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 :

BENCHIHA Abbassia

Domaine : Mathématiques Informatique
Filière : Mathématiques
Intitulé de la formation : Statistique, Mathématiques appliquées à l'économie et à la finance

Intitulée

Estimation locale linéaire fonctionnelle : missing data

Soutenue le :…………………………….

Devant le jury composé de :

Président :

Mr MECHAB Boubaker	Maitre de Conférences A	à L'Université S.B.A.

Examinateurs :

Mr BENAISSA Samir	Professeur	à L'Université S.B.A
Mr GUENDOUZI Toufik	Professeur	à L'Université de Saida

Directeur de thèse :

Mr GHERIBALLAH Abdelkader	Professeur	à L'Université S.B.A

Co-Directeur de thèse :

Mr LAKSACI Ali	Professeur	à L'Université S.B.A

Année universitaire : 2018/2019

Remerciements

Il est naturel de remercier plus ou moins directement, les personnes qui ont cru en moi et qui m'ont permis d'arriver au bout de cette thèse.

J'aimerais tout d'abord remercier mon directeur de thèse *M^r GHRIBALLAH Abdelkader*, pour m'avoir appris à être " bonne élève " et plus autonome tout au long de ce travail de recherche.

J'adresse aussi mes remerciements aux personnes que je nomme "ressources" dans ma thèse *M^r LAKSACI Ali* et *M^r MECHAB Boubaker* et qui m'ont permis de mieux comprendre le fonctionnement actuel et passé de ce secteur.
Je voudrais aussi remercier *M^r BENAISSA Samir* pour la confiance dont il me fait preuve en faisant parties de ce jury.

Je tiens à exprimer mes remerciements à *M^r GUENDOUZI Toufik* pour sa participation à mon jury. Je suis très honoré de sa présence.

Je remercie les membres du Laboratoire de Statistique et Processus Stochastiques de l'université Djillali Liabès de Sidi Bel Abbès. J'ai toujours trouvé soutien et encouragement.

les mots les plus simples étant les plus forts, j'adresse toute mon affection à ma famille, et en particulier à mon père qui m'a fait comprendre que la vie n'est pas faite que de problèmes qu'on pourrait résoudre grâce à des formules mathématiques. Malgré mon éloignement depuis de (trop) nombreuses années, leur intelligence, leur confiance, leur tendresse, leur amour me portent et me guident tous les jours. Merci pour avoir fait de moi ce que je suis aujourd'hui. Est-ce un bon endroit pour dire ce genre de choses ? Je n'en connais en tous cas pas de mauvais. Je vous aime.

Ces remerciements ne peuvent s'achever, sans une pensée pour ma première fan (et correctrice des fautes d'orthographes de cette thèse !) : ma mère. Sa présence et ses encouragements sont pour moi les piliers fondateurs de ce que je suis et de ce que je fais.
A titre plus personnel, Je remercie chaleureusement mon mari, *Zouaoui*, pour la grande patience, l'encouragement et la confiance qu'il m'a témoigné dont il a fait preuve à la relecture de mon manuscrit. Je tiens à le remercier surtout pour son soutien moral ininterrompu et ses nombreux conseils tout le long de ma thèse.

J'adresse mes plus sincères remerciements à tous ma belle famille, qui m'ont toujours soutenue et encouragée au cours de la réalisation de ce mémoire.

Je remercie mes soeurs et toutes mes amies de leurs sympathies et leurs aides de près ou de loin, sans oublier mon petit enfant *Mohamed*.

Enfin, j'adresse mes remerciements à l'ensemble des personnes présentes à cette soutenance.

Table des matières

Résumé

Dans cette thèse, nous intéressons essentiellement à l'estimation non paramétrique de l'opérateur de régression d'une variable fonctionnelle et une variable de réponse scalaire qui n'est pas totalement observée.

Dans un premiers temps, nous considérons le problème de l'analyse de covariabilité entre une variable fonctionnelle X et une variable de réponse scalaire Y qui n'est pas totalement observée. Nous utilisons l'approche linéaire locale pour modéliser cette relation en construisant un estimateur linéaire local de l'opérateur de régression lorsque des données manquantes apparaissent dans la variable réponse.

Dans un second temps, nous analysons la covariabilité entre deux variables aléatoires lorsque la réponse une n'est pas totalement observée. Nous utilisons la méthode du noyau des plus proches voisins pour construire un estimateur de la régression linéaire locale lorsque des données manquantes apparaissent dans la variable réponse. Les résultats asymptotiques, en termes de consistances ponctuelles et de consistance uniforme presque complète, sont établis pour l'estimateur construit.

Summary

In this thesis, we are mainly interested in the non-parametric estimation of the regression operator of a functional variable and a scalar response variable that is not fully observed.

First, we consider the problem of covariability analysis between a functional variable X and a scalar response variable Y that is not fully observed. We use the local linear approach to model this relationship by constructing a local linear estimator of the regression operator when missing data appears in the response variable.

In a second step, we analyze the covariability between two random variables when response one is not fully observed. We use the nearest neighbour's kernel method to construct an estimator of local linear regression when missing data appears in the response variable. Asymptotic results, in terms of point consistencies and almost complete uniform consistency, are established for the constructed estimator.

Chapitre 1

Introduction

1.1 Estimation fonctionnelle :

L'estimation est un élément fondamental de la statistique. Elle permet de généraliser, autant que faire se peut, des résultats observés. On y distingue

- l'approche paramétrique, qui considère que les modèles sont connus, avec des paramètres inconnus. La loi de la variable étudiée est supposée appartenir à une famille de lois pouvant être caractérisée par une forme fonctionnelle connue (fonction de répartition F, densité f,...) qui dépend d'un ou plusieurs paramètres inconnus à estimer.

- l'approche non paramétrique, qui ne fait aucune hypothèse sur la loi, ni sur ses paramètres. Nos connaissances sur le modèle ne sont pas précises, ce qui est souvent le cas dans la pratique. Dans cette situation, il est naturel de vouloir estimer une des fonctions décrivant le modèle, soit généralement la fonction de répartition ou la densité (pour le cas continu) : c'est l'objectif de l'estimation fonctionnelle. Si la fonction de répartition empirique F_n résout le problème statistique fondamental de la distribution de probabilité (associée à un échantillon $(T_1, T_2, ..., T_n)$ de variables aléatoires réelles indépendantes et de même loi) en fonction des valeurs numériques observées, elle est par contre limitée pour décrire visuellement les caractères de l'échantillon. Pour Deheuvels (1980), c'est l'estimation la plus naturelle dans le cas où on ne fait, extérieurement à l'échantillon, aucune hypothèse restreignant le choix de cette distribution à une famille particulière.

L'usage de F_n est justifié par divers résultats comme le théorème de Glivenko-Cantelli assurant que

$$\sup_{t \in R} |F_n(t) - F(t)| \to 0, p.s.,$$

où F est la fonction de répartition de la loi recherchée, et le théorème de Donsker pour la normalité asymptotique. De plus, elle est simple, à la fois dans sa formulation et dans son calcul.

Cependant, elle n'en est pas moins limitée pour décrire les caractères de l'échantillon (ce qui occasionne une perte d'information sur la loi recherchée).

1.1.1 Statistiques non paramétriques :

Les méthodes statistiques traditionnelles échouent dès que nous traitons des données fonctionnelles. En effet, si l'on considère par exemple un échantillon de courbes finement discrétisées, deux problèmes statistiques cruciaux apparaissent. La première provient du rapport entre la taille de l'échantillon et le nombre de variables (chaque variable réelle correspondant à un point discret). La seconde, due à l'existence de fortes corrélations entre les variables, devient un problème mal conditionné dans le contexte du modèle linéaire multivarié. Il y a donc une réelle nécessité de développer des méthodes/modèles statistiques afin de prendre en compte la structure fonctionnelle de ce type de données. La plupart des méthodes statistiques existantes traitant des données fonctionnelles utilisent la modélisation linéaire pour l'objet à estimer. les références sur les aspects méthodologiques sont celles de Ramsay, J., Silverman, B.W (1997,2002) et les implémentations sont fournies par Clarkson, Fraley, Gu, Ramsay (2005). Notons également que, pour certains problèmes plus spécifiques, certains aspects théoriques se trouve dans Bosq, D.(2000).

D'autre part, les statistiques non paramétriques ont été développées de manière intensive. En effet, depuis le début des années soixante, beaucoup d'attention a été accordée de mannequinat gratuit (à la fois dans le cadre d'une distribution gratuite et d'un paramétrage gratuit). c'est-à-dire des modèles et/ou des méthodes statistiques. La caractéristique fonctionnelle de ces provient de la nature de l'objet à estimer (comme par exemple une fonction de densité, une fonction de régression, ...) qui n'est pas supposée être paramétrable par un nombre fini de grandeurs réelles. Dans ce réglage, on est généralement en parlant de statistiques non paramétriques pour lesquelles il y a une abondance de la littérature. Par exemple, le lecteur trouvera dans "regression and statistical models" Helland, I.(1990) une monographie précédente pour appliquée, tandis que Schimek,M "Smoothing and regression ; Approaches, Computation,and Application" en(2000) et Akritas, M., Politis, D."Recent advances and trends in nonparametric statistics" en(2003) présentent l'état de la variable l'art dans ces domaines. Il apparaît clairement que ces techniques ne concernent que le cadre classique, à savoir les données réelles ou multidimensionnelles.

Cependant, les progrès récents mêlent des idées de modélisation libre non paramétrique avec des données fonctionnelles dans un double cadre dimensionnel infini.

L'appellation Statistiques non paramétriques fonctionnelles couvre l'ensemble des statistiques. Dans la terminologie Statistiques non paramétriques fonctionnelles, l'adjectif non paramétrique se réfère à la forme de l'ensemble des contraintes alors que le mot fonctionnel désigne la forme de l'ensemble des contraintes est liée à la nature des données. En d'autres termes, les données non paramétriques les aspects proviennent de l'infinie dimensionnalité de l'objet à estimer et la désignation fonctionnelle est due à la caractéristique dimensionnelle infinie des données.

Dans ces dernières décennies, une immense innovation sur les appareils de mesure est apparue permettant d'observer plusieurs données de plus en plus complexes d'une façon continue, tels les indices boursiers, la météorologie, les images satellitaires, la chimie quantitative, la biométrie, l'imagerie médicale ..., à présent. C'est possible en raison de la précision des appareils de mesures modernes et de l'importante capacité de stockage qu'offrent les systèmes informatiques actuels d'obtenir une discrétisation très précise de ces objets mathématiques (courbes, images,...) pendant toutes leurs trajectoires et qui prennent des valeurs dans des espaces de dimension infinie.

Une modélisation statistique est importante pour mieux comprendre le fonctionnement du problème modélisé. En statistiques non paramétriques la performance des outils statistiques réduit considérablement lorsque la dimension des observations augmente. Ce grand développement technologique a imposé la modernisation des méthodes statistiques comme outils d'analyse et de contrôle. Ainsi, une nouvelle branche de la statistique, dénommée statistique fonctionnelle s'est développée pour traiter des observations comme éléments aléatoires fonctionnels.

Les premiers ouvrages de référence en la matière sur le sujet ont été consacrées à l'étude des modèles paramétriques, les monographies de Ramsay et Silverman (1997) pour le cas i.i.d où pour le cas dépendant Bosq (2000) pour les aspects théoriques. Cependant, la base d'analyse statistique via les modèles linéaires est la préliminaire connaissance de la nature de la co-variabilité entre les observations, ce qui est très compliqué à vérifier en statistiques fonctionnelles, par contre dans la statistique classique on dispose d'outils graphiques, on prend comme exemple le scatter plot qui donne un aperçu sur le rapport entre les observations. Ceci justifie intérêt de la modélisation des données fonctionnelles par des méthodes non paramétriques.

Parmi les , on peut citer pour les aspects appliqués, Bosq (2000) pour les aspects théoriques, Ferraty et Vieu (2006) pour une étude non paramétrique et Ferraty et Romain (2011) pour des développements récents. Dans le même contexte, nous renvoyons à Ferraty (2010). L'objectif de ce paragraphe est de faire une étude bibliographique sur les modèles non paramétriques conditionnels considérés dans cette thèse le traitement non paramétrique des données fonctionnelles est beaucoup plus récent que l'analyse paramétrique. En effet, les premiers résultats ont été obtenus par Gasser et al.(1998). Ils se sont intéressés à l'estimation non paramétrique du mode de la distribution d'une variable fonctionnelle vérifiant un condition fractale. En considérant la même condition fractale Ferraty et Vieu (2000) ont étudié la convergence presque complète d'un estimateur à noyau de la fonction de régression, lorsque les observations sont indépendantes et identiquement distribuées. Dabo-Niang (2002) a obtenu, la convergence presque sûre et la normalité asymptotique d'un estimateur de type histogramme de la densité d'une variable aléatoire dans un espace de dimension infinie. En utilisant la propriété de concentration de la mesure de probabilité de la variable explicative fonctionnelle, Dabo-Niang et Rhomari (2004) ont étudié la convergence en norme L^p de l'estimateur à noyau de la régression non paramétrique. La convergence presque complète pour le cas fortement mélangeant a été étudié par Ferraty et al. (2004). Masry (2005) a montré la normalité asymptotique dans le cas d'observations fonctionnelles α-mélangeantes.

Les premiers résultats sur les modèles conditionnels ont été obtenus par Ferraty et al.(2006). Ils ont précisé la vitesse de convergence presque complète des estimateurs à noyau pour la fonction de répartition conditionnelle, la densité conditionnelle et ses dérivées, le mode conditionnel et les quantiles conditionnels. Nous renvoyons à Ferraty et Vieu (2006) pour un large éventail d'applications de ces modèles en statistique fonctionnelle. Dabo-Niang et Laksaci (2007) ont ajouté des résultats sur la convergence en norme L^p de l'estimateur à noyau du mode conditionnel dans le cas i.i.d. La détermination des termes dominants de l'erreur quadratique de l'estimateur à noyau de la densité conditionnelle a été obtenue par Laksaci (2007). Ferraty et al.(2008) ont abordé l'estimation de la fonction du hasard conditionnelle et ont établi la convergence presque complète d'un estimateur à noyau de ce modèle non paramétrique.

1.1.2 L'analyse des données fonctionnelles

Dans ce travail, nous nous sommes intéressés à ces deux outils statistiques mais sous un angle particulier. En effet, dans la plupart des problèmes généralement présentées dans le domaine de l'apprentissage statistique, X est une variable aléatoire à valeurs dans l'espace $\mathbb{R}^d$. Ici, nous avons étudié le cas où X vit dans un espace de Hilbert mais de dimension quelconque (éventuellement infinie). L'intérêt de ce type de données est qu'elles apparaissent fréquemment dans les problèmes concrets : puisque les appareils d'enregistrement modernes collectent des données qui se présentent sous la forme de courbes qui peuvent être considérées comme des fonctions discrétisées en certains points. Ces données sont appelées données fonctionnelles. De manière plus formelle, on considère généralement que X est à valeurs dans un espace de Hilbert séparable $\mathcal{H}$ qui est, par exemple, l'espace L^2_τ, ensemble des fonctions de l'intervalle compact τ de $\mathbb{R}$ dans $\mathbb{R}$, de carré intégrable, que l'on munit du produit scalaire :

$$\forall \;\; f, g \in L^2_\tau, \qquad \langle f, g\rangle = \int_\tau f(t)g(t)dt.$$

L'analyse statistique des données fonctionnelles conduit à des problèmes nouveaux qui sont liés au fait que la dimension de l'espace $\mathcal{H}$ est infinie. Cela implique en particulier que certains problèmes qui sont résolus simplement lorsque $\mathcal{H} = \mathbb{R}^d$ deviennent mal posés dans un espace de dimension infinie et donnent lieu, d'un point de vue pratique, à des solutions inappropriées si le caractère fonctionnel des données n'est pas pris en compte dans leur traitement. Une des raisons de cet état de fait est la difficulté d'inverser les opérateurs définis dans des espaces de dimension infinie. Rappelons d'abord que pour tout espace de Hilbert $\mathcal{H}$ (et ceci est donc valable pour son dual $\mathcal{H}$), on définit l'espérance d'une variable aléatoire à valeurs dans $\mathcal{H}, Z$, comme étant l'unique $\mathbb{E}(Z) \in \mathcal{H}$ tel que $\forall v \in \mathcal{H}, \langle \mathbb{E}(Z), v\rangle = \mathbb{E}(\langle Z, v\rangle)$ et que le produit tensoriel dans $\mathcal{H}$ est donné par $\forall u \in \mathcal{H}, u \otimes u : v \in \mathcal{H} \to \langle u, v\rangle u$. Prenons alors l'exemple de l'opérateur de variance de $X : \Gamma_X = \mathbb{E}(X \otimes X) - \mathbb{E}(X) \otimes \mathbb{E}(X)$; cet opérateur est un opérateur de Hilbert-Schmidt : il n'est donc pas bijectif dans L^2_τ. De plus, restreint à son image, l'opérateur Γ_X peut être bijectif mais n'est jamais inversible (dans l'ensemble des opérateurs linéaires continus de L^2_τ car son inverse n'est pas borné. La plupart des modèles statistiques classiques ont pourtant été étendus au cas fonctionnel, moyennant cependant quelques adaptations : tout d'abord, [Deville, 1974], [Dauxois and Pousse, 1976], [Besse and Ramsay, 1986] et [Besse, 1991] proposent diverses approches pour l'extension des analyses factorielles au cadre fonctionnel. Parallèlement, [Saporta, 1981] présente une étude synthétique des méthodes exploratoires d'analyse des processus. Plus tard, [Ramsay and Silverman, 1997] donnent une vision globale du traitement des données fonctionnelles avec des méthodes de régression, de discrimination et également des analyses factorielles. [Aguilera et al., 1997] ont également développé un modèle linéaire de prédiction de séries chronologiques basé sur l'ACP d'un processus et qui utilise une approximation spline des facteurs principaux. Par ailleurs, [Cardot et al., 1999] ont généralisé la régression linéaire au cadre fonctionnel alors que [Ferraty and Vieu, 2002] ont proposé une approche non paramétrique du problème de la régression fonctionnelle. Une alternative à ces approches est proposée par [Preda and Saporta, 2002] qui développent un modèle fonctionnel de régression PLS (Partial Least Squares) : il s'agit d'une méthode itérative de régression linéaire qui se révèle particulièrement efficace dans le cadre fonctionnel. Dans [Preda and Saporta, 2005a], les auteurs proposent une variante de ce modèle dans le cas où l'on considère une partition de l'espace des prédicteurs en K groupes distincts. Enfin, dans [Preda and Saporta, 2005b], les auteurs proposent une application de la régression PLS dans le cadre de l'analyse discriminante pour une réponse binaire. [Dauxois et al., 2001], [Ferré and Yao, 2003] et [Ferré and Yao, 2005] développent un modèle semi-paramétrique pour variable aléatoire hilbertienne qui est une version fonctionnelle de la SIR ([Li, 1991]). Ce dernier modèle peut être utilisé aussi bien à des fins de régression que de discrimination (cf [Dauxois et al., 2001] et [Ferré and Villa, 2005a], Chapitre 4). Par ailleurs, [Rossi et al., 2004], [Rossi and Conan-Guez, 2005a] et [Rossi et al., 2005] ont développé des méthodes de traitement de données fonctionnelles par réseaux de neurones. Dans le domaine de la discrimination aussi les dernières années sont riches en travaux : un certain nombre d'entre eux présentent des approches par pénalisation de l'opérateur de covariance. Cette méthode, qui permet une régularisation du problème initial, a été développée, à l'origine par [Ivanov, 1962], [Tihonov, 1963a] et [Tihonov, 1963b]. Elle a été appliquée avec succès à l'Analyse en Composantes Principales ([Pezzulli and Silverman, 1993] et [Silverman, 1996]), à l'Analyse Discriminante ([Hastie et al., 1994] et [Hastie et al., 1995]) et à l'Analyse Canonique ([Leurgans et al., 1993]). Une approche alternative aux méthodes de régularisation est le filtrage qui consiste à projeter les données sur une base de fonctions préalablement fixées : c'est ce que font, par exemple, [Biau et al., 2005] qui développent une méthode consistante de discrimination, dans des espaces de Hilbert, par plus proches voisins. Enfin, [James and Hastie, 2001] montrent, dans le cadre de l'Analyse Discriminante, la limite de ces deux approches, particulièrement lorsque les données sont échantillonnées de manière non uniforme : ils proposent alors un modèle dans lequel la fonction sous-jacente, et non les observations, est exprimée par filtrage sur une base Spline.

1.2 Méthode du noyau :

1.2.1 Densité :

La densité, lorsqu'elle existe, est plus appropriée pour caractériser la distribution et permet de mieux la visualiser : son graphe permet de voir le ou les modes, la symétrie, la dispersion, l'aplatissement,...

Pour décrire une loi, on se sert particulièrement de la courbe représentative de la densité. De plus, les fluctuations sont plus apparentes dans les courbes de densités. D'autre part, l'outil informatique a permis de lever les difficultés calculatoires pour déterminer la densité, à partir des valeurs numériques observées d'un échantillon ; d'où l'intérêt accordé durant ce dernier demi-siècle à l'estimation de la densité, particulièrement depuis les travaux fondateurs de Rosenblatt (1956) et Parzen (1962) qui ont, en généralisant la notion d'estimation par histogramme (estimateur naif), donné naissance à la méthode du noyau (grâce à laquelle a été aussi développée l'estimation de la fonction de régression par Nadaraya et Watson (1964)).
Soit $(T_1, T_2, ..., T_n)$; un échantillon de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) de densité inconnue f. On définit l'estimateur à noyau de f par :

$$f_n(t) = \frac{1}{nh} \sum_{i=1}^{n} K\Big(\frac{t - T_i}{h}\Big),$$

où $h := h(n)$ est une suite de nombres réels positifs (dépendant de n), appelés fenêtres ou largeurs de fenêtre, qui contrôlent le lissage de la courbe estimée, et K est une fonction bornée, intégrable, d'intégrale égale à 1, appelée noyau. Si de plus $\lim_{|u|\to\infty} uK(u) = 0$, K est appelé noyau de Parzen-Rosenblatt (et f_n l'estimateur à noyau de Parzen-Rosenblatt).
Un noyau est dit d'ordre r $(r \geq 1)$ si
- les fonctions $t \to t^j K(t), j = 1, ..., r$; sont intégrables et vérifient
- $\int K(t)dt = 1$; et $\int t^j K(t)dt = 0, j = 1, ..., r$;

L'estimation de la densité est devenue un problème statistique classique. Plusieurs types d'estimateurs ont été proposés (approches par les plus proches voisins, séries orthogonales, maximum de vraisemblance pénalisé, histo-splines, ondelettes,...), mais d'une manière générale, les résultats qui en découlent ne sont pas significativement meilleurs que par la méthode du noyau. Les résultats établissant les propriétés des estimateurs proposés sont, naturellement, obtenus d'abord pour des variables indépendantes. Le plus célèbre des estimateurs reste l'estimateur à noyau, qui a été largement étudié, particulièrement pour les données complètes, indépendantes et identiquement distribuées (Prakasa Rao(1983),

Silverman (1986),...).
Les performances de l'estimateurs à noyau dépendent principalement du choix de h, le choix du noyau n'ayant pas une grande influence (voir la notion d'efficacité au sens d'Epanechnikov).
L'écart quadratique moyen (Mean Square Error, MSE en abrégé) est un critère très répandu dans la littérature comme mesure d'erreur locale (pour évaluer la précision d'une valeur estimée) :

$$MSE(t) = \mathbb{E}(f_n(t) - f(t))^2.$$

On peut décomposer le MSE en somme de la variance et du carré du biais de l'estimateur :

$$MSE(t) = \mathbb{E}(f_n(t) - \mathbb{E}(f_n(t))^2 + (\mathbb{E}f_n(t) - f(t))^2 = \sigma^2(t) + b^2(t).$$

En intégrant sur $\mathbb{R}$ on obtient le risque intégré (Mean Integrated Square Error ou MISE) comme mesure globale de l'erreur :$MISE(f_n) = \int MSE(t)dt$.
Pour le cas d'un noyau d'ordre r; on a sous certaines conditions,
$b^2(t_0) = O(h^r)$; et $\sigma^2(t_0) = O(\frac{1}{nh})$,(si $nh \to \infty$) ; pour tout $t_0 \in \mathbb{R}$; et donc $MSE(t_0) \leq C(h^r + \frac{1}{nh})$.
Le minimum en h du membre de droite est $h^{MSE} = O(n^{-\frac{1}{2r+1}})$ et par conséquent, pour cette fenêtre, uniformément en t_0, $MSE(t_0) = O(n^{-\frac{2r}{2r+1}})$; lorsque $n \to \infty$. (Tsybakov (2003)).
On retrouve les mêmes résultats optimaux pour le risque intégré (fenêtre et MISE). Pour un noyau d'ordre 2, la fenêtre optimale h^{MSE} ou h^{MISE} tend vers 0 à la vitesse $n^{-1/5}$,tandis que le MSE et le MISE à la vitesse $n^{-4/5}$.
Pour la convergence dans L^∞ (la norme infinie est définie par
$\|f\|_\infty = \sup_{t\in\mathbb{R}} |f(t)|$), sous la condition que $h = O(\frac{n}{\log n})^{-\frac{1}{2r+1}}$, on a

$$\sup_{t\in\mathbb{R}} |f_n(t) - f(t)| = O\Big(\frac{n}{\log n}\Big)^{-\frac{r}{2r+1}},$$

et plus généralement si $f^{(j)}$ désigne la dérivée j^{eme} de f,

$$\sup_{t\in\mathbb{R}} |f_n^{(j)}(t) - f^{(j)}(t)| = O\Big(\frac{n}{\log n}\Big)^{-\frac{r}{2r+1}}.$$

(Voir Gyorfi et al. (1989), Vieu (1996)).
Le cas de données complètes i.i.d. représente pour les spécialistes, le cas "idéal", un cas qui ne reflète pas toujours la réalité : que se passe-t-il si l'on n'a pas les conditions classiques, pas toujours vérifiables, d'indépendance ou de complétude ? Que deviennent les théories élaborées sous ces hypothèses "parfaites", lorsque l'on s'en écarte ? Au lieu d'observer des réalisations i.i.d. de la variable d'intérêt T (des durées de vie, par exemple), on observe la réalisation de T soumise à des perturbations, indépendantes ou non du phénomène.

1.2.2 Fonction de répartition

L'estimation de la fonction de répartition conditionnelle dans un cadre fonctionnel a été introduite par Ferraty et al. (2006). Ils ont construit un estimateur à double noyau pour la fonction de répartition conditionnelle et ils ont précisé la vitesse de convergence presque complète de cet estimateur lorsque les observations sont indépendantes et identiquement distribuées. Le cas des observations $\alpha-$mélangeantes a été étudié par Ferraty et al. (2005). Un exemple d'application sur la prévision via la médiane conditionnelle, ainsi que la détermination d'intervalles de prédiction ont été considérés dans cet article. Plusieurs auteurs ont traité l'estimation de la fonction de répartition conditionnelle comme une étude préliminaire de l'estimation des quantiles conditionnels. Citons par exemple, Ezzahrioui et Ould-Said (2005,2006) qui ont étudié la normalité asymptotique de cet estimateur dans les deux cas (i.i.d. et α-mélangeant). Une autre méthode d'estimation pour les quantiles conditionnels a été proposée par Laksaci al. (2009). Les résultats asymptotiques de cet article sont la convergence presque complète et la normalité asymptotique dans le cas i.i.d. Nous renvoyons à Cardot et al. (2004) pour une approche linéaire des quantiles conditionnels en statistique fonctionnelle. La contribution de la thèse sur ce modèle est l'étude de la convergence uniforme sur les deux arguments réels et fonctionnels de l'estimateur de la fonction de répartition conditionnelle. La vitesse de convergence de cet estimateur est précisée.

1.2.3 Fonction de régression

Les premiers résultats en statistique non paramétrique fonctionnelle ont été élaborés par Ferraty et Vieu (2000) et ils concernent l'estimation de la fonction de régression à variable explicative de dimension fractale. Ils ont établi la convergence presque complète d'un estimateur à noyau de ce modèle non paramétrique dans le cas i.i.d. En s'inspirant des développements récents de la théorie des probabilités de petites boules, Ferraty et Vieu (2004) ont généralisé ces derniers résultats au cas $\alpha-$mélangeant et ils ont exploité l'importance de la modélisation non paramétrique des données fonctionnelles en appliquant leur étude à la discrimination des courbes et à la prévision. Dans le cadre d'observations fonctionnelles $\alpha-$mélangeantes, Masry (2005) a montré la normalité asymptotique de l'estimateur de Ferraty et Vieu (2004) pour la fonction de régression. Le lecteur peut trouver dans le livre de Ferraty et Vieu (2006), un large éventail des applications de la fonction de régression en statistique fonctionnelle. La convergence en moyenne quadratique a été étudiée par Ferraty et al. (2007). Plus précisément, ils ont explicité le terme asymptotique exacte de l'erreur quadratique. Ce résultat a été utilisé par Rachdi et Vieu (2007) pour déterminer un critère de sélection automatique du paramètre de lissage basé sur la validation croisée. La version locale de ce critère a été étudiée par Benhenni et al. (2007). On trouvera dans cet article une étude comparative entre l'approche locale et globale. Comme travaux bibliographiques récents en régression, nous renvoyons le lecteur à Ferraty et Vieu (2011) ainsi qu'à Delsol (2011). Des résultats sur l'uniforme intégrabilité ont été établis par Delsol (2007,2009) et Delsol et al. (2011). D'autres travaux se sont intéressés à l'estimation de la fonction de régression en utilisant différentes approches : la méthode des k plus proches voisins par Burba et al. (2008), les techniques robustes par Azzidine et al. (2008), Attouch et al. (2009) et Crambes et al. (2008), l'estimation par la méthode simplifiée de polynôme locaux par Barrientos-Marin et al. (2010).

Un problème récurrent en statistiques est celui où l'on cherche à expliquer comment se comporte une variable d'intérêt Y en fonction d'une variable explicative X. Dans cette note, nous nous proposons d'étudier ce lien lorsque les deux variables sont hilbertiennes (à valeurs dans des espaces de Hilbert, notés respectivement $\mathcal{F}$ et $\mathcal{H}$:

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

où r est un opérateur de $\mathcal{F}$ dans $\mathcal{H}$ et ε est une variable aléatoire d'erreur.
La modélisation statistique des données fonctionnelles a connu un grand essor. Ce dernier est dû à la diversité des champs d'application dans lesquels les observations se présentent sous forme de courbes, de surfaces ou d'images.
Rappelons que le modèle de régression, qu'il soit paramétrique ou non paramétrique, reste l'outil le plus privilégié pour analyser la co-variabilité entre X et Y(quand seule la co-variable est fonctionnelle, cf. Bosq(2000) et Ramsay et Silverman (2002) pour le cadre paramétrique et Ferraty et Vieu(2006) pour le cadre non paramétrique). Notons que la littérature, sur le modèle non paramétrique de régression lorsque les deux variables sont fonctionnelles, reste très restreinte. En effet, on peut citer Dabo-Niang et Rhomari(2009), qui traite de la convergence en norme L_p de l'estimateur à noyau de la régression non paramétrique, ainsi que Ferraty,Laksaci,Tadj et Vieu (2011) , qui traite de la convergence presque complète et uniforme de l'estimateur à noyau de la fonction de régression d'une variable banachique sachant une variable explicative à valeurs dans un espace semi-métrique. En outre, la généralisation de ce résultat au cas de données dépendantes a été obtenue dans Ferraty,Laksaci,Tadj,et Vieu,(2012)(cf. aussi Ferraty,Van Keilegom et Vieu,(2012)pour la normalité asymptotique).
Dans la plupart des travaux sus-cités, les auteurs ont adopté la méthode à noyau pour l'estimation de l'opérateur de régression. Cependant, il est bien connu que cette méthode produit un biais plus élevé comparée à la méthode locale linéaire (cf. Fan et Gijbels (1996) pour le cas d'une covariable non fonctionnelle).

1.2.4 Fonction au hasard

La fonction de hasard, appelée parfois fonction de risque est très fréquemment utilisée pour l'étude de la fiabilité en statistique. Elle s'est développée très rapidement, motivée par ses applications dans des domaines exigeants. Elle mesure la probabilité instantanée qu'un évènement ait lieu à une date donnée, sachant qu'il n'a pas encore eu lieu juste avant cette date. Notons que l'usage de ce modèle s'est popularisé en économétrie, particulièrement pour l'analyse des transitions (trajectoires individuelles sur le marché du travail). Ainsi, on peut chercher à mesurer, pour un chômeur, l'évolution au fil du temps de sa propension à retrouver un emploi. Voir par exemple Florens et al.(1994), Lancaster (1990), entre autres. Les actuaires s'intéressent également à ces quantités, avec le souci de repérer les clients à risques, c'est-à-dire susceptibles d'induire des pertes pour la compagnie (non remboursement d'emprunts, risque de défaillance...).
La fonction de hasard se retrouve aussi bien sous forme continue que sous forme discrétisée. Dans le premier cas, l'estimation fonctionnelle non paramétrique s'impose d'elle même lorsqu'on n'a aucune idée sur la forme à priori de la fonction de hasard (ou lorsqu'on se refuse à émettre des hypothèses sur l'appartenance à une famille de lois particulières). Dans le second cas, on estime les taux de hasard (décès, panne) comme étant des paramètres. Les précurseurs de l'analyse non paramétrique furent Watson et Leadbetter (1964a, 1964b) ont proposé un estimateur pour lequel ils établissent la propriété de convergence asymptotiquement sans biais. L'estimation non paramétrique de la fonction de hasard a été abordée par de nombreux.
La littérature sur l'estimation de la fonction de hasard conditionnelle est relativement restreinte en statistique fonctionnelle. L'article de Ferraty et al.(2008) est un travail précurseur sur le sujet. Dans cette publication les auteurs ont établi la convergence presque complète d'un estimateur à noyau de la fonction de hasard conditionnelle, lorsque les observations sont indépendantes et identiquement distribuées. Dans le même contexte, Quantela-del Rio (2008) a établi la convergence presque complète et la normalité asymptotique de l'estimateur proposé par Ferraty et al.(2008) sur la fonction de hasard conditionnelle. L'auteur a illustré ces résultats 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.

1.3 Méthode locale lineaire

L'estimation de la fonction de hasard conditionnelle est une technique statistique qui permet une meilleure compréhension De la relation entre une variable de réponse et un ensemble de co-variables, en comparaison avec les méthodes habituelles de régression. Cette technique revêt donc une grande importance chez les scientifiques où la connaissance des moyens conditionnels obtenue par des méthodes de régression ne suffit pas à tirer de précieuses conclusions sur le problème à étudié. En outre, les fonctions de hasard apparaissent dans une variété de domaines. L'une des applications les plus utiles l'étude de la fiabilité et l'analyse des durées de vie. Cependant, la densité de probabilité, et son interprétation résultante, est conditionnelle sur l'hypothèse que le modèle utilisé pour produire les prévisions est correctement spécifié.
La méthode de noyau, la fonction du noyau $K(u)$, est supposée satisfaire certaines conditions spécifiques. Les choix de $K(u)$ sont définies en termes de fonctions de densité de probabilité univariées et unimodales. De plus, Youndjé (1993 et 1996) et d'autres donnent le biais, la variance, L'erreur quadratique moyenne (MSE) et les propriétés de convergence, aussi la proposition également d'un estimateur de noyau alternatif avec une (MSE) plus petite que l'estimateur standard dans certains situations courantes. D'autre part, nous ne pouvons pas continuer notre introduction sans mentionnant l'oeuvre de Fan et al.(1996), qui propose un autre estimateur densité conditionnelle en généralisant l'estimateur de Rosenblatt en utilisant des techniques polynomiales locales.
Les premiers résultats ont été obtenus par Ferraty et Vieu (2005) et Ferraty et al. (2006). Ils ont établi la convergence presque complète, Dans les deux cas les données i.i.d. et fortement mélangeant, des estimateurs du noyau du distribution conditionnel et de la densité de probabilité conditionnelle. En outre, ils ont présenté certaines applications de leurs résultats à la fois sur le mode conditionnel et sur les quantiles conditionnels. Parmi les nombreux articles qui concernent la modélisation non paramétrique à la distribution conditionnelle d'une variable réelle donnée une variable aléatoire , nous nous référons seulement à Dabo-Niang et Laksaci (2007) pour les mode d'estimation, et à Laksaci (2007) pour l'expression asymptotique des principaux termes dans l'erreur quadratique des estimateurs de noyaux de densité conditionnelle.

D'autre part, la méthode locale linéaire présente des avantages par rapport à l'estimation par la méthode du noyau, et encore plus récente. Il est bien connu que les estimateurs avec la méthode locale linéaire ont des belles propriétés asymptotiques lorsque la malédiction de la dimensionnalité est contrôlée au moyen de considérations appropriées sur les probabilités des petites boules de la variable fonctionnelle (voir Ferraty et Vieu 2006 et Références). Cependant, il est également bien connu que, comme dans le cadre standard dimensionnel fini, le paramètre de lissage doit être choisi de manière appropriée pour assurer une bonne modélisations (Voir Laksaci, 2007). Remarquons que certains articles (voir par exemple Youndjé et al. 1993), ont traité le problème de la sélection des paramètres de lissage dans les paramètres non paramétriques pour l'estimation de la densité conditionnelle. De plus, la sélection du paramètre de lissage Dans le réglage dimensionnel infini est beaucoup plus compliqué. En particulier, Scatterplot qui est un outil graphique permettant d'explorer la relation entre les variables et la réponse scalaire n'est pas disponible, et il devient donc très difficile d'avoir quelques informations sur la forme de la relation entre la variable fonctionnelle et la réponse scalaire. Par conséquent, différentes zones présentant des concentrations faibles ou élevées peuvent apparaître Dans une telle relation, même si elle ne figure pas dans l'échantillon de données fonctionnelles.

En outre, certains auteurs ont fait une comparaison entre la méthode locale constante (la méthode de base) et L'ajustement par la méthode locale linéaire, on peut cité parmis eux le travail de Fan (1992). En 2001, Irizarry (2001) a étudié la régression linéaire dans Un cadre de paramétrage linéaire assez général. Le livre de Loader (2004) est un document pédagogique Important pour l'estimation de la méthode linéaire locale. Deux ans plus tard, le gallois et Yee ont abordé le cas où la variable de réponse est multidimensionnelle. Ils ont évalué le bias et la variance asymptotique de l'estimateur local linéaire de la fonction de régression et la densité. La contribution de Hafen (2010) à l'ajustement local linéaire fait un grand développement dans les domaines pratique et théorique.

Tous les résultats que nous avons cités précédemment ont été obtenus pour une variable explicative de vecteur, d'autre part, Le cas d'une variable explicative fonctionnelle a été abordé très récemment. A ce stade, la littérature est très limitée. L'article de Baìllo et Grané (2009) est le premier dans ce domaine. Ils ont établi la convergence en norme L_2 d'une méthode locale linéaire lorsque la variable explicative est à valeurs dans un espace de Hilbert. Une généralisation au cas d'une variable de banach aléatoire a été développé par Barrientos-Marin et al. (2010). Dans cet article, Ces auteurs nous proposent une version fonctionnelle de la méthode locale très linéaire et les utilisations dans le domaine pratique. Nous nous référons aussi Demongeot et al. (2010) se sont abordés l'estimation de la densité conditionnelle dans un cas où les observations sont indépendantes équidistribuées. de plus ils ont établi la convergence ponctuel et uniforme presque complet pour l'estimateur construit. Deux ans plus tard l'estimation de la fonction de la densité et ses applications a été récemment traitée par Demongeot et al. (2012). Afin de terminer notre propre littérature sur la méthode locale linéaire, nous mentionnons l'oeuvre de Laksasi el al. (2013) dans le but de l'estimation de la fonction de distribution pour le cas spatiale.

1.4 Données incomplétes

1.4.1 Censure

C'est le phénomène le plus couramment rencontré en analyse de survie. La variable d'intérêt T, n'est pas observée (l'individu n'a pas subi l'évènement), et est majorée ou minorée par une variable ou une valeur (de censure), notée C qui, elle, a été observée. On considère une variable d'intérêt T (une durée de vie, par exemple). Au lieu d'observer les variables $T_1, T_2, ..., T_n$, qui nous intéressent, on n'observe T_i que lorsque f $\{T_i < C\}$, sinon on sait seulement que $\{T_i > C\}$. On parle alors de censure à droite, la plus fréquente.

Remarque Si l'individu a déjà subi l'évènement avant d'être observé (on observe C et non T_i), et que l'on sait que $\{T_i < C\}$, il y a censure à gauche.
L'évènement est dit censuré par intervalle, si au lieu de l'observer, on sait seulement qu'il est observé entre 2 dates $\{C_1 < T_i < C_2\}$.

Si Y est la variable réellement observée, on utilise la notation $Y_i = T_i \wedge C = \min(T_i, C)$. Ce type de censure (à droite) est souvent rencontré en fiabilité (pour les durées de vie des pièces fabriquées pendant une période précise), médecine (pour tester l'efficacité d'un traitement), biologie,...

On en distingue trois types :

a Censure de type I
Nous sommes dans le cas décrit précédemment où C est une valeur constante fixée.

b Censure de type II
Si on observe $Y_i = T_i \wedge T_{(r)}$; où r est un entier fixé, on parle de censure de type II ou de censure jusqu'au r^{ieme} "décès", $T_{(r)}$ étant la r^{ieme} statistique d'ordre.
Exemple(Droesbeke et al. (1989)) : Pour tester la fiabilité d'un système complexe, on met en état de fonctionnement n systèmes du même type, et on s'arrête lorsque la rième panne est observée.

c Censure de type III (ou censure aléatoire de type I)
Soient $C_1, ..., C_n$ des v.a. On observe les variables $Y_i = T_i \wedge C_i = \min(T_i, C_i)$.
L'information disponible peut être résumée par :
- la variable Y_i réellement observée,
- l'indicatrice $\sigma_i = \mathbb{I}_{\{T_i \leq C_i\}}$.

$\sigma_i = 1$ si on observe les "vraies" durées (l'évènement est observé), i.e. $Y_i = T_i$, $\sigma_i = 0$ si on observe les durées incomplètes (l'individu est censuré), i.e. $Y_i = C_i$.
Exemple (Koziol et Green (1976)) : Un essai clinique est réalisé sur 211 individus atteints du cancer de la prostate (phase 4) traités par oestrogène (hormone). A la fin de l'étude, 90 meurent du cancer de la prostate, 105 meurent d'autres causes et 16 sont encore vivants. Les censurés à droite sont les $105 + 16$ ($= 121$) individus qui ne sont pas morts du cancer de la prostate (objet de l'étude).

1.4.2 Missing data

Motivé par son interaction avec d'autres domaines appliqués ; l'analyse fonctionnelle des données (FDA) est devenue un sujet majeur de la recherche au cours des dernières années. Les principales références sur ce sujet sont les monographies de Ramsay et Silverman (2005), Bosq (2000), Ferraty et Vieu (2006), Cuevas (2014), Zhang (2014) et Hsing et Eubank (2015). Dans cet article, nous sommes intéressés par l'étude de l'estimation linéaire locale de la régression d'un opérateur. Lorsque ce régresseur est de type fonctionnel et sa réponse variable présente quelques observations manquantes.

Notez que l'approche polynomiale locale présente divers avantages par rapport à la méthode du noyau (Voir FAN et Gijbels (1996), pour avoir une idée approfondie sur la comparaison entre ces 02 méthodes). La modélisation linéaire locale a été récemment introduite dans la FDA par Baìllo et Grané (2009). Ils ont prouvé la L^2-consistance de l'estimation locale linéaire de la fonction de régression lorsque la variable explicative prend des valeurs dans l'espace de Hilbert. Barrientos et al. (2010) ont introduit une version alternative et simplifiée qui peut être utilisé pour des co-variables fonctionnelles plus générales. Chouaf et Laksaci (2012) dans la mesure où ils sont concernés, ont étudié une version spatiale de l'estimation proposée par Barrientos et al. (2010). Ils ont établi la cohérence presque complète de cette estimation lorsque les observations sont spatialement dépendantes. Nous retournons à Demongeot et al. (2014) pour l'estimation fonctionnelle locale linéaire de la fonction de distribution cumulative conditionnelle. Tous ces travaux ne concernent que le cas complet des données. Notre but principal, dans cet article, est d'étudier le cas où la variable de réponse des données manquantes est subit au hasard (aléatoire) (MAR).

L'analyse de régression lorsque les données sont incomplètes a suscité une attention considérable dans la multi variété et statistique. En particulier, plusieurs modèles de régression sont considérés comme des données manquantes. Nous citons par exemple : Yates (1933) pour la régression linéaire, Cheng (1994) pour la méthode du noyau, Chen et Shao (2000) pour l'approche de voisinage le plus proche. Pèrez-Gonzalez et al. (2009) pour la régression polynomiale locale. Boente et al. (2009) pour la M-régression. Cependant, la littérature sur ce sujet est encore limitée.

Les premièrs résultats sur ce sujet sont obtenus par Farraty et al. (2013). Ils ont construit une estimation de régression lorsque le régresseur est complètement observé et que la réponse scalaire sera variable. Ils ont déclaré que l'asymptotique propriété de cette estimation est classée dans le cas i-i-d. Ling et al. (2015) ont établi la convergence dans la probabilité et la normalité asymptotique et la même estimation, mais dans le cas ergodique. Nous retournons à Ling et al. (2016) pour l'estimation de mode conditionnel dans le cas ergodique MAR.

L'objectif principal du présent travail est de construire une autre estimation paramétrique de la régression fonctionnelle lorsqu'il y a des réponses manquantes. Plus précisément, nous construisons une nouvelle estimation de la fonction non-paramétrique, par la méthode de lissage linéaire et nous montrons ses propriétés asymptotiques. En particulier, dans certaines conditions générales nous prouvons une ponctualité presque complète et cohérente de cette estimation.

Bibliographie

[1] Akritas, M., Politis, D. (2003). Recent advances and trends in nonparametric statistics. Elsevier, Amsterdam .

[2] Aguilera A.M, F.A. Ocana, and M.J. Valderrama. (1997). An approximated principal component prediction for continuous time stochastic processes. *Applied Stochastic Models and Data Analysis,* 13 , 61-72.

[3] Attouch, M., Laksaci, A., Ould-Saïd, E. (2009). Asymptotic Distribution of Robust Estimator for Functional Nonparametric Models. *Communications in Statistics : Theory and Methods.* 38, 1317-1335.

[4] Azzeddine, N., Laksaci, A., Ould-Saïd, E. (2008). On the robust nonparametric regression estimation for functional regressor. *Statistic and Probabability Letters.* 78, 3216-3221.

[5] Baìllo, A., Grané, A. (2009). Local linear regression for functional predictor and scalar response. *Journal of Multivariate Analysis,* 100, 102-111.

[6] Barrientos-Marin, J., F. Ferraty, F and Vieu, P. (2010). Locally modelled regression and functional data. *Journal of Nonparametric Statistics,* 22(5), 617-632.

[7] Benhenni, K., Ferraty, F., Rachdi, M., Vieu, P. (2007). Locally smoothing regression with functional data. *Computational. Statistics.* 22, 353-370.

[8] Berlinet, A. Elamine, A. Mas.(2011). Local linear regression for functional data. *Ann. Inst. Stat. Math.* 63, 1047-1075.

[9] Besse and J.O. Ramsay. (1986). Principal component analysis of sampled curves. *Psychometrica,* 51, 285-311.

[10] Besse. (1991). Approximation spline de l'analyse en composantes principales d'une variable alťeatoire hilbertienne. *Annales de la Facultťe des Sciences de Toulouse, XII(3)* , 329-349.

[11] Biau, F. Bunea, and M. Wegkamp. (2005). Functional classification in Hilbert spaces. *EEE Transactions on Information Theory,* 51 , 2163-2172.

[12] Boente, G., Gonzalez-Manteiga, W. and Perez-Gonzalez, A. (2009) Robust nonparametric estimation with missing data. *J. Statist. Plann. Inference,* 139, 571-592.

[13] Boj, P. Delicado, J. Fortiana.(2010). Distance-based local linear regression for functional predictors, Comput. *Stat. Data Anal.* 54, 429-437.

[14] Bosq, D.(2000). Linear Processes in Function Spaces, Theory and Applications. *Lecture Notes in Statistics,* 149, Springer-Verlag, New York .

[15] Burba, F., Ferraty, F., Vieu, P. (2008). Convergence de l'estimateur á noyau des k plus proches voisns en régression fonctionnelle non-paramétrique. *C. R. Acad. Sci. Paris.* 346, 339-342.

[16] Cardot.H, F. Ferraty, and P Sarda. (1999). Functional Linear Model. *Statistics and Probability Letter,* 45, 11-22.

[17] Cardot, H., Crambes, C., Sarda, P. (2004). Spline estimation of conditional quantiles for functional covariates, C. R. *Math. Acad. Sci. Paris.* 339, 141-144.

[18] Chen, J.H., Shao, J., (2000). Nearest neighbor imputation for survey data. *J. Official Statist.,* 16, 113-131.

[19] Cheng, P.E., (1994). Nonparametric estimation of mean functionals with data missing at random. *J. Amer. Statist. Assoc.* 89, 81-87.

[20] Chouaf, A. and Laksaci, A., (2013). On the functional local linear estimate for spatial regression. *Stat. Risk Model,* 29, 189-214.

[21] Clarkson, D.B., Fraley, C., Gu, C.C., Ramsay, J.S.(2005). S+ functional data analysis user's guide. Comp. Statist. Series, Springer, New York.

[22] Crambes, C., Delsol, L., Laksaci, A. (2008). L_p errors for robust estimators in functional nonparametric regression. *Journal of Nonparametric Statistics.* 20, 573-598.

[23] Cuevas,(2014). A partial overview of the theory of statistics with functional data, *J. Stat. Plan. Inference* 147 , 1-23.

[24] Dabo-Niang, S., Laksaci, A. (2007). Propriétés asymptotiques d'un estimateur à noyau du mode conditionnel pour variable explicative fonctionnelle. (French) [Asymptotic properties of the kernel estimator of the conditional mode when the regressor is functional]. *Ann.* I.S.U.P. 51 , 27-42.

[25] Dauxois, L. Ferré, and A.F. Yao.(2001). Un modèle semi-paramétrique pour variable aléatoire hilbertienne. *C.R. Acad. Sci. Paris,* 327(I), 947-952.

[26] Dauxois, A. Pousse. (1976) *Les analyses factorielles en calcul des probabilités et en statistique : essai d'étude synthétique.* Thèse, Université Toulouse III.

[27] Deheuvels, P., Einmahl, J. (2000). Functional limit laws for the increments of Kaplan-Meier product limit processes and applications. *Ann. Probab.* 28, 1301-1335.

[28] Delsol, L. (2007). Régression non paramétrique fonctionnelle : expression asymptotique des moments. *Ann. I.S.U.P. Vol LI,* 3, 43-67.

[29] Delsol, L. (2009). Advances on asymptotic normality in nonparametric functional Time Series Analysis Statistics. *Statistics,* 43, 13-33.

[30] Delsol, L. (2011). Nonparametric methods for mixing functional random variables. *In The Oxford Handbook of Functional Data Analysis (Ed. F. Ferraty and Y. Romain).* Oxford University Press.

[31] Demongeot, J., Laksaci, A., Madani, F. and Rachdi, M. (2010). Local linear estimation of the conditional density for functional data. C. R. *Math., Acad. Sci. Paris,* 348, 931-934.

[32] Demongeot, J. , Laksaci, A. , Madani, F. and Rachdi, M. (2012). Functional data : Local linear estimation of the conditional density and its application Statistics C. R. *Math., Acad. Sci. Paris,* 348, 931-934.

[33] Demongeot, A. Laksaci, F. Madani, M. Rachdi.(2013). Functional data : local linear estimation of the conditional density and its application, *Statistics.* 47, 26-44.

[34] Demongeot, J., Laksaci, A., Rachdi, M. and Rahmani, S. (2014). On the local linear modelization of the conditional distribution for functional data. *Sankhya A,* 76, 328-355.

[35] Deville,. J.C.(1974). Deville. Méthodes statistiques et numériques de l'analyse harmonique. *Annales de l'INSEE,* 15(Janvier-Avril), 3-97 .

[36] Droesbeke, J.J., Fichet, B., Tassi, P. (1989). Analyse statistique des durées de vie. *Economica.*

[37] Ezzahrioui, M., Ould-Saïd, E. (2005). Asymptotic normality of nonparametric estimators of the conditional mode for functional data. *Technical report, No.249, LMPA, Univ. Littoral Côte d'Opale.*

[38] Ezzahrioui, M., Ould-Saïd, E. (2006). On the asymptotic properties of a nonparametric estimator of the conditional mode for functional dependent data. *Preprint, LMPA No 277, Univ. du Littoral Côte d'Opale.*

[39] Fan, J. (1992). Design-adaptative nonparametric regression. *Journal of the American Statistical association,* 87, 998-1004.

[40] Fan, J., Gijbels, I. (1996). Local Polynomial Modelling and its Applications. *Monographs on Statistics and Applied Probability 66,* Chapman and Hall, London.

[41] Ferraty, F. Laksaci, A., Vieu, P. (2005). Functional time series prediction via conditional mode estimation. C. R. Math. Acad. Sci. Paris. 340, 389-392.

[42] Ferraty, F. Laksaci, A., Vieu, P. (2006). Estimation of some characteristics of the conditional distribution in nonparametric functional models. Statistical Inference for Stochastic Processes. 9, 47-76.

[43] Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference*, 140, 335-352.

[44] Ferraty, A. Laksaci, A. Tadj, P. Vieu.(2011). Kernel regression with functional response. *Electronic. J. Stat.* 5, 159-171.

[45] Ferraty, A. Laksaci, A. Tadj, P. Vieu.(2012). Estimation de la fonction de régression pour variable explicative et réponses fonctionnelles dépendante. *C. R. Acad. Sci. Paris, Ser.* I 350, 717-720.

[46] Ferraty, F., Rabhi, A. and Vieu, P. (2005). Conditional quantiles for dependent functional data with application to the climatic El Nino phenomenon. *Sankhya.* 67 , 378-398.

[47] 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.

[48] Ferraty, I. Van Keilegom, P. Vieu.(2012). Regression when both response and predictor are functions, J. Multivar. *Anal.* 109, 10-28.

[49] Ferraty, F., Vieu, P. (2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. C. R. *Acad. Sci., Paris.* 330, No.2, 139-142.

[50] Ferraty and P. Vieu. (2002). The functional nonparametric model and application to spectrometric data. *Computational Statistics,* 17, 515-564.

[51] Ferraty, F., Vieu, P. (2004). Nonparametric models for functional data, with application in regression times series prediction and curves discrimination. *J. Nonparametric Statist.* 16, 111-125.

[52] Ferraty, F., Vieu, P. (2006). Nonparametric functional data analysis. Theory and Practice. *Springer Series in Statistics. Springer. New* York.

[53] Ferraty, F., Vieu, P. (2007). Nonparametric regression on functional data : inference and practical aspects. *Aust. N. Z. J. Stat.* 49, 267-286.

[54] Ferraty, F., Vieu, P. (2011). Kernel regression estimation for functional data. *In the Oxford Handbook of Functional Data Analysis (Ed. F. Ferraty and Y. Romain).* Oxford University Press.

[55] Ferraty, F., Sued, M. and Vieu, P. (2013). Mean estimation with data missing at random for functional covariables. *Statistics,* 47, 688-706.

[56] Ferré and A.F. Yao,(2003). Functional sliced inverse regression analysis. *Statistics,* 37, 475-488.

[57] Ferré and A.F. Yao, (2005). Smoothed functional inverse regression. *Statistica Sinica,* 15(3), 665-683.

[58] Ferré and N. Villa. (2005a). Discrimination de courbes par régression inverse fonctionnelle. *Revue de Statistique Appliquée,* LIII(1), 39-57.

[59] Florens, J. P., Larribeau, S. and Mouchart, M. (1994). Bayesian Encompassing Tests of a Unit Root Hypothesis. *Econometric Theory.* 10, 747-763.

[60] Hafen, R. P. (2010). Local regression models : Advancements, applications, and new methods. Thesis (Ph.D.).

[61] Hastie, R. Tibshirani, and A. Buja. (1994). Flexible discriminant analysis by optimal scoring. *Journal of the American Statistical Association,* 89, 1255-1270.

[62] Hastie, A. Buja, and R. Tibshirani. (1995). Penalized discriminant analysis. *Annals of Statistics,* 23, 73-102.

[63] Hastie, R. Tibshirani, and J. Friedman. (2001). The Elements of Statistical Learning : Data Mining, Inference and Prediction. Springer-Verlag.

[64] Helland, I.(1990). Pls regression and statistical models, Scand. *J. Statist.* 17, 97-114 .

[65] Hsing, R. Eubank. (2015). Theoretical Foundations of Functional Data Analysis, With an Introduction to Linear Operators, Wiley Series in Probability and Statistics, John Wiley and Sons, Chichester, UK.

[66] Irizarry, R. A. (2001). Local Regression With Meaningful Parameters Local Regression With Meaningful Parameters. *The American Statistician Volume 55, Issue 1,* 72-79.

[67] Ivanov.V.K.(1962). On linear problems which are not well-posed. *Soviet. Math.Docl.,* 145(2), 1962.

[68] James and T.J. Hastie, (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society, Series B,* 63, 533-550.

[69] Koziol, J.A., Green, S.B. (1976). A Cramer-Von Mises statistic for randomly censored data. *Biometrika* 63, 465-474.

[70] Laksaci, A. (2007). Erreur quadratique de l'estimateur á noyau de la densit é conditionnelle à variable explicative fonctionnelle. (French) [Quadratic error of the kernel

estimator of conditional density when the regressor is functional.] *C. R. Math. Acad. Sci. Paris.* 345, 171-175.

[71] Laksaci, A., Lemdani, M., Ould-Saïd, E. (2009). A generalized L1-approach for a kernel estimator of conditional quantile with functional regressors : consistency and asymptotic normality. *Statist. Probab. Lett.* 79, 1065-1073.

[72] 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.

[73] Laksaci, A., Rachdi, M. and Rahmani, S. (2013). Spatial modelization : Local linear estimation of the conditional distribution for functional data. *Spatial Statistics.* 6, 1-23.

[74] Lancaster, T. (1990). The econometric analysis of the transition data. Cambridge University Press.

[75] Leurgans, R.A. Moyeed, and B.W. Silverman. (1993). Canonical correlation analysis when the data are curves. *Journal of the Royal Statistical Society, Series B,* 55, 725-740.

[76] Li. Ferré and A.F. Yao. (2005). Smoothed functional inverse regression. Sta- tistica Sinica, 15(3), 665-683.

[77] Li. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association,* 86, 316-342.

[78] Ling, N., Liu, Y. and Vieu, P.(2015). Nonparametric regression estimation for functional stationary ergodic data with missing at random. *J. Statist. Plann. Inference,* 162, 75-87.

[79] Ling, N., Liu, Y. and Vieu, P. (2016). Conditional mode estimation for functional stationary ergodic data with responses missing at random. *Statistics,* 50, 991-1013.

[80] Loader, C. (2004). Smoothing : local regression techniques. Handbook of computational statistics, 539-563, Springer, Berlin.

[81] Masry, E.(2005). Nonparametric regression estimation for dependent functional data : Asymptotic normality. *Stoch. Proc. and their Appl.* 115, 155-177.

[82] Nadaraja, E.A. (1965). On nonparametric estimation of density function and regression. *Theory of Probability and its applications* 10, 186-190.

[83] Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* 33, 1065-1076.

[84] Perez-Gonzalez, A., Vilar-Fernandez and J. M. Gonzalez-Manteiga, W. (2009). Asymptotic properties of local polynomial regression with missing data and correlated errors. *Ann. Inst. Statist. Math.,* 61, 85-109.

[85] Pezzulli and B.W. Silverman. (1993). Some properties of smoothed principal components analysis for functional data. *Computational Statistics,* 8, 1-16.

[86] Preda and G. Saporta, (2002). Régression PLS sur un processus stochastique. *Revue de statistique appliquée,* L(2).

[87] Preda and G. Saporta. (2005a). Clusterwise PLS regression on a stochastic process. *Computational Statistics and Data Analysis,* 49(1), 99-108.

[88] Preda and G. Saporta. (2005b). PLS discriminant analysis for functional data. In ASMDA *2005 proceedings,* 653-661, Brest, France.

[89] Quintela-Del-R`io, A. (2008). Hazard function given a functional variable : Non-parametric estimation under strong mixing conditions. *J. Nonparametr. Stat.* 20, 413-430.

[90] Rachdi, M., Vieu, P. (2007). Nonparametric regression functional data : Automatic smoothing parameter selection. *J. Statist. Plan. Inf.* 137, 2784- 2801.

[91] Ramsay and Silverman, (1997). *Functional Data Analysis.* Springer Verlag, New York.

[92] Ramsay, J., Silverman, B.W. (2002). Applied functional data analysis ; Methods and case studies. *Springer-Verlag,* New York.

[93] Ramsay, J., Silverman, B.W.(2005). Functional Data Analysis. 2nd Edition. Springer-Verlag, New York.

[94] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc.Natl. Acad. Sci. USA* 42, 43-47.

[95] Rossi, B. Conan-Guez, and A. El Golli. (2004). Clustering functional data with the som algorithm. In ESANN'2004 *proceedings,* 305-312, Bruges, Belgique.

[96] Rossi, N. Delannay, B. Conan-Guez, and M Verleysen. (2005). Representation of functional data in neural networks. *Neurocomputing,* 64, 183-210.

[97] Rossi and N. Villa. (2005a). Classification in Hilbert spaces with support vector machines. In ASMDA 2005 proceedings, 635-642, Brest, France.

[98] Saporta.G.(1981). Méthodes exploratoires d'analyses des données temporelles. *Cahiers du BURO,* 37-39.

[99] Schimek, M. (2000). Smoothing and regression ; Approaches, Computation, and Application. *Wiley Series in Probability and Statistics,* Wiley, New York .

[100] Silverman. (1996). Smoothed functional principal components analysis by choice of norm. *Annals of Statistics,* 24, 1-24.

[101] Tihonov.(1963a). Regularization of incorrectly posed problems. *Soviet Math.Docl.,* 4, 1624-1627.

[102] Tihonov. (1963b). Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Docl.,* 4, 1036-1038.

[103] Tsybakov, A. B. (2003). Introduction à l'estimation non paramétrique. *Mathématiques et applications, 41, Springer.*

[104] Vieu, P. (1996). A note on density mode estimation. Statist. probab. Lett. 26, 297-307.

[105] Watson, G. S. (1964a). Smooth regression analysis. *Sankhya.* 26, 359-372.

[106] Watson, G. S. and Leadbetter, M. R. (1964b). Hazard analysis. I. *Biometrika.* 51, 175-184.

[107] Yates, F.,(1933). *T*he analysis of replicated experiments when the field results are incomplete. *Emporium J. Exp. Agriculture,* 1, 129-142.

[108] Youndjé, E. (1993). Estimation non-paramétrique de la densité conditionnelle par la méthode du noyau. Thèse 3eme cycle, Université de Rouen.

[109] Youndjé, E., Sarda, P. and Vieu, P. (1996). Optimal smooth hazard estimates. Test 5, 2, 379-394.

[110] Zhang. (2014). Analysis of Variance for Functional Data. *Monographs on Statistics and Applied Probability,* vol.127, CRC Press, Boca Raton, FL, USA.

Chapitre 2

Local linear estimate for functional regression with missing data at random

[1] Laboratoire Statistique et Processus Stochastiques
Univ. Djillali Liab
'es, BP 89, S. B. A. 22000, Algeria.
e-mails : benchiha.abbassia@yahoo.fr,
[2] Department of Mathematics, College of Science,
King Khalid University, Abha 61413, Saudi Arabia
kaidzoulikha312@gmail.com

Abstract In this paper, we consider the problem of the co-variability analysis between a functional variable X and a scalar response variable Y which is not totally observed. We use the local linear approach to model this relationship by constructing a local linear estimator of the regression operator when missing data occur in the response variable. Asymptotic results, in term of the pointwise almost complete consistencies, is established for the constructed estimator.
Keywords Almost complete convergence · Local linear method · Missing data · Uniform consistency · Regression operator.

AMS Subject classification Primary, 62G20 ; secondary 62G08 · 62G35 · 62E20.

2.1 Introduction

Motivated by its interaction with other applied fields, the functional data analysis (FDA) became a major topic of research in recent years. A key references on this topic are the monographs of Ramsay and Silverman (2005), Bosq (2000), Ferraty and Vieu (2006), Cuevas (2014), Zhang (2014) and Hsing and Eubank (2015). In this paper, we are interested by studying the local linear estimate of the regression operator when the regressor is of functional kind and the response variable presents some missing observations.

Notice that the local polynomial approach has various advantages over the kernel method, namely in the bias term (see, Fan and Gijbels 1996, for an extensive discussion on the comparison between both these methods). The local linear modeling has been recently introduced in the FDA by Baìllo and Grané (2009). They proved the L^2-consistency of the local linear estimate of the regression function when the explanatory variable takes values in a Hilbert space. Barrientos et al. (2010) have considered an alternative and simplified version which can be used for more general functional covariates. Chouaf and Laksaci (2012) as far as they are concerned have studied a spatial version of the estimate proposed by Barrientos et al. (2010). They established the almost complete consistency of this estimate when the observations are spatially dependent. We return to Demongeot et al. (2014) for the functional local linear estimate of the conditional cumulative distribution function. All these works are only concerned with the complete data case. Our main purpose, in this paper, is to study the case when the response variable has missing data at random (MAR).

The regression analysis when data are incomplete has received considerable attentions in the multivariate statistics literature. In particular, several regression models are considered for missing data. We cite, for instance, Yates (1933) for the linear regression, Cheng (1994) for the kernel method, Chen and Shao (2000) for the nearest neighbor approach, Pérez-Gonzàlez et al. (2009) for the local polynomial regression and Boente et al. (2009) for the M-regression. However, the literature on this topis is still limited.

First results on the sus-mentioned topic, in nonparametric statistics, are obtained by Ferraty et al. (2013). They constructed an estimate of the regression operator when the functional regressor is completely observed and MAR scalar response variable. They stated the asymptotic properties of this estimate in the i.i.d. case. Ling et al. (2015) have established the convergence in probability and the asymptotic normality of the same estimate but in the ergodic case. We return to Ling et al (2016) for the conditional mode estimation in the case of functional ergodic MAR data.

The main aim of the present work is to construct an alternative nonparametric estimation of the functional regression when there are missing responses. More precisely, we construct a new estimate of the functional nonparametric regression by the local linear smoothing method and we show its asymptotic properties. Specifically, under some general conditions, we prove the pointwise almost complete consistency of this estimate.

The paper is organized as follows. We present our model in Section 2. In Section 3, we state the pointwise convergence with rate of our estimate. The uniform consistency is treated in Section 4. The proofs of different results are relegated to the last section.

2.2 Model and its estimate

Consider n independent pairs of random variables (X_i, Y_i) for $i = 1, \ldots, n$ that we assume drawn from the pair (X, Y). The latter is valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a semi-metric space and d denotes a semi-metric.

Usually, the local polynomial estimate is constructed by approximating, locally, the nonparametric model by a polynomial function. In functional statistics, when the data are complete, Barrientos et al. (2010) have proposed to approximate the regression operator $R(x) = \mathbb{E}[Y|X = x]$ in a neighborhood of x by

$$R(Z) = a + b\beta(z, x), \text{ where } z \text{ is in a neighborhood of } x.$$

The scalars a and b are estimated by

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} (Y_i - a - b\beta(X_i, x))^2 \, K(h^{-1}\varrho(x, X_i)) \tag{2.1}$$

where K is a kernel and $h = h_{K,n}$ is a sequence of positive real numbers, $\beta(.,.)$ is a known function from $\mathcal{F}^2$ into $\mathbb{R}$ such that, $\forall \xi \in \mathcal{F}$, $\beta(\xi, \xi) = 0$ and $\varrho(.,.)$ is a function of $\mathcal{F} \times \mathcal{F}$ such that $d(.,.) = |\varrho(.,.)|$.

In what follows, we consider the case where the response variables are MAR data. In this case, we introduce a Bernoulli random variable δ such that $\delta = 1$ if Y is observed and $\delta = 0$ otherwise. The MAR consideration implies that the two variables Y and δ are independent conditionally to a given X, in the sense that

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

where $P(\cdot)$ is an unknown functional operator which is called the conditional probability of observing the response variable Y given the explanatory variable X.

Then, by combining the ideas of Pérez-Gonz
'alez et al. (2009) (on local polynomial regression analysis of vectorial MAR data) and Barrientos et al. (2010) (on the functional local linear smoothing method) we construct our estimate from the sample $(X_i, \delta_i, Y_i)_{i=1,\dots,n}$ of (X, δ, Y) by estimating a and b as the minimizers of

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} (Y_i - a - b\beta(X_i, x))^2 \, \delta_i K(h^{-1}\varrho(x, X_i)). \tag{2.2}$$

Formally, $(\widehat{a}, \widehat{b})$ is a solution of the system

$$(Q'_B K Q_B) \begin{pmatrix} a \\ b \end{pmatrix} - (Q'_B K Y) = 0$$

which allows to

$$\begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix} = (Q'_B K Y)^{-1}(Q'_B K Q_B),$$

where $Q'_B = \begin{pmatrix} 1 & \cdots & 1 \\ \beta(X_1, x) & \cdots & \beta(X_n, x) \end{pmatrix}$,
$K = diag(\delta_1 K(h^{-1}\varrho(x, X_1)), \dots, \delta_n K(h^{-1}\varrho(x, X_n)))$ and $Y' = (Y_1, \dots, Y_n)$. Moreover, as $\beta(x, x) = 0$, we can write

$$\widehat{R}(x) = \frac{\sum_{i,j=1}^{n} W_{ij}(x) Y_j}{\sum_{i,j=1}^{n} W_{ij}(x)} \tag{2.3}$$

where

$$W_{ij}(x) = \beta(X_i, x) \, (\beta(X_i, x) - \beta(X_j, x)) \, \delta_i \delta_j K(h^{-1}\varrho(x, X_i)) K(h^{-1}\varrho(x, X_j))$$

with the convention $0/0 = 0$.

It is worth noting that, the present contribution generalizes both works of Ferraty et al. (2013) and Barrientos et al. (2010). Indeed, the estimate of Ferraty et al. (2013) can be obtained by taking $b = 0$, while the completely observed data treated by Barrientos et al. (2010) may be obtained by putting $\delta = 1$.

2.3 Pointwise almost complete convergence

In what follows x denotes a fixed curve in $\mathcal{F}$, N_x denotes a fixed neighborhood of x, $S_{\mathbb{R}}$ will be a fixed compact subset of $\mathbb{R}$ and $\phi_x(r_1, r_2) = \mathbb{P}(r_2 \leq \varrho(X, x) \leq r_1)$. In order to enounce our results we consider the following conditions :

(H1) For any $r > 0$, $\phi_x(r) := \phi_x(-r, r) > 0$.

(H2) There exists a constant $C > 0$ such that, for all real $m \geq 1$, $E(|Y|^m|X = x) < C$ and for all $(x_1, x_2) \in N_x \times N_x$:

$$|R(x_1) - R(x_2)| \leq C_x \, d^{\kappa}(x_1, x_2), \quad \text{for some} \kappa > 0$$

where C_x is a positive constant depending on x.

(H3) The function $\beta(.,.)$ is such that :

$$\text{for all } x' \in \ \mathcal{F}, \ C_1 \, |\varrho(x, x')| \leq |\beta(x, x')| \leq C_2 \, |\varrho(x, x')|,$$

where $C_1 > 0$, $C_2 > 0$.

(H4) K is a positive and differentiable function which is supported within $[-1, 1]$.

(H5) The bandwidth h satisfies :

$$\lim_{n \to \infty} \frac{\log n}{n\phi_x(h)} = 0.$$

and

$$\forall n > n_0, \ -\frac{1}{\phi_x(h)} \int_{-1}^{1} \phi_x(zh, h) \frac{d}{dz} \left(z^2 K(z)\right) dz > C_3 > 0$$

and

$$h \int_{B(x,h)} \beta(u, x) dP(u) = o\left(\int_{B(x,h)} \beta^2(u, x) \, dP(u)\right)$$

where $B(x, r) = \{x' \in \mathcal{F}/|\delta(x', x)| \leq r\}$

(H6) The function $P(\cdot)$ is continuous on N_x and such that $P(x) > 0$.

Notice that, all these hypotheses are very standard. In particular, assumptions (H1), (H3) and (H5) are used by Barrientos et al. (2010). While Assumption (H2) is a mild regularity condition which allows to characterize the functional space of our nonparametric model. This condition is usually used to evaluate the bias term in the asymptotic results of the estimate. Finally, we point out that Assumption (H6) is usual in functional MAR models (see, for instance Ling et al., 2015).

The following theorem gives the almost-complete convergence of the estimate.

Theorem 1 *Under assumptions (H1)-(H6), we have*

$$|\widehat{R}(x) - R(x)| = O\left(h^{\kappa}\right) + O\left(\sqrt{\frac{\log n}{n\ \ \phi_x(h)}}\right), \ a.co.$$

Proof of Theorem 1. We consider the following decomposition :

$$\widehat{R}(x) - R(x) = \widehat{B}(x) + \frac{\widehat{M}(x)}{\widehat{R}_D(x)} + \frac{\widehat{Q}(x)}{\widehat{R}_D(x)}$$

where

$$\begin{aligned}
\widehat{Q}(x) &:= \left(\widehat{R}_N(x) - \mathbb{E}[\widehat{R}_N(x)]\right) - R(x)\left(\widehat{R}_D(x) - \mathbb{E}[\widehat{R}_D(x)]\right) \\
\widehat{B}(x) &:= \frac{\mathbb{E}[\widehat{R}_N(x)]}{\mathbb{E}[\widehat{R}_D(x)]} - R(x) \quad \text{and} \quad \widehat{M}(x) := -\widehat{B}(x)\left(\widehat{R}_D(x) - \mathbb{E}[\widehat{R}_D(x)]\right)
\end{aligned}$$

with

$$\widehat{R}_N(x) = \frac{1}{n\,h^2\phi_x(h)}\sum_{i,j}^{n} W_{ij}(x)Y_j \text{ and } \widehat{R}_D(x) = \frac{1}{nh^2\phi_x(h)}\sum_{i,j}^{n} W_{ij}(x).$$

Thus, the proof of Theorem 1 is based on the following intermediate results, for which the proofs are given in the Appendix.

Lemma 1 *Under the hypotheses of Theorem 1, we have that :*

$$\left|\widehat{R}_D(x) - \mathbb{E}[\widehat{R}_D(x)]\right| = O_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right).$$

and

$$|\widehat{R}_N(x) - \mathbb{E}[\widehat{R}_N(x)]| = O_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right).$$

Corollary 1 *Under the hypotheses of Lemma 1, there exists a positive real C such that*

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\widehat{R}_D(x) < C\right) < \infty.$$

Lemma 2 *Under assumptions (H1), (H2) and (H4) we have*

$$\left|\widehat{B}(x)\right| = O(h^{\kappa}).$$

$\blacksquare$

2.4 Discussions and conclusions

— **On the scope of applicability of our approach** . The novelty of this contribution is the consideration of the missing data case. This situation is more natural than the complete data case in functional statistics. Indeed, usually the functional data are observed only over limited grids. Thus, the missing functional data is more realistic situations for functional data analysis. So, it is really necessary to develop some statistical methods/models adapted to the incomplete functional data. In this contribution we focus in the local linear approach which is an alternative method to the kernel method. The considered approach has a significative advantages over the first one, namely in the bias term (see the following remark). Moreover, the kernel method can be treated as particular case of the local linear method by taking $b = 0$. Thus, in conclusion we can say that, the main feature of the present contribution is the fact that it treat a very realistic situation by a more general method which permits to increase the the scope of applicability of our approach.

— **On the convergence rate of the estimator compared to the kernel method** It is well known that the local linear method reduces considerably the bias of the kernel method. Indeed, the principal idea of the local linear smoothing is based in the linearity assumption of the regression operator at neighborhood of the conditioning point. In the non functional case this linear approximation by a Taylor expansion, but in our functional context can be expressed by

$$\text{for any } z \text{ neighborhood at } x\, R(z) = a + b\beta(x, z) + o(\beta(x, z)). \tag{2.4}$$

In the other word if we replace (H2) by decomposition (2.5) and we use the fact that $\sum_{ij} \beta_j W_{ij} = 0$ to reformulate our bias term as follows

$$\begin{aligned} E\left[\widehat{R}(x)\right] - R(x) &= E\left[\frac{\sum_{i,j=1}^n W_{ij}(x)(Y_j - R(X_j))}{\sum_{i,j=1}^n W_{ij}(x)}\right] \\ &\quad + E\left[\frac{\sum_{i,j=1}^n W_{ij}(x)(R(X_j) - R(x) + b\beta(X_j, z))}{\sum_{i,j=1}^n W_{ij}(x)}\right]. \end{aligned}$$

So, under (2.5) the seconde term is of $o(h^2)$. Finally, the bias term is of order

$$o(h).$$

The latter is significantly better than the kernel case studied by Ferraty et al. (2013) in a missing case where we have an order of

$$O(h).$$

Moreover, if we replace the previous decomposition (2.5) by
for any z neighborhood at x,

$$R(z) = a + b\beta(x,z) + o(\beta(x,z)) + c\beta^2(x,z) + o(\beta^2(x,z)). \tag{2.5}$$

we obtain, by the same arguments, a bias term

$$C'_K h^2.$$

So, we can say that similarly to the complete data case the local linear approach improve also, the bias term of the kernel method in the missing case.

— **On the efficiency of the estimator** From theoretical point of view we have proved in the previous remark that the local local linear approach has a nice asymptotic properties compared to the kernel method. However, in practice the efficiency of this estimate is closely the selection of the parameters involved in the estimate such as the smoothing parameter, the function β and β and the kernel K. In addition to this parameters, the percentage of the missing value has also an important influence. Of course the efficiency of the estimate decrease with the the percentage of the missing value. We note that there are no theatrical result on the optimal choice of these parameters in this situation of missing data. These questions is a nice prospects of this work. linked

2.5 Appendix

In what follows, when no confusion is possible, we will denote by C and C' some strictly positive generic constants. Moreover, we put, for any $x \in \mathcal{F}$, and for all $i = 1, \ldots, n$

$$K_i(x) = K(h^{-1}d(x, X_i)) \text{ and } \beta_i(x) = \beta(X_i, x).$$

Proof of lemma 1. Similarly to Barrientos et al. (2010), we write

$$\begin{aligned}
\widehat{R}_D(x) &= \underbrace{\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j K_j(x)}{\phi_x(h)}\right)}_{S_2(x)}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i K_i(x)\beta_i^2(x)}{h^2\phi_x(h)}\right)}_{S_3(x)} \\
&- \underbrace{\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j K_j(x)\beta_j(x)}{h\phi_x(h)}\right)}_{S_4(x)}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i K_i(x)\beta_i(x)}{h\phi_x(h)}\right)}_{S_5(x)}
\end{aligned}$$

and

$$\widehat{R}_N(x) = \underbrace{\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j K_j(x) Y_j}{\phi_x(h)}\right)}_{T_2(x)} \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i K_i(x)\beta_i^2(x)}{h^2\phi_x(h)}\right)}_{T_3(x)} - \underbrace{\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j K_j(x)\beta_j(x) Y_j}{h\phi_x(h)}\right)}_{T_4(x)} \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i K_i(x)\beta_i(x)}{h\phi_x(h)}\right)}_{T_5(x)}.$$

It follows that

$$\widehat{R}_D(x) - \mathbb{E}[\widehat{R}_D(x)] = (S_2(x)S_3(x) - \mathbb{E}[S_2(x)S_3(x)]) - (S_4(x)S_5(x) - \mathbb{E}[S_4(x)S_5(x)])$$

$$\widehat{R}_N(x) - \mathbb{E}[\widehat{R}_N(x)] = (T_2(x)T_3(x) - \mathbb{E}[T_2(x)T_3(x)]) - (T_4(x)T_5(x) - \mathbb{E}[T_4(x)T_5(x)]).$$

So, by a simple algebra, we write for all $i \neq j$

$$\begin{aligned} S_i(x)S_j(x) - \mathbb{E}[S_i(x)S_j(x)] &= (S_i(x) - \mathbb{E}[S_i(x)])\,(S_j(x) - \mathbb{E}[S_j(x)]) \\ &\quad + (S_j(x) - \mathbb{E}[S_j(x)])\,\mathbb{E}[S_i(x)] \\ &\quad + (S_i(x) - \mathbb{E}[S_i(x)])\,\mathbb{E}[S_j(x)] \\ &\quad + \mathbb{E}[S_i(x)]\mathbb{E}[S_j(x)] - \mathbb{E}[S_i(x)S_j(x)] \end{aligned}$$

and

$$\begin{aligned} T_i(x)T_j(x) - \mathbb{E}[T_i(x)T_j(x)] &= (T_i(x) - \mathbb{E}[T_i(x)])\,(T_j(x) - \mathbb{E}[T_j(x)]) \\ &\quad + (T_j(x) - \mathbb{E}[T_j(x)])\,\mathbb{E}[T_i(x)] \\ &\quad + (T_i(x) - \mathbb{E}[T_i(x)])\,\mathbb{E}[T_j(x)] \\ &\quad + \mathbb{E}[T_i(x)]\mathbb{E}[T_j(x)] - \mathbb{E}[T_i(x)T_j(x)] \end{aligned}$$

So, all it remains to show is that

$$\sum_n \mathbb{P}\left\{|S_i(x) - \mathbb{E}[S_i(x)]| > \eta\sqrt{\frac{\log n}{n\,\phi_x(h)}}\right\} < \infty \quad \text{for} \quad i = 2,3,4,5,$$

$$\sum_n \mathbb{P}\left\{|T_j(x) - \mathbb{E}[T_j(x)]| > \eta\sqrt{\frac{\log n}{n\,\phi_x(h)}}\right\} < \infty \quad \text{for} \quad j = 2,3,4,5,$$

$$\mathbb{E}[S_i(x)] = O(1) \qquad \text{and} \qquad \mathbb{E}[T_i(x)] = O(1) \quad \text{for} \quad i = 2,3,4,5,$$

and

$$Cov(S_2(x), S_3(x)) = o\left(\sqrt{\frac{\log n}{n\,\phi_x(h)}}\right),\ Cov(S_4(x), S_5(x)) = o\left(\sqrt{\frac{\log n}{n\,\phi_x(h)}}\right),$$

$$Cov(T_2(x), T_3(x)) = o\left(\sqrt{\frac{\log n}{n\,\phi_x(h)}}\right) \text{ and } Cov(T_4(x), T_5(x)) = o\left(\sqrt{\frac{\log n}{n\,\phi_x(h)}}\right).$$

■

Proof of the first claimed result : In order to obtain both convergence rates we apply the unbounded Bernstein's exponential inequality (see Corollary A8 in Ferraty and Vieu (2006), p. 234). We precise that, the latter is based on the asymptotic evaluation of mth order moments of the following random variables

$$Z_i^{l,k} = \frac{1}{h^l \phi_x(h)} \left(\delta_i K_i(x) Y_i^k \beta_i^l(x) - \mathbb{E}\left[\delta_i K_i(x) Y_i^k \beta_i^l(x)\right]\right)$$

for $l = 0, 1, 2$, and $k = 0, 1$.
Notice that, by the Newton's binomial expansion, we obtain :

$$\begin{aligned}
&\mathbb{E}\left|\left(\delta_i K_i(x) Y_i^k \beta_i^l(x) - \mathbb{E}\left[\delta_i K_i(x) Y_i^k \beta_i^l(x)\right]\right)^m\right| \\
&= \mathbb{E}\left|\sum_{d=0}^{m} C_m^d \left(\delta_i K_i(x) Y_i^k \beta_i^l(x)\right)^d \left(\mathbb{E}\left[\delta_i K_i(x) Y_i^k \beta_i^l(x)\right]\right)^{m-d} (-1)^{m-d}\right| \\
&\leq \sum_{d=0}^{m} C_m^d \left(\mathbb{E}\left|\delta_i K_i(x) Y_i^k \beta_i^l(x)\right|^d\right) \left|\mathbb{E}\left[\delta_i K_i(x) Y_i^k \beta_i^l(x)\right]\right|^{m-d} \\
&\leq \sum_{d=0}^{m} C_m^d \mathbb{E}\left|\delta_1 K_1^d \beta_1^{dl}(x) Y_i^{dk}\right| \left|\mathbb{E}\left[\delta_1 K_1(x) \beta_1^l(x) Y_1^k\right]\right|^{m-d}
\end{aligned}$$

where $C_{k,m} = m!/(k!(m-k)!)$.

Since the variables δ and Y are independent given X, then under Assumption (H6) we have for all $d \leq m$

$$\mathbb{E}\left[\delta Y^{dk} | X\right] = (P(x) + o(1))\mathbb{E}\left[Y^{dk} | X\right] \leq C.$$

Now, from Barrientos et al. (2010)

$$h^{-lm} \phi_x^{-m}(h) \sum_{d=0}^{m} C_m^d \mathbb{E}\left|K_1(x) \beta_1^{dl}(x)\right| \left|\mathbb{E}\left[K_1(x) \beta_1^l(x)\right]\right|^{m-d} \leq C \phi_x(h)^{-m+1}.$$

Therefore, for $l = 0, 1, 2$, and $k = 0, 1$, we obtain

$$\mathbb{E}\left|Z_i^{l,k}\right|^m = O\left((\phi_x(h))^{-m+1}\right).$$

Consequently, it suffices to apply the Corollary A8 in Ferraty and Vieu (2006), for $a_n = (\phi_x(h))^{-1/2}$, to conclude

$$\sum_n \mathbb{P}\left\{|S_i(x) - \mathbb{E}[S_i(x)]| > \eta\sqrt{\frac{\log n}{n\,\phi_x(h)}}\right\} < \infty \quad \text{for} \quad i = 2, 3, 4, 5,$$
$$\sum_n \mathbb{P}\left\{|T_j(x) - \mathbb{E}[T_j(x)]| > \eta\sqrt{\frac{\log n}{n\,\phi_x(h)}}\right\} < \infty \quad \text{for} \quad j = 2, 3, 4, 5,$$

■

Proof of the second claimed result. The proof of these results are based on the fact that the observations (X_i, δ_i, Y_i) for $i = 1, \ldots, n$ are identically distributed. Therefore

$$\begin{cases} \mathbb{E}[S_2(x)] = \dfrac{\mathbb{E}[\delta_1 K_1(x)]}{\phi_x(h)}, \\ \\ \mathbb{E}[S_3(x)] = \mathbb{E}[T_3(x)] = \dfrac{\mathbb{E}[\delta_1 K_1(x)\beta_1^2(x)]}{h^2\phi_x(h)}, \\ \\ \mathbb{E}[S_4(x)] = \mathbb{E}[S_5(x)] = \mathbb{E}[T_5(x)] = \dfrac{\mathbb{E}[\delta_1 K_1(x)\beta_1(x)]}{h\phi_x(h)}, \\ \\ \mathbb{E}[T_2(x)] = \dfrac{\mathbb{E}[\delta_1 K_1(x) Y_1]}{\phi_x(h)} \quad \text{and} \quad \mathbb{E}[T_4(x)] = \dfrac{\mathbb{E}[\delta_1 K_1(x) Y_1 \beta_1(x)]}{h\phi_x(h)}. \end{cases} \tag{2.6}$$

Thus, everything is based on on the evaluation of the following quantities

$$\mathbb{E}\left[\delta_i K_i^j(x) Y_i^k \beta_i^l(x)\right] \text{ for } l = 0, 1, 2,\ j = 1, 2 \text{ and } k = 0, 1.$$

Once again we use the fact that the variables δ and Y are conditionally independent with respect to the functional variable X. Therefore, for all $l = 0, 1, 2$, and $k = 0, 1$, we have

$$\mathbb{E}\left[\delta_i K_i^j(x) Y_i^k \beta_i^l(x)\right] = O(\mathbb{E}\left[K_i(x)\beta_i^l(x)\right]) = O(h^l\phi_x(h)). \tag{2.7}$$

■

Proof of the last claimed result. Similarly to the previous case it suffices to evaluate

$$\begin{cases} \mathbb{E}[S_2(x)S_3(x)] = \dfrac{\mathbb{E}[\delta_1 K_1^2 \beta_1^2(x)(x)]}{nh^2\phi_x^2(h)}, \\ \\ \mathbb{E}[S_4(x)S_5(x)] = \dfrac{\mathbb{E}[\delta_1 K_1^2(x)\beta_1^2(x)]}{nh^2\phi_x^2(h)}, \\ \\ \mathbb{E}[T_2(x)T_3(x)] = \dfrac{\mathbb{E}[\delta_1 K_1^2(x)\beta_1^2(x)Y_1]}{nh^2\phi_x^2(h)}, \\ \\ \mathbb{E}[T_4(x)T_5(x)] = \dfrac{\mathbb{E}[\delta_1 K_1^2(x)\beta_1^2(x)Y_1]}{nh^2\phi_x^2(h)}. \end{cases}$$

Using the same arguments as for proving (2.7) we show that

$$\mathbb{E}\left[\delta_i K_i^2(x) Y_i^k \beta_i^l(x)\right] = O(h^l \phi_x(h)).$$

It follows that

$$\begin{cases} \mathbb{E}[S_2(x)S_3(x)] = O\left(\dfrac{1}{n\phi_x(h)}\right), \ \mathbb{E}[S_4(x)S_5(x)] = O\left(\dfrac{1}{n\phi_x(h)}\right), \\ \\ \mathbb{E}[T_2(x)T_3(x)] = O\left(\dfrac{1}{n\phi_x(h)}\right), \ \mathbb{E}[T_4(x)T_5(x)] = O\left(\dfrac{1}{n\phi_x(h)}\right). \end{cases}$$

Now, by combining this last to (2.6), we get

$$\begin{aligned} Cov(S_2(x), S_3(x)) = o\left(\sqrt{\frac{1}{n\,\phi_x(h)}}\right), \quad Cov(S_4(x), S_5(x)) = O\left(\sqrt{\frac{1}{n\,\phi_x(h)}}\right), \\ Cov(T_2(x), T_3(x)) = O\left(\sqrt{\frac{1}{n\,\phi_x(h)}}\right) \quad Cov(T_4(x), T_5(x)) = O\left(\sqrt{\frac{1}{n\,\phi_x(h)}}\right). \end{aligned} \tag{2.8}$$

which completes the proof of this lemma.

■

Proof of Corollary 1. By some simple

$$\mathbb{E}\left[W_{12}(x)\right] \geq \mathbb{E}\left[\beta_1^2(x)K_1(x)\right]\mathbb{E}[\beta_2(x)K_2(x)] \geq Ch^2\phi_x(h)$$

So, we get

$$\lim_{n\to\infty}\mathbb{E}\left[\widehat{R}_D(x)\right] > C.$$

Therefrom we obtain the following implications :

$$\widehat{R}_D(x) \leq \frac{C}{2} \implies \left|\mathbb{E}[\widehat{R}_D(x)] - \widehat{R}_D(x)\right| \geq \frac{C}{2}$$

Then, our result is a direct consequence of the previous Lemma 1.

Proof of Lemma 2. We write

$$\widehat{B}(x) = \frac{\mathbb{E}\left[\widehat{R}_N(x)\right] - R(x)\mathbb{E}\left[\widehat{R}_D(x)\right]}{\mathbb{E}\left[\widehat{R}_D(x)\right]} = \frac{1}{\mathbb{E}[W_{12}]}\mathbb{E}\left[W_{12}(x)(Y_2 - R(x))\right].$$

Since the pairs (X_i, δ_i, Y_i) are identically distributed, then

$$\begin{aligned}\mathbb{E}\left[W_{12}(x)(Y_2 - R(x))\right] &= \mathbb{E}\left[\delta_1\beta_1^2(x)K_1(x)\beta_2(x)K_2(x)P(X_2)(R(X_2) - R(x))\right. \\ &\quad \left. -(\delta_1\beta_1(x)K_1(x))(\beta_2(x)K_2(x)P(X_2)(R(X_2) - R(x))\right]\end{aligned}$$

and

$$\mathbb{E}\left[W_{12}(x)\right] = \mathbb{E}\left[\delta_1\beta_1^2(x)K_1(x)\beta_2(x)K_2(x)P(X_2) - (\delta_1\beta_1(x)K_1(x))(\beta_2(x)K_2(x)P(X_2)\right]$$

Then, As K has a compact support $(0, 1)$, then the Lipshtizian assumption (H2) allows us to write that

$$\mathbb{1}_{B(x,h)}(X_2)\left|R(X_2) - R(x)\right| \leq Cd^{\kappa}(X_2, x) \leq Ch^{\kappa}$$

Therefore

$$\widehat{B}(x) = O(h^{\kappa}).$$

■

Acknowledgment : The authors would like to thank the Editor, an Associate-Editor and an anonymous reviewer for their valuable comments and suggestions which improved substantially the quality of an earlier version of this paper.

Bibliographie

[1] Baìllo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response, *J. of Multivariate Analysis*, 100, 102-111.

[2] Barrientos-Marin, J., Ferraty, F. and Vieu, P. (2010). Locally Modelled Regression and Functional Data. *J. of Nonparametric Statistics*, 22, No. 5, 617-632.

[3] Boente, G., Gonzalez-Manteiga, W. and Perez-Gonzalez, A. (2009). Robust nonparametric estimation with missing data. *J. Statist. Plann. Inference,* 139, 571-592.

[4] Bosq, D. (2000). Linear Processes in Function Spaces : Theory and applications, *Lecture Notes in Statistics,* 149, Springer.

[5] Chen, J.H., Shao, J., (2000). Nearest neighbor imputation for survey data. *J. Official Statist.,* 16, 113-131.

[6] Cheng, P.E., (1994). Nonparametric estimation of mean functionals with data missing at random. *J. Amer. Statist. Assoc.* 89, 8-87.

[7] Chouaf, A. and Laksaci, A., (2013). On the functional local linear estimate for spatial regression. *Stat. Risk Model,* 29, 189-214.

[8] Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference*, 147, 1-23.

[9] Demongeot, J., Laksaci, A., Rachdi, M. and Rahmani, S. (2014). On the local linear modelization of the conditional distribution for functional data. *Sankhya A,* 76, 328-355.

[10] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications.* London, Chapman & Hall.

[11] Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference*, 140, 335-352.

[12] Ferraty, F. and Vieu, P. (2006). Nonparametric functional data analysis. Theory and Practice. *Springer Series in Statistics.* New York.

[13] Ferraty, F., Sued, M. and Vieu, P. (2013). Mean estimation with data missing at random for functional covariables. *Statistics,* 47, 688-706.

[14] Hsing, T. and Eubank, R. (2015). Theoretical foundations of functional data analysis, with an introduction to linear operators. *Wiley Series in Probability and Statistics.* John Wiley & Sons, Chichester.

[15] Ling, N., Liu, Y. and Vieu, P.(2015). Nonparametric regression estimation for functional stationary ergodic data with missing at random. *J. Statist. Plann. Inference,* 162, 75-87.

[16] Ling, N., Liu, Y. and Vieu, P. (2016). Conditional mode estimation for functional stationary ergodic data with responses missing at random. *Statistics,* 50, 991-1013.

[17] Perez-Gonzalez, A., Vilar-Fernandez and J. M. Gonzalez-Manteiga, W. (2009). Asymptotic properties of local polynomial regression with missing data and correlated errors. *Ann. Inst. Statist. Math.,* 61, 85-109

[18] Ramsay, J. O. and Silverman, B. W. (2002). Applied functional data analysis. Methods and case studies. *Springer Series in Statistics.* New York.

[19] Yates, F.,(1933). The analysis of replicated experiments when the field results are incomplete. *Emporium J. Exp. Agriculture,* 1, 129-142.

[20] Zhang, J. (2014). Analysis of variance for functional data. *Monographs on Statistics and Applied Probability*, 127. CRC Press, Boca Raton, FL.

Chapitre 3

Local linear estimate for functional regression with missing data at random : dependant case

3.1 Introduction

Let (X_i, Y_i) for $i = 1, ..., n$ be n pairs of random variables that we assume drawn from the pair (X, Y) which is valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a semi-metric space equipped with a semi-metric d. Assume that there exists a regular version of the conditional probability of Y given $X = x$ for a fixed $x \in \mathcal{F}$, which is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ and has bounded density, denoted by f^x. In this paper, we consider the problem of the conditional density estimation by using locally modeling approach when the explanatory variable X is functional and when the observations $(Y_i, X_i)_{i\in\mathbb{N}}$ are strongly mixing. In functional statistics, there are several ways for extending the local linear ideas.
Here, we adopt the fast functional locally modeling, introduced by [2] for the regression analysis, that is, we estimate the conditional density by ba which is obtained by minimizing the following quantity :

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} (Y_i - a - b\beta(X_i, x))^2 \, K(h^{-1}\varrho(x, X_i)) \tag{3.1}$$

where 1 $\beta(.,.)$ is a known function from $\mathcal{F}^2$ into $\mathbb{R}$ such that,$\forall\xi \in \mathcal{F}$, $\beta(\xi, \xi) = 0$, with K is a kernel and $h = h_{K,n}$ is a sequence of positive real numbers, is a sequence of positive real numbers and $\varrho(.,.)$ is a function of $\mathcal{F} \times \mathcal{F}$ such that $d(.,.) = |\varrho(.,.)|$, we get explicitly the following definition of $\widehat{R}(x)$:

$$\widehat{R}(x) = \frac{\sum_{i,j=1}^{n} W_{ij}(x)Y_j}{\sum_{i,j=1}^{n} W_{ij}(x)} \tag{3.2}$$

where

$$W_{ij}(x) = \beta(X_i, x)\left(\beta(X_i, x) - \beta(X_j, x)\right)\delta_i\delta_j K(h^{-1}\varrho(x, X_i))K(h^{-1}\varrho(x, X_j))$$

with the convention $0/0 = 0$.

the are several prediction tools in nonparametric statistic, such the conditional mode, the conditional median or the conditional quantiles, are based on the preliminary estimate of this functional parameter. Since, the contribution of [10] this subject has been widely studied in functional statistics (see, [8, 10, 9, 12] for some results on the conditional density estimation and its mode in the functional framework. All these works give some asymptotic results on the Nadaraya-Watson type estimator of this model. As direct extension of this method we focus, here, the estimation by local polynomial method. Such estimation procedure has recently investigated for the functional data. For example, we quote the realizations of [1, 2, 2, 5, 21, 22] which are concerned with the local linear type regression operator estimation for independent and identically distributed functional data.

First results on the sus-mentioned topic, in nonparametric statistics, are obtained by Ferraty et al. (2013). They constructed an estimate of the regression operator when the functional regressor is completely observed and MAR scalar response variable. They stated the asymptotic properties of this estimate in the i.i.d. case. Ling et al. (2015) have established the convergence in probability and the asymptotic normality of the same estimate but in the ergodic case. We return to Ling et al (2016) for the conditional mode estimation in the case of functional ergodic MAR data.

3.2 Main result

We begin by recalling the definition of the strong mixing property. F or this we introduce the following notations. Let $\mathcal{F}_i^k(Z)$ denote the $\sigma-$algebra generated by $\{Z_j, i \leq j \leq k\}$.

Definition 1 *Let $\{Z_i, i = 1, 2, ...\}$ be a strictly stationary sequence of random variables. Given a positive integer n, set*

$$\alpha(n) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{F}_1^k(Z) \quad and \quad B \in \mathcal{F}_{k+n}^{\infty}(Z), k \in \mathbb{N}^*\}.$$

The sequence is said to be $\alpha-$mixing (strong mixing) if the mixing coefficient $\alpha(n) \to 0$ as $n \to \infty$.

There exist many processes fulfilling the strong mixing property. We quote, here, the usual ARMA processes (with innovations satisfying some existing moment conditions) are geometrically strongly mixing, i.e., there exist $\rho \in (0,1)$ and $C > 0$ such that, for any $n \geq 1$, $\alpha(n) \leq C\rho^n$ (see, e.g., Jones (1978)). The threshold models, the EXPAR models (see, Ozaki (1979)), the simple ARCH models (see Engle (1982)), theirr GARCH extension (see Bollerslev (1986)) and bilinear Markovian models are geometrically strongly mixing under some general ergodicity conditions. For more details we refer the reader to the mmonographhs of Bradley (2007) or Dedecker et al (2007).

Throughout the paer, x denotes a fixed point in $\mathcal{F}$, N_x denotes a fixed neighborhood of x, $f^{x(j)}$ denotes the j^{th} order derivative of the conditional density f^x. S will be a fixed compact subset of $\mathbb{R}$, $B(x,r) = \{x' \in \mathcal{F}/|\varrho(x',x)| \leq r\}$ and $\phi_x(r_1,r_2) = \mathbb{P}(r_2 \leq \varrho(X,x) \leq r_1)$. Notice that, our nonparametric model will be quite general in the sense that we will just need the following assumptions :

(H1) For any $r > 0$, $\phi_x(r) := \phi_x(-r,r) > 0$.

(H2) There exists a constant $C > 0$ such that, for all real $m \geq 1$, $\mathbb{E}(|Y|^m|X = x) < C$ and for all $(x_1,x_2) \in N_x \times N_x$:

$$|R(x_1) - R(x_2)| \leq C_x\, d^\kappa(x_1,x_2), \quad \text{for some} \kappa > 0$$

where C_x is a positive constant depending on x.

(H3) The function $\beta(.,.)$ is such that :

$$\text{for all } x' \in\ \mathcal{F},\ C_1\,|\varrho(x,x')| \leq |\beta(x,x')| \leq C_2\,|\varrho(x,x')|,$$

where $C_1 > 0$, $C_2 > 0$.

(H4) The sequence $(X_i,Y_i)_{i\in\mathbb{N}}$ satisfies : $\exists a > 0, \exists c > 0 : \forall n \in \mathbb{N}$ $\alpha(n) \leq cn^{-a}$ and

$$\max_{i\neq j} \mathbb{P}((X_i,Y_j) \in B(x,h) \times B(x,h)) = \varphi_x(h) > 0$$

(H5) The conditional density of (Y_i,Y_j) given (X_i,X_j) exists and is bounded.

(H6) K is a positive and differentiable function which is supported within $[-1,1]$.

(H7) The function $P(\cdot)$ is continuous on N_x and such that $P(x) > 0$.

(H8) The bandwidth h satisfies :

$$\lim_{n\to\infty} \frac{\log n}{n\phi_x(h)} = 0.$$

and

$$\forall n > n_0, \ -\frac{1}{\phi_x(h)} \int_{-1}^{1} \phi_x(zh,h) \frac{d}{dz}\left(z^2 K(z)\right) dz > C_3 > 0$$

and

$$h \int_{B(x,h)} \beta(u,x) dP(u) = o\left(\int_{B(x,h)} \beta^2(u,x)\, dP(u) \right)$$

where $B(x,r) = \{x' \in \mathcal{F} / |\delta(x',x)| \le r\}$

(H9) $\lim_{n\to\infty} h_H = 0$ and $\exists \beta_1 > 0$ such that $\lim_{n\to\infty} n^{\beta_1} h_H = \infty$,

(H10)

$$\begin{cases} \lim_{n\to\infty} h = 0, \lim_{n\to\infty} \frac{\chi_x^{(1/2)}(h) \log n}{n h_H \phi_x^2(h)} = 0 \\ \text{and} \quad \exists \eta_0 > \frac{3\beta_1+1}{a+1}, \quad C n^{\frac{(3-a)}{(a+1)}+\eta_0} \le h_H \chi_x^{(1/2)}(h), \\ \text{where} \quad \chi_x(h) = \max(\phi_x^2(h), \varphi_x(h)). \end{cases}$$

Theorem 2 *Under assumptions (H1)-(H10), we have*

$$|\widehat{R}(x) - R(x)| = O\left(h^{\kappa}\right) + O\left(\sqrt{\frac{\chi_x^{(1/2)}(h) \log n}{n\, \phi_x^2(h)}} \right) \quad a.co.$$

Proof of Theorem 2. We consider the following decomposition :

$$\widehat{R}(x) - R(x) = \widehat{B}(x) + \frac{\widehat{M}(x)}{\widehat{R}_D(x)} + \frac{\widehat{Q}(x)}{\widehat{R}_D(x)}$$

where

$$\begin{aligned} \widehat{Q}(x) &:= \left(\widehat{R}_N(x) - \mathbb{E}[\widehat{R}_N(x)]\right) - R(x)\left(\widehat{R}_D(x) - \mathbb{E}[\widehat{R}_D(x)]\right) \\ \widehat{B}(x) &:= \frac{\mathbb{E}[\widehat{R}_N(x)]}{\mathbb{E}[\widehat{R}_D(x)]} - R(x) \quad \text{and} \quad \widehat{M}(x) := -\widehat{B}(x)\left(\widehat{R}_D(x) - \mathbb{E}[\widehat{R}_D(x)]\right) \end{aligned}$$

with

$$\widehat{R}_N(x) = \frac{1}{n\, h^2 \phi_x(h)} \sum_{i,j}^{n} W_{ij}(x) Y_j \text{ and } \widehat{R}_D(x) = \frac{1}{n h^2 \phi_x(h)} \sum_{i,j}^{n} W_{ij}(x).$$

Thus, the proof of Theorem 2 is based on the following intermediate results, for which the proofs are given in the Appendix.

Lemma 3 *Under the hypotheses (H1),(H3),(H4),(H6),(H8) and (H10), we have that :*

$$\left|\widehat{R}_D(x) - \mathbb{E}[\widehat{R}_D(x)]\right| = O_{a.co.}\left(\sqrt{\frac{\chi_x^{(1/2)}(h)\log n}{n\,\phi_x^2(h)}}\right).$$

and

$$|\widehat{R}_N(x) - \mathbb{E}[\widehat{R}_N(x)]| = O_{a.co.}\left(\sqrt{\frac{\chi_x^{(1/2)}(h)\log n}{n\,\phi_x^2(h)}}\right).$$

Corollary 2 *Under the hypotheses of Lemma* **??**, *there exists a positive real C such that*

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\widehat{R}_D(x) < C\right) < \infty.$$

Lemma 4 *Under assumptions (H1), (H2) and (H6) we have*

$$\left|\widehat{B}(x)\right| = O(h^{\kappa}).$$

■

3.3 Appendix

In what follows, when no confusion is possible, we will denote by C and C' some strictly positive generic constants. Moreover, we put, for any $x \in \mathcal{F}$, and for all $i = 1, \ldots, n$:

$$K_i(x) = K(h^{-1}d(x, X_i)),\ \beta_i(x) = \beta(X_i, x).$$

Proof of lemma 3

$$\widehat{R}_D(x) = \underbrace{\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j K_j(x)}{\phi_x(h)}\right)}_{S_1(x)}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i K_i(x)\beta_i^2(x)}{h^2\phi_x(h)}\right)}_{S_2(x)}$$

$$- \underbrace{\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j K_j(x)\beta_j(x)}{h\phi_x(h)}\right)^2}_{S_3(x)}.$$

and

$$\begin{aligned}
\widehat{R}_N(x) &= \underbrace{\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j K_j(x)Y_j}{\phi_x(h)}\right)}_{T_2(x)}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i K_i(x)\beta_i^2(x)}{h^2\phi_x(h)}\right)}_{T_3(x)} \\
&- \underbrace{\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j K_j(x)\beta_j(x)Y_j}{h\phi_x(h)}\right)}_{T_4(x)}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i K_i(x)\beta_i(x)}{h\phi_x(h)}\right)}_{T_5(x)}.
\end{aligned}$$

Thus, we ca write

$$\widehat{R}_D(x) - \mathbb{E}[\widehat{R}_D(x)] = \big(S_1(x)S_2(x) - \mathbb{E}[S_1(x)S_2(x)]\big) - \big(S_3^2(x) - \mathbb{E}[S_3^2(x)]\big).$$

$$\widehat{R}_N(x) - \mathbb{E}[\widehat{R}_N(x)] = \left(T_2(x)T_3(x) - \mathbb{E}[T_2(x)T_3(x)]\right) - \left(T_4(x)T_5(x) - \mathbb{E}[T_4(x)T_5(x)]\right).$$

we have

$$\begin{aligned}
S_1(x)S_2(x) - \mathbb{E}[S_1(x)S_2(x)] &= \left(S_2(x) - \mathbb{E}[S_2(x)]\right)\left(S_1(x) - \mathbb{E}[S_1(x)]\right) \\
&\quad + \left(S_2(x) - \mathbb{E}[S_2(x)]\right)\mathbb{E}[S_1(x)] \\
&\quad + \left(S_1(x) - \mathbb{E}[S_1(x)]\right)\mathbb{E}[S_2(x)] \\
&\quad + \mathbb{E}[S_1(x)]\mathbb{E}[S_2(x)] - \mathbb{E}[S_1(x)S_2(x)]
\end{aligned}$$

and

$$\begin{aligned}
S_3^2(x) - \mathbb{E}[S_3^2(x)] &= \left(S_3(x) - \mathbb{E}[S_3(x)]\right)^2 + 2\left(S_3(x) - \mathbb{E}[S_3(x)]\right)\mathbb{E}[S_3(x)] \\
&\quad + \mathbb{E}^2[S_3(x)] - \mathbb{E}[S_3^2(x)].
\end{aligned}$$

and

$$\begin{aligned}
T_2(x)T_3(x) - \mathbb{E}[T_2(x)T_3(x)] &= \left(T_2(x) - \mathbb{E}[T_2(x)]\right)\left(T_3(x) - \mathbb{E}[T_3(x)]\right) \\
&\quad + \left(T_3(x) - \mathbb{E}[T_3(x)]\right)\mathbb{E}[T_2(x)] \\
&\quad + \left(T_2(x) - \mathbb{E}[T_2(x)]\right)\mathbb{E}[T_3(x)] \\
&\quad + \mathbb{E}[T_2(x)]\mathbb{E}[T_3(x)] - \mathbb{E}[T_2(x)T_3(x)]
\end{aligned}$$

$$\begin{aligned} T_4(x)T_5(x) - \mathbb{E}[T_4(x)T_5(x)] &= (T_4(x) - \mathbb{E}[T_4(x)])\,(T_5(x) - \mathbb{E}[T_5(x)]) \\ &\quad + (T_5(x) - \mathbb{E}[T_5(x)])\,\mathbb{E}[T_4(x)] \\ &\quad + (T_4(x) - \mathbb{E}[T_4(x)])\,\mathbb{E}[T_5(x)] \\ &\quad + \mathbb{E}[T_4(x)]\mathbb{E}[T_5(x)] - \mathbb{E}[T_4(x)T_5(x)] \end{aligned}$$

So, the claimed result can be derived from the follpwing assertion

$$\sum_n \mathbb{P}\left\{ |S_l(x) - \mathbb{E}[S_l(x)]| > \eta\sqrt{\frac{\chi_x^{(1/2)}(h)\log n}{n\,\phi_x^2(h)}} \right\} < \infty, \text{ for } l = 2,3,4, \tag{3.3}$$

$$\sum_n \mathbb{P}\left\{ |T_l(x) - \mathbb{E}[T_l(x)]| > \eta\sqrt{\frac{\chi_x^{(1/2)}(h)\log n}{n\,\phi_x^2(h)}} \right\} < \infty, \text{ for } l = 2,3,4, \tag{3.4}$$

$$S_1 = O(1) \quad \text{and} \quad \mathbb{E}[S_l] = O(1) \quad \text{for} \quad l = 2,3,4, \tag{3.5}$$

$$T_1 = O(1) \quad \text{and} \quad \mathbb{E}[T_l] = O(1) \quad \text{for} \quad l = 2,3,4, \tag{3.6}$$

$$Cov(S_1(x), S_2(x)) = o\left(\sqrt{\frac{\chi_x^{(1/2)}(h)\log n}{n\,\phi_x^2(h)}}\right) \tag{3.7}$$

$$\text{and} \quad Var[S_3(x)] = o\left(\sqrt{\frac{\chi_x^{(1/2)}(h)\log n}{n\,\phi_x^2(h)}}\right) \tag{3.8}$$

$$Cov(T_2(x), T_3(x)) = o\left(\sqrt{\frac{\chi_x^{(1/2)}(h)\log n}{n\,\phi_x^2(h)}}\right) \tag{3.9}$$

and

$$Cov(T_4(x), T_5(x)) = o\left(\sqrt{\frac{\chi_x^{(1/2)}(h)\log n}{n\,\phi_x^2(h)}}\right). \tag{3.10}$$

These lasts results are not affected by the dependence structure of the data. It suffices to prove (3.3), (3.7) and (3.8) to fonosh the proof of the claimed result. For (3.3), we set

$$\Delta_i^k = \frac{1}{h}(\delta_i K_i(x)\beta_i^k(x) - \mathbb{E}[\delta_i K_i(x)\beta_i^k(x)]) \quad \text{for} \quad k = 0,1,2.$$

then, it can be see that

$$S_{k+2}(x)-\mathbb{E}[S_{k+2}(x)]=\frac{1}{n\phi_x(h)}\sum_{i=1}^{n}\Delta_i^k \quad \text{for} \quad k=0,1,2.$$

Moreover, under (H1),(H3) and (H5), we have

$$\begin{array}{ll}\frac{1}{h^k}\delta_i K_i(x)\beta_i^k(x) & \leq \frac{1}{h^k}\delta_i K_i(x)|\varrho(X_i,x)|^k \\ & \leq \frac{1}{h^k}\delta_i K(h^{-1}\varrho(x,X_i))|\varrho(x,X_i)|^k \mathbb{I}_{]-1,1[}(h^{-1}\varrho(x,X_i)) \\ & \leq \delta_i K(h^{-1}\varrho(x,X_i)) \\ & \leq (P(x)+o(1))K(h^{-1}\varrho(x,X_i)) \leq C\end{array} \tag{3.11}$$

So, we can apply the Fuck-Nagaev exponential inequality, to get for all $r>0$ and $\varepsilon>0$, we have

$$\begin{array}{ll}\mathbb{P}\{|S_{k+2}(x)-\mathbb{E}[S_{k+2}(x)]|>\varepsilon\} & =\mathbb{P}\Big\{\Big|\frac{1}{n\phi_x(h)}\sum_{i=1}^{n}\Delta_i^k\Big|>\varepsilon\Big\} \\ & \leq \mathbb{P}\Big\{\Big|\sum_{i=1}^{n}\Delta_i^k\Big|>\varepsilon n\phi_x(h)\Big\} \\ & \leq C(A_1+A_2)\end{array} \tag{3.12}$$

where

$$A_1=(1+\frac{\varepsilon^2 n^2(\phi_x(h))^2}{S_n^2 r})^{-r/2} \qquad \text{and} \qquad A_2=nr^{-1}\Big(\frac{r}{\varepsilon n\phi_x(h)}\Big)^{a+1}$$

and

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

with

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

Next, we evaluate the asymptotic behavior of S_n^{2*}

$$S_1=\{(i,j) \quad \text{such that} \quad 1\leq i-j\leq m_n\}$$

and

$$S_2=\{(i,j) \quad \text{such that} \quad m_n+1\leq i-j\leq n-1\}$$

where $m_n \to \infty$, as $n \to \infty$. Let $J_{1,n}$ and $J_{2,n}$ be the sum of covariance over S_1 and S_2 respectively.Because of (H1),(H4) and (3.11) we can write :

$$\begin{array}{ll} J_{1,n} & \leq Cnm_n[\mathbb{P}(X_i, X_j) \in B(x,h) \times B(x,h) \\ & \quad +\mathbb{P}(X_i \in B(x,h))\mathbb{P}(X_j \in B(x,h))] \\ & \leq Cnm_n[\varphi_x(h) + \phi_x^2(h)] \leq Cnm_n\chi_x(h). \end{array}$$

Concerning the summation over S_2, we use Davydov-Rio's inequality for bounded mixing processus. This leads, for all $i \neq j$, to

$$|Cov(\Delta_i^k, \Delta_j^k| \leq C\alpha(|i-j|).$$

Therefore, using $\sum_{j \geq x+1} j^{-a} \leq \int_{u \geq x} u^{-a} = [(a-1)x^{a-1}]^{-1}$ we get, under the first part of (H4)

$$|J_{2,n}| = \Big| \sum_{(i,j)\in E_2} Cov[\Delta_i^k, \Delta_j^k] \Big| \leq C\frac{nm_n^{-a+1}}{a-1}$$

The choice $m_n = (\chi_x(h))^{-1/a}$, permits to get

$$\sum_{i \neq j} Cov(\Delta_i^k, \Delta_j^k) = O(n\chi_x^{(a-1)/a}(h)) \quad \text{ for } \quad k = 0, 1, 2,$$

Concerning the variance term, we deduce from (H1) that

$$Var[\Delta_1(x)] \leq C(\phi_x(h) + (\phi_x(h))^2) \leq C\chi_x^{1/2}(h).$$

Finally, as $a > 1$

$$S_n^2 = O(n\chi_x^{1/2}(h)). \tag{3.13}$$

Now, we apply (3.12) to $\varepsilon = \lambda\frac{\sqrt{S_n^2 \log n}}{n\phi_x(h)}$ and $r = C(\log n)^2$. It follows that

$$\begin{array}{ll} A_2 & \leq Cn^{1-(a+1)/2}\chi_x(h)^{-(a+1)/4}(\log n)^{(3a-1)/2} \\ & \leq Cn^{1-\eta(a+1)/2}(\log n)^{(3a-1)/2} \\ & \leq Cn^{-1-\nu}. \end{array}$$

and

$$A_1 \leq C\Big(1 + \frac{\lambda^2 \log n}{r}\Big)^{-r/2} = C\exp\Big(-r/2\log(1+\frac{\lambda^2\log n}{r})\Big)$$

because of $r = C(\log n)^2$, we get

$$A_1 \leq C\exp\Big(-\lambda^2\frac{\log n}{2}\Big) = Cn^{-\lambda^2/2}.$$

Thus, for λ large enough :

$$\exists \nu\prime > 0, \qquad A_1 \leq Cn^{-\lambda^2/2} \leq Cn^{-1-\nu\prime}.$$

Hence

$$S_l(x) - \mathbb{E}[S_l(x)] = O_{a.co}\left(\sqrt{\frac{\chi_x^{(1/2)}(h)\log n}{n\,\phi_x^2(h)}}\right) \qquad \text{for} \qquad l = 2, 3, 4,$$

$$T_l(x) - \mathbb{E}[T_l(x)] = O_{a.co}\left(\sqrt{\frac{\chi_x^{(1/2)}(h)\log n}{n\,\phi_x^2(h)}}\right) \qquad \text{for} \qquad l = 2, 3, 4,$$

Finally, by following similar argumens used to prove (3.13) we get

$$Cov(S_1(x), S_2(x)) = O\left(\sqrt{\frac{\chi_x^{(1/2)}(h)\log n}{n\,\phi_x^2(h)}}\right)$$

$$\text{and} \qquad Var[S_3(x)] = o\left(\sqrt{\frac{\chi_x^{(1/2)}(h)\log n}{n\,\phi_x^2(h)}}\right)$$

$$Cov(T_2(x), T_3(x)) = O\left(\sqrt{\frac{\chi_x^{(1/2)}(h)\log n}{n\,\phi_x^2(h)}}\right)$$

$$\text{and} \qquad Cov(T_4(x), T_5(x)) = o\left(\sqrt{\frac{\chi_x^{(1/2)}(h)\log n}{n\,\phi_x^2(h)}}\right)$$

which is negligible with respect $\sqrt{\frac{\chi_x^{(1/2)}(h)\log n}{n\,\phi_x^2(h)}}$. Then, the proof of our Lemma is now finished.

$\blacksquare$

Bibliographie

[1] Baillo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response, *Journal of Multivariate Analysis,* 100, 102-111.

[2] Barrientos-Marin, J.(2007). Some Practical Problems of Recent Nonparametric Procedures : Testing, Estimation, and Application. PhD thesis (in French) from the Paul Sabatier's University (Toulouse).

[3] Barrientos-Marin, J., Ferraty, F. and Vieu, P. (2010). Locally Modelled Regression and Functional Data. *J. of Nonparametric Statistics*, 22, No. 5, 617-632.

[4] Benhenni, K., Ferraty, F., Rachdi, M. and Vieu, P. (2007). Local smoothing regression with functional data. *Computational Statistics,* 22, No. 3, 353-369.

[5] Benhenni, K., Griche-Hedli, S., Rachdi, M. (2010). Estimation of the regression operator from functional fixed-design with correlated errors. *Journal of Multivariate Analysis,* 101, 476-490.

[6] Bosq, D. (2000). *Linear Processes in Function Spaces : Theory and applications.* Lecture Notes in Statistics, 149, Springer.

[7] Chu, C.-K. and Marron, J.-S. (1991). Choosing a kernel regression estimator. With comments and a rejoinder by the authors. *Statist. Sci.*, 6 , 404-436.

[8] Dabo-Niang, S. and Laksaci, A. (2007). Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle. *Pub. Inst. Stat. Univ. Paris,* 3, 27-42.

[9] Demongeot, J., Laksaci, A., Madani, F. and Rachdi, M. (2010). Local linear estimation of the conditional density for functional data. *C. R., Math., Acad. Sci. Paris,* 348, 931-934.

[10] Demongeot, J., Laksaci, A., Madani, F. and Rachdi, M. (2011). Functional data : Local linear estimation of the conditional density and its application *Statistics, Volume 00, Pages 00-00, DOI : 10.1080/02331888.2011.568117.*

[11] El Methni, M. and Rachdi, M. (2010). Local weighted average estimation of the regression operator for functional data. *Commun. Stat., Theory and Methods, to appear.*

[12] Ezzahrioui, M. and Ould-Saïd, E. (2008). Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *J. Nonparametr. Stat.,* 20, 3-18.

[13] Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.,* 87, 998-1004.

[14] Fan, J. and Yim, T.-H. (2004). A cross-validation method for estimating conditional densities. *Biometrika,* 91, 819-834.

[15] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications.* London, Chapman and Hall.

[16] Ferraty, F., Goia, A. and Vieu, P. (2002). Functional nonparametric model for time series : a fractal approach to dimension reduction. TEST, 11, No. 2, 317-344.

[17] 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,* 140, 335-352.

[18] Ferraty, F., Laksaci, A. and Vieu, P. (2005). Functional times series prediction via conditional mode. C. R., *Math., Acad. Sci. Paris,* 340, 389-392.

[19] Ferraty, F., Laksaci, A. and Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat. Inference Stoch.Process.,* 9, 47-76.

[20] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis. Theory and Practice.* Springer Series in Statistics. New York.

[21] Hyndman, R. and Yao, Q. (2002). Nonparametric estimation and symmetry tests for conditional density functions. *J. Nonparametr. Stat.,* 14, 259-278.

[22] Laksaci, A. (2007). Convergence en moyenne quadratique de l'estimateur à noyau de la densité conditionnelle avec variable explicative fonctionnelle. *Pub. Inst. Stat. Univ.Paris,* 3, 69-80.

[23] Laksaci, A., Madani, F. and Rachdi, M. (2012). Kernel conditional density estimation when the regressor is valued in a semi-metric space *Communications in Statistics-Theory and Methods.* (to appear).

[24] Mailler, H.-G. and Stadtmailler, U. (2005). Generalized functional linear models. *Ann.Stat.,* 33, No. 2, 774-805.

[25] Ouassou, I. and Rachdi, M. (2010). Stein type estimation of the regression operator for functional data. Advances and Applications in Statistical Sciences, 1, No 2, 233-250.

[26] Ould-Said, E. and Cai Z. (2005). Strong uniform consistency of nonparametric estimation of the censored conditional mode function. *J. Nonparametr. Stat.,* 17, 797-806.

[27] Rachdi, M. and Sabre, R. (2000). Consistent estimates of the mode of the probability density function in nonparametric deconvolution problems. *Statist. Probab. Lett.,* 47, 105-114.

[28] Rachdi, M. and Vieu, P. (2007). Nonparametric regression for functional data : automatic smoothing parameter selection. *Journal of Statistical Planning and Inference,* 137, 2784-2801.

[29] Ramsay, J.-O. and Silverman, B.-W. (1997). *Functional data analysis.* Springer Series in Statistics. New York.

[30] Ramsay, J.-O. and Silverman, B.-W. (2002). *Applied functional data analysis. Methods and case studies.* Springer Series in Statistics. New York.

[31] Rio, E. (1990). *Exponential inequalities ...* Springer-Verlag, New York.

[32] Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants.* Springer, ESAIM, Collection Mathématiques et Applications

[33] Sarda, P. and Vieu, P. (2000). *Kernel Regression.* Pages 43-70, Wiley, New York.

[34] Vieu, P. (1996). A note on density mode estimation. *Statist. Probab. Lett.,* 26, Pages 297-307.

[35] Youndjé, E. (1993). Estimation non paramétrique de la densité conditionnelle par la méthode du noyau. PhD thesis (in French) from the Rouen's University.

Chapitre 4

FDA : KNN local linear regression with missing data at random

[a] Laboratoire Statistique et Processus Stochastiques
Univ. Djillali Liabès, BP 89, S. B. A. 22000, Algeria.
e-mails : benchiha.abbassia@yahoo.fr, kaidzoulikha312@gmail.com, alilak@yahoo.fr
[b] Univ. Grenoble-Alpes, Laboratoire AGEIS EA 7407, AGIM Team,
UFR SHS, BSHM, BP. 47, 38040 Grenoble Cedex 09, France.
e-mail : mustapha.rachdi@univ-grenoble-alpes.fr

Abstract In this paper, we analyse the co-variability between two random variables when the response one is not totally observed. We use the kernel nearest-neighbor method to construct an estimator of the local linear regression when missing data occur in the response variable. Asymptotic results, in term of the pointwise and the uniform almost complete consistencies, are established for the proposed estimator.

Keywords Almost complete convergence · Local linear method · Missing data · Uniform consistency · Regression operator.

AMS Subject classification Primary, 62G20 ; secondary 62G08 · 62G35 · 62E20.

4.1 Introduction

Pushed by its interaction with a various applied fields, the functional data analysis (FDA) became an important topic of research in the last decade. For a background reference on this topic, we cite, for instance, the monographs of Ramsay and Silverman (2005), Bosq (2000), Ferraty and Vieu (2006), Cuevas (2014), Zhang (2014) and Hsing and Eubank (2015). In this work, we are interested to the estimation of local linear regression by kernel nearest-neighbor method. We treat the case, where the regressor is of functional kind and the response variable presents some missing observations.

It is well known that the local polynomial approach has some advantages over the kernel method (see, Fan and Gijbels 1996, for an extensive discussion on the comparison between both these methods). In functional data analysis, the local linear modeling has been introduced in the recent years. The first contribution was established by Baìllo and Grané (2009). They study the L^2-consistency of the local linear estimate of the regression function when the regressor is Hilbertian. An alternative fast version of the local linear estimator proposed Barrientos et al. (2010). The latter consider a Banachic covariates. Both cited works consider the independent case. The first result on in functional time series are given by Chouaf and Laksaci (2012). They studied the spatial version of the estimator proposed by Barrientos et al. (2010). They obtained the almost complete consistency of the local linear estimator of the functional spatial regression. In the same context, Laksaci et al. (2013) have studied the local linear estimator of some spatial conditional models. All these cited works consider the complete data case. In this paper we focus on the case when the response variable has missing data at random (MAR).

The statistical analysis of incomplete is an old subject in statistic. We cite, for instance, Yates (1933) for the linear model, Cheng (1994) for the kernel estimator, Chen and Shao (2000) for the kernel nearest neighbor approach, Pérez-Gonzàlez et al. (2009) for the local polynomial model and Boente et al. (2009) for the robust regression model. However, the literature on this topis is still limited in functional statistic. The first result on this subject are obtained by Ferraty et al. (2013). They treated the asymptotic properties of the regression operator when the functional regressor is completely observed and MAR scalar response variable. Considering the functional ergodic time series data, Ling et al. (2015) prove the asymptotic normality of the same estimator proposed by Ferraty et al. (2013). See, also, Ling et al (2016) for the kernel conditional mode estimator in functional ergodic times series data with missing at response variable.

The third component of this contribution is the kernel nearest-neighbor (KNN) method. From practical point of view, this procedures has also more practical advantages than the classical kernel method. In particular, it permits to construct an estimator adaptable to the local structure of the data. There exist an abundant literature in multivariate statistic on the KNN method (see, Collomb, 1980 for the first results and more motivations on this approach). However, in functional statistic, the literature is still limited. The pinner work is Burba et al. (2009). They established the almost complete consistency of the KNN estimator of the nonparametric functional regression. Some alternative results can be found in Laloe (2008), Lian (2011), Kudraszow and Vieu (2013) and Kara-Zaitri et al. (2017) and reference therein
In this paper, we construct an estimator of the nonparametric functional regression by combining the ideas of the local linear approach with those of the kernel nearest-neighbor (KNN) method. This approach allows to benefit from the advantages offered by both methods. As asymptotic result we prove the almost complete consistency (pointwise and uniform) with rate of the constructed estimator. These asymptotic results are established under general conditions. In particular it covers an important case of incomplete functional data case that is the MAR case.
The paper is organized as follows. We present our model in Section 2. In Section 3, we state the pointwise almost consistency with rate of our estimate. The uniform complete convergence rate is studied in Section 4. Some simulation results are discussed in Section 5. The proofs of different results are relegated to the last section.

4.2 Model and its estimate

Let n independent pairs of random variables (X_i, Y_i) for $i = 1, \ldots, n$ that we assume drawn from the pair (X, Y). The latter is valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a semi-metric space and d denotes a semi-metric.

The local linear estimator is usually constructed by approximating, locally, the nonparametric model by a linear function. As discussed in the previous section, there exist a several version of local linear estimator of the functional regression $R(x) = E[Y|X = x]$, $x \in \mathcal{F}$. In this work, we use the same estimator of Baìllo and Grané (2009). Typically we use the following linear approximation, for all x_0 in neighborhood of x

$$R(x_0) = a_x + b_x(x_0 - x) + \rho_x(x_0 - x, x_0 - x) + o(\|x_0 - x\|^2), \tag{4.1}$$

where b_x and ρ_x are, respectively, linear operator and bilinear continuous operator from $\mathcal{F}$ to $\mathcal{R}$ (or $\mathcal{F} \times \mathcal{F}$ to $\mathcal{R}$). In complete data case, the operators a_x and b_x are estimated by the kernel nearest-neighbor method as follows

$$\min_{a,b}\sum_{i=1}^{n}\left(Y_i-a-b(X_i-x)\right)^2K\left(\frac{\|x-X_i\|}{h_k}\right),\tag{4.2}$$

where K is a kernel and $h_k=\min\{h\in\widehat{R}^+,$ such that $\sum_{i=1}^{n}\mathbb{I}_{B(x,h)}(X_i)=k\}$ with $B(x,h)=\{z\in\mathcal{F}:\|x-z\|\le h\}$ the topologically closed ball, in $\mathcal{F}$, centered at x and with radius h.
In our MAR data case we introduce a Bernoulli random variable δ such that $\delta=1$ if Y is observed and $\delta=0$ otherwise. This consideration implies that the variables Y and δ are independent given X. Specifically,

$$\mathbb{P}(\delta=1|X,Y)=\mathbb{P}(\delta=1|X)=P(X).$$

The function $P(\cdot)$ is called the conditional probability of observing Y given X. In practice, this functional operator is unknown. Now, in MAR data case, the two operators a_x and b_x are obtained by the following criterion

$$\min_{(a,b)\in\mathbb{R}^2}\sum_{i=1}^{n}\left(Y_i-a-b(X_i-x)\right)^2\delta_iK\left(\frac{\|x-X_i\|}{h_k}\right).\tag{4.3}$$

Next, similarly to Baìllo and Grané (2009) we reformulate (4.3), under the linearity of b_x and the decomposition of (X_i-x) on a orthonormal basis $(v_j)_{j\ge1}$ of $\mathcal{F}$ as follows

$$\min_{a,b}\sum_{i=1}^{n}\left(Y_i-a-\sum_{j\ge1}c_{ij}(x)b(v_j)\right)^2\delta_iK\left(\frac{\|x-X_i\|}{h_k}\right),$$

where c_{ij} are the coefficients of (X_i-x) in the basis $(v_j)_{j\ge1}$. Nevertheless, this rule is not useable in practice. Thus, we consider its truncated version to a threshold J. Finally, the practical estimate of a_x and b_x are solutions of

$$\min_{a,b_1,\dots,b_J\in\mathcal{H}}\sum_{i=1}^{n}\left(Y_i-a-\sum_{j=1}^{J}c_{ij}(x)b_j\right)^2\delta_iK\left(\frac{\|x-X_i\|}{h_k}\right),$$

where $b_j=b_x(v_j)$ for $j=1,\dots,J$. Explicitly this estimator is given by

$$(Q'_BKQ_B)\begin{pmatrix}\widehat{a}_x\\ \widehat{b}_1\\ \vdots\\ \widehat{b}_J\end{pmatrix}=(Q'_BKY)\text{ where }Q_B=\begin{pmatrix}1 & c_{11}(x) & \dots & c_{1J}(x)\\ \vdots & \vdots & \vdots & \vdots\\ 1 & c_{n1}(x) & \dots & c_{nJ}(x)\end{pmatrix}$$

$K = diag(\delta_1 K(h_k^{-1}\|x - X_1\|), \ldots, \delta_n K(h_k^{-1}\|x - X_n\|))$ and $Y' = (Y_1, \ldots, Y_n)$. Assuming that $(Q'_B K Q_B)$ is a nonsingular matrix then

$$\begin{pmatrix} \widehat{a}_x \\ \widehat{b}_1 \\ \vdots \\ \widehat{b}_J \end{pmatrix} = (Q'_B K Q_B)^{-1}(Q'_B K Y).$$

Hence, the local linear estimator of the regression operator in MAR data, when the matrix $(Q'_B K Q_B)$ is a singular, is definied by

$$\widehat{R}(x) = \widehat{a}_x = e'_1(Q'_B K Q_B)^{-1}(Q'_B K Y) \text{ where } Q_B = \begin{pmatrix} 1 & c_{11} & \ldots & c_{1J} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & c_{n1} & \ldots & c_{nJ} \end{pmatrix}$$

and e'_1 denotes the transpose vector of the first vector of the canonical basis of $\mathcal{R}^J$. It is worth noting that, the present contribution generalizes is the first that consider local linear approach under MAR data using the KNN method. In addition the estimator of Burba et al. (2009) can be obtained by taking $b = 0$, and $\delta = 1$ which correspond to complete data case.

4.3 Pointwise almost complete convergence

In what follows x denotes a fixed curve in $\mathcal{F}$, N_x denotes a fixed neighborhood of x, In order to enounce our results we consider the following conditions :

(H1) For any $r > 0$, $\phi_x(r) := \mathbb{P}(X \in B(x, r)) > 0$, is invertible function and

$$\exists\, 0 < c < 1 < c' < \infty \quad \text{such that} \quad \lim_{r\to 0} \frac{\phi_x(rc)}{\phi_x(r)} < 1 < \lim_{r\to 0} \frac{\phi_x(rc')}{\phi_x(r)}.$$

(H2) The operator ρ_x is continuous and there exists a constant $C > 0$ such that, for all real $m \geq 1$, $E(|Y|^m|X = x) < C$.

(H3) The function $P(\cdot)$ is continuous on $\mathcal{N}_x$ such that $P(x') > 0$, for all $x' \in \mathcal{N}_x$.

(H4) The kernel K is a differentiable function supported on $[0, 1]$. Its derivative K' exists and is such that there exist two constants C and C' with $-\infty < C' < K'(t) < C < 0$ for $0 \leq t \leq 1$.

(H5) The number of neighbor k such that

$$\frac{\log n}{k} \to 0 \quad \text{as} \quad n \to \infty.$$

It is clear that all these hypotheses are very standard. In particular, assumptions (H1), (H2) and (H4) are used by Baìllo and Grané (2009). While Assumption (H3) is used by Ferraty et al. (2013). Finally, we point out the last condition (H5) is also used in Bureba et al. (2009). The following theorem gives the almost-complete convergence of the estimate.

Theorem 3 *Under assumptions (H1)-(H5), we have*

$$|\widehat{R}(x) - R(x)| = O(J^{-1}) + O\left(\phi_x^{-1}\left(\frac{k}{n}\right)^2\right) + O_{a.co.}\left(\sqrt{\frac{\log n}{k}}\right) \text{ as } \min(n, J) \to \infty.$$

Proof of Theorem 3. Firstly, we introduce the following notations, for $j, j' = 1, \ldots, J$:

$$\begin{aligned}
S_{n,j',j}(x) &= \frac{1}{nh_k^2\phi_x(h_k)}\sum_{i=1}^{n} c_{ij'}(x)c_{ij}(x)\delta_i K_i(x,h_k),\\
S_{n,j',0}(x) &= \frac{1}{nh_k\phi_x(h_k)}\sum_{i=1}^{n} c_{ij'}(x)\delta_i K_i(x,h_k)\\
T_{n,j}(x) &= \frac{1}{nh_k\phi_x(h_k)}\sum_{i=1}^{n} c_{ij}(x)\delta_i K_i(x,h_k)Y_i,\\
T_{n,0}(x) &= \frac{1}{n\phi_x(h_k)}\sum_{i=1}^{n} \delta_i K_i(x,h_k)Y_i\\
T_{n,j}(x)^* &= \frac{1}{nh_k\phi_x(h_k)}\sum_{i=1}^{n} c_{ij}(x)\delta_i K_i(x,h_k)(Y_i - R(X_i)),\\
T_{n,0}(x)^* &= \frac{1}{n\phi_x(h_k)}\sum_{i=1}^{n} \delta_i K_i(x,h_k)(Y_i - R(X_i))\\
e_{n,j} &= \frac{1}{nh_k\phi_x(h_k)}\sum_{i=1}^{n} c_{ij}(x)\delta_i K_i(x,h_k)\rho_x(X_i - x, X_i - x),\\
e_{n,0}(x) &= \frac{1}{n\phi_x(h_k)}\sum_{i=1}^{n} \delta_i K_i(x,h_k)\rho_x(X_i - x, X_i - x)\\
e_{n,j}^*(x) &= S_{n,j,0}(x) = \frac{1}{nh_k\phi_x(h_k)}\sum_{i=1}^{n} c_{ij}(x)\delta_i K_i(x,h_k),\\
e_{n,0}(x)^* &= S_{n,0,0}(x) = \frac{1}{n\phi_x(h_k)}\sum_{i=1}^{n} \delta_i K_i(x,h_k)
\end{aligned}$$

where $K_i(x, h_k) = K(h^{-1}\|x - X_i\|)$. Now, we can write :

$$\begin{pmatrix} \widehat{a} \\ h\widehat{b}_1 \\ \vdots \\ h\widehat{b}_J \end{pmatrix} = (S_n)^{-1}(T_n)$$

where

$$S_n(x) = (S_{n,j',j}(x))_{j',j=0,\dots,J} \text{ and } T_n(x) = (T_{n,j}(x))_{j=0,\dots,J}.$$

The error of the projection on an orthonormal basis of a Hilbert space is of order $O(J^{-1})$. Thus, by the regularity assumption (4.1) we get :

$$R(X_i) = a_x + \sum_j^J c_{ij}(x) b_x(v_j) + \rho_x(X_i - x, X_i - x) + O(J^{-1}).$$

Next, let $T_n^*(x) = (T_{n,j}(x)^*)_{j=0,\dots,J}$, $e_n = (e_{n,j})_{j=0,\dots,J}$ and $e_n^*(x) = (e_{n,j}^*(x))_{j=0,\dots,J}$. Finally, we can write

$$\begin{aligned} T_n^*(x) &= T_n(x) - (T_n(x) - T_n^*(x)) \\ &= S_n(x) \begin{pmatrix} \widehat{a}_x \\ h\widehat{b}_1 \\ \vdots \\ h\widehat{b}_J \end{pmatrix} - S_n(x) \begin{pmatrix} a \\ hb_1 \\ \vdots \\ hb_J \end{pmatrix} + e_n(x) + O(J^{-1})e_n^* \end{aligned}$$

It follows that,

$$\begin{pmatrix} \widehat{a}_x \\ h\widehat{b}_1 \\ \vdots \\ h\widehat{b}_J \end{pmatrix} - \begin{pmatrix} a \\ hb_1 \\ \vdots \\ hb_J \end{pmatrix} = S_n^{-1}(x)T_n^*(x) - S_n^{-1}(x)e_n(x) - O(J^{-1})S_n^{-1}(x)e_n^*(x).$$

So,

$$\widehat{R}(x) - R(x) = \widehat{a}_x - a_x = e_1' \left(S_n^{-1}(x)T_n^{1*}(x) - S_n^{-1}(x)e_n(x) - O(J^{-1})S_n^{-1}(x)e_n^*(x) \right)$$

Consequently, Theorem 3's result will be a consequence of the following lemmas for which the proofs are given in the last Section.

Lemma 5 *Under conditions (H1), (H4) and (H5), we have*

$$|S_{n,j',j}(x)| = O_{a.co.}(1)\,.$$

Lemma 6 *Under the conditions of the Theorem 3, we have*

$$\mathbb{E}\left[T_{n,j}(x)\right] = 0 \quad \textit{and} \quad \mathbb{E}\left[e_{n,j}(x)\right] = O\left(\phi_x^{-1}\left(\frac{k}{n}\right)^2\right).$$

Lemma 7 *Under the conditions of the Theorem 3, we have*

$$T_{n,j}(x)^* - \mathbb{E}\left[T_{n,j}(x)^*(x)\right] = O_{a.co.}\left(\sqrt{\frac{\log n}{k}}\right)$$

and

$$e_{n,j}(x) - \mathbb{E}\left[e_{n,j}(x)\right] = O\left(\phi_x^{-1}\left(\frac{k}{n}\right)^2\right) + O_{a.co.}\left(\sqrt{\frac{\log n}{k}}\right).$$

Corollary 3 *Under the conditions of the Theorem 3, we have ,*

$$T_n^*(x) = O_{a.co.}\left(\sqrt{\frac{\log n}{k}}\right)$$

and

$$e_n(x) = O\left(\phi_x^{-1}\left(\frac{k}{n}\right)^2\right) + O_{a.co.}\left(\sqrt{\frac{\log n}{k}}\right).$$

Corollary 4 *Under the conditions of the Theorem 3, we have ,*

$$e_1' S_n^{-1} T_n^{0*} = O_{a.co.}\left(\sqrt{\frac{\log n}{k}}\right)$$

and

$$e_1' S_n^{-1} e_n = O\left(\phi_x^{-1}\left(\frac{k}{n}\right)^2\right) + O_{a.co.}\left(\sqrt{\frac{\log n}{k}}\right).$$

4.4 Strong uniform consistency

The main purpose of this section is to establish the uniform almost complete convergence of $\widehat{R}(x)$ on some subset $S_{\mathcal{F}}$ of $\mathcal{F}$, such that : $S_{\mathcal{F}} \subset \bigcup_{k=1}^{d_n} B(x_k, r_n)$, where $x_k \in \mathcal{F}$ and r_n (resp. d_n) is a sequence of positive real numbers. As discussed in Ferraty et al. (2010), the uniform consistency in functional statistics is not a simple extension of the pointwise one but it is a fundamental development which requires some additional tools and topological conditions. Specifically, we need the following assumptions.

(U1) There exists a differentiable invertible function $\phi(.)$, such that :

$$\text{for all } x \in S_{\mathcal{F}},\ 0 < C\,\phi(h) \leq \mathbb{P}(X \in B(x,h)) \leq C'\,\phi(h) < \infty$$

$$\text{and, there exists } \eta_0 > 0, \quad \text{such that for all } \eta < \eta_0,\ \phi'(\eta) < C,$$

where C and C' are strictly positive constants and where ϕ' denotes the first derivative of ϕ.

(U2) For all $x \in S_{\mathcal{F}}$ the operator ρ_x is continuous and there exists a constant $C > 0$ such that, $m \geq 1,\ \sup_{x \in S_{\mathcal{F}}} E(|Y|^m | X = x) < C$.

(U3) The function $P(\cdot)$ is continuous on $S_{\mathcal{F}}$ and satisfies

$$0 < C < \inf_{x \in S_{\mathcal{F}}} P(x) < \sup_{x \in S_{\mathcal{F}}} P(x) < C' < \infty.$$

(U4) The kernel K is a positive and differentiable function which is supported within $[0, 1]$ and satisfying the following Lipschitz's condition

$$|K(x) - K(y)| \leq C\,\big||x| - |y|\big| \text{ for some } C > 0.$$

(U5) For $r_n = O\left(\dfrac{\log n}{n}\right)$, the sequence d_n satisfies

$$\frac{(\log n)^2}{k} < \log d_n < \frac{k}{\log n}$$

and there exits $\beta > 1$ such that

$$\sum_{n=1}^{\infty} \exp\left\{(1-\beta)\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)\right\} < \infty.$$

(U6) There exist $0 < c < 1 < c' < \infty$ such that

$$\lim_{r \to 0} \frac{\phi(rc)}{\phi(r)} < 1 < \lim_{r \to 0} \frac{\phi(rc')}{\phi(r)}.$$

Theorem 4 *Under assumptions (U1)-(U6), we have*

$$\sup_{x\in S_{\mathcal{F}}} |\widehat{R}(x) - R(x)| = O(J^{-1}) + O\left(\phi_x^{-1}\left(\frac{k}{n}\right)^2\right) + O_{a.co.}\left(\sqrt{\frac{\log d_n}{k}}\right), \text{ as } \min(n, J) \to \infty. \tag{4.4}$$

Proof of Theorem 4. It is clear that, the Theorem 4's proof can be deduced directly from the following uniform versions of the previous lemmas.

Lemma 8 *Under conditions (H1), (H3), (H4) and (H6), we have*

$$\sup_{x\in S_{\mathcal{F}}} |S_{n,j',j}(x)| = O_{a.co.}(1).$$

Lemma 9 *Under the conditions of the*
.. 4, we have

$$\sup_{x\in S_{\mathcal{F}}} \mathbb{E}\left[T_{n,j}(x)^*\right] = 0, \quad \textit{and} \quad \sup_{x\in S_{\mathcal{F}}} \mathbb{E}\left[e_{n,j}(x)\right] = O\left(\phi^{-1}\left(\frac{k}{n}\right)^2\right).$$

Lemma 10 *Under the conditions of the Theorem 4, we have*

$$\sup_{x\in S_{\mathcal{F}}} |T_{n,j}(x)^* - \mathbb{E}\left[T_{n,j}(x)^*\right]| = O_{a.co.}\left(\sqrt{\frac{\log d_n}{k}}\right)$$

and

$$\sup_{x\in S_{\mathcal{F}}} |e_{n,j}(x) - \mathbb{E}\left[e_{n,j}(x)\right]| = O\left(\phi^{-1}\left(\frac{k}{n}\right)^2\right) + O_{a.co.}\left(\sqrt{\frac{\log d_n}{k}}\right).$$

Corollary 5 *Under the conditions of the Theorem 4, we have ,*

$$\sup_{x\in S_{\mathcal{F}}} T_n^*(x) = O_{a.co.}\left(\sqrt{\frac{\log d_n}{k}}\right)$$

and

$$\sup_{x\in S_{\mathcal{F}}} e_n(x) = O\left(\phi^{-1}\left(\frac{k}{n}\right)^2\right) + O_{a.co.}\left(\sqrt{\frac{\log d_n}{k}}\right).$$

Corollary 6 *Under the conditions of the Theorem 4, we have ,*

$$\sup_{x\in S_{\mathcal{F}}} e_1' S_n^{-1}(x) T_n^{0*}(x) = O_{a.co.}\left(\sqrt{\frac{\log d_n}{k}}\right)$$

and

$$\sup_{x\in S_{\mathcal{F}}} e_1' S_n^{-1}(x) e_n(x) = O\left(\phi^{-1}\left(\frac{k}{n}\right)^2\right) + O_{a.co.}\left(\sqrt{\frac{\log d_n}{k}}\right).$$

4.5 Appendix

In what follows, when no confusion is possible, we will denote by C and C' some strictly positive generic constants.

Proof of Lemma 5 We treat only the case where j and j' are different from 0. The other cases can be treated by the same arguments. Indeed, we use (H1) to define $(\alpha,\beta)\in(0,1)^2$ such that

$$\phi_x^{-1}\left(\frac{k}{\alpha n}\right) \le C\phi_x^{-1}\left(\beta\frac{k}{n}\right). \tag{4.5}$$

Next, we put $h_k^+ = \phi_x^{-1}\left(\frac{k}{\alpha n}\right)$, $h_k^- = \phi_x^{-1}\left(\beta\frac{k}{n}\right)$ and similarly to Burba et al. (2009) we show that

$$\sum_n \mathbb{P}\left(h_k \notin (h_k^-, h_k^+)\right) < \infty. \tag{4.6}$$

Therefore, we have, almost complectly,

$$S_{n,j',j}(x) \in (\widetilde{S}_{n,j',j}, \widehat{S}_{n,j',j}).$$

where

$$\widehat{S}_{n,j',j} = \frac{1}{nh_k^{-2}\phi_x(h_k^-)}\sum_{i=1}^n c_{ij'}(x)c_{ij}(x)K_i(x,h_k^+)$$

$$\text{and}\quad \widetilde{S}_{n,j',j} = \frac{1}{nh_k^{+2}\phi_x(h_k^+)}\sum_{i=1}^n c_{ij'}(x)c_{ij}(x)K_i(x,h_k^-).$$

Observe that

$$\widehat{S}_{n,j',j} = \left(\frac{nh_k^{+2}\phi_x(h_k^+)}{nh_k^{-2}\phi_x(h_k^-)}\right)\frac{1}{nh_k^{+2}\phi_x(h_k^+)}\sum_{i=1}^n c_{ij'}(x)c_{ij}(x)K_i(x,h_k^+)$$

and

$$\widetilde{S}_{n,j',j} = \left(\frac{nh_k^{-2}\phi_x(h_k^-)}{nh_k^{+2}\phi_x(h_k^+)}\right)\frac{1}{nh_k^{-2}\phi_x(h_k^-)}\sum_{i=1}^{n} c_{ij'}(x)c_{ij}(x)K_i(x,h_k^-).$$

By (4.5), we get

$$\widehat{S}_{n,j',j} \leq C\frac{1}{nh_k^{+2}\phi_x(h_k^+)}\sum_{i=1}^{n}|c_{ij'}(x)c_{ij}(x)|K_i(x,h_k^+)$$

and

$$\widetilde{S}_{n,j',j} \leq C'\frac{1}{nh_k^{-2}\phi_x(h_k^-)}\sum_{i=1}^{n}|c_{ij'}(x)c_{ij}(x)|K_i(x,h_k^-).$$

Hence, our result is consequence of the following facts

$$\ddot{S}_{n,j',j}(h) = O_{a.co.}(1) \qquad \text{for} \quad h = h^{\pm}$$

with

$$\ddot{S}_{n,j',j}(x,h) = \frac{1}{nh^2\phi_x(h)}\sum_{i=1}^{n}|c_{ij'}(x)c_{ij}(x)|\delta_i K_i(x,h).$$

To do that we prove

$$\ddot{S}_{n,j',j}(x,h) - \mathbb{E}\left[\ddot{S}_{n,j',j}(x,h)\right] = O_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right)$$

$$\text{and} \quad \mathbb{E}\left[\ddot{S}_{n,j',j}(x,h)\right] = O(1). \tag{4.7}$$

For the first claimed result, we put $\widetilde{\Delta}_i = \frac{1}{h^2\phi_x(h)}|c_{ij'}(x)c_{ij}(x)|\delta_i K_i(x,h)$ and we write

$$\left|\ddot{S}_{n,j',j}(x,h) - \mathbb{E}\left[\ddot{S}_{n,j',j}(x,h)\right]\right| = \left|\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\Delta}_i - \mathbb{E}[\widetilde{\Delta}_i]\right)\right|.$$

As $(v_j)_j$ is orthonormal basis then

$$\text{for all } j \leq J, \ |c_{1j}(x)| \leq \|v_j\|\|x - X_1\| \leq \|x - X_1\|.$$

Hence,

$$\mathbb{E}\left[|c_{1j'}(x)c_{1j}(x)|\delta_1 K_1(x,h)\right] \leq \mathbb{E}\left[\|x - X_1\|^2 K_1(x,h)\right] \leq Ch^2\phi_x(h_k).$$

We deduce by (H1) and (H4) that

$$\left|\widetilde{\Delta}_i\right| < C/\phi_x(h) \quad \text{and} \quad \mathbb{E}\left|\widetilde{\Delta}_i\right|^2 < C'/\phi_x(h).$$

Thus, we use classical Bernstein's inequality (Uspensky, 1937, p205) to deduce that for $\eta > 0$

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\Delta}_i - \mathbb{E}[\widetilde{\Delta}_i]\right)\right| > \eta\sqrt{\frac{\log n}{n\,\phi_x(h)}}\right\} \;\leq\; C' n^{-C\eta^2}.$$

Now, for the expectation term

$$\mathbb{E}[\ddot{S}_{n,j',j}(x,h)] = \frac{1}{h^2\phi_x(h)}\mathbb{E}\left[|c_{1k}c_{ij}(x)|K_1(x,h)\right],$$

Once again we use the fact that the basis $(v_j)_j$ is orthonormal to get

$$\mathbb{E}\left[c_{1j'}(x)c_{ij}(x)K_1(x,h)\right] \leq \mathbb{E}\left[\|x - X_1\|^2 K_1(x,h)\right] \leq Ch^2\phi_x(h)$$

which allows to achieve the proof of this Lemma.

■

Proof of Lemma 6. Similarly to Lemma 5, we will treat only the case where $j \neq 0$. The case $j = 0$ can be obtained by a similar proof. Using the same arguments as those used in the previous lemma it suffices to evaluate the following quantities

$$\mathbb{E}\left[\ddot{T}^*_{n,j}(x)\right] = 0, \quad \text{and} \quad \mathbb{E}\left[\ddot{e}_{n,j}(x)\right] = O\left(\phi_x^{-1}\left(\frac{k}{n}\right)^2\right)$$

where

$$\ddot{T}^*_{n,j}(x) = \frac{1}{nh\phi_x(h)}\sum_{i=1}^{n}|c_{ij}(x)|\delta_i K_i(x,h)(Y_i - R(X_i)) \qquad h = h^{\pm}$$

$$\ddot{e}_{n,j}(x) = \frac{1}{nh\phi_x(h)}\sum_{i=1}^{n}|c_{ij}(x)|\delta_i K_i(x,h)\rho_x(X_i - x, X_i - x) \qquad h = h^{\pm}.$$

The evaluation of these bias terms is based on the property of the identically distribution of the observations. We start by the last term, $\mathbb{E}[\ddot{e}_{n,j}(x)]$. For this, we use the continuity of the operator ρ_x to write that

$$\|\rho_x(X_1 - x, X_1 - x - X_1)\| \leq C\|X_1 - x\|^2 \tag{4.8}$$

and as the basis $(v_j)_j$ is orthonormal, we have

$$\text{for all } j \leq J, \quad |c_{1j}(x)| \leq \|x - X_1\| \tag{4.9}$$

Combining (4.8) and (4.9) to deduce that

$$\mathbb{E}\left[|c_{1j}(x)|\delta_1 K_1(x,h)\rho_x(X_1 - x, X_1 - x)\right] \leq \mathbb{E}\left[\|x - X_1\|^3 K_1(x,h)\right] \leq Ch^3\phi_x(h).$$

Consequently

$$\mathbb{E}[\ddot{e}_{n,j}(x)] \leq Ch^2.$$

Now, for the first term we have

$$\mathbb{E}[\ddot{T}^*_{n,j}(x)] = \frac{1}{h\phi_x(h)}\mathbb{E}\left[|c_{1j}(x)|\delta_1 K_1(x,h)(Y_1 - R(X_1))\right].$$

Therefore, by conditioning by X_1, we show that

$$\mathbb{E}[\ddot{T}^*_{n,j}(x)] = \frac{1}{h\phi_x(h)}\mathbb{E}\left[|c_{1j}(x)|K_1(x,h)P(X_1)(\mathbb{E}\left[Y_1|X_1\right] - R(X_1))\right].$$

It follows that

$$\mathbb{E}[\ddot{T}^*_{n,j}(x)] = 0.$$

Proof of Lemma 7 We will give only the demonstration for $T^*_{n,j}(x)$ with $j \neq 0$. The proofs of the other cases ($T^*_{n,0}(x)$, and $e_{n,j}(x)$ can be accomplished by the same manner). Furthermore, for $T^*_{n,j}(x)$, we use (4.6) to get, almost completely

$$T^*_{n,j}(x) - \mathbb{E}\left[T^*_{n,j}(x)\right] \in \left(\widehat{T}^*_{n,j}(x) - \mathbb{E}\left[\widetilde{T}^*_{n,j}(x)\right], \widetilde{T}^*_{n,j}(x) - \mathbb{E}\left[\widehat{T}^*_{n,j}(x)\right]\right)$$

with

$$\widehat{T}^*_{n,j}(x) = \frac{1}{nh_k^-\phi_x(h_k^-)}\sum_{i=1}^n c_{ij}(x)K_i(x,h_k^+)(Y_i - R(X_i))$$

$$\text{and} \quad \widetilde{T}^*_{n,j}(x) = \frac{1}{nh_k^+\phi_x(h_k^+)}\sum_{i=1}^n c_{ij}(x)K_i(x,h_k^-)(Y_i - R(X_i)).$$

Observe that

$$\mathbb{E}\left[\widehat{T}^*_{n,j}(x)\right] = \mathbb{E}\left[\widetilde{T}^*_{n,j}(x)\right] = 0.$$

This last is deuced from Lemma 6. Simlearly to Lemma 5 It suffices to show that

$$\ddot{T}^*_{n,j}(x) - \mathbb{E}\left[\ddot{T}^*_{n,j}(x)\right] = O_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right),$$

with

$$\ddot{T}^*_{n,j}(x) = \frac{1}{nh\phi_x(h)} \sum_{i=1}^{n} |c_{ij}(x)| K_i(x,h)(Y - R(X_i)).$$

For this aim, we apply, the unbounded Bernstein's exponential inequality (see Corollary A8 in Ferraty and Vieu (2006), p. 234). We precise that, the latter is based on the asymptotic evaluation of mth order moments of the following random variables

$$\Gamma_{i,j} = \frac{|c_{ij}(x)|\delta_i K_i(x,h)(Y_i - R(X_i))}{h\phi(h)} - \mathbb{E}\left[\frac{|c_{ij}(x)| K_i(x,h)(Y_i - R(X_i))}{h\phi(h)}\right].$$

Notice that, by the Newton's binomial expansion, we obtain :

$$\begin{aligned}
&\mathbb{E}\left|(|c_{ij}(x)|\delta_i K_i(x,h)(Y_i - R(X_i)) - \mathbb{E}\left[|c_{ij}(x)|\delta_i K_i(x,h)(Y_i - R(X_i))\right])^m\right| \\
&\quad = \mathbb{E}\Bigg|\sum_{d=0}^{m} C_m^d \left(|c_{ij}(x)|\delta_i K_i(x,h)(Y_i - R(X_i))\right)^d \\
&\qquad\qquad (\mathbb{E}\left[|c_{ij}(x)|\delta_i K_i(x,h)(Y_i - R(X_i))\right])^{m-d} (-1)^{m-d}\Bigg| \\
&\quad \leq \sum_{d=0}^{m} C_m^d \left(\mathbb{E}\left||c_{ij}(x)|\delta_i K_i(x,h)(Y_i - R(X_i))\right|^d\right) \left|\mathbb{E}\left[|c_{ij}(x)|\delta_i K_i(x,h)(Y_i - R(X_i))\right]\right|^{m-d} \\
&\quad \leq \sum_{d=0}^{m} C_m^d \mathbb{E}\left||c_{ij}(x)|^d \delta_i^d K_i^d(x,h)(Y_i - R(X_i))^d\right| \left|\mathbb{E}\left[|c_{ij}(x)|\delta_i K_i(x,h)(Y_i - R(X_i))\right]\right|^{m-d}
\end{aligned}$$

where $C_{k,m} = m!/(k!(m-k)!)$.

Since the variables δ and Y are independent given X, then under Assumption (H6) we have for all $d \leq m$

$$\mathbb{E}\left[\delta Y^d | X\right] = (P(x) + o(1))\mathbb{E}\left[Y^d | X\right] \leq C.$$

Now,

$$\begin{aligned}
&h^{-m}\phi_x^{-m}(h) \sum_{d=0}^{m} C_m^d \mathbb{E}\left||c_{ij}(x)|^d \delta_i^d K_i^d(h)(Y_i - R(X_i))^d\right| \\
&\qquad\qquad \left|\mathbb{E}\left[|c_{ij}(x)|\delta_i K_i(x,h)(Y_i - R(X_i))\right]\right|^{m-d} \leq C\phi_x(h)^{-m+1}.
\end{aligned}$$

Therefore, for $l = 0, 1, 2$, and $k = 0, 1$, we obtain

$$\mathbb{E}\left|\Gamma_{i,j}\right|^m = O\left((\phi_x(h))^{-m+1}\right).$$

Consequently, it suffices to apply the Corollary A8 in Ferraty and Vieu (2006), for $a_n = (\phi_x(h))^{-1/2}$, to conclude

$$\sum_n \mathbb{P}\left\{\left|\ddot{T}^*_{n,j} - \mathbb{E}[\ddot{T}^*_{n,j}]\right| > \eta\sqrt{\frac{\log n}{n\,\phi_x(h)}}\right\} < \infty$$

■

Proof of Lemma 8. The proof of this lemma is based on same ideas as for Lemma 5, since it suffices to prove the uniformity of its claimed results. We point out the uniform version (4.6) is shown in Kudraszow and Vieu (2013). Thus, all it remains to show that

$$\sup_{x\in S_{\mathcal{F}}} |\ddot{S}_{n,j',j}(x,h) - \mathbb{E}\left[\ddot{S}_{n,j',j}(x,h)\right]| = O_{a.co.}\left(\sqrt{\frac{\log d_n}{n\phi(h)}}\right)$$
$$\text{and} \quad \sup_{x\in S_{\mathcal{F}}} \mathbb{E}\left[\ddot{S}_{n,j',j}(x,h)\right] = O(1). \tag{4.10}$$

The proof of the first part follows the same ideas as in Ferraty et al. (2010). Indeed, we define $j(x) = \arg\min_{j\in\{1,2,\ldots,d_n\}} \|x - x_k\|$, and we consider the following decomposition

$$\begin{aligned}\sup_{x\in S_{\mathcal{F}}} |\ddot{S}_{n,j',j}(x,h) - \mathbb{E}[\ddot{S}_{n,j',j}(x,h)]| \quad \leq \quad & \underbrace{\sup_{x\in S_{\mathcal{F}}} |\ddot{S}_{n,j',j}(x,h) - \ddot{S}_{n,j',j}(x_{j(x)},h)|}_{F_1} \\ & + \underbrace{\sup_{x\in S_{\mathcal{F}}} |\ddot{S}_{n,j',j}(x_{j(x)},h) - \mathbb{E}[\ddot{S}_{n,j',j}(x_{j(x)},h)]|}_{F_2} \\ & + \underbrace{\sup_{x\in S_{\mathcal{F}}} |\mathbb{E}[\ddot{S}_{n,j',j}(x_{j(x)},h)] - \mathbb{E}[\ddot{S}_{n,j',j}(x,h)]|.}_{F_3}\end{aligned}$$

Concerning the first term F_1 we the Lipschitz conditions, on the kernel K and $< v_1, x - X_i >$, we obtain

$$F_1 \leq C \sup_{x\in S_{\mathcal{F}}} \left(F_{11} + F_{12} + F_{13} + F_{14}\right),$$

where

$$\begin{aligned}
F_{11} &= \frac{C}{n\phi(h)} \sum_{i=1}^{n} \mathbb{1}_{B(x,h)\cap\overline{B(x_{j(x)},h)}}(X_i),\\
F_{12} &= \frac{C\epsilon}{n\phi(h)} \sum_{i=1}^{n} \mathbb{1}_{B(x,h)\cap B(x_{j(x)},h)}(X_i).\\
F_{13} &= \frac{C\epsilon}{nh\phi(h)} \sum_{i=1}^{n} \mathbb{1}_{B(x,h)\cap B(x_{j(x)},h)}(X_i).\\
F_{14} &= \frac{C}{n\phi(h)} \sum_{i=1}^{n} \mathbb{1}_{B(x_{j(x)},h)\cap\overline{B(x,h)}}(X_i).
\end{aligned}$$

Now, we apply a standard inequality for sums of bounded random variables (cf. Corollary A.9 in Ferraty and Vieu, 2006) with Z_i is identified such that :

$$Z_i = \begin{cases}
\dfrac{1}{\phi(h)} \sup\limits_{x\in S_{\mathcal{F}}} \left[\mathbb{1}_{B(x,h)\cap\overline{B(x_{j(x)},h)}}(X_i)\right] & \text{for } F_{11}\\[2ex]
\dfrac{\epsilon}{h\phi(h)} \sup\limits_{x\in S_{\mathcal{F}}} \left[\mathbb{1}_{B(x,h)\cap B(x_{j(x)},h)}(X_i)\right] & \text{for } F_{12} \text{ and } F_{13}\\[2ex]
\dfrac{1}{\phi(h)} \sup\limits_{x\in S_{\mathcal{F}}} \left[\mathbb{1}_{B(x_{j(x)},h)\cap\overline{B(x,h)}}(X_i)\right] & \text{for } F_{14}
\end{cases}$$

So, under the second part of (U1), we have for the first and the last cases

$$Z_1 = O\left(\frac{1}{\phi(h)}\right),\ \mathbb{E}[Z_1] = O\left(\frac{\epsilon}{\phi(h)}\right) \text{ and } \mathrm{var}(Z_1) = O\left(\frac{\epsilon}{(\phi(h))^2}\right).$$

Hence, we get

$$F_{11} = O\left(\frac{\epsilon}{\phi(h)}\right) + O_{a.co.}\left(\sqrt{\frac{\epsilon\,\log n}{n\,\phi(h)^2}}\right).$$

In the same way, assumption (U5) allows to get, for F_{12} or F_{13} cases, that

$$Z_1 = O\left(\frac{\epsilon}{h\phi(h)}\right),\ \mathbb{E}[Z_1] = O\left(\frac{\epsilon}{h}\right) \text{ and } \mathrm{var}(Z_1) = O\left(\frac{\epsilon^2}{h^2\phi(h)}\right).$$

Then

$$F_{12} \;=\; O_{a.co.}\left(\sqrt{\frac{\log d_n}{n\,\phi(h)}}\right).$$

Finally, it suffices to put together all the intermediate results and to use assumption (U5) to obtain

$$F_1 \;=\; O_{a.co.}\left(\sqrt{\frac{\log d_n}{n\,\phi(h)}}\right). \tag{4.11}$$

Concerning F_3, we use the fact that

$$F_3 \le \mathbb{E}\left[\sup_{x\in S_{\mathcal{F}}} |\ddot{S}_{n,j',j}(x,h) - \mathbb{E}[\ddot{S}_{n,j',j}(x,h)]|\right]$$

to get

$$F_3 \;=\; O\left(\sqrt{\frac{\log d_n}{n\phi(h)}}\right).$$

Concerning the terms F_2 we apply Bernstein's inequality to

$$\Delta_i = \frac{1}{h^2\phi_x(h)}\left[|c_{ij'}(x_k)c_{ij}(x_k)|\delta_i K_i(x_k,h) - \mathbb{E}[|c_{ij'}(x_k)c_{ij}(x_k)|\delta_i K_i(x_k,h]]\right)$$

Then, by standard arguments we show, for all $m \ge 1$, that

$$|\Delta_i| < C/\phi_x(h) \quad \text{and} \quad \mathbb{E}\,|\Delta_i|^2 < C'/\phi_x(h).$$

Therefore, for $C\eta^2 = \beta$, we have

$$\begin{aligned} d_n \max_{k\in\{1,\ldots,d_n\}} \mathbb{P}&\left(|\ddot{S}_{n,j',j}(x_{j(x)},h) - \mathbb{E}[\ddot{S}_{n,j',j}(x_{j(x)},h)]| > \eta\sqrt{\frac{\log d_n}{n\phi(h)}}\right) \\ &= \mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^n \Delta_i\right| > \eta\sqrt{\frac{\log d_n}{n\phi(h)}}\right) \le d_n^{1-\beta}. \end{aligned}$$

The proof of the last part of (4.10) follows the same lines allowing to show (4.7) combined with the condition (U1)

Proof of lemma 9 and 10. The proof of these lemmas are omitted it follows the same lines as in Lemma 6 and Lemma 8 respectively

■

Bibliographie

[1] Baìllo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response, *J. of Multivariate Analysis*, 100, 102-111.

[2] Barrientos-Marin, J., Ferraty, F. and Vieu, P. (2010). Locally Modelled Regression and Functional Data. *J. of Nonparametric Statistics*, 22, No. 5, 617-632.

[3] Boente, G., Gonzalez-Manteiga, W. and Perez-Gonzalez, A. (2009). Robust nonparametric estimation with missing data. *J. Statist. Plann. Inference,* 139 , 571-592.

[4] Bosq, D. (2000). *Linear Processes in Function Spaces : Theory and applications.* Lecture Notes in Statistics, 149, Springer.

[5] Burba, F., Ferraty, F. and Vieu, P. (2009). k-nearest neighbor method in functional non- parametric regression. *J. Nonparametric Statist.,* 21, 453-469.

[6] Chen, J.H., Shao, J., (2000). Nearest neighbor imputation for survey data. *J. Official Statist.,* 16, 113-131.

[7] Cheng, P.E., (1994). Nonparametric estimation of mean functionals with data missing at random. *J. Amer. Statist. Assoc.* 89, 81–87.

[8] Chouaf, A. and Laksaci, A., (2013), On the functional local linear estimate for spatial regression. *Stat. Risk Model,* 29, 189-214.

[9] Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference*, 147, 1-23.

[10] Demongeot, J., Laksaci, A., Rachdi, M. and Rahmani, S. (2014). On the local linear modelization of the conditional distribution for functional data. *Sankhya A,* 76, 328-355

[11] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. London, Chapman & Hall.

[12] Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference*, 140, 335-352.

[13] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis. Theory and Practice.* Springer Series in Statistics. New York.

[14] Ferraty, F., Sued, M. and Vieu, P. (2013). Mean estimation with data missing at random for functional covariables. *Statistics,* 47, 688-706.

[15] Hsing, T. and Eubank, R. (2015). Theoretical foundations of functional data analysis, with an introduction to linear operators. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester.

[16] Kudraszow, N. and Vieu, P. (2013). Uniform consistency of kNN regressors for functional variables. *Statist. Probab. Lett.,* 83, 1863-1870.

[17] Ling, N., Liu, Y. and Vieu, P. (2016). Conditional mode estimation for functional stationary ergodic data with responses missing at random. *Statistics,* 50, 991-1013.

[18] Ling, N., Liu, Y. and Vieu, P.(2015). Nonparametric regression estimation for functional stationary ergodic data with missing at random. *J. Statist. Plann. Inference,* 162, 75-87.

[19] Perez-Gonzalez, A., Vilar-Fernandez and J. M. Gonzalez-Manteiga, W. (2009) Asymptotic properties of local polynomial regression with missing data and correlated errors. *Ann. Inst. Statist. Math.,* 6, 85-109

[20] Ramsay, J. O. and Silverman, B. W. (2002). *Applied functional data analysis. Methods and case studies.* Springer Series in Statistics. New York.

[21] Yates, F.,(1933). The analysis of replicated experiments when the field results are incomplete. *Emporium J. Exp. Agriculture,* 1, 129-142.

[22] Zhang, J. (2014). *Analysis of variance for functional data.* Monographs on Statistics and Applied Probability, 127. CRC Press, Boca Raton, FL.

Chapitre 5

Etude simulation

L'objectif principal de cette courte section illustrative est pour évaluer la performance de notre approche en ce qui concerne le pourcentage d'observations manquantes. Précisément, on compare son comportement à l'estimateur classique

$$\widetilde{R} = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{\|x-X_i\|}{h_k}\right)}{\sum_{i=1}^{n} K\left(\frac{\|x-X_i\|}{h_k}\right)}$$

Pour cet objectif, nous générons nos données selon le modèle suivant :

$$Y = r(x) + \varepsilon$$

où $\varepsilon \sim \mathcal{N}(0, .5)$. L'opérateur $r(\cdot)$ est défini par :

$$R(X) = 5\left(\frac{1}{1 + \int_0^1 X^2(t)dt}\right).$$

La covariable fonctionnelle $X_i(t)$ sont définis, pour tout $t \in [0,1]$, par :

$$X_i(t) = 3W_i \sin(2\pi t) + \eta_i t \quad \text{où } W_i \sim \mathcal{N}(0, 0.5) \text{ et } \eta_i \sim \mathcal{N}(0, 1).$$

Un échantillon de ces courbes est tracé dans la Figure 5.1.

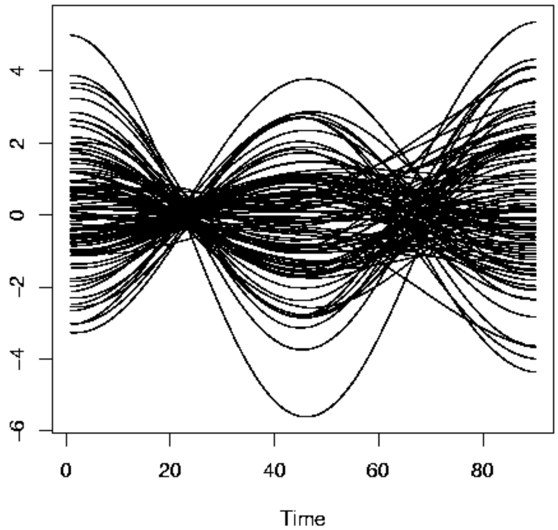

FIGURE 5.1 – Quelques courbes explicatives

Pour introduire les différents niveaux de données des observations manquantes nous générons une variable aléatoire de Bernoulli δ en ce qui concerne les exemples suivants de $P(x)$

$$\begin{cases} P(x) = 1 - \exp\left(-\left(\int_0^1 x^2(t)dt\right)^{(1/2)}\right) & \text{Cas manquant faible} \\ P(x) = \left|\sin\left(\pi * \int_0^1 x^2(t)dt\right)\right| & \text{Cas manquant moyen} \\ P(x) = \left(\sin\left(\pi * \int_0^1 x^2(t)dt\right)\right)^2 & \text{Cas manquant fort} \end{cases}$$

Nous soulignons que la proportion d'observations manquantes dans le cas faible est proche de 11 % et proche de 36 % dans le cas moyen, alors que pour le cas fort est de 55 %.

Pour cette étude comparative, nous avons mis $n = 100$, nous avons choisi la bande passante optimale h par la méthode de validation croisée sur le k -plus proche voisins (k NN) et nous utilisons la définition semi-métrique par la distance L_2 entre les courbes. Ensuite, nous choisissons un noyau quadratique K. Finalement, les résultats de l'étude de comparaison sont résumés dans le tableau 5.1. où nous donnons les erreurs quadratiques moyennes

$$\mathrm{MSE(LR)} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{Y_i} - \widehat{\mathrm{R}}_\mathrm{i}(\mathrm{X_i}))^2, \text{ et } \mathrm{MSE(CR)} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{Y_i} - \widetilde{\mathrm{R}}_\mathrm{i}(\mathrm{X_i}))^2$$

où $\widetilde{R}_i$ (resp. $\widehat{R}_i$) est la version non répétitive de $\widetilde{R}$ (resp. $\widehat{R}$) calculée en supprimant la i^{eme} donnée de l'échantillon initial. Nous obtenons les résultats suivants

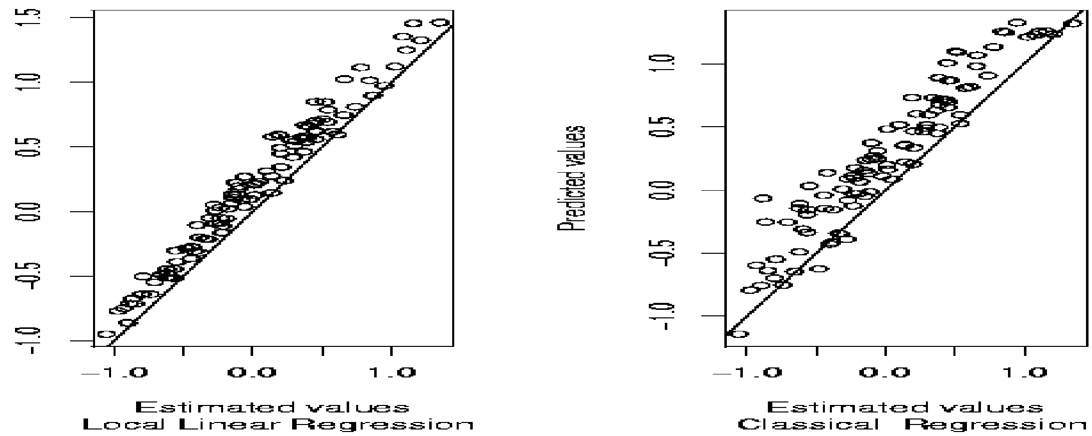

FIGURE 5.2 – Cas faible

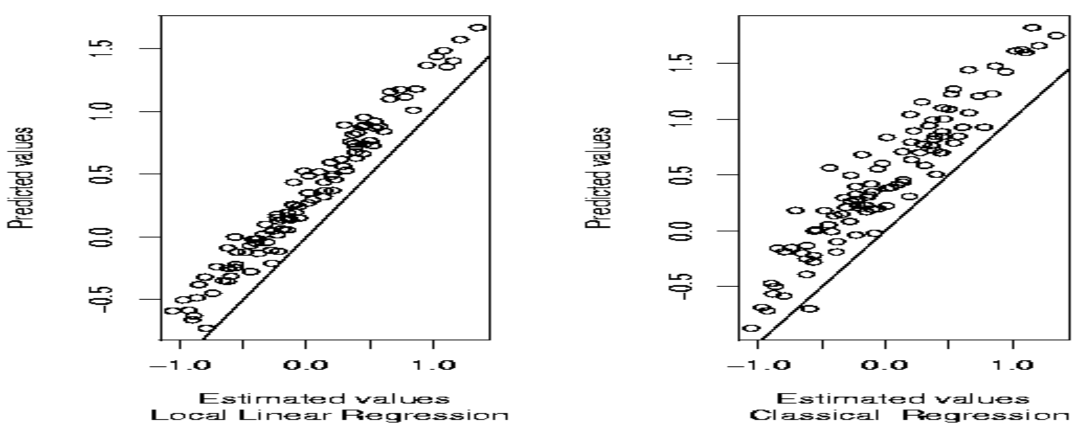

FIGURE 5.3 – Cas moyen

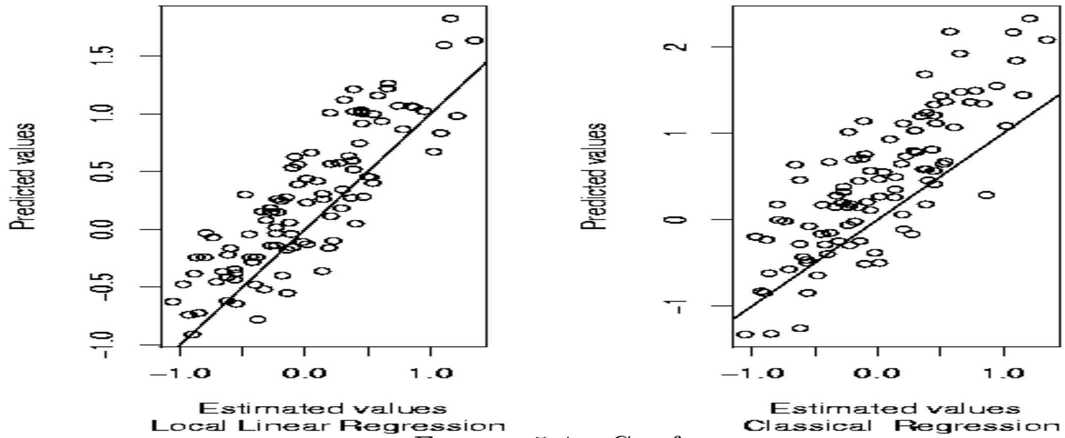

FIGURE 5.4 – Cas fort

niveau manquant	MSE(LR)	MSE(CR)
Faible	0.24	0.35
Moyen	0.74	0.91
Fort	1.78	1.92

TABLE 5.1 – Les erreurs MSE

A partir de le tableau 5.1 et des figures (2-4), on peut dire que le comportement des estimateurs est fortement influencé par le pourcentage des observations manquantes. Mais la performance du l'approche local linéaire est nettement mieux que le cas classique de régression. Globalement les trois estimateurs ont une bonne performance pour les échantillons finis pour les observations manquantes inférieures.

Conclusion et Perspectives

1. Conclusion

Dans cette thèse de doctorat par les chapitres présentés, nous avons pu contribuer au domaine de la statistique non paramétrique de la fonction de regression plus précéssement à l'estimation fonctionnelle de la fonction régression par différentes méthodes et dans différents cas de dépendance. En premier lieu, nous avons considéré un estimateur par l'approche locale linéaire qui est une généralisation de la méthode introduite par Fan et Gijbels en 1996. Comme résultats asymptotiques nous avons établi la convergence presque complète de l'estimateur de la fonction de régression en cas des données incomplètes (missing at random) avec des observations i.i.d. Une généralisation au cas des données dépendantes est aussi étudié, nous avons de même établi la vitesse de convergence presque complète ponctuelle de notre estimateur construit. Dans un second temps, nous avons considéré la méthode des plus proches voisins et nous avons établi la convergence presque complète et la normalité asymptotique de l'estimateur de la fonction régression. Il est clair que nos résultats théoriques obtenus couvrent tous les différents types de convergence stochastique comme il s'agit d'une convergence presque complète qui est la plus forte des convergences. De plus, notre estimateur possède de bonnes propriétés asymptotiques dans tous les cas de notre étude.

2. Perspectives

Les méthodes d'estimation non paramétrique ont été proposés comme alternative à la méthode de prévision. Cependant, les résultats théoriques de l'estimation (à noyau, locale linéaire et k-plus proches voisins) sont montrer et utiliser dans le domaine de statistique, mais il y a beaucoup d'étude à faire, nous proposons quelques questions ouvertes qui restent à développer pour de futures recherches. Tous les résultats obtenus dans cette thèse reposent sur l'hypothèse que les observations sont missing at random. Il pourrait être possible de considérer d'autre type comme le cas de censure et de trancature. Le travail sur les données spatiales en présence d'une variable explicative fonctionnelle n'est pas encore abordé, ce qui nous permettra d'adapter nos résultats pour le cas de missing at random pour la variable réponse. L'importance du choix de la semi métrique et le paramètre de lissage pour améliorer les vitesses de convergence et optimiser les intervalles de confiance respectivement, nous poussera à généraliser les résultats existants en utilisant d'autres familles de semi métriques et l'étude du choix de la fenêtre de lissage.

Bibliographie Générale

Bibliographie

[1] Akritas, M., Politis, D. (2003). Recent advances and trends in nonparametric statistics. *Elsevier, Amsterdam .*

[2] Aguilera A.M, F.A. Ocana, and M.J. Valderrama. (1997). An approximated principal component prediction for continuous time stochastic processes. *Applied Stochastic Models and Data Analysis,* 13, 61-72.

[3] Attouch, M., Laksaci, A., Ould-Sa`id, E. (2009). Asymptotic Distribution of Robust Estimator for Functional Nonparametric Models. *Communications in Statistics : Theory and Methods.* 38, 1317-1335.

[4] Azzeddine, N., Laksaci, A., Ould-Sa`id, E. (2008). On the robust nonparametric regression estimation for functional regressor. *Statistic and Probabability Letters.* 78, 3216-3221.

[5] Baìllo, A., Grané, A. (2009). Local linear regression for functional predictor and scalar response. *Journal of Multivariate Analysis,* 100, 102-111.

[6] Barrientos-Marin, J.(2007). Some Practical Problems of Recent Nonparametric Procedures : Testing, Estimation, and Application. PhD thesis (in French) from the Paul Sabatier's University (Toulouse).

[7] Barrientos-Marin, J., F. Ferraty, F and Vieu, P. (2010). Locally modelled regression and functional data. *Journal of Nonparametric Statistics,* 22(5), 617-632.

[8] Benhenni, K., Ferraty, F., Rachdi, M., Vieu, P. (2007). Locally smoothing regression with functional data. *Computational. Statistics.* 22, 353-370.

[9] Benhenni, K., Griche-Hedli, S., Rachdi, M. (2010). Estimation of the regression operator from functional fixed-design with correlated errors. *Journal of Multivariate Analysis,* 101, 476-490.

[10] Berlinet, A. Elamine, A. Mas.(2011). Local linear regression for functional data. *Ann. Inst. Stat. Math.* 63, 1047-1075.

[11] Besse and J.O. Ramsay. (1986). Principal component analysis of sampled curves. *Psychometrica,* 51, 285-311.

[12] Besse. (1991). Approximation spline de l'analyse en composantes principales d'une variable alťeatoire hilbertienne. *Annales de la Facultťe des Sciences de Toulouse, XII(3)* , 329-349.

[13] Biau, F. Bunea, and M. Wegkamp. (2005). Functional classification in Hilbert spaces. *EEE Transactions on Information Theory,* 51 , 2163-2172.

[14] Boente, G., Gonzalez-Manteiga, W. and Perez-Gonzalez, A. (2009) Robust nonparametric estimation with missing data. *J. Statist. Plann. Inference,* 139, 571-592.

[15] Boj, P. Delicado, J. Fortiana.(2010). Distance-based local linear regression for functional predictors, Comput. *Stat. Data Anal.* 54, 429-437.

[16] Bosq, D.(2000). Linear Processes in Function Spaces, Theory and Applications. *Lecture Notes in Statistics,* 149, Springer-Verlag, New York .

[17] Burba, F., Ferraty, F., Vieu, P. (2008). Convergence de l'estimateur á noyau des k plus proches voisns en régression fonctionnelle non-paramétrique. *C. R. Acad. Sci. Paris.* 346, 339-342.

[18] Burba, F., Ferraty, F. and Vieu, P. (2009) k-nearest neighbor method in functional non- parametric regression. *J. Nonparametric Statist.,* **21**, 453-469.

[19] Cardot.H, F. Ferraty, and P Sarda. (1999). Functional Linear Model. *Statistics and Probability Letter,* 45, 11-22.

[20] Cardot, H., Crambes, C., Sarda, P. (2004). Spline estimation of conditional quantiles for functional covariates, C. R. *Math. Acad. Sci. Paris.* 339, 141-144.

[21] Chen, J.H., Shao, J., (2000). Nearest neighbor imputation for survey data. *J. Official Statist.,* 16, 113-131.

[22] Cheng, P.E., (1994). Nonparametric estimation of mean functionals with data missing at random. *J. Amer. Statist. Assoc.* 89, 81-87.

[23] Chouaf, A. and Laksaci, A., (2013). On the functional local linear estimate for spatial regression. *Stat. Risk Model,* 29, 189-214.

[24] Chu, C.-K. and Marron, J.-S. (1991). Choosing a kernel regression estimator. With comments and a rejoinder by the authors. *Statist. Sci.*, 6 (1991), 404-436.

[25] Clarkson, D.B., Fraley, C., Gu, C.C., Ramsay, J.S.(2005). S+ functional data analysis user's guide. *Comp. Statist. Series, Springer,* New York.

[26] Crambes, C., Delsol, L., Laksaci, A. (2008). L_p errors for robust estimators in functional nonparametric regression. *Journal of Nonparametric Statistics.* 20, 573-598.

[27] Cuevas,(2014). A partial overview of the theory of statistics with functional data, *J. Stat. Plan. Inference* 147 , 1-23.

[28] Dabo-Niang, S., Laksaci, A. (2007). Propriétés asymptotiques d'un estimateur à noyau du mode conditionnel pour variable explicative fonctionnelle. (French) [Asymptotic properties of the kernel estimator of the conditional mode when the regressor is functional]. *Ann.* I.S.U.P. 51 , 27-42.

[29] Dauxois, L. Ferré, and A.F. Yao.(2001). Un modèle semi-paramétrique pour variable aléatoire hilbertienne. *C.R. Acad. Sci. Paris,* 327(I), 947-952.

[30] Dauxois, A. Pousse. (1976). Les analyses factorielles en calcul des probabilités et en statistique : essai d'étude synthétique. Thèse, Université Toulouse III.

[31] Deheuvels, P., Einmahl, J. (2000). Functional limit laws for the increments of Kaplan-Meier product limit processes and applications. *Ann. Probab.* 28, 1301-1335.

[32] Delsol, L. (2007). Régression non paramétrique fonctionnelle : expression asymptotique des moments. *Ann. I.S.U.P. Vol LI,* 3, 43-67.

[33] Delsol, L. (2009). Advances on asymptotic normality in nonparametric functional Time Series Analysis Statistics. *Statistics,* 43, 13-33.

[34] Delsol, L. (2011). Nonparametric methods for Îś-mixing functional random variables. *In The Oxford Handbook of Functional Data Analysis (Ed. F. Ferraty and Y. Romain).* Oxford University Press.

[35] Demongeot, J., Laksaci, A., Madani, F. and Rachdi, M. (2010). Local linear estimation of the conditional density for functional data. C. R. *Math., Acad. Sci. Paris,* 348, 931-934.

[36] Demongeot, J. , Laksaci, A. , Madani, F. and Rachdi, M. (2012). Functional data : Local linear estimation of the conditional density and its application Statistics C. R. *Math., Acad. Sci. Paris,* 348, 931-934.

[37] Demongeot, A. Laksaci, F. Madani, M. Rachdi.(2013). Functional data : local linear estimation of the conditional density and its application, *Statistics.* 47, 26-44.

[38] Demongeot, J., Laksaci, A., Rachdi, M. and Rahmani, S. (2014). On the local linear modelization of the conditional distribution for functional data. *Sankhya A,* 76, 328-355

[39] Deville,. J.C.(1974). Deville. Méthodes statistiques et numériques de l'analyse harmonique. *Annales de l'INSEE,* 15(Janvier-Avril), 3-97 .

[40] Droesbeke, J.J., Fichet, B., Tassi, P. (1989). Analyse statistique des durées de vie. *Economica.*

[41] El Methni, M. and Rachdi, M. (2010). Local weighted average estimation of the regression operator for functional data. *Commun. Stat., Theory and Methods, to appear.*

[42] Ezzahrioui, M., Ould-Saïd, E. (2005). Asymptotic normality of nonparametric estimators of the conditional mode for functional data. *Technical report, No.249, LMPA, Univ. Littoral Côte d'Opale.*

[43] Ezzahrioui, M., Ould-Saïd, E. (2006). On the asymptotic properties of a nonparametric estimator of the conditional mode for functional dependent data. *Preprint, LMPA No 277, Univ. du Littoral Côte d'Opale.*

[44] Fan, J. (1992). Design-adaptative nonparametric regression. *Journal of the American Statistical association,* 87, 998-1004.

[45] Fan, J., Gijbels, I. (1996). Local Polynomial Modelling and its Applications. *Monographs on Statistics and Applied Probability 66,* Chapman and Hall, London.

[46] Fan, J. and Yim, T.-H. (2004). A cross-validation method for estimating conditional densities. *Biometrika,* 91, 819-834.

[47] Ferraty, F. Laksaci, A., Vieu, P. (2005). Functional time series prediction via conditional mode estimation. *C. R. Math. Acad. Sci.* Paris. 340, 389-392.

[48] Ferraty, F. Laksaci, A., Vieu, P. (2006). Estimation of some characteristics of the conditional distribution in nonparametric functional models. Statistical Inference for Stochastic Processes. 9, 47-76.

[49] Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference*, 140, 335-352.

[50] Ferraty, A. Laksaci, A. Tadj, P. Vieu.(2011). Kernel regression with functional response. *Electronic. J. Stat.* 5 159-171.

[51] Ferraty, A. Laksaci, A. Tadj, P. Vieu.(2012). Estimation de la fonction de régression pour variable explicative et réponses fonctionnelles dépendante. *C. R. Acad. Sci. Paris, Ser.* I 350, 717-720.

[52] Ferraty, F., Rabhi, A. and Vieu, P. (2005). Conditional quantiles for dependent functional data with application to the climatic El Nino phenomenon. *Sankhya.* 67, 378-398.

[53] 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.

[54] Ferraty, I. Van Keilegom, P. Vieu.(2012). Regression when both response and predictor are functions, J. Multivar. *Anal.* 109, 10-28.

[55] Ferraty, F., Vieu, P. (2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. C. R. *Acad. Sci., Paris.* 330, No.2, 139-142.

[56] Ferraty and P. Vieu. (2002). The functional nonparametric model and application to spectrometric data. *Computational Statistics,* 17, 515-564.

[57] Ferraty, F., Vieu, P. (2004). Nonparametric models for functional data, with application in regression times series prediction and curves discrimination. *J. Nonparametric Statist.* 16, 111-125.

[58] Ferraty, F., Vieu, P. (2006). Nonparametric functional data analysis. Theory and Practice. *Springer Series in Statistics. Springer. New* York.

[59] Ferraty, F., Vieu, P. (2007). Nonparametric regression on functional data : inference and practical aspects. *Aust. N. Z. J. Stat.* 49, 267-286.

[60] Ferraty, F., Vieu, P. (2011). Kernel regression estimation for functional data. *In the Oxford Handbook of Functional Data Analysis (Ed. F. Ferraty and Y. Romain).* Oxford University Press.

[61] Ferraty, F., Sued, M. and Vieu, P. (2013). Mean estimation with data missing at random for functional covariables. *Statistics,* 47, 688-706.

[62] Ferré and A.F. Yao,(2003). Functional sliced inverse regression analysis. *Statistics,* 37, 475-488.

[63] Ferré and A.F. Yao, (2005). Smoothed functional inverse regression. *Statistica Sinica,* 15(3), 665-683.

[64] Ferré and N. Villa. (2005a). Discrimination de courbes par régression inverse fonctionnelle. *Revue de Statistique Appliquée,* LIII(1), 39-57.

[65] Florens, J. P., Larribeau, S. and Mouchart, M. (1994). Bayesian Encompassing Tests of a Unit Root Hypothesis. *Econometric Theory.* 10, 747-763.

[66] Hafen, R. P. (2010). Local regression models : Advancements, applications, and new methods. Thesis (Ph.D.).

[67] Hastie, R. Tibshirani, and A. Buja. (1994). Flexible discriminant analysis by optimal scoring. *Journal of the American Statistical Association,* 89, 1255-1270.

[68] Hastie, A. Buja, and R. Tibshirani. (1995). Penalized discriminant analysis. *Annals of Statistics,* 23, 73-102.

[69] Hastie, R. Tibshirani, and J. Friedman. (2001). The Elements of Statistical Learning : Data Mining, Inference and Prediction. *Springer-Verlag.*

[70] Helland, I.(1990). Pls regression and statistical models, Scand. *J. Statist.* 17, 97-114 .

[71] Hsing, R. Eubank. (2015). Theoretical Foundations of Functional Data Analysis, With an Introduction to Linear Operators, Wiley Series in Probability and Statistics, John Wiley and Sons, Chichester, UK.

[72] Hyndman, R. and Yao, Q. (2002). Nonparametric estimation and symmetry tests for conditional density functions. *J. Nonparametr. Stat.,* 14, 259-278.

[73] Irizarry, R. A. (2001). Local Regression With Meaningful Parameters Local Regression With Meaningful Parameters. *The American Statistician Volume 55, Issue 1,* 72-79.

[74] Ivanov.V.K.(1962). On linear problems which are not well-posed. *Soviet. Math.Docl.,* 145(2).

[75] James and T.J. Hastie, (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society, Series B,* 63, 533-550.

[76] Koziol, J.A., Green, S.B. (1976). A Cramer-Von Mises statistic for randomly censored data. *Biometrika* 63, 465-474.

[77] Kudraszow, N. and Vieu, P. (2013). Uniform consistency of kNN regressors for functional variables. *Statist. Probab. Lett.,* 83, 1863-1870.

[78] Laksaci, A. (2007). Erreur quadratique de l'estimateur á noyau de la densit é conditionnelle à variable explicative fonctionnelle. (French) [Quadratic error of the kernel estimator of conditional density when the regressor is functional.] *C. R. Math. Acad. Sci. Paris.* 345, 171-175.

[79] Laksaci, A., Lemdani, M., Ould-Sa`id, E. (2009). A generalized L1-approach for a kernel estimator of conditional quantile with functional regressors : consistency and asymptotic normality. *Statist. Probab. Lett.* 79, 1065-1073.

[80] 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.

[81] Laksaci, A., Rachdi, M. and Rahmani, S. (2013). Spatial modelization : Local linear estimation of the conditional distribution for functional data. *Spatial Statistics.* 6 1-23.

[82] Lancaster, T. (1990). The econometric analysis of the transition data. Cambridge University Press.

[83] Leurgans, R.A. Moyeed, and B.W. Silverman. (1993). Canonical correlation analysis when the data are curves. *Journal of the Royal Statistical Society, Series B,* 55, 725-740.

[84] Li. Ferré and A.F. Yao. (2005). Smoothed functional inverse regression. Statistica Sinica, 15(3), 665-683.

[85] Li. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association,* 86, 316-342.

[86] Ling, N., Liu, Y. and Vieu, P.(2015). Nonparametric regression estimation for functional stationary ergodic data with missing at random. *J. Statist. Plann. Inference,* 162, 75-87.

[87] Ling, N., Liu, Y. and Vieu, P. (2016). Conditional mode estimation for functional stationary ergodic data with responses missing at random. *Statistics,* 50, 991-1013.

[88] Loader, C. (2004). Smoothing : local regression techniques. Handbook of computational statistics, 539-563, Springer, Berlin.

[89] Mailler, H.-G. and Stadtmailler, U. (2005). Generalized functional linear models. *Ann.Stat.,* 33, No. 2, 774-805.

[90] Masry, E.(2005). Nonparametric regression estimation for dependent functional data : Asymptotic normality. *Stoch. Proc. and their Appl.* 115, 155-177.

[91] Nadaraja, E.A. (1965). On nonparametric estimation of density function and regression. *Theory of Probability and its applications* 10, 186-190.

[92] Ouassou, I. and Rachdi, M. (2010). Stein type estimation of the regression operator for functional data. Advances and Applications in Statistical Sciences, 1, No 2, 233-250.

[93] Ould-Said, E. and Cai Z. (2005). Strong uniform consistency of nonparametric estimation of the censored conditional mode function. *J. Nonparametr. Stat.,* 17, 797-806.

[94] Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* 33, 1065-1076.

[95] Perez-Gonzalez, A., Vilar-Fernandez and J. M. Gonzalez-Manteiga, W. (2009). Asymptotic properties of local polynomial regression with missing data and correlated errors. *Ann. Inst. Statist. Math.,* 61, 85-109

[96] Pezzulli and B.W. Silverman. (1993). Some properties of smoothed principal components analysis for functional data. *Computational Statistics,* 8, 1-16.

[97] Preda and G. Saporta, (2002). Régression PLS sur un processus stochastique. *Revue de statistique appliquée,* L(2).

[98] Preda and G. Saporta. (2005a). Clusterwise PLS regression on a stochastic process. *Computational Statistics and Data Analysis,* 49(1), 99-108.

[99] Preda and G. Saporta. (2005b). PLS discriminant analysis for functional data. In ASMDA *2005 proceedings,* 653-661, Brest, France.

[100] Quintela-Del-R̀io, A. (2008). Hazard function given a functional variable : Nonparametric estimation under strong mixing conditions. *J. Nonparametr. Stat.* 20, 413-430.

[101] Rachdi, M. and Sabre, R. (2000). Consistent estimates of the mode of the probability density function in nonparametric deconvolution problems. *Statist. Probab. Lett.,* 47, 105-114.

[102] Rachdi, M., Vieu, P. (2007). Nonparametric regression functional data : Automatic smoothing parameter selection. *J. Statist. Plan. Inf.* 137, 2784- 2801.

[103] Ramsay and Silverman, (1997). *Functional Data Analysis.* Springer Verlag, New York.

[104] Ramsay, J., Silverman, B.W. (2002). Applied functional data analysis ; Methods and case studies. *Springer-Verlag,* New York.

[105] Ramsay, J., Silverman, B.W.(2005). Functional Data Analysis. 2nd Edition. Springer-Verlag, New York.

[106] Rio, E. (1990). *Exponential inequalities ...* Springer-Verlag, New York.

[107] Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants.* Springer, ESAIM, Collection Mathématiques et Applications

[108] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc.Natl. Acad. Sci. USA* 42, 43-47.

[109] Rossi, B. Conan-Guez, and A. El Golli. (2004). Clustering functional data with the som algorithm. In ESANN'2004 *proceedings,* 305-312, Bruges, Belgique.

[110] Rossi, N. Delannay, B. Conan-Guez, and M Verleysen. (2005). Representation of functional data in neural networks. *Neurocomputing,* 64, 183-210.

[111] Rossi and N. Villa. (2005a). Classification in Hilbert spaces with support vector machines. In ASMDA 2005 proceedings, 635-642, Brest, France.

[112] Saporta.G.(1981). Méthodes exploratoires d'analyses des données temporelles. *Cahiers du BURO,* 37-39.

[113] Sarda, P. and Vieu, P. (2000). *Kernel Regression.* 43-70, Wiley, New York.

[114] Schimek, M. (2000). Smoothing and regression ; Approaches, Computation, and Application. *Wiley Series in Probability and Statistics,* Wiley, New York .

[115] Silverman. (1996). Smoothed functional principal components analysis by choice of norm. *Annals of Statistics,* 24, 1-24.

[116] Tihonov.(1963a). Regularization of incorrectly posed problems. *Soviet Math.Docl.,* 4, 1624-1627.

[117] Tihonov. (1963b). Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Docl.,* 4, 1036-1038.

[118] Tsybakov, A. B. (2003). Introduction à l'estimation non paramétrique. *Mathématiques et applications, 41, Springer.*

[119] Vieu, P. (1996). A note on density mode estimation. Statist. probab. Lett. 26, 297-307.

[120] Watson, G. S. (1964a). Smooth regression analysis. *Sankhya.* 26, 359-372.

[121] Watson, G. S. and Leadbetter, M. R. (1964b). Hazard analysis. I. *Biometrika.* 51, 175-184.

[122] Yates, F.,(1933). *T*he analysis of replicated experiments when the field results are incomplete. *Emporium J. Exp. Agriculture,* 1, 129-142.

[123] Youndjé, E. (1993). Estimation non-paramétrique de la densité conditionnelle par la méthode du noyau. Thèse 3eme cycle, Université de Rouen.

[124] Youndjé, E., Sarda, P. and Vieu, P. (1996). Optimal smooth hazard estimates. Test 5, 2, 379-394.

[125] Zhang. (2014). Analysis of Variance for Functional Data. *Monographs on Statistics and Applied Probability,* vol.127, CRC Press, Boca Raton, FL, USA.

ملخص

في هذه الأطروحة، نحن مهتمون أساسا في تقدير اللامعلمية المشغل تراجع متغير متغير استجابة العددية الوظيفي الذي لا يلاحظ بشكل كامل.
في البداية ، نحن نعتبر مشكلة تحليل التباين بين متغير وظيفي X ومتغير الاستجابة العددية Y غير ملاحظ بشكل كامل. نحن نستخدم منهج خطي المحلي لنموذج هذه العلاقة من خلال بناء مقدر الخطي المحلي للمشغل الانحدار عندما تظهر البيانات المفقودة في متغير الاستجابة.
في الخطوة الثانية ، نقوم بتحليل التباين المشترك بين متغيرين عشوائيين عندما لا تتم ملاحظة الإجابة بشكل كامل. نستخدم طريقة النواة المجاورة الأقرب لبناء مقدر للانحدار الخطي المحلي عندما تظهر البيانات المفقودة في متغير الاستجابة. يتم تحديد النتائج المقاربة ، من حيث اتساق النقاط والاتساق الموحد شبه الكامل ، للمقدر المُنشأ.

Résumé

Dans cette thèse, nous intéressons essentiellement à l'estimation non paramétrique de l'opérateur de régression d'une variable fonctionnelle et une variable de réponse scalaire qui n'est pas totalement observée.
Dans un premiers temps, nous considérons le problème de l'analyse de covariabilité entre une variable fonctionnelle X et une variable de réponse scalaire Y qui n'est pas totalement observée. Nous utilisons l'approche linéaire locale pour modéliser cette relation en construisant un estimateur linéaire local de l'opérateur de régression lorsque des données manquantes apparaissent dans la variable réponse.
Dans un second temps, nous analysons la covariabilité entre deux variables aléatoires lorsque la réponse une n'est pas totalement observée. Nous utilisons le noyau méthode du voisin le plus proche pour construire un estimateur de la régression linéaire locale lorsque des données manquantes apparaissent dans la variable réponse. Les résultats asymptotiques, en termes de consistances ponctuelles et de consistance uniforme presque complète, sont établis pour l'estimateur construit.

Abstract

In this thesis, we are mainly interested in the non-parametric estimation of the regression operator of a functional variable and a scalar response variable that is not fully observed.
First, we consider the problem of covariability analysis between a functional variable X and a scalar response variable Y that is not fully observed. We use the local linear approach to model this relationship by constructing a local linear estimator of the regression operator when missing data appears in the response variable.
In a second step, we analyze the covariability between two random variables when response one is not fully observed. We use the nearest neighbour's kernel method to construct an estimator of local linear regression when missing data appears in the response variable.
Asymptotic results, in terms of point consistencies and almost complete uniform consistency, are established for the constructed estimator.