac{1}{n\varphi_n(x,a_i)}\int_0^1 \left(K^2(1)\right)'\sum_{i=1}^n \phi_i(x,ua_i)du\right)\\
&= p(x)^2\left(K^2(u) - \int_0^1 \left(K^2(u)\right)'\frac{\sum_{i=1}^n \phi_i(x,a_i)}{n\varphi_n(x,a_i)}du\right)\\
&= o(1).
\end{aligned}$$

using the same argument as those used in proof of the part (1), we have

$$\frac{1}{n\varphi_n(x,a_i)} I_2 = o(1).$$

For I_3, it is the same I_3 proved in proof of lemma 3.3.1 and by the combining this results, we deduce that

$$\lim_{n\longrightarrow\infty}\sum_{i=1}^n \mathbb{E}\left(\eta_{ni}^{'2}|\mathcal{F}_{i-1}\right) = 1$$

which complete the proof of part (a).

proof 3.4.10 *of the part (b)*
The Lindeberg condition in which implies that

$$\mathbb{E}\left(\xi_{ni}^{'2}\mathbb{I}_{\xi_{ni}^{\prime}>\epsilon n}\right) \leq 4\mathbb{E}\left(\eta_{ni}^{'2}\mathbb{I}_{\eta_{ni}^{\prime}>n\epsilon/2}\right).$$

Let $a > 1$ and $b > 1$ such that $\frac{1}{a}+\frac{1}{b}=1$ making use the Holder's and Markov's inequalities one can write for all $\epsilon > 0$

$$\mathbb{E}\left(\eta_{ni}^{'2}\mathbb{I}_{\eta_{ni}^{\prime}>n\epsilon/2}\right) \quad \leq \quad \frac{\mathbb{E}\left(\eta_{ni}^{\prime}\right)^{2a}}{(n\epsilon/2)^{2a/b}}.$$

Taking $C_0 \in \mathbb{R}_+^*$ and $2a = 2+\delta$, from some $\delta > 0$, such that $\mathbb{E}\left(|Y_i(y)|^{2+\delta}\right) < \infty$ and $\mathbb{E}\left(b_i^{-(2+\delta)}\left|H_i^{\prime}(y)-F(y|x)\right|^{2+\delta}|X_i=u\right) = \overline{Z}_{2+\delta}(u)$ is a continuous function.
we obtain

$$\begin{aligned}
4\mathbb{E}\left(\eta_{ni}^{'2}\mathbb{I}_{\eta_{ni}^{\prime}>n\epsilon}\right) \quad &\leq \quad C_0\left(\frac{n\varphi_n(x,a_i)}{p(x)^2\sigma_f^2}\right)^{2+\gamma}\frac{1}{(n\varphi_n(x,a_i))^{2+\gamma}}\left(\left|H_i(y)-F(y|x)\right|^{2+\gamma}\delta_i^{2+\gamma}K_i^{2+\gamma}(x)\right)\\
&\leq \quad C_0\left(\frac{n\varphi_n(x,a_i)}{p(x)^2n\sigma_f^2}\right)^{2+\gamma}\frac{\mathbb{E}\left(\mathbb{E}\big(\left|H_i(y)-F(y|x)\right|^{2+\gamma}\delta_i^{2+\gamma}K_i^{2+\gamma}(x)|X_i\big)\right)}{(n\varphi_n(x,a_i))^{2+\gamma}}\\
&\leq \quad C_0\left(\frac{n\varphi_n(x,a_i)}{p(x)^2n\sigma_f^2}\right)^{2+\gamma}\frac{p(x)^{2+\gamma}\mathbb{E}\left(K_i^{2+\gamma}(x)\overline{Z}_{2+\gamma}(x)\right)}{\left((\varphi_n(x,a_i))^{2+\gamma}\right)}\\
&\leq \quad C_0\left(\frac{n\varphi_n(x,a_i)}{p(x)^2n\sigma_f^2}\right)^{2+\gamma}(p(x)^2)^{2+\gamma}\\
&\qquad\left(\frac{\mathbb{E}\left(K_i^{2+\gamma}(x)\left|\overline{Z}_{2+\gamma}(x)-\overline{Z}_{2+\gamma}(x)\right|+\left|\overline{Z}_{2+\gamma}(x)\right|\right)\mathbb{E}\left(K_i^{2+\gamma}(x)\right)}{(\varphi_n(x,a_i))^{2+\gamma}}\right)\\
&\leq \quad C_0\left(\frac{n}{\sigma_f^2}\right)^{2+\gamma}\mathbb{E}\left(K_i^{2+\gamma}(x)\right)\big(\left|\overline{Z}_{2+\gamma}(x)\right|+o(1)\big).
\end{aligned}$$

Consequently $\frac{1}{n}\sum_{i=1}^{n} n\mathbb{E}\left(\xi_{ni}^2\mathbb{I}_{\xi_{ni}>\epsilon n}\right) \longrightarrow 0$ as $n \longrightarrow \infty$ which completes the proof of lemma.

proof 3.4.11 *3.3.8*
We have

$$\widehat{B}_{n,2}(x,y) = \frac{\overline{f}_N(y|x)}{\overline{F}_D(x)}.$$

We write

$$\begin{aligned}
\left|\widehat{B}_{n,2}(x,y)\right| &= \frac{1}{\sum_{i=1}^{n}\mathbb{E}\big(\delta_i K_i(x)|\mathcal{F}_{i-1}\big)}\sum_{i=1}^{n}\big[\mathbb{E}\big[\delta_i K_i(x)\mathbb{E}\big[b_i^{-1}H_i'(x)|\mathfrak{B}_{i-1}\big]|\mathcal{F}_{i-1}\big] \\
&- f(y|x)\mathbb{E}\big(\delta_i K_i(x)|\mathcal{F}_{i-1}\big)\big] \\
&= \frac{p(x)}{\sum_{i=1}^{n}\mathbb{E}\big(\delta_i K_i(x)|\mathcal{F}_{i-1}\big)}\sum_{i=1}^{n}\big[\mathbb{E}\big[K_i(x)\mathbb{E}\big[b_i^{-1}H_i'(x)|X_i\big]|\mathcal{F}_{i-1}\big] \\
&- f(y|x)\mathbb{E}\big(K_i(x)|\mathcal{F}_{i-1}\big)\big] \\
&\leq \frac{p(x)}{p(x)\sum_{i=1}^{n}\mathbb{E}\big(K_i(x)|\mathcal{F}_{i-1}\big)}\sum_{i=1}^{n}\big[\mathbb{E}\big[K_i(x)\mathbb{E}\big[b_i^{-1}H_i'(x)|X_i\big] - b_i f(y|x)|\mathcal{F}_{i-1}\big]\big].
\end{aligned}$$

Next, an integration by parts and change of variable allow to get

$$\mathbb{E}\big[H_i'(y)|X_i\big] = b_i\int_{\mathbb{R}} H'(t)f(y-b_it|X_i)dt$$

thus, we have

$$\left|\mathbb{E}\big[b_i^{-1}H_i'(y)|X_i\big] - f(y|x)\right| \leq \int_{\mathbb{R}} H'(t)\left|F(y-b_it|X_i) - f(y|x)\right|dt. \quad (3.4.3)$$

Under $(H2)$*, we obtain that*

$$\mathbb{I}_{B(x,a_i)}(X_i)\left|\mathbb{E}\big[H_i'(y)|X_i\big] - b_if(y|x)\right| \leq Cb_i\int_{\mathbb{R}} H'(t)(a_i^{\beta_1} + |t|\beta_2 b_i^{\beta_2})dt \quad (3.4.4)$$

and under (H4) we prove that

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\big(K_i(x)|\mathcal{F}_{i-1}\big) = o(1).$$

We achieve the proof of lemma.

proof 3.4.12 *3.3.9*
We write

$$\begin{aligned}\widehat{R}_{n,2}(x,y) &= -\widehat{B}_{n,2}(x,y)\big(\widehat{f}_N(x,y)-\overline{f}_N(x,y)\big)\\ &= -\left(\frac{\overline{f}_N(x,y)-f(y|x)\overline{F}_D(x,y)}{\overline{F}_D(x,y)}\right)\left(\widehat{f}_N(x,y)-\overline{f}_N(x,y)\right).\end{aligned}$$

Clearly, it is suffices to show that :

(a) $\left(\frac{\overline{f}_N(x,y)-f(y|x)\overline{F}_D(x,y)}{\overline{F}_D(x,y)}\right)=o(1)$,

(b) $\left(\widehat{f}_N(x,y)-\overline{f}_N(x,y)\right)=o(1)$.

The proof of the first hand uses arguments similar to those used in the proof of lemma 3.3.5. Of the second part, will be established

(i)$\mathbb{E}\left(\widehat{f}_N(x,y)-\overline{f}_N(x,y)\right)=0$, $Var\left(\widehat{f}_N(x,y)-\overline{f}_N(x,y)\right)\longrightarrow 0 \quad as \quad n\longrightarrow \infty$

For all $i=1,...,n$, *we put*

$$\Lambda'_i(x,y)=\frac{1}{n\varphi_n(x,a_i)}\left[b_i^{-1}\delta_iK_i(x)H'_i(y)-\mathbb{E}\left(b_i^{-1}\delta_iK_i(x)H'_i(y)|\mathcal{F}_{i-1}\right)\right]$$

where $\Lambda_i(x,y)$ *is a triangular array of martingale differences according to the* $\sigma-$*algebra* $\mathcal{F}_{i-1}$ *next by* $(H1)(ii)$ *and* $(H4)$ *we obtain*

$$\widehat{f}_N(x,y)-\overline{f}_N(x,y)=\sum_{i=1}^n\Lambda'_i(x,y)$$

$\mathbb{E}\left(\Lambda'_i(x,y)\right)=0$ *by definition of* $\Lambda'_i(x,y)$,
we write

$$\sum_{i=1}^n\left(\Lambda'^2_i(x,y)\right)\leq\sum_{i=1}\mathbb{E}\left(\Lambda'^2_i(x,y)\right).$$

Furthermore, by Jensen's inequality we have

$$\begin{aligned}\mathbb{E}\left(\Lambda'^2_i(x,y)\right) &\leq \frac{1}{(n\varphi_n(x,a_i))^2}\mathbb{E}\left(b_i^{-2}\delta_i^2K_i^2(x)H'^2_i(y)|\mathcal{F}_{i-1}\right)\\ &\leq \frac{p(x)^2}{(n\varphi_n(x,a_i)^2}\mathbb{E}\left(b_i^{-2}K_i^2(x)|\mathcal{F}_{i-1}\right)\\ &\leq \frac{p(x)^2}{(n\varphi_n(x,a_i))^2}\mathbb{P}\left(X_i(x,a_i)\mathbb{E}\left(H'^2_i(y)|\mathfrak{B}_{\mathfrak{i}-\mathfrak{1}}\right)|\mathcal{F}_{i-1}\right)\\ &\leq \frac{p(x)^2Cb_i^{-1}\phi_i(x,a_i)}{(n\varphi_n(x,a_i))^2}\end{aligned}$$

so, we obtain

$$\sum_{i=1}\mathbb{E}\left(\Lambda'^2_i(x,y)\right)\leq\frac{p(x)^2C\sum_{i=1}^nb_i^{-1}\phi_i(x,a_i)}{n^2\varphi_n^2(x,a_i))}.$$

We deduce that var $\left(\widehat{f}_N(x,y) - \overline{f}_N(x,y)\right) \longrightarrow 0$ *as* $n \longrightarrow \infty$ *since (H1)(ii).*

Bibliographie

[1] Ferraty, F., Rabhi, A., and Vieu, P. (2008). Estimation non-paramétrique de la fonction de hasard avec variable explicative fonctionnelle,*Rev. Roumaine Math. Pures Appl*, **53**, 1–18.

[2] Attouch, M.k., Belabed, F.Z. (2014). The k Nearest Neighbors estimation of the conditional hazard function for functional data,

[3] Benziadi, F., Laksaci, A. and Tebboune, F. (2016). Note on conditional quantiles for functional ergodic data, *C. R. Acad.Sci.Paris,Ser.I*, **354**, 628–633.

[4] Gheriballah, A., Laksaci, A., and Sekkal, S. (2013). Nonparametric M-regression for functional ergodic data, *Statistics and Probability Letters* **83**, 902–908.

[5] Ardjoun, F.Z., Ait Hennani, L. and Laksaci, A. (2016). A recursive kernel estimate of the functional modal regression under ergodic dependence condition, *Journal of Statistical Theory and Practice*, **10**, 475–496.

[6] Laïb, N., and Louani, D. (2010). Nonparametric Kernel regression estimation for functional stationary ergodic data : Asymptotic proprties, *J. Mult. Anal.*, **101**, 2266–2281.

[7] Laïb, N., and Louani, D. (2011). Rates of strong consistencies of the regression function estimator for functional stationary ergodic data, *J. of Stat. Plan. and Inf.*, **141**, 359–372.

[8] Laksaci, A., and Mechab, B. (2014). Conditional hazard estimate for functional random fields, *Journal of Statistical Theory and Practice*, **8**, 192–200.

[9] Laksaci, A., and Mechab, B. (2010). Estimation non-paramétrique de la fonction de hasard avec variable explicative fonctionnelle : cas des données spatiales, *Rev. Roumaine Math. Pures Appl*, **55**, 35–51.

[10] Quintela-del-Río, A. (2008). Hazard function given a functional variable : Non-parametric estimation under strong mixing conditions, *J. Nonparametr. Stat.*, **20**, 413–430.

[11] Ramsay, J.O., and Silverman, B.W. (2002). Applied functional data analysis ; Methods and case studies, *Springer-Verlag*, New York.

[12] Massim, I., and Mechab, B. (2016). Local linear estimation of the conditional hazard function, *International Journal of Statistics & Economics* **17**, 1-11.

[13] Ferraty, F., and Vieu, P. (2006). Nonparametric functional data analysis, *Springer Series in Statistics*, Springer New York.

Conclusion générale

Conclusion

Nous fournissons dans cette thèse la cohérence uniforme presque complète avec les taux de la fonction de régression robuste en cas de paramètre d'échelle inconnu. Ces résultats ont été obtenus dans des conditions standard suffisantes qui permettent d'explorer différents axes structurels du sujet, la nature fonctionnelle du modèle et des données ainsi que la robustesse de la fonction de régression et la dépendance de l'observation. Les applications de données réelles (concentration maximale d'ozone, demande maximale d'électricité) ont également mis en évidence plusieurs caractéristiques attrayantes de la régression robuste avec un estimateur de paramètre d'échelle inconnu. En termes d'erreur quadratique moyenne (EQM), l'estimateur proposé se comporte de façon concurrentielle par rapport aux estimateurs existants ayant un paramètre d'échelle connu.

D'après l'expérience de ce document sur la régression robuste avec un paramètre d'échelle inconnu, nous pensons que la plupart des techniques utilisant des noyaux fonctionnels non paramétriques pourraient être étendues efficacement. Par exemple, la remise en question des questions ouvertes en ce sens pourrait concerner aussi bien les extensions à d'autres formes de prédicteurs non paramétriques (comme les variables fonctionnelles linéaires locales, les variables fonctionnelles kNN et beaucoup d'autres $\cdots$). Extensions à d'autres types de modèles de prédiction dans lesquels un stade préliminaire du noyau joue un rôle crucial (cela comprendrait de nombreux modèles de régression semi-paramétriques comme les modèles d'index simples fonctionnels, et les modèles linéaires partiels, et beaucoup d'autres modèles $\cdots$