Université Djillali Liabès (Algérie)
Université du Littoral Côte d'Opale (France)

THÈSE DE DOCTORAT

Spécialité : Mathématiques Appliquées

Option : Statistiques Mathématiques

Présentée par

SAID ATTAOUI

Sujet de la thèse

SUR L'ESTIMATION SEMIPARAMÉTRIQUE ROBUSTE EN STATISTIQUE FONCTIONNELLE

Soutenue publiquement le 09/12/2012 devant le jury composé de

Mr. AÏT SAÏDI A.	PROF.	Univ. Béjaia	Rapporteur
Mr. LAKSACI A.	PROF.	Univ. Sidi Bel Abbès	Directeur de thèse
Mr. OULD SAÏD E.	MCF-HDR	ULCO, France	Directeur de thèse
Mr. RACHDI M.	PROF.	Univ. de Grenoble 2	Rapporteur
Mr. SCHNEIDER D.	PROF.	ULCO, France	Examinateur
M. TEBOUNNE F.	M.C.A.	Univ. Sidi Bel Abbès	Président

Dédicace

À ma femme et ma fille.

À mes parents, mes sœurs et frères, et à tous ce qui me sont proches.

Remerciements

Je tiens en premier lieu à exprimer mes plus vifs remerciements à mes directeurs de thèse les Professeurs Elias Ould Saïd et Ali Laksaci de m'avoir accordé leur confiance pour travailler à leurs côtés, et de me donner l'occasion de développer mes compétences en statistique fonctionnelle dans un environnement de recherche très actif et pour leurs remarques et leurs commentaires qui m'ont permis d'améliorer cette thèse. Je suis reconnaissant pour les conseils qu'ils m'ont prodigué au cours de ces trois années, pour les nombreuses et fructueuses discussions que nous avons eues. En dehors de leurs apports scientifiques, je n'oublierai pas aussi de les remercier pour leurs qualités humaines, leur cordialité, leur patience infinie et leur soutien.

Mes remerciements s'adressent également aux Professeurs Ait Saidi et Rachdi pour avoir accepté de rapporter cette thèse. Je les remercie aussi pour le temps et l'attention qu'ils y ont consacrés pour la lecture attentive de cette thèse.

Je remercie sincèrement Monsieur Tebboune pour l'honneur qu'il me fait d'avoir accepté de présider ce jury.

Je tiens à remercier particulièrement Monsieur Schneider, pour avoir accepté de faire partie de ce jury malgré ses nombreuses occupations.

Enfin, je remercie tous les collègues de l'universités de Sidi Bel Abbès et l'USTO-Oran ainsi de l'université de Littoral qui m'ont aidé d'une façon directe ou indirecte, continue ou ponctuelle.

Je ne pourrais finir mes remerciements sans une pensée très personnelle aux membres de ma famille, plus particulièrement mes parents, pour leur amour, leur encouragement et leur soutien inconditionnel tout au long de mes années d'étude qui m'ont mené jusqu'au doctorat. Je garde aussi une pensée émue à la mémoire de mon grand-père et ma tante.

Je remercie profondément ma femme de faire partie de ma vie, sa présence, ses conseils, ses encouragements et son soutien constant et réconfortant dans les moments de doute sont pour moi très précieux.

Je tiens également à dire toute ma sympathie aux amis et personnes qui ont participé de près ou de loin à la bonne réalisation de la soutenance de cette thèse.

La recherche procède par des moments distincts et durables, intuition, aveuglement, exaltation et fièvre. Elle aboutit un jour à cette joie, et connaît cette joie celui qui a vécu des moments singuliers.

Albert Einstein : *Comment je vois le monde (1934).*

Cotutelle

Cette thèse a été préparée sous la convention de Cotutelle entre :

l'Université Djillali Liabès (Sidi Bel Abbès, Algérie)

et

l'Université du Littoral Côte d'Opale (France).

Accords de la commission des thèses et du conseil scientifique

Cette thèse est soutenue avec l'accord de la commission des thèses de la région Nord Pas de Calais (France)

et

l'accord du Conseil Scientifique de l'Université Djillali Liabès, Sidi Bel Abbès (Algérie).

Table des matières

Chapitre 1

Introduction

1.1 Résumé

Dans cette thèse, nous nous proposons d'étudier quelques paramètres fonctionnels lorsque les données sont générées à partir d'un modèle de régression à indice simple. Nous étudions deux paramètres fonctionnels.

Dans un premier temps nous supposons que la variable explicative est à valeurs dans un espace de Hilbert (dimension infinie) et nous considérons l'estimation de la densité conditionnelle par la méthode de noyau. Nous traitons les propriétés asymptotiques de cet estimateur dans les deux cas indépendant et dépendant. Pour le cas où les observations sont indépendantes identiquement distribuées (i.i.d.), nous obtenons la convergence ponctuelle et uniforme presque complète avec vitesse de l'estimateur construit. Comme application nous discutons l'impact de ce résultat en prévision non paramétrique fonctionnelle à partir de l'estimation de mode conditionnelle. La dépendance est modélisée via la corrélation quasi-associée. Dans ce contexte nous établissons la convergence presque complète ainsi que la normalité asymptotique de l'estimateur à noyau de la densité conditionnelle convenablement normalisée. Nous donnons de manière explicite la variance asymptotique. Notons que toutes ces propriétés asymptotiques ont été obtenu sous des conditions standard et elles mettent en évidence le phénomène de concentration de la mesure de probabilité de la variable fonctionnelle sur des petites boules.

Dans un second temps, nous supposons que la variable explicative est vectorielle et nous nous intéressons à un modèle de prévision assez général qui est la régression robuste. À partir d'observations quasi-associées, on construit un estimateur à noyau

pour ce paramètre fonctionnel. Comme résultat asymptotique on établit la vitesse de convergence presque complète uniforme de l'estimateur construit. Nous insistons sur le fait que les deux modèles étudiés dans cette thèse pourraient être utilisés pour l'estimation de l'indice simple lorsque ce dernier est inconnu, en utilisant la méthode d'M-estimation ou la méthode de pseudo-maximum de vraisemblance, qui est un cas particulier de la première méthode.

Autant que l'on sache le problème de l'estimation non paramétrique de la densité conditionnelle de modèle d'indice fonctionnel sur des données quasi-associées n'a pas été abordé. En général, l'estimation sur des données quasi-associées est récente dans la littérature statistique. Ce qui fait sans doute, l'originalité de cette thèse.

1.2 Summary

In this thesis, we propose to study some functional parameters when the data are generated from a model of regression to a single index. We study two functional parameters.

Firstly, we suppose that the explanatory variable takes its values in Hilbert space (infinite dimensional space) and we consider the estimate of the conditional density by the kernel method. We establish some asymptotic properties of this estimator in both independent and dependent cases. For the case where the observations are independent identically distributed (i.i.d.), we obtain the pointwise and uniform almost complete convergence with rate of the estimator. As an application we discuss the impact of this result in functional nonparametric prevision for the estimation of the conditional mode. In the dependent case we modelize the later via the quasi-associated correlation. Note that all these asymptotic properties are obtained under standard conditions and they highlight the phenomenon of concentration proprieties on small balls probability measure of the functional variable.

Secondly we suppose that the explanatory variable takes values in the finite dimensional space and we interest in a rather general prevision model which is the robust regression. From the quasi-associated data, we build a kernel estimator for this functional parameter. As an asymptotic result we establish the uniform almost complete convergence rate of the estimator. We point out by the fact that these two models studied in this thesis could be used for the estimation of the single index of the model when the latter is unknown, by using the method of M-estimation or the pseudo-maximum likelihood method which is a particular case of the first method.

As far as we know, the problem of estimating the conditional density in the functional single index parameter for quasi-associated data was not attacked. In general the nonparametric estimation under quasi-associated data is new in the statistical literature. What doubtless makes, the originality of this thesis.

1.3 Liste des travaux et communications

Publications

1. Attaoui, S., Laksaci, A. & Ould Saïd, E. (2011). *A note on the conditional density estimate in the single functional index model.* Statist. Probab. Lett. **81**, 45–53.
2. Attaoui, S., Laksaci, A. & Ould Saïd, E. (2012). *Some asymptotics results on the nonparametric conditional density estimate in the single index for quasi-associated Hilbertian processes.* Preprint, LMPA, *N°* 476, June 2012. Univ. du Littoral Côte d'Opale. Submitted.
3. Attaoui, S., Laksaci, A. & Ould Saïd, E. (2012). *Strong uniform consistency of the robust nonparametric regression estimator under quasi-associated vectorial processes.* Preprint, LMPA, *N°* 479, September 2012. Univ. du Littoral Côte d'Opale. Submitted.

Communications internationale

1. Attaoui, S., Laksaci, A. & Ould-Saïd, E. *A note on the conditional density estimate in the single functional index model.* Modélisation Stochastique et Statistique (MSS'10) ; Collogue International : USTHB, 21–23 Novembre 2010 (2ème edition).

1.4 Contexte bibliographique

1.4.1 Sur les variables associées et quasi-associées

Le concept d'association positive (PA) ou simplement d'association a été introduit indépendamment par Esary, Proschan et Walkup (1967) et Fortuyn, Kastelyn et Ginibre (FKG, 1971). Leur objectif était de trouver des applications dans la fiabilité des systèmes et dans la statistique mécanique, en basant sur l'inégalité de FKG. Ce concept est une généralisation de la dépendance positive introduite par Lehmenn

(1966) : un vecteur (X, Y) de variables aléatoire est dit positivement dépendant si pour tous réels x et y

$$\mathbb{P}(X \geq x, Y \geq y) - \mathbb{P}(X \geq x)\mathbb{P}(Y \geq y) \geq 0,$$

ou pour toutes fonctions croissantes f et g

$$Cov(f(X), g(Y)) \geq 0.$$

Des variables aléatoires $X_1, \ldots, X_n$ sont dites associées si

$$Cov(f(X_1, \ldots, X_n), g(X_1, \ldots, X_n)) \geq 0.$$

Notons aussi que, la notion de variables négativement associées (NA) a été introduit par Joag-Dev et Proschan (1983) ; ils ont donné de nombreuses propriétés et proposé plusieurs applications en Statistique.
Les variables aléatoires $X_1, \ldots, X_n$ sont dites négativement associées si pour tous sous-ensembles disjoints I et J de $\{1, 2, \ldots, n\}$

$$Cov(f(X_i,\, i \in I), g(X_j,\, j \in J)) \leq 0.$$

Une propriété fondamentale vérifiée par les variables associées et négativement associées est l'équivalence entre la non corrélation et l'indépendance. Cette propriété provient de l'inégalité de covariance suivante

$$|Cov(f(X_i,\, i \in I), g(X_j,\, j \in J))| \leq \sum_{i \in I} \sum_{j \in J} \left\| \frac{\partial f}{x_i} \right\|_\infty \left\| \frac{\partial g}{x_j} \right\|_\infty |Cov(X_i, X_j)|. \tag{1.1}$$

où I et J sont des sous-ensembles disjoints de $\mathbb{N}$, et f et g sont deux fonctions réelles ayant des dérivées bornées. Cette inégalité a été établie par Birkel (1988) pour des variables associées.

L'estimation nonparamétrique sous des données associées est largement étudiées dans la littérature dans les cas uni et multidimensionnel. Les premières études pour des données associées ont été faites au début des années soixantes par Harris pour des processus de percolation, puis par Lehmenn (1966) pour des données dépendantes. Divers propriétés asymptotiques pour des sommes variables aléatoires associées ont été étudiées par Newman (1980, 1984) et Birkel (1988) et plusieurs autres auteurs. Ils ont observé que, dans toute propriété asymptotique des variables aléatoires associés, la structure de covariance joue un rôle fondamental. Roussas (1991) a établi sous des conditions de régularité, la convergence ponctuelle et uniforme de l'estimateur de noyau de la densité sous données vectoriel associées. Il a donné également une caractérisation de la vitesse de convergence. Bagai et Prakasa Rao (1995) ont étudié l'estimation de la densité pour un processus associé stationnaire, en utilisant des propriétés précédemment citées. Ils ont étudié la convergence uniforme de l'estimateur à noyau, sans vitesse de convergence. Des rappels sur l'association (positive ou négative) et ces applications en statistique peuvent être trouvés dans Roussas (1999, 2000, 2001), Cai et Roussas (1999a, 1999b). Des résultats importants sur le théorème limite pour des données associées, en particulier négativement associées ont été obtenu par Bozorgnia *et al.* (1996), Patterson et Taylor (1997), Taylor et Patterson (1997), Taylor *et al.* (1999a, b), et la normalité asymptotique dans Roussas (1994). Bulinski (1996) ainsi Doukhan et Louhichi (1999) ont établi l'inégalité (1.1) pour des données négativement associées et gaussiennes respectivement. Masry (2001) a prouvé la convergence uniforme et la normalité asymptotique avec vitesse de convergence de la densité d'une variable vectorielle.

En se basant sur l'inégalité (1.1) Bulinski et Suquet (2001) ont introduit un nouveau concept de dependence appelé *quasi-association*, pour étudier certains champs aléatoires. Ce concept permet d'étudier une classe de variables aléatoires indépendantes par la non corrélation. Rappelons que cette classe contient en plus des variables gaussiennes, des variables associées et négativement associées et c'est un cas particulier de dépendance faible introduit par Doukhan et Louhichi (1999).

Une suite $(X_t)_{t\in\mathbb{Z}}$ de variables aléatoires réelles est dite quasi-associée si, pour tous sous-ensembles disjoints I et J de $\mathbb{Z}$ et toutes fonctions lipschitziennes $f : \mathbb{R}^I \to \mathbb{R}$ et $g : \mathbb{R}^J \to \mathbb{R}$, on a :

$$|Cov(f(X_i,\, i \in I),\, g(X_i,\, j \in J))| \leq \mathrm{Lip}(f)\mathrm{Lip}(g) \sum_{i\in I}\sum_{j\in J} |Cov(X_i, X_j)| \tag{1.2}$$

Doukhan et Louhichi (1999) ont défini une notion de dépendance faible qui permet de traiter le mélange et l'association par une approche unifiée.
Bulinski et Shabanovich (1998) ont montré que, toute collection de variables aléatoires positivement ou négativement associées admettant un second moment fini satisfait (1.2). Par conséquent, les champs aléatoires sont quasi-associés. Shashkin (2002), a prouvé que n'importe quel champ aléatoire gaussien $X = \{X_t,\, t \in T\}$ à valeurs dans $\mathbb{R}^p$, est quasi-associé. Un champ aléatoire gaussien à valeurs réelles est associé si et seulement si sa fonction de covariance est non negative (Pitt (1982)) et négativement associée si et seulement si $Cov(X_s, X_t) \leq 0$, $\forall s \neq t$ (Joag-Dev et Proschan (1983)).
Une suite $(X_t)_{t\in\mathbb{Z}}$ de variables aléatoires réelles est dite (λ, ϕ) faiblement dépendante s'il existe une suite $(\lambda_r)_{r\in\mathbb{N}}$ décroissante vers 0 quand t tend vers l'infini, et une fonction $\phi : \mathbb{R}_+^2 \times \mathbb{N}^2 \to \mathbb{R}$ telles que pour tout u-uplets $(s_1 \leq \ldots \leq s_u)$ et v-uplets

$(t_1 \leq \ldots \leq t_v)$ avec $s_1, \ldots, s_u \leq s_u + r \leq t_1, \ldots, t_v$, on a :

$$|Cov(f(X_{s_1}, \ldots, X_{s_u}),\, g(X_{t_1}, \ldots, X_{t_v}))| \leq \phi(\mathrm{Lip}(f), \mathrm{Lip}(g), u, v)\lambda_r, \tag{1.3}$$

pour toutes les fonctions $f, g \in \mathcal{L}_1$ qui sont définies respectivement sur $\mathbb{R}^u$ et $\mathbb{R}^v$, où $\mathcal{L}$ est l'ensemble de toutes les fonctions lipschitziennes définies sur $\mathbb{R}^n$, $n \in \mathbb{N}^*$, et $\mathcal{L}_1$ le sous-espace défini par $\mathcal{L}_1 = \{h \in \mathcal{L};\ \|h\|_\infty \leq 1\}$, $\|h\|_\infty$ désigne la norme infinie de la fonction h.

1.4.2 Sur la problématique des modèles à direction révélatrice

Depuis plusieurs années, un intérêt croissant est porté aux modèles qui incorporent à la fois des parties paramétriques et nonparamétriques. Ce type de modèles sont appelés modèle semiparamétrique. Cette considération est due en premier lieu aux problèmes dûs à la mauvaise spécification de certains modèles. Aborder un problème de mauvaise spécification de manière semiparamétrique consiste à ne pas spécifier la forme fonctionnelle de certaines composantes du modèle. Cette approche complète celle des modèles nonparamétrique, qui ne peuvent pas être utiles dans des échantillons de petite taille, ou avec un grand nombre de variables. A titre d'exemple, dans le cas de la régression classique, le paramètre important dont on suppose l'existence est la fonction de régression de Y connaissant la covariable X, notée $r(x) = \mathbb{E}(Y|X = x)$, $X, Y \in \mathbb{R}^p \times \mathbb{R}$. Pour ce modèle, la méthode nonparamétrique considère seulement des hypothèses de régularité sur la fonction r. Évidemment, cette méthode a certains inconvénients. On peut citer le problème de fléau de la dimension (curse of dimensionality). Ce problème apparaît dès que le nombre de régresseurs p augmente, la vitesse de convergence de l'estimateur nonparamétrique de r qui est supposé k fois différentiable est $O\left(n^{-k/(2k+p)}\right)$ se détériore. Le deuxième inconvénient est lié au manque de moyen de quantifier l'effet de chaque variable explicative. Pour pallier à ces inconvénients, une approche alternative est naturellement fournie par la modélisation semiparamétrique qui suppose l'introduction d'un paramètre sur les régresseurs en écrivant que la fonction de régression est de la forme

$$\mathbb{E}_\theta(Y|X) = \mathbb{E}(Y| < X, \theta >= x), \tag{1.4}$$

Les modèles ainsi définies sont connus dans la littérature sous le nom de modèles à indice simple.
Ces modèles permettent d'obtenir un compromis entre un modèle paramétrique, généralement trop restrictif et un modèle nonparamétrique où la vitesse de convergence des d'estimateurs se détériore vite en présence d'un grand nombre de variables explicatives. Dans ce domaine, différents types de modèles ont déjà été étudiés dans la

littérature : parmi les plus célèbres, on peut citer les modèles additifs, les modèles partiellement linéaires ou encore les modèles à direction révélatrice unique (single index model). L'idée de ces modèles, dans le cas de l'estimation de la densité conditionnelle ou de la régression consiste à se ramener à des covariables de dimension plus petite que la dimension de l'espace des variables, permettant ainsi de pallier au problème de fléau de la dimension. Par exemple, dans le modèle partiellement linéaire on décompose la quantité que l'on cherche à estimer, en une partie linéaire et une partie fonctionnelle. Cette dernière quantité ne pose pas de problème d'estimation puisqu'elle s'exprime en fonction de variables explicatives de dimension finie, évitant ainsi les problèmes liés au fléau de la dimension. Afin de traiter le problème de fléau de la dimension dans le cas des séries chronologies, plusieurs approches semiparamétriques ont été proposées. A titre d'exemple on peut citer : Xia et An (2002) pour le modèle d'indice. Une présentation générale de ce type de modèle est donnée dans Ichimura *et al.* (1993) où la convergence et la normalité asymptotique sont obtenues. Dans le cas des M-estimateurs, Delecroix et Hristache (1999) prouvent la consistance et la normalité asymptotique de l'estimateur de l'indice et ils étudient son efficacité. La littérature statistique sur ces méthodes est riche, citons Huber (1985) et Hall (1989) présentent une méthode d'estimation qui consiste à projeter la fonction densité ainsi que la régression sur un espace de dimension un pour se ramener à une estimation nonparamétrique pour des covariables unidimensionnelle. Cela revient exactement à estimer ces fonctions dans un modèle à indice simple.
On pourra trouver une présentation générale de ce modèle dans Ichimura (1993) où la consistence et la normalité asymptotique sont établies. Par ailleurs, des exemples montrent que les modèles à direction révélatrice unique sont particulièrement adaptés à l'étude des données de survies (voir Delecroix et Geenens (2006)). Newey et Stoker (1993) prouvent l'efficacité de ce modèle pour l'estimation de l'indice avec la méthode ADE (Average Derivative Estimation), dans le cas de l'estimation de la régression et l'estimation par pseudo- maximum de vraisemblance ou dans le cas de l'estimation de la densité conditionnelle. Dans le cas des M-estimateurs, Delecroix et Hristache (1999) prouvent la consistence et la normalité asymptotique de l'estimateur de l'indice et ils étudient son efficacité. Pour des rappels plus détaillés sur les modèles à direction révélatrice unique, le lecteur pourra se référer à Delecroix et Geenens (2006).

1.4.3 Sur les modèles semiparamétrique en statistique fonctionnelle

En générale la littérature sur la modélisation semiparamétrique pour des donnée fonctionnelles est très restreinte. A notre humble connaissance, la première étude est due à Ferré *et al.* (2001) dans le cadre d'une extension de la méthode SIR (sliced inverse régression) aux variables fonctionnelles et dont une partie utilise des outils similaires à l'approche théorique de Cardot *et al.* (1999). Pour un modèle d'indice fonctionnel de la fonction de régression Ferraty *et al.* (2003) ont obtenu des propriétés asymptotiques, lorsque l'indice est connu. Ils sont établis dans le cas i.i.d. la convergence presque complète de l'estimateur à noyau de la régression pour ce modèle. Leurs résultats ont été étendu au cas dépendent par Aït Saidi *et al.* (2005). Une étude générale est faite par Aït Saidi *et al.* (2008), lorsque l'indice fonctionnel est inconnu. Ils ont proposé un estimateur de ce paramètre, basant sur la technique de validation croisée. Dabo-Niang et Serge (2010), utilise un modèle semi-paramétrique fonctionnelle, où la variable réponse (à valeurs réelles) est expliquée par la somme d'une combinaison linéaire de composantes supposées inconnues d'une variable aléatoire multivariée et une transformation inconnu d'une variable aléatoire fonctionnelle. Leur étude est concentrée sur l'estimation paramétrique des coefficients dans la combinaison, ils utilisent la méthode non paramétrique pour supprimer l'effet de la variable explicative fonctionnelle, renforcé par l'approche des moindres carrés généralisés pour obtenir un estimateur de ces coefficients. Ils ont établi sous des hypothèses standards, la consistance ainsi que la normalité asymptotique de l'estimateur. Ils illustrent les propriétés caractéristiques des estimateurs par des simulations du type Monte Carlo. Bouraine *et al.* (2011) ont étendu le modèle de Aït Saidi *et al.* (2008) dans le cas d'indice multi-fonctionnel pour l'estimation de la fonction de régression.

1.5 Les méthodes robustes dans les modèles à direction révélatrice unique

Les M-estimateurs ont été introduits par Huber (1964), où il a proposé de généraliser l'estimateur du maximum de vraisemblance en le considérant comme un problème de minimisation d'une certaine fonction. On considère un échantillon i.i.d. de n variables aléatoires réelles $Y_1, \ldots, Y_n$ issu d'une même variable Y. Huber se place dans un cadre général d'estimateurs $\widehat{\theta}$ définis par

$$\widehat{\theta} = \arg\min_{\theta \in \mathbb{R}} \sum_{i=1}^{n} \rho(Y_i, \theta), \tag{1.5}$$

où la fonction $\rho : \mathbb{R}^2 \to \mathbb{R}$ est supposée mesurable. Différents choix de la fonction ρ conduisent aux différents estimateurs de fonctionnelles de la loi de Y (mode, médiane et quantile).

Intuitivement, la famille des M-estimateurs peut être vue comme une généralisation d'une des définitions de la moyenne. En effet, l'espérance $\mathbb{E}(Y)$ d'une variable aléatoire Y peut-être définie comme la solution du problème de minimisation suivant :

$$\mathbb{E}(Y) = \arg\min_{t\in\mathbb{R}} \mathbb{E}[(Y-t)^2].$$

Ce qui permet naturellement de proposer l'estimateur de la moyenne suivant

$$\widehat{\mathbb{E}(Y)} = \arg\min_{t\in\mathbb{R}} \frac{1}{n}\sum_{i=1}^{n}(Y_i - t)^2.$$

Huber dans son étude a distingué deux classes de M-estimateur selon la convexité ou non-convexité de la fonction ρ. Dans les deux cas cette fonction doit vérifier certaines conditions.
Pour le cas de l'estimation de la régression, dans un modèle à indice simple, Delecroix et Hristache (1999) proposent d'estimer θ par la méthode des M-estimateurs. L'idée de cette méthode est donc de définir un estimateur $\widehat{\theta}$ de θ, solution du problème de maximisation suivant

$$\theta = \arg\max_{\theta\in\Theta} \mathbb{E}[\psi(Y, \mathbb{E}[Y \mid X\theta]) \mid X = x]. \tag{1.6}$$

L'estimateur nonparamétrique $\widehat{\theta} =: \widehat{\theta}_\psi$ est, par analogie, une solution du problème de maximisation

$$\widehat{\theta}_\psi = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \psi(Y_i, \widehat{r}(X_i\theta)). \tag{1.7}$$

où ψ est une fonction définie sur $\mathbb{R}^2$ et à valeurs dans $\mathbb{R}$ satisfaisant un certain nombre de conditions et $r(X\theta) =: r_\theta(X)$ est la fonction de régression supposée inconnue. Plusieurs choix de la fonction ψ peuvent être considérés, pour plus de détail sur ces choix le lecteur pourra consulter Serfling (1980), pp. 243-245.
Delecroix et Hristache (1999) montrent que $\widehat{\theta}_\psi$ est presque sûrement consistant et asymptotiquement normal dans le cas où ψ est la log-vraisemblance d'une densité appartenant à une famille exponentielle. Pour l'estimation de θ_ψ par la méthode M-estimation on peut citer Sherman (1994), Xia et Li (1999) et Xia *et al.* (1999). Il existe d'autres méthodes d'estimation basées sur les M-estimateurs dans la littérature. A titre d'exemple la méthode de pseudo-maximum de vraisemblance qui essentiellement basée sur l'estimation de la densité conditionnelle (voir la section suivante).

1.6 La densité conditionnelle dans les modèles à direction révélatrice comme modèle préliminaire en estimation robuste

Comme c'est indiqué dans le paragraphe précédent que la méthode de pseudo-maximum de vraisemblance est un cas particulier de l'estimation robuste dans les modèle d'indice simple. En effet, Delecroix *et al.* (2003) propose d'estimer la densité conditionnelle en considérant le modèle suivant

$$f_{Y|X}(x, y) = f_\theta(x\theta, y). \tag{1.8}$$

où $f_\theta(t, y)$ représente la densité conditionnelle de Y sachant $x\theta = t$ évaluée au point y. L'idée consiste à prendre la fonction de vraisemblance

$$\prod_{i=1}^{n} f_\theta(X_i\theta, Y_i) f_X(X_i).$$

et la log-vraisemblance

$$\sum_{i=1}^{n} \log f_\theta(X_i\theta, Y_i) + \sum_{i=1}^{n} \log f_X(X_i).$$

Puisque le terme $\sum_{i=1}^{n} f_X(X_i)$ ne dépend pas de θ, l'estimateur du maximum de vraisemblance pourrait être défini, si f_θ était connue, en maximisant le premier terme $\sum_{i=1}^{n} \log f_\theta(X_i\theta, Y_i)$. Comme f_θ est inconnue, ils définissent le M-estimateur suivant :

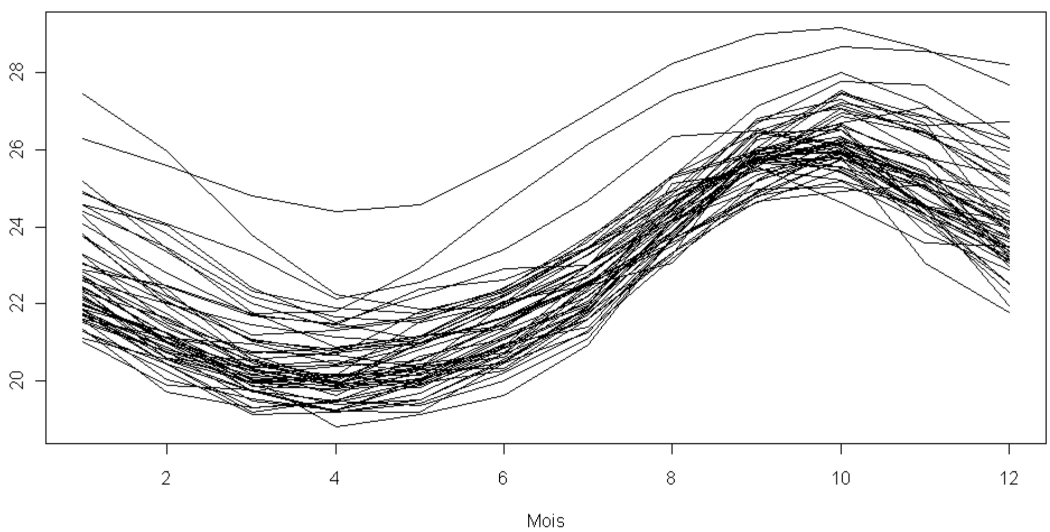
$$\widehat{\theta} = \arg\max_{\theta\in\Theta} \frac{1}{n} \sum_{i=1}^{n} \log(\widehat{f}_\theta(X_i\theta, Y_i)). \tag{1.9}$$

où $\widehat{f}_\theta$ représente l'estimateur à noyau de f_θ. Notons que dans la majorité des travaux, l'introduction d'une fonction de *trimming* dans ce type de modèle est très importante pour éviter les problèmes d'estimation quand le dénominateur de $\widehat{f}_\theta$ est nul. En effet, si on note $f_{X\theta}$ la densité de $X\theta$ alors on s'assure que l'estimateur défini en (1.9) ne prenne en compte que les X_i pour lesquels $f_{X\theta}$ est positif, sans avoir pour autant à supposer $f_{X\theta}(x) > 0$ pour tout x, y et θ. Toutefois les propriétés asymptotique reposent essentiellement sur l'estimation de la densité conditionne. Ainsi on peut dire que l'une des applications quasi-immédiate de notre travail est de construire un estimateur de l'indice dans le cas où ce dernier est inconnu en utilisant, soit les techniques robustes ou bien la méthode de pseudo-maximum de vraisemblance.

1.7 Données fonctionnelles : Applications

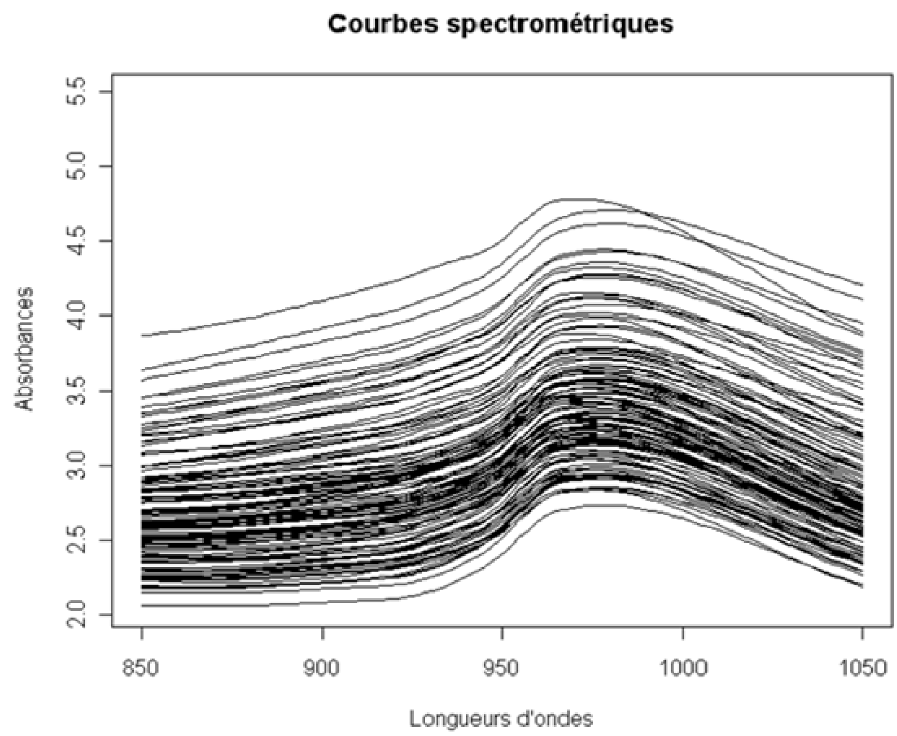
Depuis l'essor de l'analyse des données fonctionnelles, les statisticiens ne cessent de développer des applications permettant le traitement de ce type de données. Nous présentons dans ce paragraphe trois jeux de données fonctionnelles particuliers provenant de différents domaines : climatologie et chimiométrie.

- *El Niño phénomène :* Le phénomène *El Niño* a été enregistré à la fin du 19 siècle par des marins péruviens qui avaient constaté l'apparition d'un courant chaud à la période de Noël (survient une ou deux fois chaque dix ans) et qui provoque des déséquilibres climatologiques à l'échelle planètaire. Ce courant correspond à une phase plus chaude que d'habitude appelée oscillation Australe *El Niño* ou *ENSO* (sigle d'*El Niño* Southern Oscillation). Cette étude s'intéresse à un jeu de données provenant de prélèvement de température mensuelle de la surface océanique effectué depuis 1950 dans une zone située au large de nord du Pérou (aux coordonnées 0-10° Sud, 80-90° Ouest) où peut apparaître le courant marin *El Niño*. Ces données et leur description sont disponibles sur le site du centre de prévision du climat américain : *http //www.cpc.ncep.noaa.gov/data/indices/*. L'évolution des températures au cours du temps est réellement un phénomène continu. Le nombre de mesures dont nous disposons permet de prendre en considération la nature fonctionnelle des données.



57 courbes de température à la surface autour du courant marin *El Niño* par tranches de 12 mois depuis juin 1950.

– *Données spectrométriques :* On s'intéresse à des données spectrométriques couramment utilisées en chimie quantitative. On suppose qu'on dispose d'un échantillon de 215 morceaux de viandes finement hachées contenant chacun un certain taux de matière grasse. Pour chaque morceau, on observe le spectre dans le proche infrarouge pour 100 longueurs d'onde $(t_j)_{j=1,\dots,100}$ reparties entre 850 et 1050 nanomètres avec un pas constant. On observe donc pour chaque morceau i, une famille *discrète* $X_i(t_j)_{j=1,\dots,100}$ que l'on peut considérer comme une version discrétisée de la variable fonctionnelle $X_i = \{X_i(t)\}_{t\in[850;1050]}$. On parle dans ce cas de courbe spectrométrique. Ces données et leur description précise sont disponibles sur le site de StatLib (*http ://lib.stat.cmu.edu/datasets/tecator*). Les 215 spectres figurent ci-dessous montrent clairement l'aspect fonctionnel.

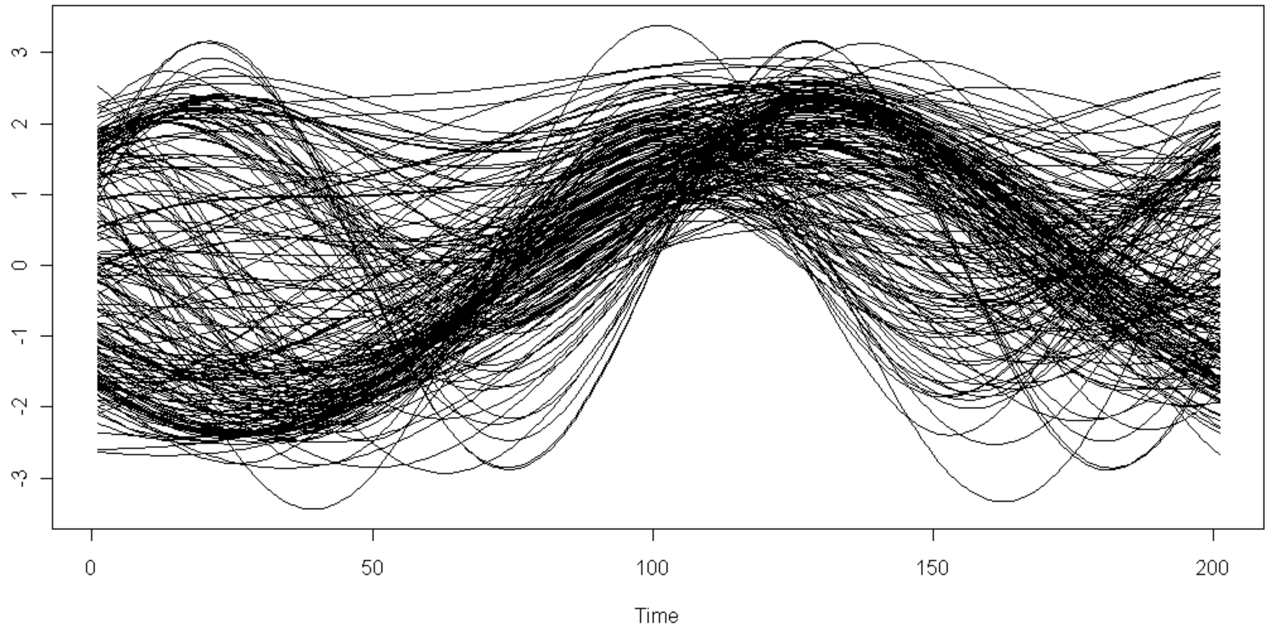


Les courbes spectrométriques : $X_{i=1,\dots,215}(t_{j=1,\dots,100})$.

– *Données quasi-associées :* On s'intéresse à des données fonctionnelles issues de mélange de deux processus stochastiques gaussiens $Z_1(t)$ et $Z_2(t)$ sur un intervalle $[-1, 1]$ définies par

$$Z_1(t) = \sqrt{-2\log(U)}\cos(2\pi(1-W)t), \; Z_2(t) = \sqrt{-2\log(1-U)}\sin(2\pi W t)$$

où U, W est une variable aléatoire uniformément distribuée sur l'intervalle $[0, 1]$. Le processus stochastiques $X(t) = Z_1(t) + Z_2(t)$ est gaussien quasi-associé. On génère un échantillon de taille 200 $\{X_i(t)\}_{i=1,...200}$ de $X(t)$, et on observe chaque variable X_i sur $(t_j)_{j=1,...,100} \in [-1, 1]$). Les courbes obtenues apparaîssent dans la figure suivante



Courbes gaussiennes quasi-associées : $X_{i=1,...,200}(t), \; t_{j=1,...,100} \in [-1, 1]$.

1.8 Plan de la thèse

Dans cette thèse on s'intéresse aux quelques paramètres fonctionnels dans les modèles à indice révélatrice. Nous traitons la densité conditionnelle et la régression robuste, en considérant deux types de corrélations à savoir le cas i.i.d. et le cas des variables quasi-associées. La variable explicative pour le premier paramètre fonctionnel qui est

la densité conditionnelle est de dimension infinie, tandis que pour le cas de régression robuste nous supposons que la variable explicative est vectorielle. La thèse se décompose en quatre chapitres.
Après avoir décrire en bref la thèse, on présentera au chapitre deux la densité conditionnelle dans le cas où les variables sont indépendantes identiquement distribuées. On construit dans ce cas un estimateur à noyau pour ce paramètre fonctionnel. Nous établissons la convergence presque complète ponctuelle et uniforme de cet estimateur. Notre étude pris en considération la concentration de la mesure de probabilité de la variable explicative via la structure de l'indice fonctionnel, dans des petites boules. L'intérêt de notre étude est de montrer comment l'estimation de la densité conditionnelle peut être utilisée pour obtenir une estimation de l'indice fonctionnel simple si ce dernier est inconnu. Plus précisément, ce paramètre peut être estimé par le pseudo-maximum de méthode de la vraisemblance qui est basé sur l'estimation préliminaire de la densité conditionnelle.
Dans le troisième chapitre on étendra nos résultats obtenus dans le deuxième chapitre au troisième sous dépendance. Le type de dépendance étudié dans ce chapitre est de quasi-association, cas spécial de dépendance faible. Nous démontrons la convergence presque complète de l'estimateur de la densité conditionnelle de l'indice fonctionnelle pour des données quasi-associées. Nous établissons aussi la normalité asymptotique de l'estimateur de la densité.
Le dernier chapitre est consacré à l'estimation robuste de la fonction de régression dans le modèle à indice révélatrice. Nous considérons le même type de dépendance que le chapitre précédant et on établit la convergence uniforme presque complète d'un estimateur à noyau pour ce paramètre fonctionnel. Nous rappelons que l'estimation de ce modèle est une étude préliminaire très utile pour l'estimation de l'indice du modèle, lorsque ce dernier est inconnu. Notons également que l'estimation de l'indice par cette dernière méthode est une généralisation de la méthode de pseudo-maximum de vraisemblance mentionnée dans les deux chapitres précédents

1.9 Résultats de la thèse

Dans cette Section on donne une brève présentation des résultats obtenus dans la thèse.

1.9.1 Notations

Soit (X, Y) un couple de variable aléatoire à valeurs dans $\mathcal{F} \times \mathbb{R}$, où $\mathcal{F}$ est un espace de Hilbert menu de la norme induite par sa produit scalaire $< \cdot, \cdot >$. Pour tout $x \in \mathcal{F}$, on suppose que la densité conditionnelle de Y sachant X existe sous structure d'indice. Une telle structure suppose que l'explication de Y sachant X se fait à travers un indice fonctionnel fixé θ. ($\theta \in \mathcal{F}$). Plus précisément, on suppose que la densité de Y sachant $X = x$, dénotée par $f(\cdot|x)$, est donnée par

$$f_\theta(y|x) =: f(y| < \theta, x >), \qquad \forall y \in \mathbb{R}.$$

Afin d'assurer l'identifiabilité de modèle, nous supposons que f est différentiable par rapport à (w.r.t.) x, et θ tel que $< \theta, e_1 >= 1$, où e_1 est le premier vecteur de la base orthonormée de $\mathcal{F}$. Clairement, sous cette condition, on a, pour tout $x \in \mathcal{F}$,

$$f_1(\cdot \,| < \theta_1, x >) = f_2(\cdot \,| < \theta_2, x >) \Rightarrow \quad f_1 \equiv f_2 \text{ et } \theta_1 = \theta_2.$$

On note par $f(\theta, y, x)$ la densité conditionnelle de Y sachant $< \theta, x >$ et on définie l'estimateur à noyau $\widehat{f}(\theta, y, x)$ de $f(\theta, y, x)$, par

$$\widehat{f}(\theta, y, x) = \frac{h_H^{-1} \sum_{i=1}^n K(h_K^{-1}(< x - X_i, \theta >))H(h_H^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}(< x - X_i, \theta >))}, \qquad \forall y \in \mathbb{R}$$

avec la convention $0/0 = 0$. Les fonctions K et H sont les noyaux et $h_K := h_{K,n}$ (resp. $h_H := h_{H,n}$) est une suite de réels positifs tend vers 0 quand n tend vers ∞.
Notons que pour $\mathcal{F} = \mathbb{R}$, notre estimateur est celle définie par Delecroix *et al.* (2003) dans le cas réel.

1.9.2 Résultats : Cas i.i.d

Dans cette partie, on suppose que les observations sont indépendantes identiquement distribuées. On suppose aussi que la densité conditionnelle $f(\theta, y, x)$ vérifie la condition de régularité. Plus précisément, pour tout point fixé x de l'espace $\mathcal{F}$, on note par $\mathcal{N}_x$ son voisinage et pour un compact $\mathcal{C}$ de $\mathbb{R}$, on a : $\forall (y_1, y_2) \in \mathcal{C}^2$, $\forall (x_1, x_2) \in \mathcal{N}_x \times \mathcal{N}_x$,

$$|f(\theta, y_1, x_1) - f(\theta, y_2, x_2)| \leq C_{\theta,x} \left(\|x_1 - x_2\|^{\beta_1} + |y_1 - y_2|^{\beta_2} \right), \quad \beta_1 > 0, \, \beta_2 > 0.$$

Sous une condition de concentration de la variable fonctionnelle X et des hypothèses techniques standards sur l'estimateur nonparamétrique définie ci-dessus, on a le résultat suivant

Theorem 1.9.1 *On a*

$$\left|\widehat{f}(\theta,y,x)-f(\theta,y,x)\right| = O(h_K^{\beta_1}) + O(h_H^{\beta_2}) + O\left(\sqrt{\frac{\log n}{nh_H\phi_{\theta,x}(h_K)}}\right), \qquad p.co.$$

où $\phi_{\theta,x}(h_K)$ est fonction de concentration de la mesure de probabilité sous structure d'indice de la variable explicative X dans la boule de centre x et de rayon h_K.
Dans le résultat suivant on a extencié le résultat de la convergence ponctuelle au cas uniforme. Le but de cette étude est motivé par le fait que la convergence uniforme est un outil indispensable pour établir toutes les propriétés asymptotiques de l'estimation de l'indice fonctionnel θ lorsqu'il est inconnu. Ainsi, en renforçant les conditions de résultat précédant par les conditions topologique suivantes : Soient $S_{\mathcal{F}}$ (resp. $\Theta_{\mathcal{F}}$, l'espace des paramètres) tels que

$$S_{\mathcal{F}} \subset \bigcup_{k=1}^{d_n^{S_{\mathcal{F}}}} B(x_k, r_n) \qquad \text{et} \qquad \Theta_{\mathcal{F}} \subset \bigcup_{j=1}^{d_n^{\Theta_{\mathcal{F}}}} B(t_j, r_n)$$

on aura le résultat

Theorem 1.9.2 *Pour tout compact $\mathcal{C}$, $S_{\mathcal{F}}$ et $\Theta_{\mathcal{F}}$*

$$\sup_{\theta\in\Theta_{\mathcal{F}}}\sup_{x\in S_{\mathcal{F}}}\sup_{y\in\mathcal{C}}\left|\widehat{f}(\theta,y,x)-f(\theta,y,x)\right| = O(h_K^{b_1})+O(h_H^{b_2})+O_{a.co.}\left(\sqrt{\frac{\log d_n^{S_{\mathcal{F}}}+\log d_n^{\Theta_{\mathcal{F}}}}{nh_H\phi(h_K)}}\right).$$

La démonstration de ces résultats, ainsi les conditions imposées seront données en détail dans le chapitre 2.

1.9.3 Résultats : Cas dépendant

Pour l'extension de nos résultats dans le cas de données dépendantes, nous modélisons la notion de dépendance faible (weak-dependance), en considérant une suite d'observations quasi-associées (voir Bulinski et Suquet (2001)). Ce type de données est exploité pour deux modèles conditionnels, un pour la densité conditionnelle présente dans le troisième chapitre, l'autre pour la régression robuste présenté dans le quatrième. Les résultats obtenus sont données dans les deux Sections suivantes

1.9.4 Résultats de cas dépendant pour la densité conditionnelle

Dans ce cas on fait une extension de résultat obtenu dans le deuxième chapitre pour des variables fonctionnelles quasi-associées. On établit dans un premier temps la convergence ponctuelle presque complète de l'estimateur de la densité conditionnelle en ajoutant aux conditions utilisées précédemment des conditions sur la concentration de la loi conjointe des couples (X_i, X_j) et par quelque conditions sur le coefficient de covariance λ_k définie dans ce cas fonctionnel par

$$\lambda_k := \sup_{s \geq k} \sum_{|i-j| \geq s} \lambda_{i,j},$$

où

$$\lambda_{i,j} = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} |Cov(X_i^k, X_j^k)| + \sum_{k=1}^{\infty} |Cov(X_i^k, Y_j)| + \sum_{l=1}^{\infty} |Cov(Y_i, X_j^l)| + |Cov(Y_i, Y_j)|.$$

X_i^k est le k ème composante de X_i définie par $X_i^k :=< X_i, e_k >$.
On montre le résultat suivant

Theorem 1.9.3

$$\left| \widehat{f}(\theta, y, x) - f(\theta, y, x) \right| = O(a_n^{\beta_1}) + O(b_n^{\beta_2}) + O_{a.co.} \left(\sqrt{\frac{\log n}{n b_n F_\theta(x, a_n)}} \right).$$

Dans un second temps on établira la normalité asymptotique de l'estimateur. On obtient le résultat

Theorem 1.9.4

(1.10)

$$\sqrt{n b_n F_\theta(x, a_n)} \left(\widehat{f}(\theta, y, x) - f(\theta, y, x) + B_H^f(\theta, y, x) b_n^2 + B_K^f(\theta, y, x) a_n + o(b_n^2) + o(a_n) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_1^2(x)).$$

où

$$\sigma_1^2(x) := \frac{C_2 f(\theta, y, x)}{C_1^2} \left(\int_{\mathbb{R}} H^2(t)\ dt \right),\ B_H^f(\theta, y, x) = \frac{1}{2} \frac{\partial^2 f(\theta, y, x)}{\partial y^2} \int t^2 H'(t) dt,$$

$$B_K^f(\theta, y, x) = a_n \Phi_0'(0) \frac{C_0}{C_1}$$

avec

$$C_0 = K(1) - \int_0^1 K'(s) \beta_\theta(x, s) ds \ \ and\ \ C_j = K(1) - \int_0^1 (K^j)'(s) \beta_\theta(x, s) ds \qquad for \quad j = 1,\ 2.$$

et $\xrightarrow{\mathcal{D}}$ *symbolise la convergence en distribution.*

1.9.5 Résultats de cas dépendant pour la régression robuste

Notation

Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathbb{R}^d \times \mathbb{R}$, où $\mathbb{R}^d$ est l'espace semi-métrique, et $d(.,.)$ désigne distance euclidienne, telle que : $\forall x_1, x_2 \in \mathbb{R}^d$, $d(x_1, x_2) = \|x_1 - x_2\|$. Pour tout $x \in \mathbb{R}^d$, on considère une fonction réelle, mesurable, notée ψ_x. Le paramètre fonctionnel à estimer ici est noté $\theta_x := \theta(x)$ est une solution par rapport à t de l'équation :

$$\Psi(x, t) = \mathbb{E}(\psi_x(Y, t)|X = x) = 0.$$

On suppose que $\forall x \in \mathbb{R}^d$, θ_x existe et vérifie $\Psi(x, \theta_x) = 0$. L'estimateur nonparamétrique de la fonction $\Psi(x, t)$ noté $\widehat{\Psi}(x, t)$ est donnée par

$$\widehat{\Psi}(x, t) = \frac{\sum_{i=1}^{n} K\Big(h_n^{-1}(x - X_i)\Big)\psi_x(Y_i, t)}{\sum_{i=1}^{n} K\Big(h_n^{-1}(x - X_i)\Big)} \quad \forall t \in \mathbb{R}$$

où K est la fonction de noyau et $h_n := h_{n,K}$ est une suite des réels positif converge à zéro pour n tend vers l'infini. L'estimateur naturel de θ_x noté $\widehat{\theta}_x$, vérifie évidemment l'équation $\widehat{\Psi}(x, \widehat{\theta}_x) = 0$.
Notons que si $\psi(y, t) = y - t$, l'estimateur $\widehat{\theta}_x$ définie l'estimateur de Nadaraya-Watson de la fonction de régression classique (voir Watson(1964)).
Sous des conditions standard, on établit la convergence uniforme presque complète de $\widehat{\theta}_x$ vers θ sur un compact $\mathcal{C}$ de $\mathbb{R}^d$ tel que : $\mathcal{C} \subset \bigcup_{k=1}^{d_n} \mathcal{B}(x_k, \tau_n)$ où $\mathcal{B}(x_k, \tau_n)$ est la boule euclidienne, centrée en x_k et de rayon $\tau_n > 0$, avec $d_n \tau_n = C$. On définie le coefficient de covariance λ_k par :

$$\lambda_k := \sup_{s \geq k} \sum_{|i-j| \geq s} \lambda_{i,j},$$

avec

$$\lambda_{i,j} = \sum_{k=1}^{d} \sum_{l=1}^{d} |Cov(X_i^k, X_j^k)| + \sum_{k=1}^{d} |Cov(X_i^k, Y_j)| + \sum_{l=1}^{d} |Cov(Y_i, X_j^l)| + |Cov(Y_i, Y_j)|.$$

et X_i^k le k-ème composante de vecteur X_i.
Sous ces conditions on a le résultat

Theorem 1.9.5 *Pour tout compact $\mathcal{C}$*

$$\sup_{x\in\mathcal{C}}\left|\widehat{\theta}_x-\theta_x\right|=O\left(h_n^p+\sqrt{\frac{\log n}{nh_n^d}}\right)\qquad a.co$$

1.10 Bibliographie générale

Bibliographie

[1] Aït Saidi, A., Ferraty, F., Kassa, R., (2005). *Single functional index model for a time series.* R. Roumaine Math. Pures et Appl. **50**, 321–330.

[2] Aït Saidi, A., Ferraty, F., Kassa, R., Vieu, P., (2008). *Cross-validated estimation in the single functional index model.* Statistics. **42**, 475–494.

[3] Ango Nze, P., Bühlmann, P. & Doukhan, P. (1993) *Functional estimation for mixing time series.* C. Acad. Sci. Paris **317**, 405–408.

[4] Ango Nze, P., Bühlmann, P. & Doukhan, P. (2002) *Weak dependence beyond mixing and asymptotics for nonparametric regression.* Ann. Statist. **30**, N. **2**.

[5] Attouch, M., Laksaci, A. & Ould-Saïd, E. (2010). *Asymptotic Distribution of Robust Estimator for Functional Nonparametric Models.* Communications in Statistics : Theory and Methods. **38**, 1317–1335.

[6] Azzedine, N., Laksaci, A. & Ould-Saïd, E. (2008). *On robust nonparametric regression estimation for functional regressor.* Statist. Probab. Lett.**78**, 3216–3221.

[7] Bagai, L. & Prakasa Rao, L. S. (1995). *Kernel-type density and failure rate estimation for associated sequences.* Ann. Inst. Statist. Math. **47**, 253–266.

[8] Billingsley, P. (1995). *Probability and Measure.* 3rd edition, Wiley, NY.

[9] Birkel, T. (1988). *On the convergence rate in Central Limit Theorem for associated processes.* Ann. Probab., **16**, 1685–1698.

[10] Boente, G. & Fraiman, R. (1989). *Robust nonparametric regression estimation.* J. Multivariate Anal. **29**, 180–198.

[11] Bouraine M., Ait Saidi, A., Ferraty, F., & Vieu Ph. (2010). *Choix optimal de l'indice multi-fonctionnel : Méthode de validation croisée.* Rev. Roumaine de Math. Pures et Appliquées Vol 55, **5**, 355–367.

[12] Bozorgnia, A., Patterson, R.F., & Taylor, R.L. (1996). *Limit theorems for negatively dependent random variables.* Proceedings of the First World Congress of Nonlinear Analysis, Tampa, Florida, August 19–26, 1992.

[13] Bulinski, A.V. (1996). *On the convergence rates in the CLT for positively and negatively dependent random fields.* In Probability Theory and Mathematical Statistics (LA. Ibragimov and A.Yu. Zaitsev, Eds), 3–14. Gordon and Breach Publishers, Amsterdam.

[14] Bulinski A. & Suquet, C. (2001). *Normal approximation for quasi-associated random fields.* Statist. Probab. Lett. **54**, 215–226.

[15] Bulinski A. & Shashkin, A. (2007). *Limit theorems for associated random fields and related systems.*

[16] Cai, Z. W. & Roussas, G. G. (1999a). *Weak convergence for smooth estimator of a distribution function under negative association.* Stochast. Analys. and Applic. **17**, 245–268.

[17] Cai, Z. W. & Roussas, G. G., (1999b). *Berry–Esseen bounds for a smooth estimate of a distribution function under association.* J. Nonparametric Statist. **11,** 79–106

[18] Cardot, H., Ferraty, F. & Sarda, P. (1999). *Functional Linear Model.* Statist. & Prob. Letters, **45**, 11–22.

[19] Cardot, H., Ferraty, F., & Sarda, P. (2003). *Spline Estimators for the Functional Linear Model.* Statistica Sinica **13**, 571–591.

[20] Cardot, H., Ferraty, F. & Sarda, P. (2004). *Testing for no effect in functional linear regression models, some computational approaches.* Comm. Statist. Simulation Comput. **33**, 171–199.

[21] Cardot, H. Crambes, C. & Sarda, P. (2005). *Quantiles regression when the covariates are functions.* J. of Nonpar. Statist, **17**, 541–856.

[22] Collomb, G. & Härdle, W. (1986). *Strong uniform convergence rates in robust nonparametric time series analysis and prediction : Kernel regression estimation from dependent observations.* Stoch. Proc. Appl. **23**, 77–89.

[23] Crambes, C., Delsol, L. & Laksaci, A. (2009). *Robust nonparametric estimation for functional data.* In : S, Dabo-Niang, F, Ferraty (Ed.) Functional and Operatorial Statistics. Contribution to statistics. Physica-Verlag Heidelberg.

[24] Dabo-Niang, D. & Serge, G. (2010). *Functional semiparametric partially linear model with autoregressive errors.* J. Multivariate Analysis, Vol 101, **2** 307–315.

[25] Delecroix, M, Härdle, W. & Hristache, M., (1999). *M-estimateurs semiparamétriques dans les modèles à direction révélatrice unique.* Bull. Belg. Math. Soc. Simon Stevin, **6**, No 2, 161–185.

[26] Delecroix, M, Härdle, W.& Hristache, M., (2003). *Efficient estimation in conditional single-index regression.* J. Multivariate Anal. **86**, 213–226.

[27] Delecroix, M, Geenens, G. (2006). *A survey about single-index models theory.* J. Statist. Plann. Inference, **1** No 2, 203–230.

[28] Delecroix, M, Geenens, G. (2006). *On semiparametric M-estimation in single index regression.* J. Statist. Plann. Inference, **136** No 3, 730–769.

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

[30] Dedecker, J. & Prieur, C. (2004). *Coupling for τ-dependent sequences and applications.*, J. Theor Proba. **4**, 861-885.

[31] Dedecker, J., Doukhan, P., Lang, G., León R., J. R., Louhichi, S., and Prieur, C. (2007). *Weak dependence : with examples and applications.*, Lecture Notes in Statistics.

[32] Dedecker, J. & Prieur, C. (2004). *Coupling for τ-dependent sequences and applications.*, J. Theor Proba. **4**, 861-885.

[33] Doukhan, P. (1994). *Mixing : Properties and Examples.*, Lecture Notes in Statistics.

[34] Doukhan, P., Louhichi, S. (1999). *A new weak dependence condition and applications to moment inequalities.* Stoch. Proc. Appl. **84**, 313–342.

[35] Doukhan, P., Lang, G., Surgailis D., Tyssière, G. (2010). *Dependence in Probability and Statistics.*, Lecture Notes in Statistics.

[36] Esary, J., Proschan, F. & Walkup, D. (1967). *Association of random variables with applications.* Ann. Math. Statist. **38**, 1466–1476.

[37] Ezzahrioui, M., Ould Saïd, E., (2010). *Some asymptotic results of a nonparametric conditional mode estimator for functional time series data.* Statist. Neerlandica, Vol 64, **2**, pages 171–201.

[38] Ferraty, F., Peuch, A., Vieu, P., (2003). *Modèle à indice fonctionnel simple.* C. R. Mathématiques Paris **336**, 1025–1028.

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

[40] Ferraty, F. Mas, A. & Vieu, P., (2007). *Advances in nonparametric regression for functional variables.* Aust. and New Zeal. J. of Statist., **49**, 1-20.

[41] Ferraty, F., Laksaci, A., Vieu, Ph., (2006). *Estimating some characteristics of the conditional distribution in nonparametric functional models.* Statist. Inf. for Stoch. Proc. **9**, 47–76.

[42] Ferraty, F., Laksaci, A., Tadj, A., Vieu, Ph., (2009). *Rate of uniform consistency for nonparametric estimates with functional variables.* J. Statist. Plann. and Inf. 140, 335–352.

[43] Ferraty, F., Park, J., Vieu, Ph. (2011). *Functional average derivative regression.* Priprint.

[44] Ferraty, F., Park, J., Vieu, Ph. (2011). *Estimation of a Functional Single Index Model* In : F, Ferraty (Ed.) Recent advences in functional data analysis and related topics. Contribution to statistics. Physica.

[45] Ferré, L, Dauxois, J. & Yao, A. F. (2001). *Un modèle semi-paramétrique pour variables aléatoires hilbertiennes (in french).* C. R. Acad. Sci. Paris Série. I Math. **333**, No. 10, 947–952.

[46] Fortuyn, C., Kastelyn, P. & Ginibre, J. (1971). *Correlation inequalities in some partially ordered sets.* Comm. Math. Phys. **74**, 119–127.

[47] Kallabis, R. S & Neumann, M. H. (2006). *An exponential inequality under weak dependence.* Bernoulli **12**, 333–350.

[48] Hall, P. (1989). *On projection pursuit regression.* Ann. Statist., **17** No 2 : 573–588.

[49] Hall, P., Müller, H. G. & Yao, F. (2009). *Estimation of functional derivatives.* Ann. Statist. **37**, 3307–3329.

[50] Härdle, W. & Tsybakov, A. (1988). *Robust nonparametric regression with simultaneous scale curve estimation.* Ann. Statist.**16**, 120–135.

[51] Härdle, W., Hall, P., Ichumira, H., (1993). *Optimal smoothing in single-index models,* Ann. Statist. **21**, 157–178.

[52] Harris, T. E. (1960). *A lower bound for the critical probability in a certain percolation process.* Proc. Cambridge Philos. Soci. **56**, 13–20.

[53] Hristache, M., Juditsky, A., Spokoiny, V. (2001). *Direct estimation of the index coefficient in the single-index model.* Ann. Statist. **29**, 595–623.

[54] Huber, P. J., (1964). *Robust estimation of a location parameter.* Ann. Math. Statist. **35**, 73–101.

[55] Huber, P. J. (1985). *Projection pursuit.* Ann. Statist., **13** No 2 : 435–475.

[56] Jong-Dev, K., & Proschan, F. (1983). *Negative association of random variables, with applications.* Ann. Statist. **11**, 286–295.

[57] Lehmann, E. L. (1966). *Some concepts of dependence.* Ann. Math. Statist., **37**, 1137–1153.

[58] Leibscher, E. (2001). *Estimation of the density and the regression function under mixing conditions.* Statist. & Decisions, **19**, 9–26.

[59] Li, W. V. & Shao, Q.M. *Gaussian processes : Inequalities, Small Ball Probabilites and Applications.* In Stochastic processes : Theory and Methods, Ed. C.R. Rao and D. Shanbhag.

[60] Masry, E. (1986). *Recursive probability density estimation for weakly dependent stationary processes* IEEE. Trans. Inform. Theory **32**, 254–267.

[61] Masry, E. (2002). *Multivariate probability density estimation for associated processes : strong consistency and rates.* Statist. Probab. Lett. **58**, 205–219.

[62] Newey, W. K. & Stoker, T. M. (1993). *Efficiency of weighted average derivative estimators and index models.* Econometrica, Vol 61, **2**, 1199–1223.

[63] Newman, C. M. (1980). *Normal fluctuations and the FKG inequalities.* Comm. Math. Phys. Vol 74, **2**, 119–128.

[64] Newman, C. M. (1984). *Asymptotic independence and limit theorems for positively and negatively dependent random variables :* in Inequalities in Statistics and Probability, IMS Lect. Notes–Monographs Series,**5**, 127–140.

[65] Patterson, R. F., & Taylor, R. L. (1997). *Strong laws of large numbers for negatively dependent random elements.* Nonlinear Analysis, Theory, Methods and Applications **30**, 4229–4235.

[66] Ramsay, J. O. (1982). *When the data are functions.* J. Psychometrika. **47**, 379–396.

[67] Ramsay, J. O & Dalzell, C. (1991). *Some tools for functional data analysis.* J. R. Statist. Soc. B. **53** ; 539–572.

[68] Ramsay, J. O & Silverman, B. W. (1997). *Functional data analysis.* Springer-Verlag, New York.

[69] Ramsay, J. O & Silverman, B. W. (2002). *Applied functional data analysis : Methode and Case Studies.* Springer-Verlag, New York.

[70] Ramsay, J. O & Silverman, B. W. (2005). *Functional data analysis.* Seconde edition

[71] Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants (in french).* Springer, Mathématiques & Applications, **31**.

[72] Roussas, G. G. (1991). *Kernel estimates under association : strong uniform consistency.* Statist. Probab. Lett. **12**, 215–224.

[73] Roussas, G. G. (1994). *Asymptotic normality of random fields of positively or negatively associated processes.* J of Multivar. Analys. **50**, 152–173.

[74] Roussas, G. G. (1999). *Positive and negative dependence with some statistical applications.* In Asymptotics, Nonparametrics and Time Series (S. Ghosh, Ed.), 757–788. Marcell Dekker, Inc., New York.

[75] Roussas, G. G. (2000) *Prediction under association.* In : Limnios, N., Nikulin, M.S. (Eds.), Recent Advances in Reliability Theory : Methodology, Practice and Inference. BirkhSauser, Boston.

[76] Roussas, G. G. (2001) *Fixed design regression under association,* Lecture Notes-Monograph Series. 71–89.

[77] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley.

[78] Sherman, R. P. (1994). *Maximal inequalities for degenerate U-processes with applications to optimization estimators.* Ann. Statist., **22**, N 1 : 439–459.

[79] Taylor, R. L., Patterson, R. F., & Bozorgnia, A. (2001). *Weak laws of large numbers for arrays of rowwise negatively dependent random variables.* Vol 14, **3**, 227–236.

[80] Taylor, R. L., Patterson, R. F., & Bozorgnia, A. (2002). *A strong law of large numbers for arrays of rowwise negatively dependent random variables.* Vol 20, **3**, 643–656.

[81] Xia, X, Tong, H & Li, W. K. (1999). *On extended partially linear single-index models.* Biometrika, **86**, 831–842.

[82] Xia, X & Li, W. K. (1999). *On the estimation and testing of functional-coefficient linear models.* Stat. Sinica, **9**, 735–757

[83] Xia, X & An H. Z. (2002). *An projection pursuit autoregression in time series.* J. of Time Series Analysis, Vol 20, **6**, 693–714.

[84] Youndjé, E., (1993). *Estimation non paramétrique de la densité conditionnelle par la méthode du noyau.* Thèse de Doctorat, Université de Rouen.

Chapitre 2

Single index model : i.i.d process case

Ce chapitre a fait l'objet d'une publication dans *Statistic and Probability Letters,* **81***, (2011), 45–53*

A note on the conditional density estimate in single functional index model

Said Attaoui[1], Ali Laksaci[2] and Elias Ould Said[3]

Abstract In this paper, we consider estimation of the conditional density of a scalar response variable Y given a Hilbertian random variable X when the observations are linked with a single-index structure. We establish the pointwise and the uniform almost complete convergence (with the rate) of the kernel estimate of this model. As an application, we show how our result can be applied in the prediction problem via the conditional mode estimate. Finally, the estimation of the functional index via the pseudo-maximum likelihood method is also discussed but not attacked.

Key words : Conditional single-index, conditional density, nonparametric estimation, semiparametric estimation, semi-metric choice.

2.1 Introduction

For the past two decades, the single-index model, a special case of projection pursuit regression, has proven to be an efficient way of coping with the high dimensional problem in nonparametric regression. Here we deal with single-index modeling when the explanatory variable is functional. More precisely, we consider the problem of estimating the conditional density of a real variable Y given a functional variable X when the explanation of Y given X is done through its projection on one functional direction.

The conditional density plays an important role in nonparametric prediction, because the several prediction tools in nonparametric statistic, such as the conditional mode, the conditional median or the conditional quantiles, are based on the preliminary estimate of this functional parameter. Nonparametric estimation of the conditional density has been widely studied, when the data is real The first related result in nonparametric functional statistic was obtained by Ferraty *et al.* (2006). They established the almost complete consistency in the independent and identically distributed (i.i.d.)

1. Université Djillali Liabes, Sidi Bel Abbès
2. Université Djillali Liabes, Sidi Bel Abbès
3. Université Littoral Côte d'Opale

random variables of the kernel estimator of the conditional probability density. These result have been extend to dependent data by Ezzahrioui and Ould Saïd (2010).

The single-index approach is widely applied in econometrics as a reasonable compromise between nonparametric and parametric models. Such kind of modelization is intensively studied in the multivariate case. Without pretend to exhaustivity, we quote for example Härdle *et al.* (1993), Hristache *et al.* (2001). Based on the regression function, Delecroix *et al.* (2003) studied the estimation of the single-index and established some asymptotic properties. The literature is strictly limited in the case where the explanatory variable is functional (that is a curve). The first asymptotic properties in the fixed functional single-model were obtained by Ferraty *et al.* (2003). They established the almost complete convergence, in the i.i.d. case, of the link regression function of this model. Their results were extended to dependent case by Aït Saidi *et al.* (2005). Aït Saidi *et al.* (2008) studied the case where the functional single-index is unknown. They proposed an estimator of this parameter, based on the cross-validation procedure.

The goal of this paper is to study the functional single-index model via its conditional density estimation. This extend, in different way, the works of Delecroix *et al.* (2003) and Ferraty *et al.* (2006). We construct an estimator of the conditional density of Y given X by a kernel method and we prove, under general conditions, its pointwise and uniform almost complete convergence (with rate). In practice, this study has great importance, because, it permit us to construct a prediction method based on the conditional mode estimator. Moreover, in the case where the functional single index is unknown, our estimate can be used to estimate this parameter via the pseudo-maximum likelihood estimation method. Noting that the estimation of the functional single-index has great interest on the semi-metric choice in nonparametric functional data analysis but it has been not attacked in this paper.

The paper is organized as follows. We present our model in Section 2. In Section 3 we introduce notations, assumptions and state the main results. Section 4 is devoted to some discussions and comments on the impact of our study in the prediction problem. The proofs of the results are relegated to the last section.

2.2 Model

Let (X, Y) be a couple of random variables taking its values in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a Hilbertian space with scalar product $< \cdot, \cdot >$. Let $(X_i, Y_i)_{1 \leq i \leq n}$, be n copies of independent vectors each having the same distribution as (X, Y). Assume that the conditional density of Y given X exists and has a single-index structure. Such structure supposes that the explanation of Y given X is done through a fixed functional index θ in $\mathcal{F}$. More precisely, we suppose that the conditional density of Y given $X = x$, denoted by $f(\cdot|x)$, is given by

$$\forall y \in \mathbb{R} \qquad f_\theta(y|x) =: f(y| < \theta, x >).$$

To ensure the identifiability of the model, we consider the same conditions as those in Ferraty *et al.* (2003) on the regression operator. In other words, we assume that f is differentiable with respect to (w.r.t.) x, and θ such that
$< \theta, e_1 >= 1$, where e_1 is the first vector of an orthonormal basis of $\mathcal{F}$. Clearly, under this condition, we have, for all $x \in \mathcal{F}$,

$$f_1(\,\cdot\,| < \theta_1, x >) = f_2(\,\cdot\,| < \theta_2, x >) \Rightarrow \quad f_1 \equiv f_2 \text{ and } \theta_1 = \theta_2.$$

In what follows we denote by $f(\theta, \cdot, x)$, the conditional density of Y given $< \theta, x >$ and we define the kernel estimator $\widehat{f}(\theta, y, x)$ of $f(\theta, y, x)$ by

$$\widehat{f}(\theta, y, x) = \frac{h_H^{-1} \sum_{i=1}^n K(h_K^{-1}(< x - X_i, \theta >))H(h_H^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}(< x - X_i, \theta >))}, \qquad \forall y \in \mathbb{R}$$

with the convention $0/0 = 0$. The functions K and H are kernels and $h_K := h_{K,n}$ (resp. $h_H := h_{H,n}$) is a sequence of positive real numbers which goes to zero as n tends to infinity. Note that a similar estimate was already introduced in the special case where X is a real random variable by Delecroix *et al.* (2003).

2.3 Main results

All along the paper, when no confusion is possible, we will denote by C and C' some strictly positive generic constants. In the following, we put, for any $x \in \mathcal{F}$, and $i = 1, \dots, n$,
$K_i(\theta, x) := K(h_K^{-1} < x - X_i, \theta >)$ and, for all $y \in \mathbb{R}$, $H_i(y) := H(h_H^{-1}(y - Y_i))$

2.3.1 Pointwise almost complete convergence

In he following x is a fixed point in $\mathcal{F}$, $\mathcal{N}_x$ is a fixed neighborhood of x and $\mathcal{C}$ is a fixed compact subset of $\mathbb{R}$. In order to establish the almost complete (a.co.) [4]convergence

4. We say that a sequence Z_n converges *a.co.* to Z if and only if, for any $\epsilon > 0, \sum_n \mathbb{P}(|Z_n - Z| > \epsilon) < \infty$.

of our estimate we need the following assumptions :

(H1) $\mathbb{P}(|<X-x,\theta>|<h) =: \phi_{\theta,x}(h) > 0$.

(H2) The conditional density $f(\theta,y,x)$ satisfies the Hölder condition, that is : $\forall(y_1,y_2)\in\mathcal{C}^2$, $\forall(x_1,x_2)\in\mathcal{N}_x\times\mathcal{N}_x$,

$$|f(\theta,y_1,x_1)-f(\theta,y_2,x_2)| \leq C_{\theta,x}\left(\|x_1-x_2\|^{b_1}+|y_1-y_2|^{b_2}\right), \quad b_1>0,\, b_2>0.$$

(H3) K is a positive bounded function with support $[-1,1]$.

(H4) H is bounded function, such that

$$\int H(t)dt = 1, \qquad \int |t|^{b_2}H(t)dt < \infty \text{ and } \int H^2(t)dt < \infty,$$

(H5) The bandwidths h_K and h_H satisfy

$$\lim_{n\to\infty} h_H = 0,\ \lim_{n\to\infty} h_K = 0 \ \text{ and } \ \lim_{n\to\infty}\frac{\log n}{nh_H\phi_{\theta,x}(h_K)} = 0.$$

Comments on the assumptions

Our assumptions are very standard for this kind of model. Assumptions (H1) and (H3) are the same as those used in Ferraty *et al.* (2003). Assumption (H2) is a regularity conditions which characterize the functional space of our model and is needed to evaluate the bias term of our asymptotic results. Assumptions (H4) and (H5) are technical conditions and are also similar to those in Ferraty *et al.* (2006).

Our first main result is given in the following theorem.

Theorem 2.3.1 *Under Assumptions (H1)-(H5), and for any fixed y, we have, as n goes to infinity*

$$\left|\widehat{f}(\theta,y,x)-f(\theta,y,x)\right| = O(h_K^{b_1})+O(h_H^{b_2})+O_{a.co.}\left(\sqrt{\frac{\log n}{nh_H\phi_{\theta,x}(h_K)}}\right). \tag{2.1}$$

In the particular case, where the real random variable $Z :=<X,\theta>$ has continuous density we can reformulate the general result given in Theorem 2.3.1 in the following way

Corollary 2.3.1 *Under Assumptions (H2)-(H5) and if the density of Z does not vanish and for any fixed y, we have, as n goes to infinity*

$$\left|\widehat{f}(\theta,y,x)-f(\theta,y,x)\right| = O(h_K^{b_1})+O(h_H^{b_2})+O_{a.co.}\left(\sqrt{\frac{\log n}{nh_Hh_K}}\right).$$

Proof of Theorem 2.3.1.
The proof is based on the following decomposition

$$\begin{aligned}\widehat{f}(\theta,y,x)-f(\theta,y,x) &= \frac{1}{\widehat{f}_D(\theta,x)}\left\{\left(\widehat{f}_N(\theta,y,x)-\mathbb{E}\left[\widehat{f}_N(\theta,y,x)\right]\right)+\left(\mathbb{E}\left[\widehat{f}_N(\theta,y,x)\right]-f(\theta,y,x)\right)\right\} \\ &\quad -\frac{f(\theta,y,x)}{\widehat{f}_D(\theta,x)}\left\{\widehat{f}_D(\theta,x)-1\right\}\end{aligned}$$

where

$$\widehat{f}_N(\theta,y,x)=\frac{1}{nh_H\mathbb{E}\left[K_1(\theta,x)\right]}\sum_{i=1}^{n}K_i(\theta,x)H_i(y), \qquad \widehat{f}_D(\theta,x)=\frac{1}{n\mathbb{E}\left[K_1(\theta,x)\right]}\sum_{i=1}^{n}K_i(\theta,x).$$

So, the proof is a direct consequence of the following results

Lemma 2.3.1 *(see Aït-Saidi et al., 2005) Under Assumptions (H1), (H3) and (H5), as n goes to infinity, we have*

$$\left|\widehat{f}_D(\theta,x)-1\right|=O_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_{\theta,x}(h_K)}}\right). \tag{2.2}$$

Furthermore, we have

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\widehat{f}_D(\theta,x)\right|\leq 1/2\right)<\infty. \tag{2.3}$$

Lemma 2.3.2 *Under Assumptions (H1)-(H5), as n goes to infinity, we have*

$$\left|\widehat{f}_N(\theta,y,x)-\mathbb{E}\left[\widehat{f}_N(\theta,y,x)\right]\right|=O_{a.co.}\left(\sqrt{\frac{\log n}{nh_H\phi_{\theta,x}(h_K)}}\right). \tag{2.4}$$

Lemma 2.3.3 *Under Assumptions (H1), (H2) and (H4), as n goes to infinity, we have*

$$\left|\mathbb{E}\left[\widehat{f}_N(\theta,y,x)\right]-f(\theta,y,x)\right|=O(h_K^{b_1})+O(h_H^{b_2}). \tag{2.5}$$

2.3.2 Uniform almost complete convergence

This section is devoted to derivation of the the uniform version of Theorem 2.3.1. The study of the uniform consistency is motivated by the fact that the latter is an indispensable tool for studying the asymptotic properties of all estimate of the functional index if is unknown. Noting that, in the multivariate case, the uniform consistency is a standard extension of the pointwise one, however, in our functional case, it requires some additional tools and topological conditions (see Ferraty *et al.*, 2009, for more discussion on the uniform convergence in nonparametric functional statistics). Thus, in addition to the conditions introduced in the previous section, we need the following ones. Firstly, we suppose that $\mathcal{C}$ is subset compact of $\mathbb{R}$ and $S_{\mathcal{F}}$ (resp. $\Theta_{\mathcal{F}}$, the space of parameters) are such that

$$S_{\mathcal{F}} \subset \bigcup_{k=1}^{d_n^{S_{\mathcal{F}}}} B(x_k, r_n) \qquad \text{and} \quad \Theta_{\mathcal{F}} \subset \bigcup_{j=1}^{d_n^{\Theta_{\mathcal{F}}}} B(t_j, r_n) \tag{2.6}$$

with x_k (resp. t_j) $\in \mathcal{F}$ and $d_n^{S_{\mathcal{F}}}, d_n^{\Theta_{\mathcal{F}}}$ are sequences of positive real numbers which tend to infinity as n goes to infinity.

Furthermore, we need the following assumptions :

(U1) There exists a differentiable function $\phi(.)$ such that $\forall x \in S_{\mathcal{F}}$, and $\forall \theta \in \Theta_{\mathcal{F}}$,

$$0 < C\phi(h) \leq \phi_{\theta,x}(h) \leq C'\phi(h) < \infty \ \text{ and } \ \exists \eta_0 > 0,\ \forall \eta < \eta_0,\ \phi'(\eta) < C,$$

(U2) The conditional density is such that $\forall (y_1, y_2) \in \mathcal{C} \times \mathcal{C},\ \forall (x_1, x_2) \in S_{\mathcal{F}} \times S_{\mathcal{F}}$, and $\forall \theta \in \Theta_{\mathcal{F}}$,

$$|f(\theta, y_1, x_1) - f(\theta, y_2, x_2)| \leq C\left(\|x_1 - x_2\|^{b_1} + |y_1 - y_2|^{b_2}\right),$$

(U3) The kernel K satisfy (H3) and Lipschitz's condition holds

$$|K(x) - K(y)| \leq C\|x - y\|,$$

(U4) H is a bounded Lipschitz continuous function, such that

$$\int H(t)dt = 1, \qquad \int |t|^{b_2} H(t)dt < \infty \text{ and } \int H^2(t)dt < \infty,$$

(U5) For some $\gamma \in (0,1)$, $\lim_{n\to+\infty} n^{\gamma} h_H = \infty$, and for $r_n = O\left(\frac{\log n}{n}\right)$ the sequences $d_n^{S_{\mathcal{F}}}$ and $d_n^{\Theta_{\mathcal{F}}}$ satisfy :

$$\frac{(\log n)^2}{n h_H\, \phi(h_K)} < \log d_n^{S_{\mathcal{F}}} + \log d_n^{\Theta_{\mathcal{F}}} < \frac{n h_H\, \phi(h_K)}{\log n},$$

and

$$\sum_{n=1}^{\infty} n^{(3\gamma+1)/2}(d_n^{S_\mathcal{F}} d_n^{\Theta_\mathcal{F}})^{1-\beta} < \infty, \text{ for some } \beta > 1.$$

Remark 2.3.1 *Note that Assumptions (U1) and (U2) are, respectively, the uniform version of (H1) and (H2). Assumption (U4) is condition (H4) added by Lipschitz condition. Assumptions (U1) and (U5) are linked with the the topological structure of the functional variable. For examples of subsets such as (2.6) see Ferraty et al. (2009).*

Theorem 2.3.2 *Under Assumptions (U1)-(U5), we have, as n goes to infinity*
(2.7)

$$\sup_{\theta\in\Theta_\mathcal{F}}\sup_{x\in S_\mathcal{F}}\sup_{y\in\mathcal{C}}\left|\widehat{f}(\theta,y,x)-f(\theta,y,x)\right| = O(h_K^{b_1})+O(h_H^{b_2})+O_{a.co.}\left(\sqrt{\frac{\log d_n^{S_\mathcal{F}}+\log d_n^{\Theta_\mathcal{F}}}{nh_H\phi(h_K)}}\right).$$

In the particular case, where the functional single-index is fixed we get the following result.

Corollary 2.3.2 *Under Assumptions (U1)-(U5), we have, as n goes to infinity*

$$\sup_{x\in S_\mathcal{F}}\sup_{y\in\mathcal{C}}\left|\widehat{f}(\theta,y,x)-f(\theta,y,x)\right| = O(h_K^{b_1})+O(h_H^{b_2})+O_{a.co.}\left(\sqrt{\frac{\log d_n^{S_\mathcal{F}}}{nh_H\phi(h_K)}}\right).$$

Proof. Clearly Theorem 2.3.2 and Corollary 2.3.2 can be deduced from the following intermediate results which are uniform version of Lemmas **3.1-3.3**.

Lemma 2.3.4 *Under Assumptions (U1), (U3) and (U5), we have as* $n \to \infty$

$$\sup_{\theta\in\Theta_\mathcal{F}}\sup_{x\in S_\mathcal{F}}\left|\widehat{f}_D(\theta,x)-1\right| = O_{a.co.}\left(\sqrt{\frac{\log d_n^{S_\mathcal{F}}+\log d_n^{\Theta_\mathcal{F}}}{n\phi(h_K)}}\right).$$

Corollary 2.3.3 *Under the assumptions of Lemma 2.3.4, we have,*

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\inf_{\theta\in\Theta_\mathcal{F}}\inf_{x\in S_\mathcal{F}}\widehat{f}_D(\theta,x)<\frac{1}{2}\right)<\infty.$$

Lemma 2.3.5 *Under Assumptions (U1), (U2) and (H4), we have, as n goes to infinity*

$$\sup_{\theta\in\Theta_{\mathcal{F}}}\sup_{x\in S_{\mathcal{F}}}\sup_{y\in\mathcal{C}}\left|f(\theta,y,x)-\mathbb{E}\left[\widehat{f}_N(\theta,y,x)\right]\right| = O(h_K^{b_1})+O(h_H^{b_2}).$$

Lemma 2.3.6 *Under the assumptions of Theorem 2.3.2, we have, as n goes to infinity*

$$\sup_{\theta\in\Theta_{\mathcal{F}}}\sup_{x\in S_{\mathcal{F}}}\sup_{y\in\mathcal{C}}\left|\widehat{f}_N(\theta,y,x)-\mathbb{E}\left[\widehat{f}_N(\theta,y,x)\right]\right| = O_{a.co.}\left(\sqrt{\frac{\log d_n^{S_{\mathcal{F}}}+\log d_n^{\Theta_{\mathcal{F}}}}{nh_H\phi(h_K)}}\right).$$

2.4 Some applications and comments

2.4.1 The conditional mode in functional single-index model

Let us now study the estimation of the conditional mode in the functional single-index model. Our main aim, here, is to establish the a.co. convergence of the kernel estimator of the conditional mode of Y given θX, denoted by $M_\theta(x)$, uniformly on fixed subset $S_{\mathcal{F}}$ of $\mathcal{F}$. For this, we assume that $M_\theta(x)$ satisfies on S the following uniform uniqueness property

(U6) $\forall\epsilon_0>0\ \exists\eta>0,\ \forall\upsilon: S_{\mathcal{F}}\to\mathcal{C}$,

$$\sup_{x\in S_{\mathcal{F}}}|M_\theta(x)-\upsilon(x)|\ge\epsilon_0\ \Rightarrow\ \sup_{x\in S_{\mathcal{F}}}|f(\theta,\upsilon(x),x)-f(\theta,M_\theta(x),x)|\ge\eta.$$

Moreover, we also suppose that there exists some integer $j>1$ such that $\forall x\in S$ the function $f(\theta,\cdot,x)$ is j times continuously differentiable w.r.t. y on $\mathcal{C}$ and

(U7)

$$\begin{cases} f^{(l)}(\theta,M_\theta(x),x)=0 \text{ if } 1\le l<j \\ \text{and}\ \ f^{(j)}(\theta,\cdot,x)\ \text{ is uniformly continuous on}\,\mathcal{C} \\ \text{such that}\ \ |f^{x(j)}(\theta,\cdot,\,x)|>C>0 \end{cases}$$

where $f^{(j)}(\theta,\cdot,x)$ is the j^{th} order derivative of the conditional density $f(\theta,\cdot,x)$.

We estimate the conditional mode $M_\theta(x)$ with a random variable $\widehat{M_\theta(x)}$ such that

$$\widehat{M_\theta(x)}=\arg\sup_{y\in\mathcal{C}}\widehat{f}(\theta,y,x).$$

From Corollary 2.3.2 we derive the following result.

Corollary 2.4.1 *Under the assumptions of Theorem 2.3.2 and if the conditional density $f(\theta,\cdot,x)$ satisfies (U6) and (U7), we have*

$$\sup_{x\in S_{\mathcal{F}}}|\widehat{M_\theta(x)} - M_\theta(x)| = O(h_K^{b_1}) + O(h_H^{b_2}) + O_{a.co.}\left(\sqrt{\frac{\log d_n^{S_{\mathcal{F}}}}{n^{1-\gamma}\phi(h_K)}}\right).$$

2.4.2 Application to prediction

Let us now define the application framework of our results to prediction problem. For each $n \in \mathbb{N}^*$, let $(X_i(t))_{t\in\mathbb{R}}$ $i = 1,\dots,n$ be a Hilbertian random variable. For each curve $(X_i(t))_{t\in\mathbb{R}}$, we have a real response variable Y_i. We suppose that the observations $(X_i, Y_i)_{1\le i\le n}$ are generated with single-index structure. The prediction aim is to evaluate y_{new} given $(X_{n+1}(t))_{t\in\mathbb{R}} = x_{new}$. The estimation of the conditional mode in functional single-index model shows that the random variable $\widehat{M_\theta(x_{new})}$, is the best approximation of y_{new}, given x_{new}. Applying the result in the above corollary, we obtain the following result.

Corollary 2.4.2 *Under the assumptions of Corollary 2.4.1, we have as n goes to infinity*

$$\widehat{M_\theta(x_{new})} - M_\theta(x_{new}) \to 0 \quad a.co.$$

2.4.3 On the estimation of the functional single index

Another way to highlight the interest of our study is to show how the conditional density estimate can be used to derive an estimate of the functional single index if the latter is unknown. The estimation of the functional single index has been extensively studied in the multivariate case. In the functional case, Aït-Saidi *et al.* (2009) adopt the leave-out-one-curve cross-validation procedure. Alternatively, this parameter can be estimated via the pseudo-maximum likelihood method which is based on the preliminary estimate of the conditional density of Y given X by

$$\widehat{\theta} = \arg\max_{\theta\in\Theta_{\mathcal{F}}} \widehat{L}(\theta)$$

where

$$\widehat{L}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\log\widehat{f}(\theta, Y_i, X_i).$$

This method has been studied by Delecroix *et al.* (2003) in the real case where they showed that this technique has minimal variance among all estimators. The asymptotic optimality of this procedure in functional statistic, is an important prospect of the present work.
As an application, this approach can be used for answering the semi-metric choice question. Indeed, it is well known that, in nonparametric functional statistic, the projection-type semi-metric is very interesting for increasing the concentration property. The functional index model is a particular case of this family of semi-metric, because it is based on the projection on one functional direction. So, the estimation procedures of this direction permit us to compute adaptive semi-metrics in the general context of nonparametric functional data analysis.

2.5 Appendix

Proof of Lemma 2.3.2 We put

$$\Lambda_i := \frac{1}{\mathbb{E}\left[K_1(\theta,x)\right] h_H} \left\{K_i(\theta,x)H_i(y) - \mathbb{E}\left[K_1(\theta,x)H_1(y)\right]\right\}.$$

Because H is bounded and is L_2 integrable, a simple algebraic calculations give, for all $i \leq n$

$$\mathbb{E}\left[\Lambda_i^2\right] \leq h_H^{-1}\phi_{\theta,x}^{-1}(h_K).$$

Thus, using the classical Bernstein inequality, we get the result. ■

Proof of Lemma 2.3.3 Conditioning w.r.t. X_1, the hypothesis (H1) allows us to write

$$\mathbb{E}\left[\widehat{f}_N(\theta,y,x)\right] - f(\theta,y,x) = \frac{1}{h_H \mathbb{E}\left[K_1(\theta,x)\right]} \mathbb{E}\left[\mathbb{I}_{B(x,h_K)} K_1(\theta,x) \mathbb{E}\left[H_1(y)|\langle\theta,X\rangle\right] - f(\theta,y,x)\right]$$

where $\mathbb{I}_A$ is the indicator function of the set A. Then, by (H2) and change of variable we get

$$\begin{aligned}\mathbb{I}_{B(x,h_K)}\left|\mathbb{E}\left[H_1(y)|\langle\theta,X\rangle\right] - f(\theta,y,x)\right| &\leq \mathbb{I}_{B(x,h_K)} \int_{\mathbb{R}} H(t)\left|f(\theta,y-h_H t,X) - f(\theta,y,x)\right| dt \\ &\leq C(h_K^{b_1} + h_H^{b_2}).\end{aligned}$$

The result follows immediately. ■

Proof of Lemma 2.3.4 The proof of this Lemma follows the same ideas as in Ferraty *et al.* (2009). Indeed, for all $x \in S_{\mathcal{F}}$, we set

$$k(x) = \arg\min_{k\in\{1,\dots r_n\}} \|x - x_k\| \quad \text{and } j(\theta) = \arg\min_{j\in\{1,\dots \ell_n\}} \|\theta - t_j\|.$$

We consider the following decomposition

$$\begin{aligned}
\sup_{x\in S_{\mathcal{F}}}\sup_{\theta\in\Theta_{\mathcal{F}}}\left|\widehat{f}_D(\theta,x) - \mathbb{E}\Big[\widehat{f}_D(\theta,x)\Big]\right| \leq\ & \underbrace{\sup_{x\in S_{\mathcal{F}}}\sup_{\theta\in\Theta_{\mathcal{F}}}\left|\widehat{f}_D(\theta,x) - \widehat{f}_D(\theta,x_{k(x)})\right|}_{T_1} \\
&+ \underbrace{\sup_{x\in S_{\mathcal{F}}}\sup_{\theta\in\Theta_{\mathcal{F}}}\left|\widehat{f}_D(\theta,x_{k(x)}) - \widehat{f}_D(t_{j(\theta)},x_{k(x)})\right|}_{T_2} + \underbrace{\sup_{x\in S_{\mathcal{F}}}\sup_{\theta\in\Theta_{\mathcal{F}}}\left|\widehat{f}_D(t_{j(\theta)},x_{k(x)}) - \mathbb{E}\left[\widehat{f}_D(t_{j(\theta)},x_{k(x)})\right]\right|}_{T_3} \\
&+ \underbrace{\sup_{x\in S_{\mathcal{F}}}\sup_{\theta\in\Theta_{\mathcal{F}}}\left|\mathbb{E}\left[\widehat{f}_D(t_{j(\theta)},x_{k(x)})\right] - \mathbb{E}\left[\widehat{f}_D(\theta,x_{k(x)})\right]\right|}_{T_4} + \underbrace{\sup_{x\in S_{\mathcal{F}}}\sup_{\theta\in\Theta_{\mathcal{F}}}\left|\mathbb{E}\Big[\widehat{f}_D(\theta,x_{k(x)})\Big] - \mathbb{E}\Big[\widehat{f}_D(\theta,x)\Big]\right|}_{T_5}.
\end{aligned}$$

Firstly, let us study T_1, T_2, T_4 and T_5. For theses terms we use the Hölder continuity condition on K and the Cauchy-Schwartz's inequality.
With theses arguments we get for T_1

$$\begin{aligned}
T_1 &\leq \frac{C}{\phi(h_K)}\sup_{x\in S_{\mathcal{F}}}\sup_{\theta\in\Theta_{\mathcal{F}}}\frac{1}{n}\sum_{i=1}^{n}\left|K_i(\theta,x) - K_i(\theta,x_{k(x)})\right| \\
&\leq \frac{Cr_n}{h_K\phi(h_K)},
\end{aligned}$$

and analogously, for T_2 we obtain

$$T_2 \leq \frac{Cr_n}{h_K\phi(h_K)}\mathbb{1}_{B(x,h_K)\cup B(x_{k(x)},h_K)}(X_i).$$

The latter can be treated by Bernstein's inequality, with

$$Z_i := \frac{\epsilon}{h_K\,\phi(h_K)}\mathbb{1}_{B(x,h_K)\cup B(x_{k(x)},h_K)}(X_i),$$

which gives, for n tending to infinity

$$(2.8)\quad T_1 = O\left(\sqrt{\frac{\log d_n^{S_{\mathcal{F}}} + \log d_n^{\Theta_{\mathcal{F}}}}{n\phi(h_K)}}\right) \quad \text{and} \quad T_2 = O\left(\sqrt{\frac{\log d_n^{S_{\mathcal{F}}} + \log d_n^{\Theta_{\mathcal{F}}}}{n\phi(h_K)}}\right).$$

Moreover, using the fact that $T_4 \leq T_1$ and $T_5 \leq T_2$ to get, for n tending to infinity

$$(2.9)\ T_4 = O\left(\sqrt{\frac{\log d_n^{S_\mathcal{F}} + \log d_n^{\Theta_\mathcal{F}}}{n\phi(h_K)}}\right) \quad \text{and} \quad T_5 = O\left(\sqrt{\frac{\log d_n^{S_\mathcal{F}} + \log d_n^{\Theta_\mathcal{F}}}{n\phi(h_K)}}\right).$$

Now, we deal with T_3. For all $\eta > 0$, we have

$$\begin{aligned}
&\mathbb{P}\left(T_3 > \eta\sqrt{\frac{\log d_n^{S_\mathcal{F}} + \log d_n^{\Theta_\mathcal{F}}}{n\phi(h_K)}}\right)\\
&\leq d_n^{S_\mathcal{F}} d_n^{\Theta_\mathcal{F}} \max_{k\in\{1\ldots d_n^{S_\mathcal{F}}\}} \max_{j\in\{1\ldots d_n^{\Theta_\mathcal{F}}\}} \mathbb{P}\left(\left|\widehat{f}_D(t_{j(\theta)}, x_{k(x)}) - \mathbb{E}\left[\widehat{f}_D(t_{j(\theta)}, x_{k(x)})\right]\right| > \eta\sqrt{\frac{\log d_n^{S_\mathcal{F}} + \log d_n^{\Theta_\mathcal{F}}}{n\phi(h_K)}}\right)\\
&= \mathcal{L}.
\end{aligned}$$

Putting

$$\Delta_i = \frac{1}{\phi(h_K)}\left\{K_i(\theta, x_{k(x)}) - \mathbb{E}\left[K_i(\theta, x_{k(x)})\right]\right\}$$

Applying Bernstein's exponential inequality to Δ_i, under (H3), to get $\forall j \leq d_n^{\Theta_\mathcal{F}}$ and $\forall k < d_n^{S_\mathcal{F}}$

$$\begin{aligned}
\mathbb{P}\left(\left|\widehat{f}_D(t_{j(\theta)}, x_{k(x)}) - \mathbb{E}\left[\widehat{f}_D(t_{j(\theta)}, x_{k(x)})\right]\right| > \eta\sqrt{\frac{\log(d_n^{S_\mathcal{F}} d_n^{\Theta_\mathcal{F}})}{n\phi(h_K)}}\right) &= \mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}\Delta_i\right| > \eta\sqrt{\frac{\log(d_n^{S_\mathcal{F}} d_n^{\Theta_\mathcal{F}})}{n\phi(h_K)}}\right)\\
&\leq 2(d_n^{S_\mathcal{F}} d_n^{\Theta_\mathcal{F}})^{-C\eta^2}.
\end{aligned}$$

Therefore, by choosing $C\eta^2 = \beta$ and under (U7), we have, $\mathcal{L} = O\left((d_n^{S_\mathcal{F}} d_n^{\Theta_\mathcal{F}})^{1-C\eta^2}\right)$. Finally the result can be easily deduced from the latter together with (2.8) and (2.9).
■

Proof of Lemma 2.3.5 It suffices to combine the proof of Lemma 2.3.4 with Assumption (U2) where the Lipschitz condition is assumed uniformly on x, y and θ.
■

Proof of Lemma 2.3.6 We keep the same notations as in Lemma 2.3.4 and we use the compactness of $S_\mathcal{F}$. We can write

$$S_\mathcal{F} \subset \bigcup_{k=1}^{z_n}(y_j - \ell_n, y_j + \ell_n)$$

with $\ell_n = n^{-\frac{3}{2}\gamma-\frac{1}{2}}$ and $z_n \leq C\, n^{\frac{3}{2}\gamma+\frac{1}{2}}$. Taking

$$j(y) = \arg \min_{j\in\{1,2,\ldots,z_n\}} |y - t_j|.$$

We get the following decomposition :

$$\begin{aligned}
\left|\widehat{f}_N(\theta,y,x) - \mathbb{E}\left[\widehat{f}_N(\theta,y,x)\right]\right| \leq & \underbrace{\left|\widehat{f}_N(\theta,y,x) - \widehat{f}_N(\theta,y,x_{k(x)}))\right|}_{F_1} + \underbrace{\left|\widehat{f}_N(\theta,y,x_{k(x)}) - \widehat{f}_N(t_{j(\theta)},y,x_{k(x)})\right|}_{F_2} \\
& + \underbrace{\left|\widehat{f}_N(t_{j(\theta)},y,x_{k(x)}) - \widehat{f}_N(t_{j(\theta)},y_{j(y)},x_{k(x)})\right|}_{F_3} + \underbrace{\left|\widehat{f}_N(t_{j(\theta)},y_{j(y)},x_{k(x)}) - \mathbb{E}\left[\widehat{f}_N(t_{j(\theta)},y_{j(y)},x_{k(x)})\right]\right|}_{F_4} \\
& + \underbrace{\left|\mathbb{E}\left[\widehat{f}_N(t_{j(\theta)},y_{j(y)},x_{k(x)})\right] - \mathbb{E}\left[\widehat{f}_N(t_{j(\theta)},y,x_{k(x)})\right]\right|}_{F_5} + \underbrace{\left|\mathbb{E}\left[\widehat{f}_N(t_{j(\theta)},y,x_{k(x)})\right] - \mathbb{E}\left[\widehat{f}_N(\theta,y,x_{k(x)})\right]\right|}_{F_6} \\
& + \underbrace{\left|\mathbb{E}\left[\widehat{f}_N(\theta,y,x_{k(x)})\right] - \mathbb{E}\left[\widehat{f}_N(\theta,y,x)\right]\right|}_{F_7}.
\end{aligned}$$

Using the same ideas as for T_1, T_2, T_4 and T_5, permit us to get, , for n tending to infinity

$$(2.10)\quad F_1 = F_7 = O\left(\sqrt{\frac{\log n}{nh_H\phi(h_K)}}\right), \text{ and } F_2 = F_6 = O\left(\sqrt{\frac{\log n}{nh_H\phi(h_K)}}\right).$$

Concerning the terms F_3 and F_5, using Lipschitz's condition on the kernel H, permits us to write,

$$\left|\widehat{f}_N(t_{j(\theta)},y,x_{k(x)}) - \widehat{f}_N(t_{j(\theta)},y_{j(y)},x_{k(x)})\right| \leq \frac{\ell_n}{h_H^2\,\phi(h_K)}.$$

Now, the fact that $\lim_{n\to+\infty} n^\gamma h_H = \infty$ and choosing $\ell_n = n^{-\frac{3}{2}\gamma-\frac{1}{2}}$ imply that

$$\frac{\ell_n}{h_H^2\,\phi(h_K)} = o\left(\sqrt{\frac{\log d_n^{S_\mathcal{F}}}{nh_H\phi(h_K)}}\right).$$

Hence, for n large enough, we have

$$(2.11)\qquad F_3 = F_5 = O_{a.co.}\left(\sqrt{\frac{\log d_n^{S_\mathcal{F}}}{n^{1-\gamma}\phi(h_K)}}\right).$$

Finally, the evaluation of the term (F_4) is very close to (T_3) in Lemma 2.3.4. applying Bernstein's exponential inequality to

$$\Gamma_i = \frac{1}{h_H\,\phi(h_K)}\left[K_i(x_k)H_i(t_j) - \mathbb{E}\left(K_i(x_k)H_i(t_j)\right)\right],$$

Firstly, it follows from and the fact that the kernels K and H are bounded, we get $\forall j \leq z_n$,

$$\mathbb{P}\left(\left|\widehat{f}_N(t_{j(\theta)}, y_{j(y)}, x_{k(x)}) - \mathbb{E}\left[\widehat{f}_N(t_{j(\theta)}, y_{j(y)}, x_{k(x)})\right]\right| > \eta\sqrt{\frac{\log d_n^{S_{\mathcal{F}}}}{n\,h_H\,\phi(h_K)}}\right) \leq 2\exp\{-C\eta^2\log d_n^{S_{\mathcal{F}}}\}.$$

Choosing $z_n = O\left(\ell_n^{-1}\right) = O\left(n^{\frac{3}{2}\gamma+\frac{1}{2}}\right)$, we get

$$\begin{aligned}
\mathbb{P}\left(F_4 > \eta\sqrt{\frac{\log d_n^{S_{\mathcal{F}}}}{n\,h_H\,\phi(h_K)}}\right) &\leq z_n d_n^{S_{\mathcal{F}}}\,\mathbb{P}\left(\left|\widehat{f}_N(t_{j(\theta)}, y_{j(y)}, x_{k(x)}) - \mathbb{E}\left[\widehat{f}_N(t_{j(\theta)}, y_{j(y)}, x_{k(x)})\right]\right| > \eta\sqrt{\frac{\log d_n^{S_{\mathcal{F}}}}{n\,h_H\,\phi(h_K)}}\right)\\
&\leq C' z_n (d_n^{S_{\mathcal{F}}})^{1-C\eta^2}.
\end{aligned}$$

Putting $C\eta^2 = \beta$ and using (U5), to get

$$F_4 = O_{a.co.}\left(\sqrt{\frac{\log d_n^{S_{\mathcal{F}}}}{nh_H\,\phi(h_K)}}\right). \tag{2.12}$$

So, the Lemma can be easily deduced from (2.10)-(2.12). ■

Bibliographie

[1] Aït Saidi, A., Ferraty, F., Kassa, R., (2005). *Single functional index model for a time series.* R. Roumaine Math. Pures et Appl. **50**, 321–330.

[2] Aït Saidi, A., Ferraty, F., Kassa, R., Vieu, P., 2008). *Cross-validated estimation in the single functional index model.* Statistics **42**, 475–494.

[3] Delecroix, M, Härdle, W., Hristache, M., (2003). *Efficient estimation in conditional single-index regression.* J. Multivariate Anal. **86**, 213–226.

[4] Ezzahrioui, M., Ould Saïd, E., (2010). *Some asymptotic results of a nonparametric conditional mode estimator for functional time series data.* To appear in Statist. Neerlandica.

[5] Ferraty, F., Laksaci, A., Vieu, P., (2006). *Estimating some characteristics of the conditional distribution in nonparametric functional models.* Statist. Inf. for Stoch. Proc. **9**, 47–76.

[6] Ferraty, F., Laksaci, A., Tadj, A., Vieu, P., (2009.) *Rate of uniform consistency for nonparametric estimates with functional variables.* J. Statist. Plann. and Inf. 140, 335–352.

[7] Ferraty, F., Peuch, A., Vieu, P., (2003). *Modèle à indice fonctionnel simple.* C. R. Mathématiques Paris **336**, 1025–1028.

[8] Ferraty, F., Vieu, P., (2006). *NonParametric Functional Data Analysis : Theory and Practice.* Springer Series in Statistics, Springer, New York.

[9] Härdle, W., Hall, P., Ichumira, H., (1993). *Optimal smoothing in single-index models,* Ann. Statist. **21**, 157–178.

[10] Hristache, M., Juditsky, A., Spokoiny, V. (2001). *Direct estimation of the index coefficient in the single-index model.* Ann. Statist. **29**, 595–623.

[11] Laksaci, A., Lemdani, M., Ould Saïd, E., (2009). *A generalized L^1-approach for a kernel estimator of conditional quantile with functional regressors : Consistency and asymptotic normality.* Statist. & Probab. Lett. **79**, 1065–1073.

[12] Rosenblatt, M., (1969). *Conditional probability density and regression estimators. In Multivariate Analysis II, Ed. P.R. Krishnaiah.* Academic Press, New York and London.

Chapitre 3

Single index model : Quasi-associated process case

Some asymptotics results on the nonparametric conditional density estimate in the single index for quasi-associated Hilbertian processes[1]

SAID ATTAOUI, ALI LAKSACI AND ELIAS OULD SAID

Abstract We study the nonparametric estimation of the conditional density of a scalar response variable given a random variable taking values in separable Hilbert space. The goal of the paper is to establish, under general condition the almost complete convergence rates and the asymptotic normality of the conditional density estimator, when the variables are functional quasi-associated, based on the single-index structure.

keywords : Asymptotic normality · Conditional model · Functional Hilbert space · Kernel density estimation · Quasi-associated variables · Single-index model · Weak dependence.
Subject classifications : 62G07 ; 62G08.

1. Preprint, ULCO, LMPA, N° 476, June 2012, Submitted.

3.1 Introduction

For several years, increasing interest has been focused on models that incorporate both parametric and nonparametric components. Such type of model is called semiparametric and is frequently used in statistics and econometric modeling. Indeed the flexibility of semiparametric models provides important new ways to investigate problems in substantive economics. As a particular case, single-index modeling has been successfully offering such flexibility, and providing an optimal approach to compromise between nonparametric and parametric models.

Dealing with single-index models when the explanatory variable is an element of a finite-dimensional space have been studied extensively in both statistical and econometric literatures. Among various papers devoted to the study of the single-index, we quote, for example Härdle *et al.* (1993), Horowitz and Härdle (1996), Hristache *et al.* (2001). Based on the regression function, Delecroix *et al.* (2003) studied the estimation of the single-index and established some asymptotic properties. Hence, the main aim of this work is to contribute to the single-index modeling literature by study the problem of estimating the conditional density of a real variable Y given a functional covariable X when the explanation of Y given X is done through its projection on functional direction.

For the functional single-index models, the literature is closely limited, the first asymptotic properties in the fixed functional single-model were obtained by Ferraty *et al.* (2003). They established the almost complete convergence, in both independent and dependent cases, of the link regression function of this model. Based on the cross-validation procedure Aït Saïdi *et al.* (2008) proposed an estimator of this parameter, where the functional single-index is unknown. Ferraty *et al.* (2011) introduced a new estimator of the single-index based on the idea of functional derivative estimation. Recently Attaoui *et al.* (2011), studied the functional single-index model via its conditional density kernel estimator, and they established its pointwise and uniform almost complete convergence (a.co.) [2] rates.

In this paper, we are interested in the study of the functional single-index model via its conditional density estimation, and extending the results of Attaoui *et al.* (2011) in independent to dependent case. The kind of dependence discussed in this

2. We say that the sequence $(W_n)_n$ converges a.co. to zero, if and only if $\forall \eta > 0$, $\sum_{n\geq 1} \mathbb{P}(|W_n| > \eta) < \infty$. Furthermore, we say that $W_n = O_{a.co.}(w_n)$, if there exists $\eta_0 > 0$, such that $\sum_{n\geq 1} \mathbb{P}(|W_n| > \eta_0 w_n) < \infty$.

paper is that of quasi-associated data. Recall that the concept of quasi-association is a special case of weak dependence introduced by Doukhan and Louhichi (1999) for real-valued stochastic processes. It was applied by Bulinski and Suquet (2001) to real-valued random fields, and it provides a unified approach to studying families of both positively and negatively dependent random variables. Many works were treated data under positive and negative dependant random variables, one can quote, Newman (1984), Matula (1992) and Roussas (1999, 2000).

To the best of our knowledge, there is few papers dealing with the nonparametric estimation for quasi-associated random variables. We quote, Douge (2008) studied a limit theorem for quasi-associated Hilbertian random variables. Our purpose is to establish the almost complete convergence rate of the kernel estimator and the asymptotic normality of the conditional density based on the single-index structure by using a Bernstein-type inequality given by Kallabis and Neumann (2006) for dependent variables. In order to define the quasi-association, we need to introduce the following notations.
A sequence $(X_n)_{n\in\mathbb{N}}$ of real random vector variables is said to be quasi-associated, if for any disjoint subsets I and J of $\mathbb{N}$ and all bounded Lipschitz functions $f : \mathbb{R}^{|I|d} \to \mathbb{R}$ and $g : \mathbb{R}^{|J|d} \to \mathbb{R}$ satisfying :
(3.1)

$$|Cov(f(X_i,\, i \in I),\, g(X_j,\, j \in J))| \leq \text{Lip}(f)\text{Lip}(g) \sum_{i\in I} \sum_{j\in J} \sum_{k=1}^{d} \sum_{l=1}^{d} \left|Cov(X_i^k, X_j^l)\right|$$

(here and in the sequel $|I|$ denotes the cardinality of a finite set I)
where X_i^k denotes the k^{th} component of X_i, and

$$\text{Lip}(f) = \sup_{x\neq y} \frac{|f(x) - f(y)|}{\|x-y\|_1}, \ \text{ with } \ \|(x_1, \ldots, x_k)\|_1 = |x_1| + \ldots + |x_k|.$$

The paper is organized as follows. We present our model in the next section. In Section 3 we introduce notations, assumptions and state the main results. Some discussion is given in Section 4. The last section is devoted to the proof of results.

3.2 Model

Let $\{Z_i := (X_i, Y_i),\, 1 \leq i \leq n\}$ be n quasi-associated random processes, identically distributed as the random pair $Z = (X, Y)$ with values in $\mathcal{H} \times \mathbb{R}$, where $\mathcal{H}$ is a separable real Hilbert space with the norm $\|.\|$ generated by an inner product $< \cdot, \cdot >$.

We consider the semi-metric d_θ, associated to the single-index $\theta \in \mathcal{H}$ defined by $\forall x_1, x_2 \in \mathcal{H} : \, d_\theta(x_1, x_2) := \, | < x_1 - x_2, \theta > | \,$. Under such topological structure and for a fixed functional θ, we suppose that the conditional density of Y given $X = x$ denoted by $f(.|x)$ exists and is given by

$$\forall y \in \mathbb{R}, \quad f_\theta(y|x) =: f(y| < \theta, x >). \tag{3.2}$$

Clearly, the identifiability of the model is assured, and we have for all $x \in \mathcal{H}$,

$$f_1(\cdot \mid < \theta_1, x >) = f_2(\cdot \mid < \theta_2, x >) \Rightarrow \quad f_1 \equiv f_2 \text{ and } \theta_1 = \theta_2.$$

For more detail see Ferraty *et al.* (2003) In what follows we denote by $f(\theta, \cdot, x)$, the conditional density of Y given $< \theta, x >$ and we define the kernel estimator $\widehat{f}(\theta, \cdot, x)$ of $f(\theta, \cdot, x)$ by :

$$\widehat{f}(\theta, y, x) = \frac{b_n^{-1} \sum_{i=1}^n K(a_n^{-1} d_\theta(x, X_i)) H(b_n^{-1}(y - Y_i))}{\sum_{i=1}^n K(a_n^{-1} d_\theta(x, X_i)))}, \quad \forall y \in \mathbb{R} \tag{3.3}$$

with the convention $0/0 = 0$, where K and H are kernels function and a_n (resp. b_n) is a sequence of bandwidths decreasing to zero as n goes to infinity.

3.3 Main results

First we give a definition of quasi-association for random variables with values in a separable Hilbert space.

Definition 1 *Let* $(\mathcal{H}, < \cdot, \cdot >)$ *a separable Hilbert space with a orthonormal basis* $\{e_k, \, k \geq 1\}$. *A sequence* $(X_n)_{n \in \mathbb{N}}$ *of real random variables taking values in* $\mathcal{H}$ *is said to be quasi-associated, with respect to the basis* $\{e_k\}$ *if for any* $d \geq 1$, *the* d*-dimensional sequence* $\{(< X_i, e_{j_1} >, \ldots, < X_i, e_{j_d} >), \, i \in \mathbb{N}\}$ *is quasi-associated.*

Observe that the definition of quasi-association in the Hilbert space depends on the choice of the basis.

3.3.1 Pointwise almost complete convergence

All along the paper, when no confusion will be possible, we will denote by C or/and C' some strictly positive generic constants whose values are allowed to change. The

variable x is a fixed point in $\mathcal{H}$, $\mathcal{N}_x$ is a fixed neighborhood of x. We assume that the random pair $Z_i = \{(X_i, Y_i),\ i \in \mathbb{N}\}$ is stationary quasi-associated processes.
Let λ_k the covariance coefficient defined as :

$$\lambda_k := \sup_{s\geq k} \sum_{|i-j|\geq s} \lambda_{i,j},$$

where

$$\lambda_{i,j} = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} |Cov(X_i^k, X_j^l)| + \sum_{k=1}^{\infty} |Cov(X_i^k, Y_j)| + \sum_{l=1}^{\infty} |Cov(Y_i, X_j^l)| + |Cov(Y_i, Y_j)|,$$

X_i^k denotes the k^{th} component of X_i defined as $X_i^k :=< X_i, e_k >$.

Now, we introduce our main assumptions.

(H1) $0 < \mathbb{P}(| < x - X_i, \theta > | < r) := F_\theta(x, r) > 0$ and the function $F_\theta(x, \cdot)$ is a differentiable at 0.

(H2) The conditional density $f(\theta, y, x)$ satisfies the Hölder condition, that is : $\forall (y_1, y_2) \in \mathcal{C}^2$, $\forall (x_1, x_2) \in \mathcal{N}_x \times \mathcal{N}_x$,

$$|f(\theta, y_1, x_1) - f(\theta, y_2, x_2)| \leq C_{\theta,x}\left(d_\theta^{\beta_1}(x_1, x_2) + |y_1 - y_2|^{\beta_2}\right), \quad \beta_1 > 0,\ \beta_2 > 0$$

where $\mathcal{C}$ is a fixed compact subset of $\mathbb{R}$.

(H3) H is a bounded continuous Lipschitz function, such that

$$\int H(t)dt = 1, \qquad \int |t|^{\beta_2} H(t)dt < \infty \ \text{ and } \int H^2(t)dt < \infty.$$

(H4) K is a bounded continuous Lipschitz function such that

$$C\mathbb{1}_{[0,1]}(t) < K(t) < C'\mathbb{1}_{[0,1]}(t).$$

(H5) The sequence of random pairs $\{(X_i, Y_i),\ i \in \mathbb{N}\}$ is quasi-associated with covariance coefficient λ_k, $k \in \mathbb{N}$ satisfying

$$\exists \alpha > 0, \quad \exists C > 0, \text{ such that } \lambda_k \leq Ce^{-\alpha k}.$$

(H6) $0 < \sup_{i\neq j} \mathbb{P}\Big((X_i, X_j) \in B(x, a_n) \times B(x, a_n)\Big) =$
$\max_{i\neq j}\left\{\mathbb{P}\Big(d_\theta(x, X_i) < a_n), \mathbb{P}(d_\theta(x, X_j) < a_n\Big)\right\} = O(F_\theta^2(x, a_n))$.

(H7) The bandwidths a_n and b_n satisfy : $\lim_{n\to\infty} \dfrac{\log^5 n}{nb_nF_\theta(x, a_n)} = 0$.

Comments on the hypotheses.
Assumptions (H1) and (H2) are the same as used by Attaoui *et al* (2011). Condition (H1) characterizes the measure concentration of the functional variable in the topological structure of the single-index, whereas (H2) is the regularity condition of the functional space of our model. These assumptions needed to evaluate the bias term. The boundedness of the kernel K in (H4) are standard, the lipschitzianity intervenes in the proof under the quasi-association property. In order to establish the almost complete convergence rate of our model for quasi-associated data, we need Hypothesis (H5). The hypothesis (H6) describe the asymptotic behavior of the joint distribution of the couple (X_i, X_j). Assumptions (H3) and (H7) are technical conditions imposed for brevity of proofs.

Discussion about quasi-associated processes
Note that any family of independent real-valued random variables is automatically associated (Newman (1984)) and negatively associated (Joag-Dev and Proschan (1983)). There are interesting stochastic models in mathematical statistics, reliability theory, percolation theory and statistical mechanics described by families of positively or negatively associated random variables, see Bulinski and Suquet (2001). Bulinski and Shabanovich (1998) showed that, any positively or negatively associated collections of random variables with finite second moment satisfies (4.1). Consequently, such fields are quasi-associated. In Shashkin (2002), it was proved that any Gaussian random field $X = \{X_t,\, t \in T\}$ with values in $\mathbb{R}^p$, is quasi-associated. A real-valued Gaussian random field is associated if and only if its covariance function is nonegative (Pitt (1982)) and negatively associated if and only if $Cov(X_s, X_t) \leq 0$, $\forall s \neq t$ (Joag-Dev and Proschan (1983)).
With results of Shashkin (2002), we can conclude that the concept of quasi-association is strictly wider than that of positive or negative association for random fields with finite second moments.

Our main result is given in the following theorem.

Theorem 3.3.1 *Under Hypotheses (H1)-(H7), and for any fixed y, as n goes to infinity, we have,*

$$\left|\widehat{f}(\theta, y, x) - f(\theta, y, x)\right| = O(a_n^{\beta_1}) + O(b_n^{\beta_2}) + O_{a.co.}\left(\sqrt{\frac{\log n}{n b_n F_\theta(x, a_n)}}\right).$$

3.3.2 Proof of Theorem 3.3.1.

Denote, for $i = 1, \dots n$, $K_i := K(a_n^{-1} d_\theta(x, X_i))$ and $H_i := H(b_n^{-1}(y - Y_i))$. Similarly to Attaoui *et al.* (2011), the proof is based on the following decomposition

$$\begin{aligned}\widehat{f}(\theta, y, x) - f(\theta, y, x) &= \frac{1}{\widehat{f}_D(x)} \Big\{ \Big(\widehat{f}_N(\theta, y, x) - \mathbb{E}\Big[\widehat{f}_N(\theta, y, x)\Big]\Big) \\ &\quad + \Big(\mathbb{E}\Big[\widehat{f}_N(\theta, y, x)\Big] - f(\theta, y, x)\Big)\Big\} - \frac{f(\theta, y, x)}{\widehat{f}_D(\theta, x)} \Big\{ \widehat{f}_D(x) - 1 \Big\}\end{aligned}$$

where

$$\widehat{f}_N(\theta, y, x) = \frac{1}{n b_n \mathbb{E}[K_1]} \sum_{i=1}^{n} H_i K_i, \qquad \widehat{f}_D(x) = \frac{1}{n \mathbb{E}[K_1]} \sum_{i=1}^{n} K_i$$

and the following results :

Lemma 3.3.1 *(See, Attaoui et al. (2010)) Under Hypotheses (H1)–(H3), and for any fixed y, as n goes to infinity, we have*

$$\Big| \mathbb{E}\Big[\widehat{f}_N(\theta, y, x)\Big] - f(\theta, y, x) \Big| = O(a_n^{\beta_1}) + O(b_n^{\beta_2}).$$

Lemma 3.3.2 *Under the hypotheses (H1) and (H4)-(H7), and for any fixed y, as n goes to infinity, we have*

$$\Big| \widehat{f}_N(\theta, y, x) - \mathbb{E}\Big[\widehat{f}_N(\theta, y, x)\Big] \Big| = O_{a.co.}\left(\sqrt{\frac{\log n}{n b_n F_\theta(x, a_n)}} \right).$$

Lemma 3.3.3 *Under Hypotheses of Lemma 3.3.2, as n goes to infinity, we have*

$$\Big| \widehat{f}_D(\theta, x) - 1 \Big| = O_{a.co.}\left(\sqrt{\frac{\log n}{n F_\theta(x, a_n)}} \right).$$

Furthermore, we have

$$\sum_{n=1}^{\infty} \mathbb{P}\Big(\Big| \widehat{f}_D(\theta, x) \Big| \le 1/2 \Big) < \infty. \tag{3.4}$$

3.3.3 Asymptotic normality

This section contains results on the asymptotic normality of the estimator $\widehat{f}(\theta, y, x)$. For that we need the following additional hypotheses.

(H1') The concentration property (H1) holds. Moreover, there exists a function $\beta_\theta(x, \cdot)$ such that

$$\forall s \in [0,1], \quad \lim_{n\to\infty} F_\theta(x, sa_n)/F_\theta(x, a_n) = \beta_\theta(x, s).$$

(H2') For $l \in \{0, 2\}$, the functions $\Phi_l(s) = \mathbb{E}\left[\frac{\partial^l f(\theta,y,X)}{\partial y^l} - \frac{\partial^l f(\theta,y,x)}{\partial y^l}\Big| d_\theta(x, X) = s\right]$ are differentiable at $s = 0$.

(H3') The kernel H is an even bounded function and satisfies (H3).

(H4') The kernel K is a differentiable function satisfying (H4) and its derivative K' exists and is such that there exist two constants C and C' with $-\infty < C < K'(t) < C' < 0$ for $0 \le t \le 1$.

(H7') The bandwidths a_n and b_n satisfy
$i)$ $\lim_{n\to\infty} \dfrac{1}{nb_n F_\theta(x, a_n)} = 0,$
$ii)$ $\lim_{n\to\infty} nb_n^5 F_\theta(x, a_n) = 0$ and $\lim_{n\to\infty} nb_n a_n^2 F_\theta(x, a_n) = 0.$

Comments on the hypotheses
Assumptions (H1)'-(H4)' are technicals and permit to give an explicit asymptotic variance. Assumption (H7) implies (H7)' $i)$, but we keep the latter because they intervenes in different places. Finally (H7)' $ii)$ permit to remove the bias term in the asymptotic normality result.

Now we are in position to give our second main result.

Theorem 3.3.2 *Assume that $(H1') - (H4')$, $(H5)$, $(H6)$ and $(H7')\, i)$ hold, as n goes to infinity, we have*

$$\sqrt{nb_n F_\theta(x, a_n)}\left(\widehat{f}(\theta, y, x) - f(\theta, y, x) + B_H^f(\theta, x, y) b_n^2 + B_K^f(\theta, x, y) a_n + o(b_n^2) + o(a_n)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_1^2(x))$$

where

$$\sigma_1^2(x) := \frac{C_2 f(\theta, y, x)}{C_1^2}\left(\int_{\mathbb{R}} H^2(t)\, dt\right), \; B_H^f(\theta, x, y) = \frac{1}{2}\frac{\partial^2 f(\theta, y, x)}{\partial y^2}\int t^2 H'(t) dt,$$

$$B_K^f(\theta, x, y) = \Phi_0'(0)\frac{C_0}{C_1}a_n$$

with

$$C_0 = K(1) - \int_0^1 sK'(s)\beta_\theta(x,s)ds \ \ \text{and} \ \ C_j = K(1) - \int_0^1 (K^j)'(s)\beta_\theta(x,s)ds \qquad \text{for} \quad j = 1, 2$$

and $\xrightarrow{\mathcal{D}}$ *means the convergence in distribution.*

In order to remove the bias term, we need an additional condition on the bandwidth parameters a_n and b_n.

Corollary 3.3.1 *Under the hypotheses of Theorem 3.3.2 and if* $(H7')\,ii)$ *holds, as* n *goes to infinity, we have*

$$\sqrt{nb_nF_\theta(x,a_n)}\left(\widehat{f}(\theta,y,x) - f(\theta,y,x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_1^2(x))$$

where $\sigma_1^2(x)$ *is as in Theorem 3.3.2.*

3.3.4 Proof of Theorem 3.3.2.

The proof is based on following decomposition

$$\begin{aligned}
\widehat{f}(\theta,y,x) - f(\theta,y,x) &= \frac{\widehat{f}_N(\theta,y,x) - f(\theta,y,x)\widehat{f}_D(\theta,x)}{\widehat{f}_D(\theta,x)} \\
&= \frac{1}{\widehat{f}_D(\theta,x)}\left\{\widehat{f}_N(\theta,y,x) - \mathbb{E}\left[\widehat{f}_N(\theta,y,x)\right] + \mathbb{E}\left[\widehat{f}_N(\theta,y,x)\right]\right. \\
&- \left. f(\theta,y,x) - f(\theta,y,x)(\widehat{f}_D(\theta,x) - 1)\right\}.
\end{aligned}$$

Finally, to state the asymptotic normality of $f(\theta,y,x)$, we show that the numerator suitably normalized is asymptotically normally distributed (with law $\mathcal{N}(0,\sigma_1^2(x))$) and that the denominator converges in probability to 1.
Then, the proof of Theorem 3.3.2 can be deduced from the following lemmas :

Lemma 3.3.4 *Under the hypothesis* $(H1')$, $(H4')$, $(H5)$, $(H6)$ *and* $(H7')\,i)$, *as* n *goes to infinity, we have*

$$\sqrt{nb_nF_\theta(x,a_n)}\left[\widehat{f}_D(\theta,x) - 1\right] \to 0, \ \text{in probability as} \quad n \to \infty.$$

Observe that, the result of this lemma permits to write

$$\widehat{f}_D(\theta, x) - 1 \to 0, \text{ in probability as } \quad n \to \infty.$$

Lemma 3.3.5 *Under the hypotheses of Theorem 3.3.2, as n goes to infinity, we have*

$$\mathbb{E}\left[\widehat{f}_N(\theta, y, x)\right] - f(\theta, y, x) = B_H^f(\theta, x, y)b_n^2 + B_K^f(\theta, x, y)a_n + o(b_n^2) + o(a_n)$$

Lemma 3.3.6 *Under the hypotheses of Theorem 3.3.2, as n goes to infinity, we have*

$$\sqrt{nb_nF_\theta(x, a_n)}\left(\widehat{f}_N(\theta, y, x) - \mathbb{E}\left[\widehat{f}_N(\theta, y, x)\right]\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_1^2(x)).$$

3.4 Auxillary results and proofs

First of all, we state the following lemmas.

Lemma 3.4.1 *(See, Douge (2010)) Let $(X_n)_{n\in\mathbb{N}}$ be a quasi-associated sequence of random variables with values in $\mathcal{H}$. Let $f \in BL(\mathcal{H}^{|I|}) \cap \mathbb{L}^\infty$ and $g \in BL(\mathcal{H}^{|J|}) \cap \mathbb{L}^\infty$, for some finite disjoint subsets $I, J \in \mathbb{N}$. Then*

$$|Cov(f(X_i,\, i \in I),\, g(X_j,\, j \in J))| \leq \text{Lip}(f)\text{Lip}(g) \sum_{i\in I}\sum_{j\in J}\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\left|Cov(X_i^k, X_j^l)\right|$$

where $(BL(\mathcal{H}^u)\,;\, u > 0)$ is the set of bounded Lipschitz functions $f : \mathcal{H}^u \to \mathbb{R}$ and $\mathbb{L}^\infty$ is the set of bounded functions.

Lemma 3.4.2 *(See, Kallabis and Neumann (2006))*
Let $X_1, \ldots, X_n$ the real random variables such that $\mathbb{E}X_j = 0$ and $\mathbb{P}(|X_j| \leq M) = 1$, for all $j = 1, \ldots, n$ and some $M < \infty$. Let $\sigma_n^2 = Var\left(\sum_{j=1}^n X_j\right)$. Assume, furthermore, that there exist $K < \infty$ and $\beta > 0$ such that, for all u-tuplets $(s_1, \ldots, s_u)$ and all v-tuplets $(t_1, \ldots, t_v)$ with $1 \leq s_1 \leq \ldots \leq s_u \leq t_1 \leq \ldots \leq t_v \leq n$, the following inequality is fulfilled :

$$\left|Cov(X_{s_1} \ldots X_{s_u}, X_{t_1} \ldots X_{t_v})\right| \leq K^2 M^{u+v-2} v e^{-\beta(t_1 - s_u)}.$$

Then,

$$\mathbb{P}\Big(\Big|\sum_{j=1}^{n} X_j\Big| > t\Big) \leq \exp\left\{-\frac{t^2/2}{A_n + B_n^{\frac{1}{3}} t^{\frac{5}{2}}}\right\}$$

for some $A_n \leq \sigma_n^2$ *and* $B_n = \left(\frac{16nK^2}{9A_n(1-e^{-\beta})} \vee 1\right)\frac{2(K \vee M)}{1-e^{-\beta}}$.

Proof of Lemma 3.3.2. We use Lemma 5.3.2 on the variables

$$\Delta_i = \frac{1}{nb_n\mathbb{E}\left[K_1\right]}\chi(X_i, Y_i), \qquad 1 \leq i \leq n$$

where

$$\chi(z, w) = H(b_n^{-1}(y-w))K(a_n^{-1}d_\theta(x,z)) - \mathbb{E}\left[H_1K_1\right], \, z \in \mathcal{H}, \, w \in \mathbb{R}.$$

Clearly, we have $\mathbb{E}(\Delta_i) = 0$ and

$$\widehat{f}_N(\theta, y, x) - \mathbb{E}\Big[\widehat{f}_N(\theta, y, x)\Big] = \sum_{i=1}^{n} \Delta_i.$$

Moreover, we can write

$$\|\chi\|_\infty{}^3 \leq 2C\|K\|_\infty\|H\|_\infty \text{ and } \mathrm{Lip}(\chi) \leq C(a_n^{-1}\|H\|_\infty\mathrm{Lip}(K) + b_n^{-1}\|K\|_\infty\mathrm{Lip}(H)).$$

Now, to apply Lemma 5.3.2 we have to evaluate the variance term $Var\left(\sum_{i=1}^{n} \Delta_i\right)$ and the covariance term $Cov(\Delta_{s_1} \ldots \Delta_{s_u}, \Delta_{t_1} \ldots \Delta_{t_v})$, for all $(s_1, \ldots, s_u) \in \mathbb{N}^u$ and $(t_1, \ldots, t_v) \in \mathbb{N}^v$ with $1 \leq s_1 \leq \ldots \leq s_u \leq t_1 \leq \ldots \leq t_v \leq n$. Firstly, for the covariance term, we consider the following cases

– If $t_1 = s_u$. By using the fact that $\mathbb{E}\left[|H_1K_1|\right] = O(b_nF_\theta(x, a_n))$, and $\mathbb{E}\left[|K_1|\right] = O(F_\theta(x, a_n))$, we have

$$\begin{aligned}
|Cov(\Delta_{s_1} \ldots \Delta_{s_u}, \Delta_{t_1} \ldots \Delta_{t_v})| &\leq \left(\frac{C}{nb_n\mathbb{E}\left[K_1\right]}\right)^{u+v} \mathbb{E}\left|\chi(X_1, Y_1)\right|^{u+v} \\
&\leq \left(\frac{C\|H\|_\infty\|K\|_\infty}{nb_n\mathbb{E}\left[K_1\right]}\right)^{u+v} \mathbb{E}\left[|K_1H_1|\right] \\
&\leq b_nF_\theta(x, a_n)\left(\frac{C}{nb_nF_\theta(x, a_n)}\right)^{u+v}. \qquad (3.5)
\end{aligned}$$

3. For any function f we denote by $\|f\|_\infty$ the supremun norm

– If $t_1 > s_u$, we use the quasi-association, under (H5), we get :

$$\begin{aligned}
|Cov(\Delta_{s_1}\dots\Delta_{s_u},\Delta_{t_1}\dots\Delta_{t_v})| &\leq \left(\left(a_n^{-1}\text{Lip}(K)+b_n^{-1}\text{Lip}(H)\right)(nb_n\mathbb{E}\left[K_1\right])^{-1}\right)^2 \\
&\quad \left(\frac{C}{nb_n\mathbb{E}\left[K_1\right]}\right)^{u+v-2}\sum_{i=1}^{u}\sum_{j=1}^{v}\lambda_{s_i,t_j} \\
&\leq \left(a_n^{-1}\text{Lip}(K)+b_n^{-1}\text{Lip}(H)\right)^2\left(\frac{C}{nb_n\mathbb{E}\left[K_1\right]}\right)^{u+v} v\lambda_{t_1-s_u} \\
&\leq \left(a_n^{-1}\text{Lip}(K)+b_n^{-1}\text{Lip}(H)\right)^2\left(\frac{C}{nb_nF_\theta(x,a_n)}\right)^{u+v} ve^{-\alpha(t_1-s_u)}. \qquad (3.6)
\end{aligned}$$

On the other hand, by (H6) we have,

$$\begin{aligned}
|Cov(\Delta_{s_1}\dots\Delta_{s_u},\Delta_{t_1}\dots\Delta_{t_v})| &\leq \left(\frac{C\|H\|_\infty\|K\|_\infty}{nb_n\mathbb{E}\left[K_1\right]}\right)^{u+v-2}\left(|\mathbb{E}[\Delta_{s_u}\Delta_{t_1}]|+\mathbb{E}\left|\Delta_{s_u}\right|\mathbb{E}\left|\Delta_{t_1}\right|\right) \\
&\leq \left(\frac{C\|H\|_\infty\|K\|_\infty}{nb_n\mathbb{E}\left[K_1\right]}\right)^{u+v-2}\left(\frac{C}{nb_n\mathbb{E}\left[K_1\right]}\right)^2\times \\
&\quad b_n^2\Big(\sup_{i\neq j}\mathbb{P}((X_i,X_j)\in B(x,a_n)\times B(x,a_n)) + \ (\mathbb{P}(X_1\in B(x,a_n))^2\Big) \\
&\leq \left(\frac{C}{nb_nF_\theta(x,a_n)}\right)^{u+v}(b_nF_\theta(x,a_n))^2. \qquad (3.7)
\end{aligned}$$

Furthermore, taking a γ-power of (3.6), $(1-\gamma)$-power of (3.7), with $0<\gamma<1$, we obtain an upper-bound of the tree terms as follows : for $1\leq s_1\leq \dots\leq s_u\leq t_1\leq\dots\leq t_v\leq n$:

$$|Cov(\Delta_{s_1}\dots\Delta_{s_u},\Delta_{t_1}\dots\Delta_{t_v})| \leq b_nF_\theta(x,a_n)\left(\frac{C}{nb_nF_\theta(x,a_n)}\right)^{u+v} ve^{-\gamma\alpha(t_1-s_u)}.$$

Secondly, for the variance term $Var\left(\sum_{i=1}^n\Delta_i\right)$, we put, for all $1\leq i\leq n,\ \Gamma_i(\theta,y,x)= H_iK_i$. Thus,

$$\begin{aligned}
Var\left(\sum_{i=1}^n\Delta_i\right) &= \frac{1}{(nb_n\mathbb{E}\left[K_1\right])^2}\sum_{i=1}^n\sum_{j=1}^n Cov(\Gamma_i(\theta,y,x),\Gamma_j(\theta,y,x)) \\
&= \frac{1}{n(b_n\mathbb{E}\left[K_1\right])^2}Var\left(\Gamma_1(\theta,y,x)\right) \\
&+ \frac{1}{(nb_n\mathbb{E}\left[K_1\right])^2}\sum_{i=1}^n\sum_{\substack{j=1\\ j\neq i}}^n Cov(\Gamma_i(\theta,y,x),\Gamma_j(\theta,y,x)). \qquad (3.8)
\end{aligned}$$

For the first term,

$$Var\left(\Gamma_1(\theta,y,x)\right) = \mathbb{E}\left[H_1^2K_1^2\right] - \left(\mathbb{E}\left[H_1K_1\right]\right)^2. \tag{3.9}$$

Then,

$$\mathbb{E}\left[H_1^2K_1^2\right] = \mathbb{E}\left[K_1^2\mathbb{E}\left[H_1^2|X_1\right]\right].$$

Thus, under (H2) and (H3), and by integration on the real component y we get

$$\mathbb{E}[H_1^2|X_1] = O(b_n).$$

As, for all $j \geq 1$, $\mathbb{E}\left[K_1^j\right] = O(F_\theta(x,a_n))$, then

$$\mathbb{E}\left[H_1^2K_1^2\right] = O(b_nF_\theta(x,a_n)).$$

It follows that

$$\left(\frac{1}{n\left(b_n\mathbb{E}[K_1]\right)^2}\right)Var(\Gamma_1(\theta,y,x)) = O\left(\frac{1}{nb_nF_\theta(x,a_n)}\right). \tag{3.10}$$

Concerning the covariance term in (4.19), we us the following decomposition

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} Cov(\Gamma_i(\theta,y,x),\Gamma_j(\theta,y,x)) &= \sum_{i=1}^{n}\sum_{\substack{j=1\\ 0<|i-j|\leq m_n}}^{n} Cov(\Gamma_i(\theta,y,x),\Gamma_j(\theta,y,x))\\
&\quad + \sum_{i=1}^{n}\sum_{\substack{j=1\\ |i-j|>m_n}}^{n} Cov(\Gamma_i(\theta,y,x),\Gamma_j(\theta,y,x))\\
&=: I + II.
\end{aligned}$$

where (m_n) is a sequence of positive integer which goes to infinity as $n \to \infty$.
From Assumptions (H1), (H3) and (H6), we have, for $i \neq j$

$$\begin{aligned}
I &\leq nm_n\left(\max_{i\neq j}\left|\mathbb{E}[\Gamma_i(\theta,y,x)\Gamma_j(\theta,y,x)]\right| + \left(\mathbb{E}[\Gamma_1(\theta,y,x)]\right)^2\right)\\
&\leq nm_n\Big(\max_{i\neq j}\left|\mathbb{E}[H_iK_iH_jK_j]]\right| + \left(\mathbb{E}\left[H_1K_1\right]\right)^2\Big)\\
&\leq Cnm_n\Big(b_n^2F_\theta^2(x,a_n) + \left(b_nF_\theta(x,a_n)\right)^2\Big)\\
&\leq Cnm_nb_n^2F_\theta^2(x,a_n). \qquad (3.11)
\end{aligned}$$

Since both kernels H and K are bounded and Lipschitz, we get

$$\begin{aligned}
II &\leq \left(a_n^{-1}\mathrm{Lip}(K)+b_n^{-1}\mathrm{Lip}(H)\right)^2 \sum_{i=1}^{n} \sum_{\substack{j=1\\|i-j|>m_n}}^{n} \lambda_{i,j} \\
&\leq C\left(a_n^{-1}\mathrm{Lip}(K)+b_n^{-1}\mathrm{Lip}(H)\right)^2 \sum_{i=1}^{n} \sum_{\substack{j=1\\|i-j|>m_n}}^{n} \lambda_{i,j} \\
&\leq C\left(a_n^{-1}\mathrm{Lip}(K)+b_n^{-1}\mathrm{Lip}(H)\right)^2 \sum_{i=1}^{n} \sum_{\substack{j=1\\|i-j|>m_n}}^{n} \lambda_{i,j} \\
&\leq Cn\left(a_n^{-1}\mathrm{Lip}(K)+b_n^{-1}\mathrm{Lip}(H)\right)^2 \lambda_{m_n} \\
&\leq Cn\left(a_n^{-1}\mathrm{Lip}(K)+b_n^{-1}\mathrm{Lip}(H)\right)^2 e^{-\alpha m_n}. \qquad (3.12)
\end{aligned}$$

Then, by (3.11) and (3.12), we get

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n} Cov(\Gamma_i(\theta,y,x),\Gamma_j(\theta,y,x)) \;\leq\; C\left(nm_nb_n^2F_\theta^2(x,a_n)+n\left(a_n^{-1}\mathrm{Lip}(K)+b_n^{-1}\mathrm{Lip}(H)\right)^2 e^{-\alpha m_n}\right).$$

By choosing $m_n=\log\left(\dfrac{\left(a_n^{-1}\mathrm{Lip}(K)+b_n^{-1}\mathrm{Lip}(H)\right)^2}{\alpha b_n^2F_\theta^2(x,a_n)}\right)$, we get

$$\frac{1}{nb_nF_\theta(x,a_n)}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n} Cov(\Gamma_i(\theta,y,x),\Gamma_j(\theta,y,x)) \;\to\; 0, \text{ as } n\to\infty. \qquad (3.13)$$

Finally, by combining results (4.23), (3.10) and (4.20), we get

$$Var\left(\sum_{i=1}^{n}\Delta_i\right)=O\left(\frac{1}{nb_nF_\theta(x,a_n)}\right). \qquad (3.14)$$

So, the variables Δ_i, $i=1,\ldots,n$ satisfy the conditions of Lemma 5.3.2 for

$$K_n=\frac{C}{nb_n\sqrt{F_\theta(x,a_n)}},\; M_n=\frac{C}{nb_nF_\theta(x,a_n)} \text{ and } A_n=Var\left(\sum_{i=1}^{n}\Delta_i\right).$$

Thus,

$$\begin{aligned}
&\mathbb{P}\left(\left|\widehat{f}_N(\theta,y,x)-\mathbb{E}\Big[\widehat{f}_N(\theta,y,x)\Big]\right|>\eta\sqrt{\frac{\log n}{nb_nF_\theta(x,a_n)}}\right)\\
&\qquad = \mathbb{P}\left(\Big|\sum_{i=1}^{n}\Delta_i\Big|>\eta\sqrt{\frac{\log n}{nb_nF_\theta(x,a_n)}}\right)\\
&\qquad \leq \exp\left\{-\frac{\eta^2\log n}{(nb_nF_\theta(x,a_n))\left(Var\left(\sum_{i=1}^{n}\Delta_i\right)+\frac{\log^{5/6}n}{(nb_nF_\theta(x,a_n))^{(7/6)}}\right)}\right\}\\
&\qquad \leq \exp\left\{-\frac{\eta^2\log n}{\left(C+\frac{\log^{5/6}n}{(nb_nF_\theta(x,a_n))^{(1/6)}}\right)}\right\}\\
&\qquad \leq C'\exp\left\{-C\eta^2\log n\right\}
\end{aligned}$$

by (H7). Finally, for a suitable choice of η, Borel-Cantelli's lemma allows to finish the proof of this Lemma. ■

Proof of Lemma 3.3.2. The proof is a direct application of Lemma 3.3.3 case when we replace $\chi(\cdot,\cdot)$ by

$$\chi'(z)=K(a_n^{-1}d_\theta(x,z))-\mathbb{E}\left[K_1\right],\ z\in\mathcal{H}.$$

Thus, we obtain

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\widehat{f}_D(\theta,x)-\mathbb{E}\Big[\widehat{f}_D(\theta,x)\Big]\right|>\eta\sqrt{\frac{\log n}{nF_\theta(x,a_n)}}\right)<\infty. \tag{3.15}$$

Moreover, we have

$$\begin{aligned}
\mathbb{P}\left(\left|\widehat{f}_D(\theta,x)\right|\leq 1/2\right) &\leq \mathbb{P}\left(\left|\widehat{f}_D(\theta,x)-1\right|>1/2\right)\\
&\leq \mathbb{P}\left(\left|\widehat{f}_D(\theta,x)-\mathbb{E}\left[\widehat{f}_D(\theta,x)\right]\right|>1/2\right),
\end{aligned}$$

and using (3.15) we can write

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\widehat{f}_D(\theta,x)\right|\leq 1/2\right)<\infty$$

which completes the proof. ■

Proof of Lemma 3.3.4. First recall that $\mathbb{E}\left[\widehat{f}_D(\theta,x)\right]=1$. Then using the same arguments used in (3.14) we get

$$Var\left(\widehat{f}_D(\theta,x)\right)=O\left(\frac{1}{nF_\theta(x,a_n)}\right)$$

that is

$$Var\left(\sqrt{nb_nF_\theta(x,a_n)}\widehat{f}_D(\theta,x)\right)=O\left(b_n\right)$$

which finishes the proof ■

Proof of Lemma 3.3.5. Firstly, for $\mathbb{E}[\widehat{f}_N(\theta,y,x)]$, we start by writing

$$\mathbb{E}[\widehat{f}_N(\theta,y,x)]=\frac{1}{\mathbb{E}[K_1]}\mathbb{E}\left[K_1\mathbb{E}[b_n^{-1}H_1|X]\right] \text{ with } b_n^{-1}\mathbb{E}\left[H_1|X\right]=\int_{\mathbb{R}}H(t)f(\theta,y-b_nt,X)dt.$$

The latter can be re-written, by using a Taylor expansion under $(H3')$, as follows

$$b_n^{-1}\mathbb{E}[H_1|X]=f(\theta,y,X)+\frac{b_n^2}{2}\left(\int t^2H(t)dt\right)\frac{\partial^2f(\theta,y,X)}{\partial^2y}+o(b_n^2).$$

Thus, we get

$$\mathbb{E}\left[\widehat{f}_N(\theta,y,x)\right]=\frac{1}{\mathbb{E}[K_1]}\left(\mathbb{E}\left[K_1f(\theta,y,X)\right]+\left(\int t^2H(t)dt\right)\mathbb{E}\left[K_1\frac{\partial^2f(\theta,y,X)}{\partial^2y}\right]+o(b_n^2)\right).$$

Let $\psi_l(\cdot,y):=\frac{\partial^lf(\cdot,y,\cdot)}{\partial^ly}$: for $l\in\{0,2\}$, since $\Phi_l(0)=0$, we have

$$\begin{aligned}\mathbb{E}\left[K_1\psi_l(X,y)\right]&=\psi_l(x,y)\mathbb{E}[K_1]+\mathbb{E}\left[K_1\left(\psi_l(X,y)-\psi_l(x,y)\right)\right]\\&=\psi_l(x,y)\mathbb{E}[K_1]+\mathbb{E}\left[K_1\left(\Phi_l(d_\theta(x,X)\right)\right]\\&=\psi_l(x,y)\mathbb{E}[K_1]+\Phi_l'(0)\mathbb{E}\left[d_\theta(x,X)K_1\right]+o(\mathbb{E}\left[d_\theta(x,X)K_1\right]).\end{aligned}$$

So,

$$\begin{aligned}\mathbb{E}\left[\widehat{f}_N(\theta,y,x)\right]&=f(\theta,y,x)+\frac{b_n^2}{2}\frac{\partial^2f(\theta,y,x)}{\partial y^2}\int t^2H(t)dt+o\left(b_n^2\frac{\mathbb{E}\left[d_\theta(x,X)K_1\right]}{\mathbb{E}[K_1]}\right)\\&\quad+\Phi_0'(0)\frac{\mathbb{E}\left[d_\theta(x,X)K_1\right]}{\mathbb{E}[K_1]}+o\left(\frac{\mathbb{E}\left[d_\theta(x,X)K_1\right]}{\mathbb{E}[K_1]}\right).\end{aligned}$$

Similarly to Ferraty *et al.* (2007) we show that

$$\frac{1}{F_\theta(x,a_n)}\mathbb{E}\left[d_\theta(x,X)K_1\right] = C_0 a_n + o(a_n)$$

and

$$\frac{1}{F_\theta(x,a_n)}\mathbb{E}\left[K_1\right] \to C_1.$$

Hence,

$$\mathbb{E}\left[\widehat{f}_N(\theta,y,x)\right] = f(\theta,y,x) + \frac{b_n^2}{2}\frac{\partial^2 f(\theta,y,x)}{\partial y^2}\int t^2 H(t)dt + \Phi_0'(0)\frac{C_0}{C_1}a_n + o(b_n^2) + o(a_n).$$

■

Proof of Lemma 3.3.6. Define

$$Z_{ni}(\theta,y,x) = \frac{\sqrt{F_\theta(x,a_n)}}{\sqrt{nb_n}\mathbb{E}\left[K_1\right]}\left(\Gamma_i(\theta,y,x) - \mathbb{E}\left[\Gamma_i(\theta,y,x)\right]\right),$$

and

$$S_n := \sum_{i=1}^{n} Z_{ni}(\theta,y,x).$$

Therefore,

$$S_n = \sqrt{nb_nF_\theta(x,a_n)}\left(\widehat{f}_N(\theta,y,x) - \mathbb{E}\left[\widehat{f}_N(\theta,y,x)\right]\right).$$

Thus, our claimed result is, now

$$S_n \to \mathcal{N}(0,\sigma_1^2(x)). \tag{3.16}$$

To do that, we use the basic technique of Doob (1953, pages 228-231). Indeed, we consider $p = p_n$ and $q = q_n$ two sequences of natural numbers tending to infinity, such that

$$p = o\left(\sqrt{nb_nF_\theta(x,a_n)}\right), \qquad q = o(p)$$

and we split S_n into

$$S_n = T_n + T_n' + \zeta_k \text{ with } T_n = \sum_{j=1}^{k}\eta_j, \text{ and } T_n' = \sum_{j=1}^{k}\xi_j$$

where

$$\eta_j := \sum_{i\in I_j} Z_{ni}(\theta,y,x), \quad \xi_j := \sum_{i\in J_j} Z_{ni}(\theta,y,x), \quad \zeta_k := \sum_{i=k(p+q)+1}^{n} Z_{ni}(\theta,y,x)$$

with

$$\begin{aligned} I_j &= \{(j-1)(p+q)+1,\ldots,(j-1)(p+q)+p\}, \\ J_j &= \{(j-1)(p+q)+p+1,\ldots,j(p+q)\}. \end{aligned}$$

Observe that, for $k = \left[\frac{n}{p+q}\right]$, (where [.] stands for the integral part), we have $\frac{kq}{n} \to 0$, and $\frac{kp}{n} \to 1$, $\frac{q}{n} \to 0$, which imply that $\frac{p}{n} \to 0$ as $n \to \infty$. Now, our asymptotic result is based on

$$\mathbb{E}(T'_n)^2 + \mathbb{E}(\zeta_k)^2 \to 0 \tag{3.17}$$

and

$$T_n \to \mathcal{N}(0, \sigma_1^2(x)). \tag{3.18}$$

Proof of (3.17) :
By stationarity we get

$$\mathbb{E}(T'_n)^2 = kVar(\xi_1) + 2 \sum_{1\le i<j\le k} |Cov(\xi_i, \xi_j)| \tag{3.19}$$

and

$$kVar(\xi_1) \le qkVar(Z_{n1}(\theta, y, x)) + 2k \sum_{1\le i<j\le q} Cov(Z_{ni}(\theta, y, x), Z_{nj}(\theta, y, x)). \tag{3.20}$$

From (3.10) and the fact that $\frac{kq}{n} \to 0$. we get

$$\begin{aligned} qkVar(Z_{n1}(\theta, y, x)) &= F_\theta(x, a_n) kq \frac{1}{nb_n \left(\mathbb{E}[K_1]\right)^2} Var(\Gamma_1(\theta, y, x)) \\ &= O\left(\frac{kq}{n}\right) \to 0, \quad \text{as } n \to \infty. \end{aligned}$$

On the other hand, we have

$$k \sum_{1\le i<j\le q} |Cov(Z_{ni}(\theta, y, x), Z_{nj}(\theta, y, x))| = \frac{k}{nb_n} \frac{F_\theta(x, a_n)}{\left(\mathbb{E}[K_1]\right)^2} \sum_{1\le i<j\le q} |Cov(\Gamma_i(\theta, y, x), \Gamma_j(\theta, y, x))|.$$

Similarly to (3.11), we obtain,

$$\sum_{1\le i<j\le q} |Cov(\Gamma_i(\theta, y, x), \Gamma_j(\theta, y, x))| = o\left(qb_n F_\theta(x, a_n)\right).$$

Then

$$k\sum_{1\le i<j\le q}|Cov(Z_{ni}(\theta,y,x),Z_{nj}(\theta,y,x))| = o\left(\frac{kq}{n}\right) \quad\rightarrow\quad 0, \quad \text{as } n\to\infty. \tag{3.21}$$

From (3.20)-(3.21) we obtain

$$kVar(\xi_1)\longrightarrow \quad 0, \ \text{as} \quad n\to\infty. \tag{3.22}$$

We use the stationarity, to evaluate the second term in the right-hand side of (3.19)

$$\begin{aligned}\sum_{1\le i<j\le k}|Cov(\xi_i,\xi_j)| &= \sum_{l=1}^{k-1}(k-l)|Cov(\xi_1,\xi_{l+1})|\\ &\le k\sum_{l=1}^{k-1}|Cov(\xi_1,\xi_{l+1})|\\ &\le k\sum_{l=1}^{k-1}\sum_{(i,j)\in J_1\times J_{l+1}} Cov(Z_{ni}(\theta,y,x),Z_{nj}(\theta,y,x)).\end{aligned}$$

It is clear that, for all $(i,j)\in J_1\times J_j$, we have $|i-j|\ge p+1>p$, then

$$\begin{aligned}\sum_{1\le i<j\le k}|Cov(\xi_i,\xi_j)| &\le k\frac{CF_\theta(x,a_n)\left(a_n^{-1}\mathrm{Lip}(K)+b_n^{-1}\mathrm{Lip}(H)\right)^2}{nb_n\left(\mathbb{E}[K_1]\right)^2}\sum_{i=1}^{p}\sum_{j=2p+q+1,|i-j|>p}^{k(p+q)}\lambda_{i,j}\\ &\le \frac{CkpF_\theta(x,a_n)\left(a_n^{-1}\mathrm{Lip}(K)+b_n^{-1}\mathrm{Lip}(H)\right)^2}{nb_n\left(\mathbb{E}[K_1]\right)^2}\lambda_p\\ &\le \frac{Ckp\left(a_n^{-1}\mathrm{Lip}(K)+b_n^{-1}\mathrm{Lip}(H)\right)^2}{nb_nF_\theta(x,a_n)}e^{-\alpha p}\\ &\le \frac{Ckp}{nb_n^3F_\theta^3(x,a_n)}e^{-\alpha p}\to 0.\end{aligned}$$

Finally, by combining this last result and (3.22) we can write

$$\mathbb{E}(T_1')^2\to 0 \ \text{ as } n\to\infty. \tag{3.23}$$

Since $(n-k(p+q)) \leq p$, we have

$$\begin{aligned}
\mathbb{E}(\zeta_k)^2 &\leq (n-k(p+q))Var(Z_{n1}(\theta,y,x)) + 2\sum_{1\leq i<j\leq n} |Cov(Z_{ni}(\theta,y,x), Z_{nj}(\theta,y,x))| \\
&\leq pVar(Z_{n1}(\theta,y,x)) + 2\sum_{1\leq i<j\leq n} |Cov(Z_{ni}(\theta,y,x), Z_{nj}(\theta,y,x))| \\
&\leq \frac{pF_\theta(x,a_n)}{nb_n\left(\mathbb{E}[K_1]\right)^2} Var(\Gamma_1(\theta,y,x)) + \underbrace{\frac{CF_\theta(x,a_n)}{nb_n\left(\mathbb{E}[K_1]\right)^2}\sum_{1\leq i<j\leq n} |Cov(\Gamma_i(\theta,y,x),\Gamma_j(\theta,y,x))|}_{o(1)} \\
&\leq \frac{Cp}{n} + o(1).
\end{aligned}$$

Then,

$$\mathbb{E}(\zeta_k)^2 \to 0, \quad \text{as } n\to\infty$$

which combining with (3.22) completes the proof of (3.17).

Proof of (3.18) :
The proof of convergence in (3.18) is based in the following two results

$$\left|\mathbb{E}\left(e^{it\sum_{j=1}^k \eta_j}\right) - \prod_{j=1}^k \mathbb{E}\left(e^{it\eta_j}\right)\right| \to 0, \tag{3.24}$$

and

$$kVar(\eta_1) \to \sigma_1^2(x), \quad k\mathbb{E}(\eta_1^2 \mathbb{I}_{\{\eta_1>\epsilon\sigma_1(x)\}}) \to 0. \tag{3.25}$$

Proof of (3.24) :

$$\left|\mathbb{E}\left(e^{it\sum_{j=1}^k \eta_j}\right) - \prod_{j=1}^k \mathbb{E}\left(e^{it\eta_j}\right)\right| \leq$$

$$\left|\mathbb{E}\left(e^{it\sum_{j=1}^k \eta_j}\right) - \mathbb{E}\left(e^{it\sum_{j=1}^{k-1} \eta_j}\right)\mathbb{E}\left(e^{it\eta_k}\right)\right| + \left|\mathbb{E}\left(e^{it\sum_{j=1}^{k-1} \eta_j}\right) - \prod_{j=1}^{k-1} \mathbb{E}\left(e^{it\eta_j}\right)\right|$$

$$= \left|Cov\left(e^{it\sum_{j=1}^{k-1} \eta_j}, e^{it\eta_k}\right)\right| + \left|\mathbb{E}\left(e^{it\sum_{j=1}^{k-1} \eta_j}\right) - \prod_{j=1}^{k-1} \mathbb{E}\left(e^{it\eta_j}\right)\right|, \tag{3.26}$$

and successively, we have

$$\left|\mathbb{E}\left(e^{it\sum_{j=1}^{k}\eta_j}\right)-\prod_{j=1}^{k}\mathbb{E}\left(e^{it\eta_j}\right)\right|\leq$$

(3.27)

$$\left|Cov\left(e^{it\sum_{j=1}^{k-1}\eta_j},e^{it\eta_k}\right)\right|+\left|Cov\left(e^{it\sum_{j=1}^{k-2}\eta_j},e^{it\eta_{k-1}}\right)\right|+\ldots+\left|Cov\left(e^{it\eta_2},e^{it\eta_1}\right)\right|.$$

Once again we apply Lemma 3.4.1, to write

$$\left|Cov\left(e^{it\eta_2},e^{it\eta_1}\right)\right| \leq Ct^2\left(a_n^{-1}\text{Lip}(K)+b_n^{-1}\text{Lip}(H)\right)^2\frac{F_\theta(x,a_n)}{nb_n\left(\mathbb{E}[K_1]\right)^2}\sum_{i\in I_1}\sum_{j\in I_2}\lambda_{i,j}.$$

Applying this inequality to each term on the right-hand side of (3.27) we obtain

$$\left|\mathbb{E}\left(e^{it\sum_{j=1}^{k}\eta_j}\right)-\prod_{j=1}^{k}\mathbb{E}\left(e^{it\eta_j}\right)\right|\leq Ct^2\left(a_n^{-1}\text{Lip}(K)+b_n^{-1}\text{Lip}(H)\right)^2\frac{F_\theta(x,a_n)}{nb_n\left(\mathbb{E}[K_1]\right)^2}$$

$$\times\left[\sum_{i\in I_1}\sum_{j\in I_2}\lambda_{i,j}+\sum_{i\in I_1\cup I_2}\sum_{j\in I_3}\lambda_{i,j}+\ldots+\sum_{i\in I_1\cup\ldots\cup I_{k-1}}\sum_{j\in I_k}\lambda_{i,j}\right].$$

Observe that for every $2\leq l\leq k-1$, $(i,j)\in I_l\times I_{l+1}$, we have $|i-j|\geq q+1>q$, then

$$\sum_{i\in I_1\cup\ldots\cup I_{l-1}}\sum_{j\in I_l}\lambda_{i,j}\leq p\lambda_q.$$

Therefore, inequality (3.26) becomes

$$\begin{aligned}\left|\mathbb{E}\left(e^{it\sum_{j=1}^{k}\eta_j}\right)-\prod_{j=1}^{k}\mathbb{E}\left(e^{it\eta_j}\right)\right| &\leq Ct^2\left(a_n^{-1}\text{Lip}(K)+b_n^{-1}\text{Lip}(H)\right)^2\frac{F_\theta(x,a_n)}{nb_n\left(\mathbb{E}[K_1]\right)^2}kp\lambda_q\\ &\leq Ct^2\left(a_n^{-1}\text{Lip}(K)+b_n^{-1}\text{Lip}(H)\right)^2\frac{F_\theta(x,a_n)}{b_n\left(\mathbb{E}[K_1]\right)^2}kpe^{-\alpha q}\\ &\leq Ct^2\left(a_n^{-1}\text{Lip}(K)+b_n^{-1}\text{Lip}(H)\right)^2\frac{1}{nb_nF_\theta(x,a_n)}kpe^{-\alpha q}\\ &\leq Ct^2\frac{kp}{nb_n^3F_\theta^3(x,a_n)}e^{-\alpha q}\to 0,\end{aligned}$$

Proof of (3.25) :
By the same arguments used in (3.19), we have

$$\begin{aligned}\lim_{n\to\infty}kVar(\eta_1) &= \lim_{n\to\infty}kpVar(Z_{n1}(\theta,y,x))\\ &= \lim_{n\to\infty}\frac{kp}{nb_n}\frac{F_\theta(x,a_n)}{(\mathbb{E}[K_1])^2}Var(\Gamma_1(\theta,y,x)).\end{aligned}$$

So, by using the same arguments as those used by Ferraty *et al.* (2007) we get

$$\begin{aligned}
\frac{1}{F_\theta(x,a_n)}\mathbb{E}\left[K_1^2\right] &\longrightarrow K^2(1)-\int_0^1 (K^2(s))'\beta_\theta(x,s))ds+o(1),\\
\frac{\mathbb{E}\left[K_1^2H_1^2\right]}{b_n\mathbb{E}\left[K_1^2\right]} &\longrightarrow f(\theta,y,x)\int H^2(t)dt,\\
\frac{\mathbb{E}[K_1H_1]}{b_n\mathbb{E}\left[K_1\right]} &\longrightarrow f(\theta,y,x)
\end{aligned}$$

which imply that

$$\frac{F_\theta(x,a_n)}{b_n(\mathbb{E}[K_1])^2}Var(\Gamma_1(\theta,y,x))\longrightarrow \sigma_1^2(x). \tag{3.28}$$

Hence

$$kVar(\eta_1)\to\sigma_1^2(x).$$

For the second part of (3.25), we use the fact that $|\eta_1|\leq Cp|Z_{n1}(\theta,y,x)|\leq \dfrac{Cp}{\sqrt{nb_nF_\theta(x,a_n)}}$, and Tchebychev's inequality to get

$$\begin{aligned}
k\mathbb{E}(\eta_1^2\mathbb{1}_{\{\eta_1>\epsilon\sigma_1(x)\}}) &\leq \frac{Cp^2k}{nb_nF_\theta(x,a_n)}\mathbb{P}(\eta_1>\epsilon\sigma_1(x))\\
&\leq \frac{Cp^2k}{nb_nF_\theta(x,a_n)}\frac{Var(\eta_1)}{\epsilon^2\sigma_1^2(x)}\\
&= O\left(\frac{p^2}{nb_nF_\theta(x,a_n)}\right),
\end{aligned}$$

which completes the proof of Lemma 3.3.6. ∎

Bibliographie

[1] Aït Saidi, A., Ferraty, F. & Kassa, R. (2008). *Cross-validated estimation in the single functional index model.* Statistics **42**, 475–494.

[2] Ango Nze, P., Bühlmann, P. & Doukhan, P. (2002). *Weak dependence beyond mixing and asymptotics for non parametric regression.* Ann. Statist. **30-2**, 397–430.

[3] Attaoui, S., Laksaci, A. & Ould-Saïd, E. (2011). *A note on the conditional density estimate in the single functional index model.* Statist. Probab. Lett. **81**, 45–53.

[4] Bulinski A. & Suquet, C. (2001). *Normal approximation for quasi-associated random fields.* Statist. Probab. Lett. **54**, 215–226.

[5] Bulinski A. & Shabanovich, E. (1998). *Asymptotical behaviour of some functionals of positively and negatively dependent random fields. (in Russian).* Fundam. Appl. Math., **4**, 479–492

[6] Dedecker, J., Doukhan, P., Lang, G., León R., J. R., Louhichi, S., and Prieur, C. (2007) *Weak dependence : with examples and applications.*, Lecture Notes in Statistics.

[7] Delecroix, M., Härdle, W., Hristache, M., (2003). *Efficient estimation in conditional single-index regression.* J. Multivariate Anal. **86**, 213–226.

[8] Doob, J.L. (1953). *Stochastic Processes.* John Wiley & Sons, New York.

[9] Douge, L. (2010). Théorèmes limites pour des variables quasi-associées hilbertiennes. Ann. I.S.U.P., **54**, 51–60.

[10] Doukhan, P., Louhichi, S. (1999). *A new weak dependence condition and applications to moment inequalities.* Stoch. Proc. Appl. **84**, 313–342.

[11] Esary, J., Proschan, F. & Walkup, D. (1967). *Association of random variables with applications.* Ann. Math. Statist. **38**, 1466–1476.

[12] Ferraty, F., Peuch, A. & Vieu, P., (2003). *Modèle à indice fonctionnel simple.* C. R. Mathématiques Paris 336, 1025–1028.

[13] Ferraty, F. Mas, A. & Vieu, P., (2007). *Advances in nonparametric regression for functional variables.* Aust. and New Zeal. J. of Statist., **49**, 1-20.

[14] Ferraty, F., Park, J., Vieu, P. (2011). *Estimation of a functional single index model.* In : F, Ferraty (Ed.) Recent advences in functional data analysis and related topics. Contribution to statistics. Physica.

[15] Fortuyn, C., Kastelyn, P. & Ginibre, J. (1971). *Correlation inequalities in some partially ordered sets.* Comm. Math. Phys. **74**, 119–127.

[16] Gasser, T., Hall, P. & Presnell, B. (1998). *Nonparametric estimation of the mode of a distribution of random curves.* J. Roy. Statist. Soc. Ser. B, **60**, 681–691.

[17] Härdle W., Hall P. & Ichimura H. (1993). *Optimal smoothing in single-index models.* Ann. Statist. **21**, 157-178.

[18] Horowitz, J. & Härdle, W. (1996). *Direct semiparametric estimation of single index models with discrete covariates.* J. Assoc. Statist. Amer. **91**, 16321641.

[19] Hristache, M., Juditsky, A. & Spokoiny, V. (2001). *Direct estimation of the index coefficient in the single-index model.* Ann. Statist. **29**, 595–623.

[20] Jong-Dev, K., & Proschan, F. (1983). *Negative association of random variables, with applications.* Ann. Statist. **11**, 286–295.

[21] Kallabis, R. S & Neumann, M. H. (2006). *An exponential inequality under weak dependence.* Bernoulli **12**, 333–350.

[22] Masry, E. (2005). *Nonparametric regression estimation for dependent functional data : asymptotic normality.* Stoch. Process. Appl. **115**, 155–177.

[23] Matula, P. (1992). *A note on the almost sure convergence of sums of negatively dependent random variables.* Statist. Probab. Lett. **15**, 209–213.

[24] Newman, C. M. (1984). *Asymptotic independence and limit theorems for positively and negatively dependent random variables :* In Inequalities in Statistics and Probability, IMS Lect. Notes–Monographs Series, **5**, 127–140.

[25] Pitt L. (1982). *Positively correlated normal variables are associated.* Ann. Probab., **10**, 496–499.

[26] Roussas, G. G. (1999). *Positive and negative dependence with some statistical applications.* In : Asymptotics, Nonparametrics and Time Series (S. Ghosh, Ed.)., pp. 757–788. Marcell Dekker, Inc., New York.

[27] Roussas, G. G. (2000). *Prediction under association.* In : Limnios, N., Nikulin, M.S. (Eds.)., Recent Advances in Reliability Theory : Methodology, Practice and Inference. BirkhSauser, Boston.

[28] Shashkin, A. P (2002). *Quasi-association for system of Gaussian vectors.* Uspehi Ma. Nauk.

Chapitre 4

Robust regression

Strong uniform consistency of the robust nonparametric regression estimation for quasi-associated vectorial processes [1]

Said Attaoui, Ali Laksaci and Elias Ould Said

Abstract In this work we study a family of robust nonparametric estimators for a regression function based on a kernel method when the data exhib the quasi-associated dependency condition. Under general conditions and based on the exponential inequality given by Kallabis and Neumann, we establish the uniform almost complete convergence of these estimators.

Key words : Almost complete convergence. Uniform convergence. Robust nonparametric estimation. Exponential inequality. Weak dependence. Quasi-associated variables.

1. Preprint, ULCO, LMPA, N^o 479, September 2012, Submitted.

4.1 Introduction

Since a long time, mixing processes has the dominated type of dependency. However, mixing conditions may be difficult to be checked for particular models or are too strong to be true. In other words, some classes of process have many applications in statistics, have not the mixing property. The framework describing dependence structure discussed in this paper is of type quasi-association. Nothing that the concept of quasi-association was introduced by Bulinski and Suquet (2001) for real-valued random fields, and it generalizes the positively associated variables introduced by Esary *et al.* (1967) applied to reliability of systems (see the book by Barlow and Proschan (1975)) and the negatively associated random introduced by Jong-Dev and Proschan (1983), this latter type of association has also found a number of applications in certain fields.

Nonparametric estimation involving (positively and negatively) associated random variables has been studied extensively and more asymptotic property was obtained. Many papers was devoted to the nonparametric kernel estimation, for further motivation and background material, the reader is referred to Roussas (1991, 1993, 2000), Matula (1992), Masry (2002), ...
The main contribution of Doukhan and Louhichi (1999) is by considering of a new concept of weak dependence and developed some applications in Statistics, Stochastic Algorithms, Resampling, and Econometry (we refer the reader to Dedecker *et al* (2007) or Doukhan *et al* (2010) monograph's). In our work we investigate the idea that quasi-associated variables can be considered as a special case of weak dependence.

When dealing with the prediction problem, the regression function is an interesting problem in statistical inference, and it plays an important role in many applications. The literature, dealing with regression for independent and dependent cases, is quite extensive. For a comprehensive review, see, for example, Györfi *et al.* (1989), Roussas (1990), and Fan and Gijbels (1996). It is well known, that there are others alternatives to the classical regression prediction in presence of outliers. The main idea is to use the robust estimator. Since the pioneer paper of Huber (1964), there are many papers, we quote without pretend to the exhaustivity, Collomb and Härdle (1986), Härdle and Tsybakov(1988), Boente and Fraiman (1989) and Laïb and Ould Saïd (2000) for a finite dimensional space and Azzedine *et al.* (2008) for infinite dimensional space. As far as we know, there is no contribution on the nonparametric robust regression for quasi-associated variables.

The goal of this paper, is to study of a family of nonparametric robust estimator of the regression function, based on the kernel method. Under general conditions, we establish the uniform almost complete convergence of these estimators by using the exponential inequality. Recall that, for the weak dependency, a first exponential inequality was obtained by Doukhan and Louhichi (1999), a Bennett inequality by Dedecker and Prieur (2004), and a Bernstein-type inequality by Kallabis and Neumann (2006). In our proof, we use the latest. Recall that, for the real-valued data, Mack and Silverman (1982) obtained the weak and strong uniform convergence rates of a kernel regression function. For the multivariate case, Liebscher (2001) extended the results of Ango Nze *et al.* (1993), and Bosq (1998) obtained rates of the strong uniform convergence, under the α-mixing and β-mixing sequences for the density and the regression function. Later Ango Nze *et al.* (2002) study the nonparametric regression estimation when the data exhib the weak dependency condition.

In order to define the quasi-associated processes, we need to introduce the following notations.
A sequence $(X_n)_{n\in\mathbb{N}}$ of real random vectors variables is said to be quasi-associated, if for any disjoint subsets I and J of $\mathbb{N}$ and all bounded Lipschitz functions $f : \mathbb{R}^{|I|d} \to \mathbb{R}$ and $g : \mathbb{R}^{|J|d} \to \mathbb{R}$ satisfying :
(4.1)

$$|Cov(f(X_i,\, i\in I),\, g(X_j,\, j\in J))| \le \mathrm{Lip}(f)\mathrm{Lip}(g)\sum_{i\in I}\sum_{j\in J}\sum_{k=1}^{d}\sum_{l=1}^{d}\left|Cov(X_i^k, X_j^l)\right|$$

(here and in the sequel $|I|$ denotes cardinality of a finite set I.)
where X_i^k denotes the k^{th} component of X_i, and

$$\mathrm{Lip}(f) = \sup_{x\neq y}\frac{|f(x)-f(y)|}{\|x-y\|_1}, \ \text{ with } \ \|(x_1,\ldots,x_k)\|_1 = |x_1| + \ldots + |x_k|.$$

The paper is as follows : The model considered here will be given and described in the next Section. In Section 3, we give some notations and hypotheses which permit to give our main result and some Lemmata. In Section 4, we give application to prediction for Markov process. Finally, the last Section is devoted to the proof of Lemmata.

4.2 Model

Let $\{Z_i = (X_i, Y_i),\ 1 \le i \le n\}$ be n stationary quasi-associated random processes, identically distributed as the random pair $Z = (X, Y)$ with values in $\mathbb{R}^d \times \mathbb{R}$. For any $x \in \mathbb{R}^d$, we consider ψ_x a real-valued Borel function satisfying some regularity conditions to be stated below. The nonparametric model studied in this paper, denoted by θ_x, is implicitly defined as a zero with respect to (w.r.t) t of the following equation

$$\Psi(x,t) = \mathbb{E}(\psi_x(Y,t)|X=x) = 0. \tag{4.2}$$

We suppose that, for all $x \in \mathbb{R}^d$, θ_x exists and is the unique zero w.r.t t of (4.2). The kernel estimator of the function $\Psi(x,t)$ noted $\widehat{\Psi}(x,t)$ is given by :

$$\widehat{\Psi}(x,t) = \frac{\sum_{i=1}^{n} K\Big(h_n^{-1}(x-X_i)\Big)\psi_x(Y_i,t)}{\sum_{i=1}^{n} K\Big(h_n^{-1}(x-X_i)\Big)} \quad \forall t \in \mathbb{R} \tag{4.3}$$

where K is a kernel function and $h_n := h_{n,K}$ (for simplicity) is a sequence of bandwidths converging to zero as n goes to infinity. A natural estimator of θ_x denoted by $\widehat{\theta}_x$, is a zero w.r.t t of

$$\widehat{\Psi}(x,t) = 0. \tag{4.4}$$

4.3 Main results

Our purpose in this Section, is to establish the uniform almost complete convergence on some compact set $\mathcal{C}$ of $\mathbb{R}^d$. Let λ_k the covariance coefficient defined as :

$$\lambda_k := \sup_{s \geq k} \sum_{|i-j| \geq s} \lambda_{i,j},$$

where

$$\lambda_{i,j} = \sum_{k=1}^{d}\sum_{l=1}^{d} |Cov(X_i^k, X_j^l)| + \sum_{k=1}^{d} |Cov(X_i^k, Y_j)| + \sum_{l=1}^{d} |Cov(Y_i, X_j^l)| + |Cov(Y_i, Y_j)|.$$

X_i^k denotes the k^{th} component of X_i.
In the following, we fixed x a point in $\mathbb{R}^d$, $\mathcal{N}_x$ a fixed neighborhood of x, and throughout the paper, when no confusion will be possible, we will denote by C or/and C' some strictly positive generic constants whose values are allowed to change, and we considers the following assumptions :

(U1) The density f of the variable X is positif function, of class C^p such that :

$$\sup_{|i-j|\geq 1} \|f_{(X_i,X_j)}\|_\infty < \infty.$$

(U2) The random pair $\{(X_i, Y_i),\ i \in \mathbb{N}\}$ is quasi-associated with covariance coefficient λ_k, $k \in \mathbb{N}$ checked

$$\exists a > 0, \quad \exists C > 0, \text{ such that } \lambda_k \leq Ce^{-ak}.$$

Note that any family of independent real-valued random variables is automatically associated (Newman (1984)) and negatively associated (Joag-Dev and Proschan (1983)). There are interesting stochastic models in mathematical statistics, reliability theory, percolation theory and statistical mechanics described by families of positively or negatively associated random variables, see Bulinski and Suquet (2001). Bulinski and Shabanovich (1998) showed that, any positively or negatively associated collections of random variables with finite second moment satisfy (4.1). Consequently, such fields are quasi-associated. In Shashkin (2002), it was proved that any Gaussian random field $X = \{X_t,\ t \in T\}$ with values in $\mathbb{R}^p$, is quasi-associated. A real-valued Gaussian random field is associated if and only if its covariance function is nonegative (Pitt (1982)) and negatively associated if and only if $Cov(X_s, X_t) \leq 0$, $\forall s \neq t$ (Joag-Dev and Proschan (1983)).

With results of Shashkin (2002), we can conclude that the concept of quasi-association is strictly wider than that of positive or negative association for random fields with finite second moments.

(U3) The function Ψ is such that :

$$\begin{cases} \text{(U3a) The function } \Psi(x,.) \text{ is of class } C^p \text{ on } [\theta_x - \delta_0, \theta_x + \delta_0],\ \delta_0 > 0, \\ \text{(U3b) For each fixed } t \in [\theta_x - \delta_0, \theta_x + \delta_0], \text{ the function } \Psi(.,t) \text{ is continous at the point } x. \end{cases}$$

(U4) K is continuous lipschitz function of order p, with compact support.

$$\text{(U5)} \begin{cases} \text{(U5a)} \lim_{n\to\infty} h_n = 0,\ \lim_{n\to\infty} \dfrac{\log^5 n}{nh_n^d} = 0. \\ \text{(U5b)} \lim_{n\to\infty} n^\delta h_n = \infty, \text{ for some } 0 < \delta < 1. \end{cases}$$

Comments on the hypotheses.

Our assumptions are classical in a multivariate nonparametric estimation. The condition (U1) relate the density of the explanatory variable X, is indispensable to have a good majoration of covariance's and which enables us to obtain our convergence rate. In order to establish the almost complete convergence rate of our model under quasi-association properties, we need to the hypotheses (U2), when our convergence rate is the same convergence obtained in multivariate estimation for independence or mixing data (see Bosq (1998)). The robustness of our model is linked to the function ψ_x in hypotheses (U3) when we kept the same condition given by Collomb and Härdle (1986) in the multivariate case. The rest of hypotheses are technical conditions imposed for concision of proofs.

Our main result is given in the following theorem.

Theorem 4.3.1 *Under the hypotheses (U1)-(U5), the estimator $\widehat{\theta}_x$ exist and unique, for n enough large, we have,*

$$\sup_{x\in\mathcal{C}}\left|\widehat{\theta}_x-\theta_x\right| = O\left(h_n^p+\sqrt{\frac{\log n}{nh_n^d}}\right) \qquad a.co. \tag{4.5}$$

Proof.
For the proof of Theorem 4.3.1, we use the fact ψ is strictly monotone and continuous w.r.t. to the second component, therefore, we can write, under this consideration, for all $\epsilon>0$

$$\Psi(x,\theta_x-\epsilon)\le\Psi(x,\theta_x)=0\le\Psi(x,\theta_x+\epsilon) \text{ and } \widehat{\Psi}(x,\widehat{\theta}_x-\epsilon)\le\widehat{\Psi}(x,\widehat{\theta}_x)=0\le\widehat{\Psi}(x,\widehat{\theta}_x+\epsilon).$$

Hence, for all $\epsilon>0$, we have

$$\begin{aligned}\mathbb{P}\left(\left|\widehat{\theta}_x-\theta_x\right|\ge\epsilon\right) \;\le\; & \mathbb{P}\left(\left|\Psi(x,\theta_x+\epsilon)-\Psi(x,\theta_x+\epsilon)\right|\ge\Psi(x,\theta_x+\epsilon)\right)\\ & +\mathbb{P}\left(\left|\Psi(x,\theta_x-\epsilon)-\Psi(x,\theta_x-\epsilon)\right|\ge-\Psi(x,\theta_x-\epsilon)\right).\end{aligned}$$

Moreover, under (U3a), we get for all $x\in\mathcal{C}$

$$\widehat{\theta}_x-\theta_x=\frac{\widehat{\Psi}(x,\widehat{\theta}_x)-\Psi(x,\widehat{\theta}_x)}{\widehat{\Psi}'(x,\theta_x^*)},$$

for some $\theta_x^*\in(\widehat{\theta}_x,\theta_x)$. Thus, we can write

$$\exists\delta_0>0,\ \text{ such that } \sup_{x\in\mathcal{C}}|\widehat{\theta}_x-\theta_x|\mathbb{1}_{\{|\widehat{\theta}_x-\theta_x|\le\delta_0\}}=O\left(\sup_{x\in\mathcal{C}}\sup_{t\in[-\delta_0,\delta_0]}|\widehat{\Psi}(x,t)-\Psi(x,t)|\right)$$

So, to show Theorem 4.3.1 it suffices to prove that

$$\sup_{x\in\mathcal{C}}\sup_{t\in[-\delta_0,\delta_0]}|\widehat{\Psi}(x,t)-\Psi(x,t)|=O\left(h_n^p+\sqrt{\frac{\log n}{nh_n^d}}\right) \qquad a.co. \tag{4.6}$$

The proof of (4.6) is based on the following decomposition, thus one has for all $t \in [\theta_x - \delta_0, \theta_x + \delta_0]$:

$$\begin{aligned}
\sup_{x\in\mathcal{C}} \sup_{t\in[-\delta_0,\delta_0]} \left|\widehat{\Psi}(x,t) - \Psi(x,t)\right| &\leq \frac{1}{\inf_{x\in\mathcal{C}} \left|\widehat{f}_D(x)\right|} \Big\{ \sup_{x\in\mathcal{C}} \sup_{t\in[-\delta_0,\delta_0]} \left|\widehat{f}_N(x,t) - \mathbb{E}\left[\widehat{f}_N(x,t)\right]\right| \\
&+ \sup_{x\in\mathcal{C}} \sup_{t\in[-\delta_0,\delta_0]} \left|\mathbb{E}\left[\widehat{f}_N(x,t)\right] - H(x,t)\right| \\
&+ \sup_{x\in\mathcal{C}} \sup_{t\in[-\delta_0,\delta_0]} \left|\Psi(x,t)\left(f(x) - \mathbb{E}\left[\widehat{f}_D(x)\right]\right)\right| \\
&+ \sup_{x\in\mathcal{C}} \sup_{t\in[-\delta_0,\delta_0]} \left|\Psi(x,t)\left(\mathbb{E}\left[\widehat{f}_D(x)\right] - \widehat{f}_D(x)\right)\right| \Big\}.
\end{aligned}$$

where

$$H(x,t) = \Psi(x,t)f(x) \text{ and } \sup_{x\in\mathcal{C}} |H(x,t)| < \infty.$$

Finally, we need the following Lemmas and Corollaries which their proofs are relegated to the Proofs Section.

Lemma 4.3.1 *Under hypotheses (U1)-(U5). Then we have for all* $t \in [\theta_x - \delta_0, \theta_x + \delta_0]$:

$$\sup_{x\in\mathcal{C}} \sup_{t\in[-\delta_0,\delta_0]} \left|\widehat{f}_N(x,t) - \mathbb{E}\left[\widehat{f}_N(x,t)\right]\right| = O\left(\sqrt{\frac{\log n}{nh_n^d}}\right) \qquad a.co.$$

Lemma 4.3.2 *Under hypothesis (U1) and (U3b) we have :*

$$\sup_{x\in\mathcal{C}} \sup_{t\in[-\delta_0,\delta_0]} \left|\mathbb{E}\left[\widehat{f}_N(x,t)\right] - H(x,t)\right| = O(h_n^p).$$

Lemma 4.3.3 *Under hypotheses of Lemma 4.3.1, we have for all* $t \in [\theta_x - \delta_0, \theta_x + \delta_0]$:

$$\sup_{x\in\mathcal{C}} \left|\mathbb{E}\left[\widehat{f}_D(x)\right] - \widehat{f}_D(x)\right| = O\left(\sqrt{\frac{\log n}{nh_n^d}}\right) \qquad a.co.$$

Lemma 4.3.4 *Under Lemma 4.3.2, we have for all* $t \in [\theta_x - \delta_0, \theta_x + \delta_0]$:

$$\sup_{x\in\mathcal{C}} \left|f(x) - \mathbb{E}\left[\widehat{f}_D(x)\right]\right| = O(h_n^p).$$

Corollary 4.3.1 *Under Lemma 4.3.3, we have :*

$$\exists c > 0 \qquad \sum_{n=1}^{\infty} \mathbb{P}\Big(\inf_{x\in\mathcal{C}} \Big| \widehat{f}_D(x) \Big| \leq c \Big) < \infty.$$

Lemma 4.3.5 *Under the hypotheses of Theorem 4.3.1, $\widehat{\theta}_x$ exist and unique almost sure, for all n enough large.*

Remark 4.3.1 *Under the hypotheses of Theorem 4.3.1, if in addition $h_n \sim \left(\frac{\log n}{n}\right)^{1/(2p+d)}$, then we have,*

$$\sup_{x\in\mathcal{C}} |\widehat{\theta}_x - \theta_x| = O\left(\left(\frac{\log n}{n} \right)^{p/(2p+d)} \right) \qquad a.co. \tag{4.7}$$

4.4 Application to Markov process prediction

Let ζ_i, $i \in \mathbb{Z}$ be a $\mathbb{R}^{d_0}$-valued strictly stationary process. Suppose that (ζ_i) is a Markov process of order k, namely

$$\mathbb{E}(G(\zeta_i)|\zeta_{i-s},\, s \geq 1) = \mathbb{E}(G(\zeta_i)|\zeta_{i-1}, \dots, \zeta_{i-k}), \qquad a.s \tag{4.8}$$

for each Borelian real function G such that $\mathbb{E}(|G(\zeta_i)|) < +\infty$.
Given the data $\zeta_1, \dots, \zeta_N$. The goal is to predict the non-observed real random variable

$$\xi_{N+N'} = \theta(\zeta_{N+N'}),$$

where $1 \leq N' \leq N-k$ and θ is defined as solution w.r.t of t for the following equation :

$$\Psi(x,t) = \mathbb{E}(\psi_x(\zeta_{k-1+N'}, t)|(\zeta_0, \dots, \zeta_{k-1}) = x),\; x \in \mathbb{R}^d.$$

For that purpose let us construct the quasi-associated process

$$Z_i = (X_i, Y_i) = ((\zeta_i, \dots, \zeta_{i+k-1}), \theta(\zeta_{i+k-1+N'})),\; t \in \mathbb{Z}$$

and consider the robust estimator $\widehat{\theta}$ based on the data $(Z_i,\, 1 \leq i \leq n)$ where $n = N - (k - 1 + N')$. In the present case $d = kd_0$, $d_0 \geq 1$.
From $\widehat{\theta}$ we construct the predictor

$$\widehat{\xi}_{N+N'} = \widehat{\theta}(X_{n+N'}); \tag{4.9}$$

where $\widehat{\theta}$ is defined as solution w.r.t of t for the estimator $\widehat{\Psi}(x,t)$ of $\Psi(x,t)$.
Nothing that when $\psi_x(y,t) = y - t$, the estimator $\widehat{\theta}$ present the kernel estimator of the classical regression. Bosq (1998) was establish the quadratic mean convergence and the almost convergence rate of the predictor $\widehat{\xi}_{N+N'}$ under α-mixing data. For more detail of other nonparametric asymptotic results of $\widehat{\xi}_{N+N'}$, see Bosq and Blanke (2007).

4.5 Auxillary results and proofs

Proof of Lemma 4.3.1. By the compactness of $\mathcal{C}$ we write

$$\mathcal{C} \subset \bigcup_{k=1}^{d_n} B(x_k, \tau_n),$$

with $d_n \leq n^{\beta}$, where $\beta = \frac{\delta(d+2)}{2} + \frac{1}{2}$ and $\tau_n \leq d_n^{-1}$. Now, for all $x \in \mathcal{C}$, we pose $k(x) = \arg\min_{k\in\{1,\dots d_n\}} \|x - x_k\|$, and we can consider the following decomposition

$$\begin{aligned}
\sup_{x\in\mathcal{C}} \sup_{t\in[-\delta_0,\delta_0]} \left|\widehat{f}_N(x,t) - \mathbb{E}\Big[\widehat{f}_N(x,t)\Big]\right| \leq & \underbrace{\sup_{x\in\mathcal{C}} \sup_{t\in[-\delta_0,\delta_0]} \left|\widehat{f}_N(x,t) - \widehat{f}_N(x_{k(x)},t)\right|}_{T_1} \\
& + \underbrace{\sup_{x\in\mathcal{C}} \sup_{t\in[-\delta_0,\delta_0]} \left|\widehat{f}_N(x_{k(x)},t) - \mathbb{E}\left[\widehat{f}_N(x_{k(x)},t)\right]\right|}_{T_2} \\
& + \underbrace{\sup_{x\in\mathcal{C}} \sup_{t\in[-\delta_0,\delta_0]} \left|\mathbb{E}\Big[\widehat{f}_N(x_{k(x)},t)\Big] - \mathbb{E}\Big[\widehat{f}_N(x),t)\Big]\right|}_{T_3}.
\end{aligned}$$

• Concerning T_1 and T_3 : The Lipschitz's condition of the kernel K in (U4) and the boundedness condition of the function ψ in (U3) allows to write directly, for all $x \in \mathcal{C}$, and $\forall t \in [\theta_x - \delta_0, \theta_x + \delta_0]$

$$\begin{aligned}
\left|\widehat{f}_N(x,t) - \widehat{f}_N(x_{k(x)},t)\right| &= \frac{1}{nh_n^d}\left|\sum_{i=1}^n \psi(Y_i,t)K_i(x) - \sum_{i=1}^n \psi(Y_i,t)K_i(x_{k(x)})\right| \\
&\leq \frac{C}{h_n^{d+1}}\|x - x_{k(x)}\| \\
&\leq \frac{C\tau_n}{h_n^{d+1}}
\end{aligned}$$

Since $\tau_n = O(n^{-\beta})$, we have

$$\sup_{x\in\mathcal{C}} \sup_{t\in[-\delta_0,\delta_0]} \left| \widehat{f}_N(x,t) - \widehat{f}_N(x_{k(x)},t) \right| = O\left(\sqrt{\frac{\log n}{nh_n^d}}\right) \qquad a.co. \tag{4.10}$$

and

$$\sup_{x\in\mathcal{C}} \sup_{t\in[-\delta_0,\delta_0]} \left| \mathbb{E}\left[\widehat{f}_N(x_{k(x)},t)\right] - \mathbb{E}\left[\widehat{f}_N(x,t)\right] \right| = O\left(\sqrt{\frac{\log n}{nh_n^d}}\right) \qquad a.co. \tag{4.11}$$

• Concerning T_2 : Using the compactness of $[-\delta_0,\, \delta_0]$, we can write

$$[-\delta_0,\, \delta_0] \subset \bigcup_{j=1}^{z_n} (t_j - l_n, t_j + l_n) \tag{4.12}$$

with $l_n = n^{-1/2p}$ and $z_n = O\left(n^{1/2p}\right)$. We consider the intervals extremities grid

$$\mathcal{G}_n = \{t_j - l_n, t_j + l_n, 1 \le j \le z_n\}\,. \tag{4.13}$$

Then the monotony of $\mathbb{E}[\widehat{f}_N(x,\cdot)]$ and $\widehat{f}_N(x,\cdot)$ gives, for $1 \le j \le d_n$

$$\mathbb{E}\left[\widehat{f}_N(x_{k(x)},t_j - l_n)\right] \le \sup_{t\in(t_j-l_n,t_j+l_n)} \mathbb{E}\left[\widehat{f}_N(x_{k(x)},t)\right] \le \mathbb{E}\left[\widehat{f}_N(x_{k(x)},t_j + l_n)\right]$$

$$\widehat{f}_N(x_{k(x)},t_j - l_n) \le \sup_{t\in(t_j-l_n,t_j+l_n)} \widehat{f}_N(x_{k(x)},t) \le \widehat{f}_N(x_{k(x)},t_j + l_n). \tag{4.14}$$

Now, from (U3a) we have, for any $t_1, t_2 \in [-\delta_0, \delta_0]$

$$\left| \mathbb{E}\left[\widehat{f}_N(x_{k(x)},t_1)\right] - \mathbb{E}\left[\widehat{f}_N(x_{k(x)},t_2)\right] \right| \le C|t_1 - t_2|^p. \tag{4.15}$$

So, we deduce from (4.12)–(4.15) that

$$\sup_{x\in\mathcal{C}} \sup_{t\in[-\delta_0,\delta_0]} \left| \widehat{f}_N(x_{k(x)},t) - \mathbb{E}\left[\widehat{f}_N(x_{k(x)},t)\right] \right| \le \max_{1\le k\le d_n} \max_{1\le j\le z_n} \max_{z\in\{t_j-l_n,t_j+l_n\}} \left| \widehat{f}_N(x,z) - \mathbb{E}\left[\widehat{f}_N(x,z)\right] \right| + 2Cl_n^p. \tag{4.16}$$

It is clear that

$$l_n^p = o\left(\sqrt{\frac{\log n}{nh_n^d}}\right). \tag{4.17}$$

Then, it suffices to prove that for some positive real η sufficiently large

$$\mathbb{P}\left(\max_{1\le k\le d_n}\max_{1\le j\le z_n}\max_{z\in\{t_j-l_n,t_j+l_n\}}\left|\widehat{f}_N(x,z)-\mathbb{E}\left[\widehat{f}_N(x,z)\right]\right|>\eta\sqrt{\frac{\log n}{nh_n^d}}\right) \quad < \quad \infty$$

Observe that

$$\mathbb{P}\left(\max_{1\le k\le d_n}\max_{1\le j\le z_n}\max_{z\in\{t_j-l_n,t_j+l_n\}}\left|\widehat{f}_N(x_{k(x)},z)-\mathbb{E}\left[\widehat{f}_N(x_{k(x)},z)\right]\right|>\eta\sqrt{\frac{\log n}{nh_n^d}}\right)$$

$$\le 2d_nz_n\max_{k\in\{1,\ldots d_n\}}\max_{j\in\{1,\ldots z_n\}}\max_{z\in\{t_j-l_n,t_j+l_n\}}\mathbb{P}\left(\left|\widehat{f}_N(x_{k(x)},z))-\mathbb{E}\left[\widehat{f}_N(x_{k(x)},z))\right]\right|>\eta\sqrt{\frac{\log n}{nh_n^d}}\right)$$

To prove the probability in the right side of the previous inequality, we need to the Bernstein-type inequality for a dependent random variables (see Kallabis and Neumann (2006)). Indeed, we put for all $i=1,\ldots,n$: $\Gamma_i(x_{k(x)})=\psi(Y_i,z)K_i(x_{k(x)})$. Then we have :

$$\begin{aligned}
\widehat{f}_N(x_{k(x)},z))-\mathbb{E}\left[\widehat{f}_N(x_{k(x)},z))\right] &= \frac{1}{nh_n^d}\sum_{i=1}^n\Gamma_i(x_{k(x)})-\mathbb{E}\Big[\frac{1}{nh_n^d}\sum_{i=1}^n\Gamma_i(x_{k(x)})\Big]\\
&= \frac{1}{nh_n^d}\sum_{i=1}^n\chi(X_i,Y_i)\\
&= \sum_{i=1}^n\Delta_i.
\end{aligned}$$

Where

$$\Delta_i=\frac{1}{nh_n^d}\chi(X_i,Y_i),$$

and

$$\chi(u,v)=\psi(v,z)K(h_n^{-1}(x_{k(x)}-u))-\mathbb{E}\left[\psi(Y,z)K(h_n^{-1}(x_{k(x)}-X))\right],\ u\in\mathbb{R}^d,\ v\in\mathbb{R}.$$

Clearly, we have $\mathbb{E}(\Delta_i)=0$, and moreover, we can write $\|\chi\|_\infty\le 2C\|K\|_\infty$, $\mathrm{Lip}(\chi)\le 2C(\mathrm{Lip}(\psi)+h_n^{-1}\mathrm{Lip}(K))$.

Now, by the fact that the terms of variance and covariance intervening in the application of inequality of Kallabis and Newman, we need to evaluate the variance term $Var\left(\sum_{i=1}^n\Delta_i\right)$ and the covariance term $Cov(\Delta_{s_1}\ldots\Delta_{s_u},\Delta_{t_1}\ldots\Delta_{t_v})$, for all $(s_1,\ldots,s_u)\in\mathbb{N}^u$ and $(t_1,\ldots,t_v)\in\mathbb{N}^v$ with $1\le s_1\le\ldots\le s_u\le t_1\le\ldots\le t_v\le n$.

Firstly for the variance term, we have

(4.18)

$$nh_n^dVar\left(\sum_{i=1}^n\Delta_i\right)=\frac{1}{h_n^d}Var\left(\Gamma_1(x_{k(x)})\right)+\frac{1}{nh_n^d}\sum_{i=1}^n\sum_{\substack{j=1\\j\ne i}}^n Cov(\Gamma_i(x_{k(x)}),\Gamma_j(x_{k(x)})).$$

We have under (U1) and (U4)

$$\begin{aligned} Var\left(\Gamma_1(x_{k(x)})\right) &\leq \mathbb{E}[|\psi(Y_1,z)K_1(x_{k(x)})|^2] \leq C\mathbb{E}[|K_1(x_{k(x)})|^2] \\ &\leq Ch_n^d \int_{\mathbb{R}^d} K^2(u) f(x_{k(x)} - h_n u) du \\ &\to Ch_n^d f(x_{k(x)}) \int_{\mathbb{R}^d} K^2(u) du; \quad \text{as } n \text{ go to } \infty. \end{aligned} \tag{4.19}$$

Now, let us evaluate the asymptotic behavior of the sum in the right hand side of (4.18). For this we can us the technique of Masry (1986). Indeed, we will need the following decomposition

$$\begin{aligned} \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} Cov(\Gamma_i(x_{k(x)}),\Gamma_j(x_{k(x)})) &= \sum_{i=1}^{n}\sum_{\substack{j=1\\ 0<|i-j|\leq m_n}}^{n} Cov(\Gamma_i(x_{k(x)}),\Gamma_j(x_{k(x)})) \\ &+ \sum_{i=1}^{n}\sum_{\substack{j=1\\ |i-j|>m_n}}^{n} Cov(\Gamma_i(x_{k(x)}),\Gamma_j(x_{k(x)})). \end{aligned}$$

where (m_n) is a sequence of positive integer go to ∞ as $n \to \infty$.
It follows from Assumptions (U1) and (U4) for $i \neq j$, that

$$\begin{aligned} \mathbb{E}\left[\left|\Gamma_i(x_{k(x)})\Gamma_j(x_{k(x)})\right|\right] &= \mathbb{E}\left[\left|\psi(Y_i,z)K_i(x_{k(x)})\psi(Y_j,z)K_j(x_{k(x)})\right|\right] \\ &\leq C\mathbb{E}\left[\left|K_i(x_{k(x)})K_j(x_{k(x)})\right|\right] \leq \mathbb{E}\left|K\left(h_n^{-1}(x_{k(x)}-X_i)\right)K\left(h_n^{-1}(x_{k(x)}-X_j)\right)\right| \\ &\leq Ch_n^{2d}\int_{\mathbb{R}^d} K(u)K(v)\left|f_{(X_i,X_j)}(x_{k(x)}-h_n u, x_{k(x)}-h_n v)\right| dudv \\ &\leq Ch_n^{2d}, \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}\left[\left|\Gamma_i(x_{k(x)})\right|\right] &= \mathbb{E}\left[\left|\psi(Y_i,z)K_i(x_{k(x)})\right|\right] \\ &\leq C\mathbb{E}\left[\left|K_i(x_{k(x)})\right|\right] \leq C\mathbb{E}\left|K\left(h^{-1}(x_{k(x)}-X_i)\right)\right| \\ &\leq Ch_n^d \int_{\mathbb{R}^d} K(u)\left|f_{X_i}(x_{k(x)}-h_n u)\right| du \\ &\leq Ch_n^d. \end{aligned}$$

Then, we get

$$\begin{aligned} \sum_{i=1}^{n}\sum_{\substack{j=1\\ 0<|i-j|\leq m_n}}^{n} Cov(\Gamma_i(x_{k(x)}),\Gamma_j(x_{k(x)})) &\leq nm_n\left(\mathbb{E}[\Gamma_i(x_{k(x)})\Gamma_j(x_{k(x)})] + \left(\Gamma_1(x_{k(x)})\right)^2\right). \\ &\leq Cnm_n h_n^{2d}. \end{aligned}$$

By quasi-association of the sequence (X_i, Y_i) and (U2), we can write

$$\begin{aligned}\sum_{i=1}^{n}\sum_{\substack{j=1\\|i-j|>m_n}}^{n} Cov(\Gamma_i(x_{k(x)}),\Gamma_j(x_{k(x)})) &\leq Ch_n^{-2}\sum_{i=1}^{n}\sum_{\substack{j=1\\|i-j|>m_n}}^{n}\lambda_{i,j}\\ &\leq Cnh_n^{-2}\lambda_{m_n}\\ &\leq Cnh_n^{-2}e^{-am_n},\end{aligned}$$

So,

$$\frac{1}{nh_n^d}\sum_{i=1}^{n}\sum_{\substack{j=1\\i\neq j}}^{n} Cov(\Gamma_i(x_{k(x)}),\Gamma_j(x_{k(x)})) \leq C\left(m_n h_n^d + h_n^{-(d+2)}e^{-am_n}\right)$$

Now, taking $m_n = \frac{1}{a}\log\left(ah_n^{-2(d+1)}\right)$. Then

$$\frac{1}{nh_n^d}\sum_{i=1}^{n}\sum_{\substack{j=1\\i\neq j}}^{n} Cov(\Gamma_i(x_{k(x)}),\Gamma_j(x_{k(x)})) \to 0. \tag{4.20}$$

Finally, by combining (4.19) and (4.20), we get :

$$nh_n^d Var\left(\sum_{i=1}^{n}\Delta_i\right) \sim f(x_{k(x)})\int_{\mathbb{R}^d}K^2(u)du + o(1).$$

Secondly, concerning the covariance term, we consider the following cases

– If $t_1 = s_u$, we obtain

$$\begin{aligned}|Cov(\Delta_{s_1}\ldots\Delta_{s_u},\Delta_{t_1}\ldots\Delta_{t_v})| &\leq \left(\frac{C}{nh_n^d}\right)^{u+v}\mathbb{E}\,|\chi(X_1,Y_1)|^{u+v}\\ &\leq \left(\frac{C\|K\|_\infty}{nh_n^d}\right)^{u+v}\mathbb{E}\left|K_1(x_{k(x)})\right|\\ &\leq h_n^d\left(\frac{C}{nh_n^d}\right)^{u+v}.\end{aligned} \tag{4.21}$$

– If $t_1 > s_u$, by using the quasi-association of the variables Δ_i, then we have :

$$\begin{aligned}|Cov(\Delta_{s_1}\ldots\Delta_{s_u},\Delta_{t_1}\ldots\Delta_{t_v})| &\leq \left(\frac{2(\mathrm{Lip}(\psi)+h_n^{-1}\mathrm{Lip}(K))}{nh_n^d}\right)^2\left(\frac{2C\|K\|_\infty}{nh_n^d}\right)^{u+v-2}\sum_{i=1}^{u}\sum_{j=1}^{v}\lambda_{s_i,t_j}\\ &\leq h_n^{-2}\left(\frac{C}{nh_n^d}\right)^{u+v}v\lambda_{t_1-s_u}\\ &\leq h_n^{-2}\left(\frac{C}{nh_n^d}\right)^{u+v}ve^{-a(t_1-s_u)}.\end{aligned} \tag{4.22}$$

On the other hand, we have

$$\begin{aligned}|Cov(\Delta_{s_1}\dots\Delta_{s_u},\Delta_{t_1}\dots\Delta_{t_v})| &\leq \left(\frac{2C\|K\|_\infty}{nh_n^d}\right)^{u+v-2}(|\mathbb{E}\Delta_{s_u}\Delta_{t_1}|+\mathbb{E}\,|\Delta_{s_u}|\,\mathbb{E}\,|\Delta_{t_1}|)\\ &\leq \left(\frac{C}{nh_n^d}\right)^{u+v} h_n^{2d}. \qquad (4.23)\end{aligned}$$

Now, taking a $\frac{d}{2d+2}$-power of (4.22), $\frac{d+2}{2d+2}$-power of (4.23) to obtain an upper-bound of the tree terms as follows

$$|Cov(\Delta_{s_1},\dots\Delta_{s_u},\Delta_{t_1},\dots\Delta_{t_v})| \leq h_n^d\left(\frac{C}{nh_n^d}\right)^{u+v} ve^{-\frac{ad}{2d+2}(t_1-s_u)}.$$

So, the variables Δ_i, $i=1,\dots,n$ checked the conditions of Theorem 2.1 of Kallabis and Newman (2006, pp. 2), for

$$K_n=\frac{C}{n\sqrt{h_n^d}},\ M_n=\frac{C}{nh_n^d}\ \text{ and }\ A_n=Var\left(\sum_{i=1}^n\Delta_i\right)=O\left(\frac{1}{nh_n^d}\right).$$

The valued of K_n and M_n implies that $B_n=O\left(\frac{1}{nh_n^d}\right)$.

Thus, applying (U5a), we obtain

$$\begin{aligned}\mathbb{P}\left(\left|\widehat{f}_N(x_{k(x)},z))-\mathbb{E}\left[\widehat{f}_N(x_{k(x)},z))\right]\right|>\eta\sqrt{\frac{\log n}{nh_n^d}}\right) &= \mathbb{P}\left(\left|\sum_{i=1}^n\Delta_i\right|>\eta\sqrt{\frac{\log n}{nh_n^d}}\right)\\ &\leq \exp\left\{-\frac{\eta^2\log n/(2nh_n^d)}{\left(Var\left(\sum_{i=1}^n\Delta_i\right)+C(nh_n^d)^{-\frac{1}{3}}\left(\frac{\log n}{nh_n^d}\right)^{\frac{5}{6}}\right)}\right\}\\ &\leq \exp\left\{-\frac{\eta^2\log n}{C+\left(\frac{\log^5 n}{nh_n^d}\right)^{\frac{1}{6}}}\right\}\\ &\leq C'\exp\left\{-C\eta^2\log n\right\}, \qquad (4.24)\end{aligned}$$

By using the fact that, $d_nz_n\leq n^{\beta+\frac{1}{2p}}$, we can write

$$\mathbb{P}\left(\max_{1\leq k\leq d_n}\max_{1\leq j\leq z_n}\max_{z\in\{t_j-l_n,t_j+l_n\}}\left|\widehat{f}_N(x,z)-\mathbb{E}\left[\widehat{f}_N(x,z)\right]\right|>\eta\sqrt{\frac{\log n}{nh_n^d}}\right) \leq C'n^{\beta+1/2p-C\eta^2}.$$

The suitable choice of η such that, $\beta + 1/2p - C\eta^2 < -1$, combined by equations (4.10) and (4.11), allows to finish the proof of this lemma.
Proof of Lemma 4.3.2. Since both f and Ψ are of class C^p, then H is of class C^p. By equidistribution of the variables, we get one conditioning to X :

$$\begin{aligned}
\mathbb{E}[\widehat{f}_N(x,t)] &= \frac{1}{h_n^d}\int_{\mathbb{R}^d} \mathbb{E}[\psi(Y,t)|X=u)]K\left(\frac{x-u}{h_n}\right) f(u)\,du \\
&= \frac{1}{h_n^d}\int_{\mathbb{R}^d} \Psi(u,t)K\left(\frac{x-u}{h_n}\right) f(u)\,du \\
&= \int_{\mathbb{R}^d} H(x-h_n z,t)K(z)\,dz.
\end{aligned}$$

By using a development of Taylor of $H(x-h_n z,t)$ in neighbor of $x=(x_1,\dots,x_d)$ and the fact that K is of order p we can write

$$\begin{aligned}
\mathbb{E}[\widehat{f}_N(x,t)] - H(x,t) &= \int_{\mathbb{R}^d} K(z_1,\dots,z_d)\left\{\frac{(-h_n)^j}{j!}\sum_{j=i_1+\ldots+i_d=p}\frac{\partial^j H}{\partial x_1^{i_1},\dots,\partial x_d^{i_d}}\Big((h_n z_1)^{i_1}\cdot\ldots\cdot(h_n z_d)^{i_d}\Big)\right\} \\
&\quad dz_1,\dots,dz_d + o(h_n^p).
\end{aligned}$$

So,

$$\sup_{x\in\mathcal{C}}\left|\mathbb{E}[\widehat{f}_N(x,t)] - H(x,t)\right| = O(h_n^p).$$

Proof of Lemma 4.3.3. We have

$$\begin{aligned}
\sup_{x\in\mathcal{C}}\left|\widehat{f}_D(x)-\mathbb{E}\left[\widehat{f}_D(x)\right]\right| &\le \underbrace{\sup_{x\in\mathcal{C}}\left|\widehat{f}_D(x)-\widehat{f}_D(x_{k(x)})\right|}_{T_1'} + \underbrace{\sup_{x\in\mathcal{C}}\left|\widehat{f}_D(x_{k(x)})-\mathbb{E}\left[\widehat{f}_D(x_{k(x)})\right]\right|}_{T_2'} \\
&\quad + \underbrace{\sup_{x\in\mathcal{C}}\left|\mathbb{E}\left[\widehat{f}_D(x_{k(x)})\right]-\mathbb{E}\left[\widehat{f}_D(x)\right]\right|}_{T_3'}.
\end{aligned}$$

Again, by using the compactness of $\mathcal{C}$, and replying that $k(x) = \arg\min_{k\in\{1,\ldots d_n\}}\|x-x_k\|$, with $d_n \le n^\beta$, where $\beta = \dfrac{\delta(d+2)}{2} + \dfrac{1}{2}$, we get
• Concerning T_1' and T_3' : The Lipschitz's condition of the kernel K permit to write,

for all $x \in \mathcal{C}$

$$\begin{aligned}
\left|\widehat{f}_D(x) - \widehat{f}_D(x_{k(x)})\right| &= \frac{1}{nh_n^d}\left|\sum_{i=1}^{n} K_i(x) - \sum_{i=1}^{n} K_i(x_{k(x)})\right| \\
&\leq \frac{C}{h_n^{d+1}}\|x - x_{k(x)}\| \\
&\leq \frac{C\tau_n}{h_n^{d+1}}
\end{aligned}$$

The results declines directly, by the fact that $\frac{\tau_n}{h_n^{d+1}} = O\left(\sqrt{\frac{\log n}{nh_n^d}}\right)$. Thus

$$\sup_{x\in\mathcal{C}}\left|\widehat{f}_D(x) - \widehat{f}_D(x_{k(x)})\right| = O\left(\sqrt{\frac{\log n}{nh_n^d}}\right) \qquad a.co. \tag{4.25}$$

and

$$\sup_{x\in\mathcal{C}}\left|\mathbb{E}\left[\widehat{f}_D(x_{k(x)})\right] - \mathbb{E}\left[\widehat{f}_D(x)\right]\right| = O\left(\sqrt{\frac{\log n}{nh_n^d}}\right) \qquad a.co. \tag{4.26}$$

• Concerning T_2' : For all real $\eta > 0$, we have

$$\begin{aligned}
\mathbb{P}\left(T_2' > \eta\sqrt{\frac{\log n}{nh_n^d}}\right) &\leq \mathbb{P}\left(\max_{k\in\{1,\ldots d_n\}}\left|\widehat{f}_D(x_{k(x)}) - \mathbb{E}\left[\widehat{f}_D(x_{k(x)})\right]\right| > \eta\sqrt{\frac{\log n}{nh_n^d}}\right) \\
&\leq d_n \max_{k\in\{1,\ldots d_n\}}\mathbb{P}\left(\left|\widehat{f}_D(x_{k(x)}) - \mathbb{E}\left[\widehat{f}_D(x_{k(x)})\right]\right| > \eta\sqrt{\frac{\log n}{nh_n^d}}\right) \qquad (4.27)
\end{aligned}$$

The proof of the last probability is a very close to Lemma 4.3.1 it suffices to replace ψ by 1. Thus,

$$\mathbb{P}\left(\sup_{x\in\mathcal{C}}\left|\widehat{f}_D(x) - \mathbb{E}\left[\widehat{f}_D(x)\right]\right| > \eta\sqrt{\frac{\log n}{nh_n^d}}\right) \leq n^{\beta - C\eta^2}.$$

The suitable choose of η, such that $\beta - C\eta^2 < -1$, combining with (4.25) and (4.26) allows to write

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\sup_{x\in\mathcal{C}}\left|\mathbb{E}\left[\widehat{f}_D(x)\right] - \widehat{f}_D(x)\right| > \eta\sqrt{\frac{\log n}{nh_n^d}}\right) < \infty.$$

Proof of Lemma 4.3.4. Using the same arguments as those used in Lemma 4.3.2, it suffices to write

$$\begin{aligned} \mathbb{E}\left[\widehat{f}_D(x)\right] &= \frac{1}{h_n^d}\int_{\mathbb{R}^d} K\left(\frac{x-u}{h_n}\right) f(u)\,du \\ &= \int_{\mathbb{R}^d} f(x-h_n z)K(z)\,dz. \end{aligned}$$

Next, we obtain by a development of Taylor of f

$$\sup_{x\in\mathcal{C}}\left|f(x)-\mathbb{E}\left[\widehat{f}_D(x)\right]\right| = O(h_n^p).$$

Proof of Corollary 4.3.1. We can write, $\forall x \in \mathcal{C}$:

$$\inf_{x\in\mathcal{C}} \widehat{f}_D(x) < \frac{\sup_{x\in\mathcal{C}} f(x)}{2} \Rightarrow \sup_{x\in\mathcal{C}} |f(x) - \widehat{f}_D(x)| > \frac{\sup_{x\in\mathcal{C}} f(x)}{2}$$

Then,

$$\mathbb{P}\left(\inf_{x\in\mathcal{C}} |\widehat{f}_D(x)| \le \frac{\sup_{x\in\mathcal{C}} f(x)}{2}\right) \quad \le \quad \mathbb{P}\left(\sup_{x\in\mathcal{C}} |f(x) - \widehat{f}_D(x)| \ge \frac{\sup_{x\in\mathcal{C}} f(x)}{2}\right)$$

The use of the results of Lemma 4.3.3 and Lemma 4.3.4 complete the proof of the corollary.

Proof of Lemma 4.3.5. The strict monotony of ψ imply :

$$\Psi(x,\theta_x-\epsilon) \le \Psi(x,\theta_x) \le \Psi(x,\theta_x+\epsilon)$$

The Lemmas 4.3.1, 4.3.2, 4.3.3 and the corollary 4.3.1 show that for all real fixed t.

$$\widehat{\Psi}(x,t) - \Psi(x,t) \to 0 \qquad a.co.$$

So, for all n sufficiently large, we have :

$$\widehat{\Psi}(x,\theta_x-\epsilon) \le 0 \le \widehat{\Psi}(x,\theta_x+\epsilon) \qquad a.co.$$

As, ψ and K are continuous functions, then $\widehat{\Psi}(x,t)$ is a continuous function for all t. There exist $t_0 = \widehat{\theta}_x$ in interval $[\theta_x-\epsilon, \theta_x+\epsilon]$ such that $\widehat{\Psi}(x,\widehat{\theta}_x) = 0$. Finally, the unicity of $\widehat{\theta}_x$, will be a direct consequence of the strict monotony of ψ and the positivity of K.

Bibliographie

[1] Ango Nze, P., Bühlmann, P. & Doukhan, P. (1993) *Functional estimation for mixing time series.* C. Acad. Sci. Paris **317**, 405–408.

[2] Ango Nze, P., Bühlmann, P. & Doukhan, P. (2002) *Weak dependence beyond mixing and asymptotics for nonparametric regression.* Ann. Statist. **30**, N. **2**.

[3] Azzedine, N., Laksaci, A. & Ould-Saïd, E. (2008). *On robust nonparametric regression estimation for functional regressor.* Statist. Probab. Lett.**78**, 3216–3221.

[4] Barlow, R.E. & Proschan, F. (1981). *Statistical Theory of Reliability and Life Testing. Probability Models.* Holt, Rinehart & Winston, New York.

[5] Boente, G. & Fraiman, R. (1989). *Robust nonparametric regression estimation.* J. Multivariate Anal. **29**, 180–198.

[6] Bosq, D. (1998). *Nonparametric statistics for stochastic process : Estimation and prediction. Springer-Verlag, New York.*

[7] Bosq, D. & Blanke, D. (2007). *Inference and prediction in large dimensions.* Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester.

[8] Bulinski A. & Shabanovich, E. (1998). *Asymptotical behaviour of some functionals of positively and negatively dependent random fields. (in Russian).* Fundam. Appl. Math., **4**, 479–492

[9] Bulinski A. & Suquet, C. (2001). *Normal approximation for quasi-associated random fields.* Statist. Probab. Lett. **54**, 215–226.

[10] Bulinski A. & Shashkin, A. (2007). *Limit theorems for associated random fields and related systems.*

[11] Collomb, G. & Härdle, W. (1986). *Strong uniform convergence rates in robust nonparametric time series analysis and prediction : Kernel regression estimation from dependent observations.* Stoch. Proc. Appl. **23**, 77–89.

[12] Dedecker, J. & Prieur, C. (2004). *Coupling for τ-dependent sequences and applications.*, J. Theor Proba. **4**, 861-885.

[13] Dedecker, J., Doukhan, P., Lang, G., León R., J. R., Louhichi, S., and Prieur, C. (2007). *Weak dependence : with examples and applications.*, Lecture Notes in Statistics.

[14] Doukhan, P., Lang, G., Surgailis D., Tyssière, G. (2010). *Dependence in Probability and Statistics.*, Lecture Notes in Statistics.

[15] Doukhan, P., Louhichi, S. (1999). *A new weak dependence condition and applications to moment inequalities.* Stoch. Proc. Appl. **84**, 313–342.

[16] Esary, J., Proschan, F. & Walkup, D. (1967). *Association of random variables with applications.* Ann. Math. Statist. **38**, 1466–1476.

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

[18] Györfi, L., Härdle, W., Sarda, P. & Vieu, P. (1989). *Nonparametric curve estimation from time series.* Lecture Notes in Statistics, vol. 60, Springer, New York.

[19] Härdle, W. & Tsybakov, A. (1988). *Robust nonparametric regression with simultaneous scale curve estimation.* Ann. Statist.**16**, 120–135.

[20] Huber, P. (1964). *Robust Estimation of a Location Parameter.* Ann. Math. Statist. Vol. **35**, 73–101.

[21] Jong-Dev, K. & Proschan, F. (1983). *Negative association of random variables, with applications.* Ann. Statist. **11**, 286–295.

[22] Kallabis, R. S & Neumann, M. H. (2006). *An exponential inequality under weak dependence.* Bernoulli **12**, 333–350.

[23] Leibscher, E. (2001). *Estimation of the density and the regression function under mixing conditions.* Statist. & Decisions, **19**, 9–26.

[24] Mack, Y. P. & Silverman, B. W. (1992). *Weak and strong uniform consistency of kernel regression estimates.* Springer-Verlag, **61**, 105–115.

[25] Masry, E. (1986). *Recursive probability density estimation for weakly dependent stationary processes* IEEE. Trans. Inform. Theory **32**, 254–267.

[26] Masry, E. (2002). *Multivariate probability density estimation for associated processes : strong consistency and rates.* Statist. Probab. Lett. **58**, 205–219.

[27] Newman, C. M. (1984). *Asymptotic independence and limit theorems for positively and negatively dependent random variables :* in Inequalities in Statistics and Probability, IMS Lect. Notes–Monographs Series,**5**, 127–140.

[28] Pitt L. (1982). *Positively correlated normal variables are associated.* Ann. Probab., **10**, 496–499.

[29] Roussas, G. G. (1991). *Kernel estimates under association : strong uniform consistency.* Statist. Probab. Lett. **12**, 215–224.

[30] Roussas, G. G. (1993). *Curve estimation in random fields of associated processes.* J. Nonparametric Statist. **2**, 215–224.

[31] Roussas, G. G. (2000). *Prediction under association.* In : Limnios, N., Nikulin, M.S. (Eds.)., Recent Advances in Reliability Theory : Methodology, Practice and Inference. BirkhSauser, Boston.

[32] Shashkin, A. P (2002). *Quasi-association for system of Gaussian vectors.* Uspehi Ma. Nauk.

Chapitre 5

Conclusion, Perspectives et outils

5.1 Conclusion

Dans cette thèse, nous avons abordé l'estimation de deux classes de fonctions, la densité conditionnelle et la fonction de régression. Dans un contexte fonctionnelle, on a estimé la densité conditionnelle, en utilisant l'approche semiparamétrique de l'indice fonctionnelle, tandis dans le cas vectoriel on a estimé la fonction de régression, en utilisant l'approche robuste. Dans la première partie et pour la densité conditionnelle, nous avons étudié le modèle de l'indice fonctionnelle lorsque les covariables sont indépendantes. Dans cette étude, nous avons répondu à une question cruciale dans l'estimation non paramétrique par la méthode du noyau. Il s'agit du problème de choix de la métrique. Nous avons proposé une méthode de sélection automatique pour l'indice fonctionnelle par la méthode de pseudo-maximum de vraisemblance de la densité conditionnelle. Pour la régression, nous avons considéré la méthode de M-estimation pour une famille d'estimateurs robustes quand les données sont vectorielles et quasi-associées. L'objectif de cette étude est de faire une combinaison entre les deux approches (semiparamétrique et robuste) pour estimer l'indice fonctionnel.
Dans le cas i.i.d., pour la densité conditionnelle, nous avons obtenu des résultats de convergence presque complète. Pour des variables quasi-associées, nous avons établi la vitesse de convergence presque complète ponctuelle et la normalité asymptotique. Ces résultats sont obtenus sous des conditions standard en statistique non paramétrique. Notons que nos estimateurs sont faciles à mettre en oeuvre pour la simulation des données réelles ou fonctionnelle.

5.2 Perspectives

Le travail abordé dans cette thèse, nous permet d'envisager à traiter plusieurs extensions possibles.

1. On peut étendre le modèle robuste ainsi que l'indice fonctionnel sous quasi-association pour des données fonctionnelles, le principal objectif, c'est de prouver la quasi-associativité des variables dans des espaces métriques.
2. Un point qui nous préoccupe est l'estimation de l'indice fonctionnel en se basant sur la méthode de pseudo-maximum de vraisemblance.
3. Un autre point qui nous paraît important est l'obtention d'un résultat de normalité asymptotique pour des variables quasi-associées hilbertiennes du modèle robuste, la convergence sous ce type de dépendance a été déjà obtenue.
4. Pour aller plus loin, plusieurs problèmes peuvent être envisagés. Par exemple le cas de données spatiales pour les modèles définies ci-dessus, et également le cas où la variable réponse est fonctionnelle pour le modèle d'indice fonctionnel. Ce dernier point, à notre humble avis, est extrêmement intéressant.

5.3 Inégalités type exponentiel

5.3.1 Inégalité exponentielle pour le cas i.i.d.

Lemma 5.3.1 *Let $W_1, \dots, W_n$ a sequence of random variables which are independent identically distributed, and $\sigma_n^2 = \mathbb{E}(W_j^2)$. If there exists $M := M_n < \infty$ such that $W_1 \leq M$. Then we have*

$$\forall \epsilon > 0,\ \mathbb{P}\left(\frac{1}{n}\left|\sum_{j=1}^{n} W_j\right| > \epsilon\right) \leq 2\exp\left\{-\frac{\epsilon^2 n/2}{\sigma^2 + \epsilon M}\right\}$$

If $w_n = n^{-1}\sigma_n^2 \log n$, such that $\lim_{n\to\infty} w_n = 0$, with $M/\sigma_n^2 < \infty$, then we have :

$$\frac{1}{n}\sum_{j=1}^{n} W_j = O_{a.co}(\sqrt{w_n}).$$

5.3.2 Inégalité exponentielle pour le cas dépendant

Lemma 5.3.2 *(Kallabis and Neumann (2006)) Let $W_1, \dots, W_n$ the real random variables such that $\mathbb{E}W_j = 0$ and $\mathbb{P}(|W_j| \leq M) = 1$, for all $j = 1, \dots, n$ and some*

$M < \infty$. Let $\sigma_n^2 = Var\left(\sum_{j=1}^n W_j\right)$. Assume, furthermore, that there exist $K < \infty$ and $\beta > 0$ such that, for all u-tuplets $(s_1, \dots, s_u)$ and all v-tuplets $(t_1, \dots, t_v)$ with $1 \le s_1 \le \dots \le s_u \le t_1 \le \dots \le t_v \le n$, the following inequality is fulfilled :

$$\left|Cov(W_{s_1} \dots W_{s_u}, W_{t_1} \dots W_{t_v})\right| \le K^2 M^{u+v-2} v e^{-\beta(t_1 - s_u)}.$$

Then,

$$\mathbb{P}\Big(\Big|\sum_{j=1}^n W_j\Big| > t\Big) \le \exp\left\{-\frac{t^2/2}{A_n + B_n^{\frac{1}{3}} t^{\frac{5}{3}}}\right\}$$

for $A_n \le \sigma_n^2$ and $B_n = \left(\frac{16nK^2}{9A_n(1-e^{-\beta})} \vee 1\right) \frac{2(K \vee M)}{1-e^{-\beta}}$.

Summary In this thesis, we propose to study some functional parameters when the data are generated from a model of regression to a single index. We study two functional parameters.

Firstly, we suppose that the explanatory variable takes its values in Hilbert space (infinite dimensional space) and we consider the estimate of the conditional density by the kernel method. We establish some asymptotic properties of this estimator in both independent and dependent cases. For the case where the observations are independent identically distributed (i.i.d.), we obtain the pointwise and uniform almost complete convergence with rate of the estimator. As an application we discuss the impact of this result in functional nonparametric prevision for the estimation of the conditional mode. In the dependent case we modelize the later via the quasi-associated correlation. Note that all these asymptotic properties are obtained under standard conditions and they highlight the phenomenon of concentration proprieties on small balls probability measure of the functional variable.

Secondly we suppose that the explanatory variable takes values in the finite dimensional space and we interest in a rather general prevision model which is the robust regression. From the quasi-associated data, we build a kernel estimator for this functional parameter. As an asymptotic result we establish the uniform almost complete convergence rate of the estimator. We point out by the fact that these two models studied in this thesis could be used for the estimation of the single index of the model when the latter is unknown, by using the method of M-estimation or the pseudo-maximum likelihood method which is a particular case of the first method.

As far as we know, the problem of estimating the conditional density in the functional single index parameter for quasi-associated data was not attacked. In general the nonparametric estimation under quasi-associated data is new in the statistical literature. What doubtless makes, the originality of this thesis.