



*République Algérienne Démocratique et Populaire*  
*Ministère de l'enseignement supérieur et de la recherche scientifique*

# THÈSE

En vue de l'obtention du

Doctorat de l'Université Djillali Liabès de Sidi  
Bel Abbès (S.B.A.), Algérie

Discipline : Mathématiques Appliquées

Spécialité : Statistique

---

*présentée et soutenue par*

**Amina Nacéri**

Mai 2017

---

*sous le titre :*

**Sur la modélisation locale linéaire en  
statistique fonctionnelle**

---

*Devant le jury composé de :*

<b>Mohamed Attouch</b>	Univ. Djillali Liabès, S.B.A. (Algérie)	Président
<b>Abdelkader Gheriballah</b>	Univ. Djillali Liabès, S.B.A. (Algérie)	Rapporteur
<b>Saâdia Rahmani</b>	Univ. Saida (Algérie)	Rapporteur
<b>Ali Laksaci</b>	Univ. Djillali Liabès, S.B.A. (Algérie)	Directeur de thèse
<b>Mustapha Rachdi</b>	Univ. Grenoble Alpes, Grenoble (France)	Co-directeur de thèse

**Résumé.**

L'objectif principal de cette thèse est l'estimation locale linéaire de l'opérateur de régression lorsque la variable réponse et la variable explicative appartiennent à un espace de Hilbert.

Nous considérons, dans un premier temps, le cas où la variable réponse est réelle et la variable explicative est fonctionnelle. Nous démontrons, dans ce cas, la convergence en moyenne quadratique d'un estimateur local linéaire inspiré de celui de Barrientos et al. (2010). Le cas où les deux variables sont hilbertiennes est traité dans la deuxième partie de cette thèse. Nous y construisons et démontrons la convergence presque-complète d'un estimateur local linéaire. Notons que la vitesse de convergence obtenue garde la même structure que dans le cadre classique. Dans la troisième partie de ce manuscrit, nous y généralisons les résultats obtenus au cas où la variable explicative appartient à un espace semi-métrique. De même que dans le cas précédent, nous établissons la convergence presque-complète de l'estimateur local linéaire construit en précisant sa vitesse de convergence. Afin de mettre en évidence l'importance de ces résultats théoriques, nous exposons à la fin de ce manuscrit, une étude applicative sur des données simulées puis sur des données biomédicales.

**Abstract.**

The main aim of this thesis is to study the local linear estimation of the regression operator when both the response and the explanatory variables belong to an Hilbert space. First, we consider the case when the response is scalar and the covariable is functional. The used estimator, in this case, is constructed from same ideas as in Barrientos et al. (2010). Then, we establish its mean square convergence with rates. Secondly, we assume that the two variables (the response and the explanatory variables) belong to an Hilbert space. In this last case, we construct a local linear estimator of the regression operator and we prove its almost-complete consistency. This asymptotic

result is given with the same convergence rate as in the classical case. The last part, of this manuscript, may be considered as a generalization of the above one. In fact, it concerns the case where the response variable is in an Hilbert space and the explanatory variable belongs to a semi-metric space. In this case, we construct a local linear estimator and we establish its asymptotic properties with giving the corresponding rates of convergence. Finally, we illustrate our theoretical studies by some applications on some simulated and biomedical real data.

#### Liste de publications.

1. Laksaci, A., Nacéri, A. and Rachdi, M. (2015). *Exact Quadratic Error of the Local Linear Regression Operator Estimator for Functional Covariates*. *Functional Statistic and Application*, Springer (2015), 79-90
2. Demongeot, J., Laksaci, A., Nacéri, A. and Rachdi, M. (2016). Estimation locale linéaire de la fonction de régression pour des variables hilbertiennes. *C.R. Math. Acad. Sci. Paris*, **354**, 847-850.
3. Demongeot, J., Laksaci, A., Nacéri, A. and Rachdi, M. (2017). Local linear regression modelization when all variables are curves. *Statistics and Probability Letters*, **121**, 37-44.
4. Nacéri, A. (2015). Functional local linear regression model with functional response. *Journal of Probability and Statistical Sciences*, **27**, 1-11.



# Table des matières

<b>1</b>	<b>Introduction générale</b>	<b>7</b>
1.1	Etat de l'art . . . . .	7
1.1.1	Analyse non paramétrique des données fonctionnelles (NPFDA) . . . . .	10
1.1.2	Plan de la thèse . . . . .	14
1.1.3	Contributions originales de la thèse . . . . .	15
1.2	Modèle local linéaire fonctionnel . . . . .	18
<b>2</b>	<b>Exact quadratic error of the local linear regression operator estimator for functional covariates</b>	<b>21</b>
2.1	Introduction . . . . .	21
2.2	Functional local linear modeling : estimation and comments .	23
2.3	Main results . . . . .	25
2.4	Appendix : Proofs . . . . .	28
<b>3</b>	<b>Local linear regression modelization when all variables are curves</b>	<b>35</b>
3.1	Introduction . . . . .	35
3.2	The model and the estimator . . . . .	38

3.3	Estimation error . . . . .	39
3.4	Uniform almost-complete convergence . . . . .	41
3.5	Some concluding remarks . . . . .	43
3.6	Proofs . . . . .	45
<b>4</b>	<b>Functional local linear regression model with functional response</b>	<b>53</b>
4.1	Introduction . . . . .	53
4.2	Model . . . . .	55
4.3	The main result . . . . .	57
4.4	Appendix . . . . .	60
<b>5</b>	<b>Application sur des données réelles</b>	<b>71</b>
5.1	introduction . . . . .	71
5.2	Modèle et Données . . . . .	72
5.3	Débruitage par régression locale constante . . . . .	74
5.4	Débruitage par régression locale linéaire . . . . .	77
<b>6</b>	<b>Conclusion et perspectives</b>	<b>81</b>
6.1	Conclusion générale . . . . .	81
6.2	Perspectives . . . . .	82

# Chapitre 1

## Introduction générale

### 1.1 Etat de l'art

La modélisation locale linéaire est une approche statistique très importante surtout dans la modélisation non paramétrique. Il s'agit d'une méthode d'estimation alternative à la méthode à noyau et qui a des avantages certains par rapport à cette dernière, notamment dans la partie biais de l'estimateur ou les problèmes aux bords. Les premiers résultats sur cette méthode reviennent à Stone (1977). Signalons que la contribution de Stone (1977) est déterminante non seulement sur la modélisation locale linéaire mais sur l'analyse non paramétrique d'une manière plus générale. En fait, Stone (1977) a déterminé la vitesse de convergence minimax d'une famille d'estimateurs construits à partir d'une fonction de poids. Il a considéré pour cela, une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) et il a donné les conditions nécessaires et suffisantes permettant d'obtenir la consistance de ses estimateurs. Il a également étudié plusieurs modèles non paramétriques tels que l'estimation de l'espérance conditionnelle, des covariances, des écarts-types, des corrélations et des quantiles. Depuis ce temps, plusieurs travaux ont vu le jour, dans cette direction, prenant en compte l'influence locale des

observations.

On peut considérer que depuis Stone (1977), l'estimation locale linéaire s'est bien développée et le travail de Fan (1992) peut être considéré comme révélateur dans cette direction. Ce dernier a prouvé les avantages qu'à l'estimation locale linéaire sur plusieurs méthodes d'estimation. En effet, il a montré que le lissage local linéaire est asymptotiquement plus efficace que tout lissage, que ce soit dans la méthode d'estimation à noyau, par projection orthogonale ou par des splines. Par suite, Fan et Gijbels (1992) ont donné la forme actuelle de l'estimateur local linéaire de la régression, en particulier, ils ont construit celui-ci en utilisant une pondération issue de la méthode à noyau. En outre, Ils ont donné la forme explicite de l'erreur quadratique moyenne (MSE) et de l'erreur quadratique moyenne intégrée (MISE). Comme nous l'avons mentionné plus haut, Fan et Gijbels (1992) ont prouvé que (i) la méthode locale linéaire produit un biais moindre que celui de la méthode à noyau et surtout (ii) la résistance de la nouvelle méthode aux effets de bord. Par ailleurs, concernant les propriétés minimax de l'estimateur local linéaire de la fonction de régression, nous renvoyons à Fan (1993). Par la suite, Ruppert (1994) a généralisé les résultats de Fan et Gijbels (1992) au cas où la variable explicative (covariable) est multidimensionnelle. Ce dernier a explicité le terme dominant du biais et de la variance conditionnelle de l'estimateur local linéaire de la régression. De plus, Il a montré que ces termes sont identiques au cas uni-varié développé par Fan et Gijbels (1992). Cependant, Ruppert (1994) a montré que toutes ses propriétés asymptotiques sont conditionnées par des hypothèses de régularité sur la fonction de régression. Ensuite, Fan et Gijbels (1994) ont étudié l'estimation locale linéaire de la fonction de régression lorsque la variable réponse est observée avec censure, et ils ont appliqué leurs résultats sur des données médicales. L'approche locale linéaire a été généralisée en une approche polynomiale en introduisant une pondération dans la méthode à noyau, par Fan et al. (1995). Dans Fan et al. (1995), on



a montré que le gain sur la partie biais augmente avec le degré du polynôme dans l'approximation locale. Après ceci, Masry (1996), quant à lui, a repris ce travail et a étudié la convergence forte et la normalité asymptotiques quand les observations sont fortement mélangeantes. Dans Masry (2001), il a obtenu des résultats de convergence forte du même estimateur, de la régression et de ses dérivées, mais en considérant une corrélation alternative au mélange fort. Nous référons aussi à Masry et Mielniczuk (1999) pour l'estimation locale linéaire des dérivées de la fonction de régression. Dans la lignée de ces travaux, Cai et Masry (2000) ont adapté l'approche locale linéaire pour les modèles additifs. En effet, sous des conditions de mélange fort, ils ont montré la normalité asymptotique des estimateurs des différentes composantes du modèle.

Par ailleurs, plusieurs auteurs se sont intéressés aux mêmes problématiques quand les variables sont fonctionnelles et spatiales (cf. Dabo-Niang et al. (2011), Laïb et Louani (2011) pour des données dépendantes et Douge (2010) pour le cas de variables quasi-associées).

Nous avons cité, ci-dessus, quelques travaux précurseurs sur l'estimation locale linéaire de la régression, mais notons que cette approche a été utilisée aussi pour d'autres modèles tels que les modèles conditionnels. A titre d'exemple, dans Yu et Jones (1998) on s'est intéressé à l'estimation des quantiles conditionnels. Ces derniers ont construit deux estimateurs des quantiles conditionnels (i) comme inverse de l'estimateur local linéaire de la fonction de répartition et (ii) en utilisant l'approche  $L_1$ .

Nous rappelons que le lecteur pourrait consulter Wang (2015) ou Wang, L. (2016) pour une étude bibliographique récente sur ce sujet.

D'autres auteurs se sont intéressés à des versions alternatives à la méthode du noyau classique comme l'estimation locale linéaire par la méthode des  $k$  plus proche voisins  $k$ NN,... Nous citons pour cela, l'article de Burba et al. (2009) dont on trouve la vitesse de convergence presque-complète de l'estimateur de

la fonction de régression, lorsque les observations sont i.i.d. Cette méthode d'estimation a été adoptée ensuite pour d'autres paramètres conditionnels, à savoir le mode conditionnel (cf. Attouch et Bouabça, 2013), ainsi que la fonction de hasard conditionnelle (cf. Attouch et Belabed, 2014). Nous renvoyons d'ailleurs, à Kudraszow et Vieu (2013) pour la convergence uniforme de la fonction de lissage et à Kara-Zaitri et al. (2016 et 2017) pour l'uniformité par rapport au paramètre de lissage de l'estimateur  $k$ NN de plusieurs paramètres statistiques.

### 1.1.1 Analyse non paramétrique des données fonctionnelles (NPFDA)

L'analyse statistique, non paramétrique, des données fonctionnelles (NPFDA : Non Parametric Functional Data Analysis) est un axe de recherche assez ancien mais dont la constatation de ses intérêts pour les applications pratiques est relativement récente, avec l'avènement de l'informatique. En outre, la modélisation des variables aléatoires à trajectoires continues est un problème très important en pratique comme en atteste les nombreux ouvrages édités sur ce sujet, dont on peut citer les plus populaires tels que, Ramsay et Silverman (1997, 2000), Ferraty et Vieu (2006). Comme la NPFDA consiste à modéliser des phénomènes continûment observés dans le temps et/ou dans l'espace, on se trouve donc face à un problème de modernisation des techniques statistiques usuelles (cf. Cuevas (2014), Zhang (2014), Goia et View (2015) et Hsing et Enbank (2015)).

En voici quelques travaux spécifiques (cf. Damon et Guillas (2002) et Fernandez et al., (2003) pour la météorologie, puis Besse et al. (2000) et Hall et Heckman (2002) pour la reconnaissance vocale, et Hall et al. (2001) pour le traitement de données médicales). Rappelons que les premiers modèles dans l'Analyse des Données Fonctionnelles (FDA : Functional Data Analy-

sis), sont basés sur la discrétisation des trajectoires afin de pouvoir utiliser les outils de la statistique vectorielle sur les nouvelles variables discrétisées obtenues. Cependant, cette procédure a montré ses limites surtout au niveau computationnel et les mauvais résultats obtenus avec le fléau de la dimension. Ceci est dû au fait que l'ensemble des points de discrétisation est de plus en plus dense/important et le volume des observations devient très élevé avec le développement technologique actuel. Ainsi, la modernisation des outils statistiques classiques est devenue plus que nécessaire.

Dans cette direction les premières pierres dans l'édifice de la NPFDA ont été introduites par Ferraty et Vieu (2000) (cf. Ferraty et Vieu, 2006). Ces auteurs ont donné la version fonctionnelle de l'estimateur à noyau de l'opérateur de régression, et ont étudié ses propriétés asymptotiques dans le cas de données à dimension fractale. En considérant des observations dépendantes, Ferraty et al. (2002) ont utilisé le même estimateur comme modèle de prédiction pour des données se présentant comme un processus à temps continu. Quand la variable explicative appartient à un espace métrique, Dabo-Niang et Rhomari (2012) ont montré la convergence en norme  $L_p$  de l'estimateur de Ferraty et Vieu, puis Masry (2005) a établi sa normalité asymptotique. Ensuite, d'autres modèles non paramétriques ont été considérés dans la NPFDA. Nous trouvons dans Ferraty et al. (2006), une collection importante de prédicteurs fonctionnels tels que le mode conditionnel, la médiane conditionnelle et les quantiles conditionnels. Ces auteurs ont montré la convergence presque-complète des estimateurs à noyau de ces paramètres statistiques en donnant leurs vitesses de convergence dans le cas où les observations sont i.i.d. Après ceci, Ezzahroui et Ould-Said (2007 et 2008) ont établi la normalité asymptotique de l'estimateur à noyau pour le mode et les quantiles conditionnels. Mentionnons que la convergence en moyenne quadratique (m.q.) de l'estimateur à noyau de l'opérateur de régression a été montrée par Ferraty et al. (2007) et ensuite Laksaci (2007) a donné la formulation exacte de son MSE.

Il est remarquable, et comme c'était prévu, que dans les résultats obtenus de la convergence en m.q. précitée, figure l'influence de la largeur de fenêtre sur la vitesse de convergence de l'estimateur. Ce qui a été également constaté lors de son implémentation. Un bon choix de la largeur de fenêtre est donc crucial. Dans le but de résoudre ce problème, Rachdi et Vieu (2007) puis Benhenni et al. (2007) ont construit deux règles (locale et globale) de choix de ce paramètre de lissage. Ces techniques sont basées sur les idées de validation croisée.

La question du choix du paramètre de lissage pour l'estimation de la densité conditionnelle a été également posée et traitée par Laksaci et al. (2012). Dans ce dernier article, on a montré également que la technique de validation croisée fonctionnelle reste valable pour plusieurs modèles non paramétriques conditionnels.

Pour des raisons d'utilisation des résultats de convergence en pratique l'uniformité de la convergence se pose naturellement. Pour ce faire, Ferraty et al. (2010) ont montré que contrairement au cas vectoriel, la convergence uniforme de la version fonctionnelle de l'estimateur à noyau n'est pas une simple extension de la convergence ponctuelle, mais elle est liée à l'entropie de l'espace fonctionnel des observations. Ils ont donc montré que la vitesse de convergence du même modèle peut différer d'un espace à l'autre. Ceci, serait constaté dans les chapitres suivant lorsque nous allons modifier l'espace des données (changement de structure des données).

Rappelons que l'estimation à noyau permet d'avoir des estimateurs réguliers et faciles à utiliser, mais cette technique reste très sensible aux observations aberrantes. Dans le but de contourner ce problème, Azzedine et al. (2008) ont introduit la régression non paramétrique robuste en construisant et montrant la consistance forte des M-estimateurs, quand les données sont i.i.d. La normalité asymptotique de cet estimateur a été obtenue par Attouch et al. (2009). Nous renvoyons à Crambes et al. (2008) pour la convergence en

norme  $L_p$  de cet estimateur. Notons aussi que plusieurs résultats ont été obtenus pour l'estimation robuste de la régression quand on dispose de séries chronologiques fonctionnelles. Nous citons, à titre d'exemple, Attouch et al. (2010, 2012) et Gheriballah et al. (2012) pour le cas de données spatiales et Gheriballah et al. (2013) quand les données sont ergodiques. Très récemment, une approche d'estimation alternative à la régression robuste et à la régression classique a été développée en NPFDA. Il s'agit de la régression relative. Cette procédure d'estimation regroupe les avantages des deux types de régressions, à savoir la résistance aux présences d'observations aberrantes ainsi que les propriétés asymptotiques de l'estimation classique de la régression. L'article de Demongeot et al. (2016), sur ce modèle, est une contribution fondamentale sur ce sujet. On y trouve une compilation importante de vitesses de convergence avec différents modes de convergence tels que la convergence presque-complète, la convergence en moyenne quadratique et la normalité asymptotique.

Venons en maintenant à l'estimation de la régression non paramétrique où les deux variables (réponse et explicative) sont fonctionnelles. Celle-ci a été traitée en NPFDA, mais on est au début des investigations sur ce thème. Nous citons pour cela Ferraty et al. (2011) (resp. Dabo-Niang et Rhomari, 2012) qui ont obtenu la convergence presque-complète (resp. en norme  $L_p$ ) de l'estimateur à noyau, puis Ferraty et al. (2013) qui ont établi la convergence en loi et Heng (2011) qui a étudié la convergence forte d'un estimateur construit par la méthode des  $k$  plus proches voisins. Notons que parmi les contributions les plus récentes sur ce sujet, nous pouvons citer Ameri (2016) pour l'estimation avec la méthode récursive, Ardjoun et al. (2016) pour l'estimation du mode conditionnel dans le cas ergodique, Kara-Zaitri et al. (2017) pour la convergence uniforme par rapport au paramètre de lissage dans l'estimation non paramétrique de plusieurs statistiques telles que la régression, la densité conditionnelle, la fonction de hasard et la fonction de répartition.

Dans notre thèse, nous nous sommes concentrés sur la méthode d'estimation locale linéaire de quelques indicateurs statistiques lorsque les deux variables sont de type fonctionnel. Afin de mieux positionner notre contribution dans la vaste littérature de la NPFDA, nous allons consacrer le paragraphe suivant aux travaux originaux que nous avons réalisés sur ce sujet.

### 1.1.2 Plan de la thèse

Ce manuscrit est composé de cinq chapitres.

Nous avons commencé par un chapitre introductif où nous avons exposé un bref descriptif historique des données fonctionnelles, du modèle de régression opératoire ainsi que des techniques d'estimation nous intéressant.

Dans le deuxième chapitre, nous présentons notre première contribution qui consiste en la convergence en m.q. de l'estimateur local linéaire de l'opérateur de régression quand la variable réponse est réelle et la covariable est fonctionnelle. Cette partie a été publiée comme chapitre d'ouvrage (cf. Laksaci et al., 2015).

Le troisième chapitre est consacré au cas où les deux variables réponse et covariable sont hilbertiennes. Nous y construisons un estimateur local linéaire adapté à cette situation pour lequel nous avons montré la convergence presque-complète. Nous y avons montré également que la vitesse de convergence de cet estimateur garde la forme usuelle de la convergence presque-complète en NPFDA. Cependant, celle-ci contient un terme additionnel qui garantit l'équilibre entre la rapidité et l'efficacité de l'estimateur. Ces résultats ont été obtenus sous des conditions générales et classiques dans le contexte de la NPFDA. Ce travail a fait l'objet d'une note parue aux *C.R. Acad. Sci. Math.* et une publication parue au journal international *Statistics and Probability Letters*.

Dans le quatrième chapitre, nous avons considéré le cas général. Autrement

dit, le cas où la variable explicative est dans un espace semi-métrique et la variable réponse est dans un espace de Hilbert. Nous y avons construit un estimateur de la régression dont nous avons étudié les propriétés asymptotiques. Plus précisément, nous avons montré sa convergence presque-complète et uniforme dans le cas où les observations sont i.i.d. Nos vitesses de convergence sont obtenues sous des conditions usuelles et elles sont exprimées en fonction de la régularité du modèle, de la concentration de la variable fonctionnelle, ainsi que de l'entropie de l'espace fonctionnel. Les résultats de ce chapitre ont été publiés dans *Journal of Probability and Statistical Science*. Dans le dernier chapitre, nous avons mis en évidence l'importance de nos résultats théoriques pour des applications pratiques. Nous avons utilisé notre modèle comme outil de prévision et de classification sur des données biomédicales. Enfin, ce manuscrit, se termine par une conclusion et une liste de perspectives.

### 1.1.3 Contributions originales de la thèse

Sous des hypothèses classiques en NPFDA (cf. les chapitres suivants), on a :

**Théorème 1.1.1.** *Sous les hypothèses :*

(H1) *Pour tout  $r > 0$ ,  $\phi_x(r) := \phi_x(-r, r) > 0$ , et il existe une fonction  $\chi_x(\cdot)$  telle que :*

$$\forall t \in (-1, 1), \lim_{h \rightarrow 0} \frac{\phi_x(th, h)}{\phi_x(h)} = \chi_x(t),$$

(H2) *Le paramètre de lissage  $h := h(n)$  est tel que :*

$$\lim_{n \rightarrow \infty} h = 0 \quad \text{et} \quad \lim_{n \rightarrow \infty} n \phi_x(h) = +\infty,$$

*on obtient :*

$$\mathbb{E} [\widehat{m}(x) - m(x)]^2 = B_K^2(x)h^4 + \frac{V_K(x)}{n\phi_x(h)} + o(h^4) + o\left(\frac{1}{n\phi_x(h)}\right) \quad (1)$$

où

$$m(x) = \mathbb{E}[Y|X = x] \quad \text{et} \quad \widehat{m}(x) = \frac{\sum_{i,j=1}^n W_{ij}(x) Y_j}{\sum_{i,j=1}^n W_{ij}(x)}$$

avec

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

$$B_K(x) = \frac{1}{2} \Psi''(0) \left[ \frac{K(1) - \int_{-1}^1 (u^2 K(u))' \chi_x(u) du}{K(1) - \int_{-1}^1 K'(u) \chi_x(u) du} \right]$$

et

$$V_K(x) = (m_2(x) - m^2(x)) \left[ \frac{K^2(1) - \int_{-1}^1 (K^2(u))' \chi(u) du}{\left( K(1) - \int_{-1}^1 (K(u))' \chi(u) du \right)^2} \right].$$

où  $\Psi''$  désigne la dérivée seconde de  $\Psi$  avec la convention  $0/0 = 0$ .

Une discussion sur les hypothèses et les détails de la démonstration du Théorème 1.1.1 sont exposés au Chapitre 2.

**Théorème 1.1.2.** *Sous les hypothèses suivantes :*

(H1) *Il existe une fonction différentiable  $\phi(\cdot)$ , telle que :*

$$\forall x \in S_{\mathcal{F}}, \quad 0 < C \phi(h) \leq \phi_x(h) \leq C' \phi(h) < \infty \quad \text{et} \quad \exists \eta_0 > 0, \quad \forall \eta < \eta_0, \quad \phi'(\eta) < C$$

où  $\phi'$  désigne la première dérivée de  $\phi$ .

(H2) *Posons  $r_n = O(\log n/n)$ . Soit la suite  $d_n$  vérifiant :*

$$\frac{(\log n)^2}{n \phi(h)} < \log(d_n) < \frac{n \phi(h)}{\log n} \quad \text{et} \quad \exists \beta > 1, \quad \sum_{n=1}^{\infty} d_n^{(1-\beta)} < \infty,$$



on obtient :

$$\sup_{x \in S_{\mathcal{F}}} \|\widehat{a}(x) - a(x)\|_{\mathcal{H}} = O(J^{-1}) + O(h^2) + O_{p.co.} \cdot^1 \quad (2)$$

où

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

$K = \text{diag}(K(h^{-1}\|x - X_1\|_{\mathcal{F}}), \dots, K(h^{-1}\|x - X_n\|_{\mathcal{F}}))$ ,  $Y' = (Y_1, \dots, Y_n)$  et  $e'_1$  désigne le vecteur transposé du premier vecteur de la base canonique de  $\mathbb{R}^J$  et  $\log$  désigne la fonction logarithme népérien

Une discussion sur les hypothèses et les détails de la démonstration du Théorème 1.1.2 sont exposés au Chapitre 3.

**Théorème 1.1.3.** *Sous les hypothèses suivantes :*

(H1) *Il existe une fonction différentiable  $\phi(\cdot)$ , telle que :*

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

et

$$\exists \eta_0 > 0, \forall \eta < \eta_0, \phi'(\eta) < C,$$

où  $C$  et  $C'$  sont des constantes strictement positives .

(H2) *Le paramètre de lissage  $h := h(n)$  satisfait le fait qu'il existe un entier naturel  $n_0$ , tel que :*

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

---

1. Soit  $(z_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires réelles. On dit que  $(z_n)$  converge presque complètement (p.co.) vers zéro si, et seulement si, pour tous  $\epsilon > 0$ ,  $\sum_{n=1}^{\infty} \mathbb{P}(|z_n| > \epsilon) < \infty$ . De plus, on dit que la vitesse de convergence presque-complète de  $(z_n)$  vers zéro est de l'ordre de  $u_n$  (avec  $n \rightarrow 0$ ) et on écrit  $z_n = O_{p.co.}(u_n)$  si, et seulement s'il existe un  $\epsilon > 0$  tel que  $\sum_{n=1}^{\infty} \mathbb{P}(|z_n| > \epsilon u_n) < \infty$ . Ce type de convergence implique la convergence presque sûre et la convergence en probabilité.

et

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

on obtient :

$$\sup_{x \in \mathcal{S}_{\mathcal{F}}} \|\widehat{R}(x) - R(x)\|_{\mathcal{H}} = O(h^{b_1}) + O_{p.co.} \left( \sqrt{\frac{\log d_n}{n\phi(h)}} \right). \quad (3)$$

où  $R$  est un opérateur défini sur  $\mathcal{F}$  et à valeurs dans  $\mathcal{H}$  et

$$\widehat{R}(x) = \frac{\sum_{i,j=1}^n W_{ij}(x) Y_i}{\sum_{i,j=1}^n W_{ij}(x)}. \quad (4)$$

et  $B(x, r) = \{x' \in F \text{ tel que } d(x', x) \leq r\}$  désigne la boule fermée de centre  $x$  et de rayon  $r$ , et  $dP(x)$  est la distribution cumulative de  $X$ .

Une discussion sur les hypothèses et les détails de la démonstration du Théorème 1.1.3 sont exposés au Chapitre 4.

## 1.2 Modèle local linéaire fonctionnel

L'approche locale linéaire fonctionnelle a fait l'objet d'une question ouverte dans la monographie Ferraty et Vieu (2006), et plusieurs auteurs ont essayé d'y apporter des réponses.

Rappelons que l'approche locale linéaire est basée sur une approximation linéaire du modèle au voisinage de la variable observée. De plus, l'estimation non paramétrique en NPFDA repose sur une pondération des observations les plus proches de la variable observée.

La première réponse a été donnée par Baïllo et Grané (2009), qui ont donné la première version fonctionnelle de l'estimation locale linéaire. Celle-ci est basée sur la structure hilbertienne de l'espace. Baïllo et Grané (2009) ont déterminé le biais et la variance de leur estimateur. Ensuite, Barrientos et

al. (2010) ont donné une version alternative, facile à utiliser même quand la covariable appartient à un espace semi-métrique. La version de Barrientos et al. (2010) exploite bien la richesse de la structure topologique de l'espace semi-métrique et elle considère deux approches différentes pour quantifier la notion de voisinage. Donc, on dispose de deux types de voisinages à fixer. Ainsi, la modélisation de Barrientos et al. (2010) est plus générale dans ce sens. Comme propriétés asymptotiques de Barrientos et al. (2010), la démonstration de la convergence presque-complète de l'estimateur construit et l'obtention de sa vitesse de convergence. Celle-ci se présente comme fonction du degré de régularité du modèle non paramétrique et de la fonction de concentration de la variable fonctionnelle. Tous les résultats précités ont été obtenus dans le cas où les observations sont i.i.d. Une troisième version a été abordée par Berlinet et al. (2011). Ces derniers ont utilisé l'inverse locale de l'opérateur de covariance pour construire un estimateur local linéaire de la fonction de régression ainsi que sa dérivée. Leurs résultats asymptotiques portent sur la convergence en m.q. de l'estimateur construit. Chouaf et Laksaci (2012), quant à eux, ont généralisé les résultats dus à Barrientos et al. (2010) au cas où les observations sont spatio-fonctionnelles. Ils ont également obtenu la convergence presque-complète et la vitesse de convergence de la version spatiale de l'estimateur de Barrientos et al. (2010). En ce qui concerne les paramètres statistiques conditionnels, Demongeot et al. (2013) ont étudié l'estimation locale linéaire de la densité conditionnelle dans le cas où les observations sont i.i.d. Ils ont montré la convergence presque-complète ponctuelle puis uniforme de l'estimateur construit en précisant, pour chaque mode, sa vitesse de convergence pour les deux types de convergence : la première est exprimée en fonction de l'hypothèse de concentration tandis que la deuxième est donnée en fonction de l'entropie de l'espace fonctionnel des observations. La convergence en m.q. du même estimateur a été obtenue par Rachdi et al. (2014). Dans ce dernier, les auteurs montrent la supériorité de

l'estimateur local linéaire sur la méthode à noyau. Notons que tous ces résultats ont été obtenus dans le cas où les observations sont i.i.d. et sous des conditions, considérées maintenant, classiques en NPFDA. Notons également que le cadre spatial a été abordé relativement à ce thème de la NPFDA. En effet, Laksaci et al. (2013) ont donné la version spatiale de l'estimateur local linéaire de plusieurs modèles non paramétriques conditionnels, tels que la fonction de répartition conditionnelle, la densité conditionnelle, le mode conditionnel et la fonction de hasard conditionnelle. Ils ont montré également la convergence presque-complète de ces estimateurs (cf. Demongeot et al. (2014) pour la convergence en m.q. de l'estimateur local linéaire de la fonction de répartition conditionnelle). Récemment, Messaci et al. (2015) ont étudié l'estimation locale linéaire des quantiles conditionnels. Ils ont construit et étudié les propriétés asymptotiques d'un estimateur basé sur l'inverse de l'estimateur local linéaire de la fonction de répartition conditionnelle.

Notons, enfin, que sur ce sujet, qui est l'estimation locale linéaire en NPFDA, notre apport se quantifie en quatre contributions qui sont publiées dans des revues à comités de lecture. Dans la première contribution, nous avons traité la convergence en m.q. de l'estimateur de Barrientos et al. (2010) dont nous avons déterminé le terme dominant de la MSE. Ce terme explicite la partie asymptotiquement exacte du biais et de la variance. Notre deuxième contribution focalise l'estimation locale linéaire de l'opérateur de régression lorsque les deux variables sont fonctionnelles. On y considère dans un premier temps, la version de Baillo et Grané (2009) dont nous montrons la convergence presque-complète de cet estimateur dans le cas où la variable réponse et la variable explicative appartiennent à un espace de Hilbert. Le cas où la variable explicative est dans un espace semi-métrique est traité dans le dernier chapitre, dans lequel nous utilisons la version de Barrientos et al. (2010) et nous y établissons la convergence presque-complète et uniforme de l'estimateur.

# Chapitre 2

## Exact quadratic error of the local linear regression operator estimator for functional covariates

**Abstract.** In this chapter, it is studied the asymptotic behavior of the non-parametric local linear estimation of the regression operator when the covariates are curves. Under some general conditions we give the exact expression involved in the leading terms of the quadratic error of this estimator. The obtained results affirm the superiority of the local linear modeling over the kernel method, in functional statistics framework.

### 2.1 Introduction

This chapter deals with the nonparametric regression operator estimation by the local linear modeling (cf. Barrientos et al., 2010). This subject is motivated by the fact that the local polynomial estimation method has various advantages over the kernel method (cf. Fan and Gijbels 1996, Barrientos et al. 2010 and Demongeot et al. 2013 and references therein, for finite/infinite

dimensional frameworks). Moreover, the classical Nadaraya-Watson kernel method can be viewed as a particular case of this procedure.

Notice that, the statistical analysis of infinite dimensional data (FDA) has become a major topic of research in the last decade, as evidenced by several special issues of various statistical journals dedicated to this topic (cf. for instance, Davidian et al. (2004), Gonzalez-Manteiga and Vieu (2007), Valderama (2007) and Ferraty (2010)). Furthermore, the nonparametric treatment of such data has also been widely developed in the last few years (cf. Ferraty and Vieu (2006) and Ferraty and Romain (2011) for recent advances and references). Concerning the local linear estimation technique in the functional setup, the first results were given by Baïllo and Grané (2009). They obtained the local linear estimation of the regression function when the explanatory variable takes values in a Hilbert space. Barrientos et al. (2010), meanwhile, established the almost-complete convergence (with rate) of the proposed estimator. We refer, also, to Barrientos et al. (2010) for an other alternative version of the functional local linear modeling. Elsewhere, El Methni and Rachdi (2011) developed this smoothing local linear estimation of the regression operator for functional fixed design data in both Hilbert and semi-metric spaces. More recently, Demongeot et al. (2013 and 2014), Laksaci et al. (2013) and Rachdi et al. (2014) considered the local polynomial modeling of the conditional density function when the explanatory variable is of functional kind. Among the lot of papers on the local linear modeling in the nonfunctional case, we refer to the papers by Chu and Marron (1991), Fan (1992) and Fan and Yao (2003), among others.

The main aim of this chapter is then to determine, under general conditions, the exact rates, in the mean squared error of the local linear estimation of the regression operator, in the i.i.d. case, as proposed in Barrientos et al. (2010). Specifically, we give the exact expression involved in the leading terms of the quadratic error. We point out that our study highlight the structural axes of

this subject, namely the “dimensionality” of the model. Moreover, our results confirm the superiority of the local linear smoothing over the kernel method in the bias terms.

This chapter is organized as follows. In Section 2, we present our model. Then, we give some notations, hypotheses and the presentation of the main results in Section 3. The proofs of the results are relegated to the last Section.

## 2.2 Functional local linear modeling : estimation and comments

Consider  $n$  pairs of independent random variables  $(X_i, Y_i)$  for  $i = 1, \dots, n$  that we assume drawn from the pair  $(X, Y)$  which is valued in  $\mathcal{F} \times \mathbb{R}$ , where  $\mathcal{F}$  is a semi-metric space equipped with a semi-metric  $d$ . Our main purpose is to estimate the regression function  $m(x) = \mathbb{E}[Y|X = x]$ . For this purpose, it is well known that the main idea, in the local linear smoothing, is based on the fact that the function  $m(x)$  admits a linear approximation in the neighborhood of the conditioning point. Recall that, in the non-functional case, this linear approximation is due to a Taylor expansion of  $m(\cdot)$  but, in the functional setup, such approximation can be expressed, for any  $z$  in the neighbourhood of  $x$  by :

$$m(z) = m(x) + b\beta(z, x) + o(\beta(z, x)). \quad (1)$$

Then, the quantities  $a = m(x)$  and  $b$  are estimated by minimizing the following quantity :

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

where the function  $K$  is a kernel,  $h = h_{K,n}$  is a sequence of positive real numbers and  $\delta(\cdot, \cdot)$  and  $\beta(\cdot, \cdot)$  are two known bi-functionals defined from

$\mathcal{F} \times \mathcal{F}$  into  $\mathbb{R}$  such that :

$$\forall \xi \in \mathcal{F}, \beta(\xi, \xi) = 0 \text{ and } d(., .) = |\delta(., .)|.$$

Such fast version of functional local linear estimation has been proposed by Barrientos et al. (2010). They showed, by a simple algebra, that  $\widehat{m}(x) = \widehat{a}$  can be explicitly expressed by :

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

where

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

with the convention  $0/0 = 0$ .

Obviously, if  $b = 0$  then we obtain the Nadaraya-Watson estimator studied, in the functional case, in Ferraty et al. (2010) and the references therein. It is worth to noting that, under the fact that  $\sum_{i,j=1}^n \beta_j W_{ij} = 0$ , we can reformulate the bias term of the estimator  $\widehat{m}(x)$  as follows :

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

Clearly, under the expansion (1), we obtain a bias term of order  $o(h)$ . Undoubtedly, this bias term is significantly better than in the kernel case, studied in Ferraty et al. (2007), which is of order  $O(h)$ . Thus, we can say that the functional local linear modeling is not a simple generalization of the kernel



method, rather more than that it is an alternative approach that has important advantages, in particular in the rate of convergence of the bias term (cf. Theorem 2.3.2 and Remark 2.3.3 below).

## 2.3 Main results

In the remainder of this chapter, we suppose that the bi-functional  $\delta(x, X)$  is function of  $\beta(x, X)$  and we set :

$$\phi_x(r_1, r_2) = \mathbb{P}(r_2 \leq \delta(x, X) \leq r_1)$$

and

$$\Psi(s) = \mathbb{E}[m(X) - m(x) | \beta(x, X) = s].$$

We assume the following hypotheses :

(H1) For any  $r > 0$ ,  $\phi_x(r) := \phi_x(-r, r) > 0$  and there exists a function  $\chi_x(\cdot)$  such that :

$$\forall t \in (-1, 1), \lim_{h \rightarrow 0} \frac{\phi_x(th, h)}{\phi_x(h)} = \chi_x(t).$$

(H2) The first (resp. the second) derivative of  $\Psi$  at 0 exists.

(H3) The bi-functional operator  $\beta$  is such that :

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

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

$$\sup_{u \in B(x, r)} |\beta(u, x) - \delta(x, u)| = o(r)$$

and

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

where  $B(x, r) = \{z \in \mathcal{F} : |\delta(x, z)| \leq r\}$  and  $dP_X(x)$  is the probability distribution of  $X$ .

(H4) The kernel  $K$  is a positive, differentiable function which is supported within  $(-1, 1)$ . Moreover, its derivative  $K'$  satisfies  $K'(t) < 0$ , for  $-1 \leq t < 1$  and  $K(1) > 0$ .

(H5) The function  $m_2(\cdot) = \mathbb{E}[Y^2|X = \cdot]$  is continuous in a neighborhood of  $x$ .

(H6) The bandwidth  $h$  satisfies :

$$\lim_{n \rightarrow \infty} h = 0, \quad \text{and} \quad \lim_{n \rightarrow \infty} n \phi_x(h) = \infty.$$

**Remark 2.3.1.** *Remark that assumptions (H1) and (H2) are an adaptations of conditions  $H_1$  and  $H_3$  in Ferraty et al. (2007), when one replaces the semi-metric  $d$  by some bi-functional operator  $\delta$ . The second part of the assumption (H3) has been introduced and commented in Barrientos et al. (2010). Readers would find, in this last, several examples of bi-functional operators  $\delta$  and  $\beta$  which satisfy this condition. Finally, conditions (H4), (H5) and (H6) are classically used and are standard in the context of the quadratic error determination in functional statistics.*

**Theorem 2.3.2.** *Under assumptions (H1)-(H6), we have that :*

$$\mathbb{E}[\widehat{m}(x) - m(x)]^2 = B_K^2(x)h^4 + \frac{V_K(x)}{n\phi_x(h)} + o(h^4) + o\left(\frac{1}{n\phi_x(h)}\right) \quad (2)$$

where

$$B_K(x) = \frac{1}{2}\Psi''(0) \left[ \frac{K(1) - \int_{-1}^1 (u^2 K(u))' \chi_x(u) du}{K(1) - \int_{-1}^1 K'(u) \chi_x(u) du} \right]$$

and

$$V_K(x) = (m_2(x) - m^2(x)) \left[ \frac{K^2(1) - \int_{-1}^1 (K^2(u))' \chi(u) du}{\left(K(1) - \int_{-1}^1 (K(u))' \chi(u) du\right)^2} \right].$$

**Remark 2.3.3.** We point out that Theorem 2.3.2 shows that, the gain in the bias term is more important than  $o(h)$  which is obtained under assumption (1) (cf. the discussion in Section 2.2). Undoubtedly, this gain is due to the additional regularity condition (H2). However, the variance term is exactly the same as for the classical kernel estimator.

**Proof of Theorem 2.3.2.** It is known that :

$$\mathbb{E} [\widehat{m}(x) - m(x)]^2 = [E(\widehat{m}(x)) - m(x)]^2 + \text{Var} [\widehat{m}(x)].$$

Then the proof of this Theorem is based on the separate calculations of the bias and the variance terms of the estimator  $\widehat{m}(x)$ . For both quantities, we put :

$$\widehat{g}(x) = \frac{1}{n(n-1)EW_{12}(x)} \sum_{i \neq j, 1}^n W_{ij}(x) Y_j$$

and

$$\widehat{f}(x) = \frac{1}{n(n-1)EW_{12}(x)} \sum_{i \neq j, 1}^n W_{ij}(x).$$

So

$$\widehat{m}(x) = \frac{\widehat{g}(x)}{\widehat{f}(x)}.$$

Elsewhere, we have the following decomposition, for  $z \neq 0$  and  $p \in \mathbb{N}^*$  :

$$\frac{1}{z} = 1 - (z-1) + \dots + (-1)^p (z-1)^p + (-1)^{p+1} \frac{(z-1)^{p+1}}{z}.$$

Particularly, for  $z = \widehat{f}(x)$  and  $p = 1$  we get :

$$\begin{aligned} \widehat{m}(x) - m(x) &= (\widehat{g}(x) - m(x)) - (\widehat{g}(x) - \mathbb{E}\widehat{g}(x)) (\widehat{f}(x) - 1) \\ &\quad - \mathbb{E}\widehat{g}(x) (\widehat{f}(x) - 1) + (\widehat{f}(x) - 1)^2 \widehat{m}(x). \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{E} [\widehat{m}(x)] - m(x) &= (\mathbb{E}\widehat{g}(x) - m(x)) - \text{Cov}(\widehat{g}(x), \widehat{f}(x)) \\ &\quad + E \left( \widehat{f}(x) - \mathbb{E}\widehat{f}(x) \right)^2 \widehat{m}(x). \end{aligned}$$

Consequently, the bias term can be expressed by :

$$\mathbb{E}[\widehat{m}(x)] - m(x) = (\mathbb{E}\widehat{g}(x) - m(x)) - \text{Cov}(\widehat{g}(x), \widehat{f}(x)) + O\left(\text{Var}\left(\widehat{f}(x)\right)\right).$$

Finally, the proof of Theorem 2.3.2 is based on Lemmas 2.3.4 and 2.3.5, bellow.

**Lemma 2.3.4.** *Under Assumptions (H1)-(H4) and (H6), we have :*

$$\mathbb{E}[\widehat{g}(x)] - m(x) = B_K(x, y)h^2 + o(h^2) + O\left(\frac{1}{n\phi_x(h)}\right).$$

**Lemma 2.3.5.** *Under Assumptions (H1) and (H4)-(H6), we have :*

$$\text{Var}[\widehat{m}(x)] = V_K(x) \left[ \frac{K^2(1) - \int_{-1}^1 (K^2(u))' \chi(u) du}{\left(K(1) - \int_{-1}^1 (K(u))' \chi(u) du\right)^2} \right] + o\left(\frac{1}{n\phi_x(h)}\right).$$

Furthermore,

$$\text{Cov}(\widehat{g}(x), \widehat{f}(x)) = O\left(\frac{1}{n\phi_x(h)}\right)$$

and

$$\text{Var}[\widehat{f}(x)] = O\left(\frac{1}{n\phi_x(h)}\right).$$

## 2.4 Appendix : Proofs

In what follows, when no confusion is possible, we will denote by  $C$  and  $C'$  some strictly positive generic constants. Moreover, we set for all  $i, j = 1, \dots, n$  and for a fixed  $(x, y) \in \mathcal{F} \times \mathbb{R}$  :

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

**Proof of Lemma 2.3.4.** We have :

$$\begin{aligned} \mathbb{E}[\widehat{g}(x)] &= \mathbb{E}\left[\frac{1}{n(n-1)\mathbb{E}[W_{12}]} \sum_{j \neq i, 1}^n W_{ij} Y_j\right] \\ &= \frac{\mathbb{E}[W_{12} Y_2]}{\mathbb{E}[W_{12}]} = \frac{1}{\mathbb{E}[W_{12}]} \mathbb{E}[W_{12} \mathbb{E}[Y_2 | X_2]]. \end{aligned} \quad (3)$$

Then, it follows from (3) and the definition of the operator  $m$  that :

$$\mathbb{E} [\widehat{g}(x)] = \frac{1}{\mathbb{E}[W_{12}]} \mathbb{E} [W_{12}m(X_2)].$$

Now, by the same arguments as those used in Ferraty et al. (2007), for the regression operator estimation, we show that :

$$\begin{aligned} \mathbb{E} [W_{12}m(X_2)] &= m(x)\mathbb{E}[W_{12}] + \mathbb{E} [W_{12} (m(X_2) - m(x))] \\ &= m(x)\mathbb{E}[W_{12}] + \mathbb{E} [W_{12}\mathbb{E} [m(X_2) - m(x)|\beta(x, X_2)]] \\ &= m(x)\mathbb{E}[W_{12}] + \mathbb{E} [W_{12}\Psi (\beta(x, X_2))] \end{aligned}$$

and since  $\mathbb{E} [\beta(x, X_2)W_{12}] = 0$  and  $\Psi(0) = 0$ , we obtain :

$$\mathbb{E} [W_{12}\Psi (\beta(x, X_2))] = \frac{1}{2}\Psi''(0)\mathbb{E} [\beta^2(x, X_2)W_{12}] + o(\mathbb{E} [\beta^2(x, X_2)W_{12}]).$$

Then :

$$\mathbb{E} [\widehat{g}(x)] = m(x) + \Psi''_0(0) \frac{\mathbb{E} [\beta^2(x, X_2)W_{12}]}{2\mathbb{E}[W_{12}]} + o\left(\frac{\mathbb{E} [\beta^2(x, X_2)W_{12}]}{\mathbb{E}[W_{12}]}\right).$$

Moreover, it is clear that :

$$\begin{aligned} \mathbb{E} [\beta(x, X_2)^2W_{12}] &= (\mathbb{E} [K_1\beta_1^2])^2 - \mathbb{E}[K_1\beta_1]\mathbb{E}[K_1\beta_1^3] \\ \mathbb{E} [W_{12}] &= \mathbb{E}[K_1\beta_1^2]EK_1 - (\mathbb{E}[K_1\beta_1])^2 \end{aligned}$$

and, under the assumption (H4), we obtain that :

$$\text{for all } a > 0, \mathbb{E}[K_1^a\beta_1] \leq C \int_{B(x,h)} \beta(u, x)dP_X(u).$$

So, by using the last part of the assumption (H3), we get :

$$h\mathbb{E}[K_1^a\beta_1] = o\left(\int_{B(x,h)} \beta^2(u, x)dP_X(u)\right) = o(h^2\phi_x(h))$$

which allows to write :

$$\mathbb{E}[K_1^a\beta_1] = o(h\phi_x(h)). \quad (4)$$

Moreover, for all  $b > 1$ , we can write :

$$\mathbb{E}[K_1^a \beta_1^b] = \mathbb{E}[K_1^a \delta^b(x, X_1)] + \mathbb{E} [K_1(\beta^b(X_1, x) - \delta^b(x, X_1))].$$

Then, the second part of the assumption (H3) implies that :

$$\begin{aligned} & \mathbb{E} [K_1^a(\beta^b(X_1, x) - \delta^b(x, X_1))] \\ &= \mathbb{E} \left[ K_1^a I_{B(x,h)}(\beta(X_1, x) - \delta(x, X_1)) \sum_{l=0}^{b-1} (\beta(X_1, x))^{b-1-l} (\delta(x, X_1))^l \right] \\ &\leq \sup_{u \in B(x,h)} |\beta(u, x) - \delta(x, u)| \sum_{l=0}^{b-1} \mathbb{E} [K_1^a I_{B(x,h)} |\beta(X_1, x)|^{b-1-l} |\delta(x, X_1)|^l] \end{aligned}$$

whereas the first part of the assumption (H3) gives :

$$I_{B(x,h)} |\beta(X_1, x)| \leq I_{B(x,h)} |\delta(x, X_1)|.$$

Thus, it follows :

$$\begin{aligned} \mathbb{E} [K_1^a(\beta^b(X_1, x) - \delta^b(x, X_1))] &\leq b \sup_{u \in B(x,h)} |\beta(u, x) - \delta(x, u)| \mathbb{E}[K_1^a |\delta|^{b-1}(x, X_1)] \\ &\leq b \sup_{u \in B(x,h)} |\beta(u, x) - \delta(x, u)| h^{b-1} \mathbb{E}[K_1^a] \\ &\leq b \sup_{u \in B(x,h)} |\beta(u, x) - \delta(x, u)| h^{b-1} \phi_x(h) \end{aligned}$$

which allows to write :

$$\mathbb{E}[K_1^a \beta_1^b] = \mathbb{E}[K_1^a \delta^b(x, X_1)] + o(h^b \phi_x(h)).$$

Concerning the term  $\mathbb{E}[K_1^a \delta^b]$ , we write :

$$\begin{aligned} h^{-b} \mathbb{E}[K_1^a \delta^b] &= \int v^b K^a(v) dP_X^{h^{-1}\delta(x, X_1)}(v) \\ &= \int_{-1}^1 \left[ K^a(1) - \int_v^1 (u^b K^a(u))' du \right] dP_X^{h^{-1}\delta(x, X_1)}(v) \\ &= K^a(1) \phi_x(h) - \int_{-1}^1 (u^b K^a(u))' \phi_x(uh, h) du \\ &= \phi_x(h) \left( K^a(1) - \int_{-1}^1 (u^b K^a(u))' \frac{\phi_x(uh, h)}{\phi_x(h)} du \right). \end{aligned}$$

Finally, under the assumption (H1), we get :

$$\mathbb{E}[K_1^a \beta_1^b] = h^b \phi_x(h) \left( K^a(1) - \int_{-1}^1 (u^b K^a(u))' \chi_x(u) du \right) + o(h^b \phi_x(h)). \quad (5)$$

It follows that :

$$\frac{\mathbb{E}[\beta^2(x, X_2) W_{12}]}{\mathbb{E}[W_{12}]} = h^2 \left[ \frac{K(1) - \int_{-1}^1 (u^2 K(u))' \chi_x(u) du}{K(1) - \int_{-1}^1 K'(u) \chi_x(u) du} \right] + o(h^2).$$

Consequently :

$$\mathbb{E}[\widehat{g}(x)] = m(x) + \frac{h^2}{2} \Psi_0''(0) \left[ \frac{K(1) - \int_{-1}^1 (u^2 K(u))' \chi_x(u) du}{K(1) - \int_{-1}^1 K'(u) \chi_x(u) du} \right] + o(h^2).$$

■

**Proof of Lemma 2.3.5.** For this Lemma, we use the same ideas of Sarda and Vieu (2000) to show that

$$\begin{aligned} \text{Var}[\widehat{m}(x)] &= \text{Var}[\widehat{g}(x)] - 2(\mathbb{E}\widehat{g}(x))\text{Cov}(\widehat{g}(x), \widehat{f}(x)) \\ &\quad + (\mathbb{E}\widehat{g}(x))^2 \text{Var}(\widehat{f}(x)) + o\left(\frac{1}{n\phi_x(h)}\right). \end{aligned}$$

It is clear that :

$$\begin{aligned} \text{Var}(\widehat{g}(x)) &= \frac{1}{(n(n-1)\mathbb{E}[W_{12}])^2} \text{Var}\left(\sum_{i \neq j=1}^n W_{ij} Y_j\right) \\ &= \frac{1}{(n(n-1)(\mathbb{E}W_{12})^2)} \left( n(n-1)\mathbb{E}[W_{12}^2 Y_2^2] + n(n-1)\mathbb{E}[W_{12}W_{21}Y_2Y_1] \right. \\ &\quad \left. + n(n-1)(n-2)\mathbb{E}[W_{12}W_{13}Y_2Y_3] + n(n-1)(n-2)\mathbb{E}[W_{12}W_{23}Y_2Y_3] \right. \\ &\quad \left. + n(n-1)(n-2)\mathbb{E}[W_{12}W_{31}Y_2Y_1] + n(n-1)(n-2)\mathbb{E}[W_{12}W_{32}Y_2^2] \right. \\ &\quad \left. - n(n-1)(4n-6)(\mathbb{E}[W_{12}Y_2])^2 \right). \end{aligned}$$

Observe that the terms of the first line are negligible compared to others terms which are multiplied by  $n(n-1)(n-2)$ . Furthermore,

$$\begin{aligned}
\mathbb{E}[W_{12}^2 Y_2^2] &= O(h^4 \phi_x^2(h)), \\
\mathbb{E}[W_{12} W_{21} Y_1 Y_2] &= O(h^4 \phi_x^2(h)), \\
\mathbb{E}[W_{12} W_{13} Y_2 Y_3] &= (m(x))^2 \mathbb{E}[\beta_1^4 K_1^2] (\mathbb{E}[K_1])^2 + o(h^4 \phi_x^3(h)), \\
\mathbb{E}[W_{12} W_{23} Y_2 Y_3] &= (m(x))^2 \mathbb{E}[\beta_1^2 K_1] (\mathbb{E}[\beta_1^2 K_1^2] \mathbb{E}[K_1]) + o(h^4 \phi_x^3(h)), \\
\mathbb{E}[W_{12} W_{31} Y_2 Y_1] &= (m(x))^2 \mathbb{E}[\beta_1^2 K_1] (\mathbb{E}[\beta_1^2 K_1^2] \mathbb{E}[K_1]) + o(h^4 \phi_x^3(h)), \\
\mathbb{E}[W_{12} W_{32} Y_2^2] &= (m_2(x)) (\mathbb{E}[\beta_1^2 K_1])^2 (\mathbb{E}[K_1^2]) + o(h^4 \phi_x^3(h)), \\
\mathbb{E}[W_{12} Y_2] &= O(h^2 \phi_x^2(h)).
\end{aligned}$$

Therefore, the leading term in the expression of  $\text{Var}(\widehat{g}(x))$  is :

$$\begin{aligned}
&\frac{n(n-1)(n-2)}{(n(n-1)\mathbb{E}[W_{12}])^2} \left( (m(x))^2 (\mathbb{E}[\beta_1^4 K_1^2] (\mathbb{E}[K_1])^2 + 2BB\mathbb{E}[\beta_1^2 K_1] (\mathbb{E}[\beta_1^2 K_1^2] \mathbb{E}[K_1])) \right. \\
&\quad \left. + (m_2(x)) (\mathbb{E}[\beta_1^2 K_1])^2 (\mathbb{E}[K_1^2]) + o(h^4 \phi_x^3(h)) \right)
\end{aligned}$$

Concerning the covariance term, we have by the same fashion :

$$\begin{aligned}
\text{Cov}(\widehat{g}(x), \widehat{f}(x)) &= \frac{1}{(n(n-1)\mathbb{E}[W_{12}])^2} \text{Cov} \left( \sum_{\substack{i,j=1 \\ i \neq j}}^n W_{ij} Y_j, \sum_{\substack{i',j'=1 \\ i' \neq j'}}^n W_{i'j'} \right) \\
&= \frac{1}{(n(n-1)\mathbb{E}[W_{12}])^2} \left[ n(n-1)\mathbb{E}[W_{12}^2 Y_2] + n(n-1)\mathbb{E}[W_{12} W_{21} Y_2] \right. \\
&\quad + n(n-1)(n-2)\mathbb{E}[W_{12} W_{13} Y_2] + n(n-1)(n-2)\mathbb{E}[W_{12} W_{23} Y_2] \\
&\quad + n(n-1)(n-2)\mathbb{E}[W_{12} W_{31} Y_2] + n(n-1)(n-2)\mathbb{E}[W_{12} W_{32} Y_2] \\
&\quad \left. - n(n-1)(4n-6)(\mathbb{E}[W_{12} Y_2] \mathbb{E}[W_{12}]) \right]
\end{aligned}$$



with

$$\begin{aligned}
\mathbb{E}[W_{12}^2 Y_2] &= O(h^4 \phi_x^2(h)), \\
\mathbb{E}[W_{12} W_{21} Y_2] &= O(h^4 \phi_x^2(h)), \\
\mathbb{E}[W_{12} W_{13} Y_2] &= (m(x)) \mathbb{E}[\beta_1^4 K_1^2] (\mathbb{E}[K_1])^2 + o(h^4 \phi_x^3(h)), \\
\mathbb{E}[W_{12} W_{23} Y_2] &= (m(x)) \mathbb{E}[\beta_1^2 K_1] (\mathbb{E}[\beta_1^2 K_1^2] \mathbb{E}[K_1]) + o(h^4 \phi_x^3(h)), \\
\mathbb{E}[W_{12} W_{31} Y_2] &= (m(x)) \mathbb{E}[\beta_1^2 K_1] (\mathbb{E}[\beta_1^2 K_1^2] \mathbb{E}[K_1]) + o(h^4 \phi_x^3(h)), \\
\mathbb{E}[W_{12} W_{32} Y_2] &= (m(x)) (\mathbb{E}[\beta_1^2 K_1])^2 (\mathbb{E}[K_1^2]) + o(h^4 \phi_x^3(h)), \\
\mathbb{E}[W_{12} Y_1] &= O(h^2 \phi_x^2(h)).
\end{aligned}$$

Therefore, the leading term in the expression of  $\text{Cov}(\widehat{g}(x), \widehat{f}(x))$  is :

$$\begin{aligned}
&\frac{n(n-1)(n-2)}{(n(n-1)\mathbb{E}[W_{12}])^2} (m(x) (\mathbb{E}[\beta_1^4 K_1^2] (\mathbb{E}[K_1])^2 + 2\mathbb{E}[\beta_1^2 K_1] (\mathbb{E}[\beta_1^2 K_1^2] \mathbb{E}[K_1]) \\
&\quad + (\mathbb{E}[\beta_1^2 K_1])^2 (\mathbb{E}[K_1^2])) + o(h^4 \phi_x^3(h))).
\end{aligned}$$

Finally, for  $\text{Var}(\widehat{f}(x))$

$$\begin{aligned}
\text{Var}(\widehat{f}(x)) &= \frac{1}{(n(n-1)\mathbb{E}[W_{12}])^2} \left[ n(n-1)\mathbb{E}[[W_{12}^2] + n(n-1)\mathbb{E}[[W_{12}W_{21}] \right. \\
&\quad + n(n-1)(n-2)\mathbb{E}[[W_{12}W_{13}] + n(n-1)(n-2)\mathbb{E}[[W_{12}W_{23}] \\
&\quad + n(n-1)(n-2)\mathbb{E}[[W_{12}W_{31}] + n(n-1)(n-2)\mathbb{E}[[W_{12}W_{32}] \\
&\quad \left. - n(n-1)(4n-6)(\mathbb{E}[[W_{12}])^2] \right]
\end{aligned}$$

and similarly to the previous cases :

$$\begin{aligned}
\mathbb{E}[W_{12}^2] &= O(h^4 \phi_x^2(h)), \\
\mathbb{E}[[W_{12}W_{21}] &= O(h^4 \phi_x^2(h)), \\
\mathbb{E}[[W_{12}W_{13}] &= \mathbb{E}[[\beta_1^4 K_1^2] (\mathbb{E}[[K_1])^2 + o(h^4 \phi_x^3(h)), \\
\mathbb{E}[[W_{12}W_{23}] &= \mathbb{E}[[\beta_1^2 K_1] (\mathbb{E}[[\beta_1^2 K_1^2] \mathbb{E}[[K_1]) + o(h^4 \phi_x^3(h)), \\
\mathbb{E}[[W_{12}W_{31}] &= \mathbb{E}[[\beta_1^2 K_1] (\mathbb{E}[[\beta_1^2 K_1^2] \mathbb{E}[[K_1]) + o(h^4 \phi_x^3(h)), \\
\mathbb{E}[[W_{12}W_{32}] &= (\mathbb{E}[[\beta_1^2 K_1])^2 (\mathbb{E}[[K_1^2]) + o(h^4 \phi_x^3(h)), \\
\mathbb{E}[[W_{12}] &= O(h^2 \phi_x^2(h)).
\end{aligned}$$

Therefore,

$$\begin{aligned} & \text{Var}(\widehat{m}(x)) \\ &= \frac{(m_2(x) - m^2(x))}{n\phi_x(h)} \left[ \frac{\left( K^2(1) - \int_{-1}^1 (K^2(u))' \chi(u) du \right)}{\left( K(1) - \int_{-1}^1 (K(u))' \chi(u) du \right)^2} \right] + o\left(\frac{1}{n\phi_x(h)}\right). \end{aligned}$$

which completes the proof. ■

# Chapitre 3

## Local linear regression modelization when all variables are curves

**Abstract.** A nonparametric local linear estimator of the regression function when both the response and the explanatory variables are of the functional kind is constructed. Then, its rate of uniform almost-complete convergence is stated. This theoretical result will be a key tool for many further developments in functional data analysis.

### 3.1 Introduction

Let  $(X, Y)$  be a couple of random variables which is valued in  $\mathcal{F} \times \mathcal{H}$  where  $\mathcal{F}$  (respectively  $\mathcal{H}$ ) is a Hilbert space endowed with the inner product  $\langle \cdot, \cdot \rangle_{\mathcal{F}}$  (respectively  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ ). A recurring problem in statistics is the prediction of a response variable according to an explanatory one. To this aim, we are interested in the link between two Hilbertian variables i.e. :

$$Y = r(X) + \varepsilon, \tag{1}$$

where  $r$  is an operator from  $\mathcal{F}$  to  $\mathcal{H}$  and  $\varepsilon$  is an error random variable with mean zero.

Recall that statistical modeling of functional data has experienced great developments and has generated much interest among statisticians during the last decade. This importance comes from the diversity of application fields in which observations are curves, surfaces or images (cf. [18]).

The regression model, either parametric or nonparametric, remains the most favored tool for analyzing the co-variability between two variables  $X$  and  $Y$ . Many works have treated this kind of problematic. To cite just some popular ones, consider [4] and [18] for the parametric case and [36] for the nonparametric case. There, the authors studied, among others, the regression model when only the covariate is of functional type. However, the literature on the nonparametric regression model when the two variables are curves remains limited. Among some of the contributions in this literature, we point out [18] which deals with the  $L_p$  convergence of the kernel estimator of the nonparametric regression and [40], which investigates the strong consistency for independent and identically distributed random variables. In [40], the authors also address the case where the response variable belongs to a Banach space and the explanatory variable belongs to a semi-metric space. A generalization of the results obtained in [40] is provided in [42] by the same authors. We return to [14] and [17] for some alternative approaches in nonparametric functional data analysis. While [17] establish the uniform almost complete of the functional kNN estimator of the regression function, [14] consider the nonparametric estimation for conditional copula to analyze the relationship between real variable and functional regressors. Among the lot of recent contributions in FDA we refer only to the general works of [71] [47], [17] and [46].

In most of the aforementioned works, the authors used the regression kernel estimator. However, it is known that the latter produces high bias compared to the local linear method (cf. [8] for a thorough discussion on the comparison between the two methods). Local linear regression for functional data was introduced with several versions. The first version is considered by [1]. They studied the mean quadratic consistency of the constructed estimator when the covariates are of Hilbertian nature. Another fast version of the functional local linear estimate was given by [2]. Such a version is applied for functional nonparametric regression with banachic regressor. Contribution [3] state the asymptotic mean square error of a functional local linear estimate of the regression operator constructed by inverting the local covariance operator of the functional explanatory variable. Recently, [7] treat the functional local linear estimate of the conditional density function. These authors showed the almost complete consistency of this estimate.

Our goal is to generalize [1], considering both the response and the explanatory variables are of functional kind. In this direction, we construct and study the almost-complete uniform convergence of the local linear estimator of the regression operator. Then, the obtained rate of convergence confirms the superiority of this estimation method compared to the kernel method because it provides a lower bias.

In the following section, we present our model and the steps that allow us to construct our local linear estimator. In Section 3, we give an evaluation of the estimation error. The needed hypotheses, the main result and a sketch of its proof are given in Section 4. Section 5 is devoted to some discussions on the model, method and results. The proofs of Lemmas are deferred to the Section 6.

### 3.2 The model and the estimator

Let  $\|\cdot\|_{\mathcal{F}}$  (respectively  $\|\cdot\|_{\mathcal{H}}$ ) be the induced norm of the inner product  $\langle \cdot, \cdot \rangle_{\mathcal{F}}$  (respectively  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ ). We consider a sample of  $n$  pairs  $(X_i, Y_i)$  which follow the same law as  $(X, Y)$ .

Generally, the local polynomial estimation of the regression operator  $r$  is based on a regularity condition of this last which allows us to approximate it locally by a polynomial. We proceed by approximating the operator :

$$r(x_0) = a(x) + b_x(x_0 - x) + \rho_x(x_0 - x, x_0 - x) + o(\|x_0 - x\|_{\mathcal{F}}^2), \quad (2)$$

for every  $x_0$  in a neighborhood of  $x$  in  $\mathcal{F}$ , where  $b_x$  is a linear operator (respectively  $\rho_x$  is a bilinear operator) from  $\mathcal{F}$  to  $\mathcal{H}$  (respectively from  $\mathcal{F} \times \mathcal{F}$  to  $\mathcal{H}$ ). The two operators  $a(x)$  and  $b_x$  are the minimizers of :

$$\min_{a,b} \sum_{i=1}^n \|Y_i - a - b(X_i - x)\|_{\mathcal{H}}^2 K(h^{-1}\|x - X_i\|_{\mathcal{F}}), \quad (3)$$

where  $K$  is a kernel and  $h = h_{K,n}$  is a sequence of positive real numbers. Using the linearity of  $b$  and the decomposition of  $(X_i - x)$  on a orthonormal basis  $(v_j)_{j \geq 1}$  of  $\mathcal{F}$ , we replace the criterion (2) by :

$$\min_{a,b} \sum_{i=1}^n \left\| Y_i - a - \sum_{j \geq 1} c_{ij} b(v_j) \right\|_{\mathcal{H}}^2 K(h^{-1}\|x - X_i\|_{\mathcal{F}}),$$

where  $c_{ij}$  are the coefficients of  $(X_i - x)$  in the basis  $(v_j)_{j \geq 1}$ . As this criterion can not be explicitated completely, in practice, we consider its truncated version to a threshold  $J$ . Thus, we estimate  $a(x)$  and  $b_x$  by minimizing the following criterion :

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

where  $b_j = b(v_j)$  for  $j = 1, \dots, J$ . Then, the estimators are explicitly given by :

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

$K = \text{diag}(K(h^{-1}\|x - X_1\|_{\mathcal{F}}), \dots, K(h^{-1}\|x - X_n\|_{\mathcal{F}}))$  and  $Y' = (Y_1, \dots, Y_n)$ .

### 3.3 Estimation error

In order to establish the almost-complete convergence of our estimator, we evaluate the estimation error which is very useful even to establish other asymptotic properties such as convergence, mean square consistency or asymptotic normality. Before proceeding, we introduce the following notations, for  $k, j = 1, \dots, J$  :

$$\begin{aligned} S_{n,k,j}(x) &= \frac{1}{nh^2\phi_x(h)} \sum_{i=1}^n c_{ik}c_{ij}K_i(x) & , & \quad S_{n,k,0}(x) = \frac{1}{nh\phi_x(h)} \sum_{i=1}^n c_{ik}K_i(x) \\ T_{n,j}(x) &= \frac{1}{nh\phi_x(h)} \sum_{i=1}^n c_{ij}K_i(x)Y_i & , & \quad T_{n,0}(x) = \frac{1}{n\phi_x(h)} \sum_{i=1}^n K_i(x)Y_i \\ T_{n,j}^*(x) &= \frac{1}{nh\phi_x(h)} \sum_{i=1}^n c_{ij}K_i(x)(Y_i - r(X_i)) & , & \quad T_{n,0}^*(x) = \frac{1}{n\phi_x(h)} \sum_{i=1}^n K_i(x)(Y_i - r(X_i)) \\ e_{n,j}(x) &= \frac{1}{nh\phi_x(h)} \sum_{i=1}^n c_{ij}K_i(x)\rho_x(X_i - x, X_i - x) & , & \quad e_{n,0}(x) = \frac{1}{n\phi_x(h)} \sum_{i=1}^n K_i(x)\rho_x(X_i - x, X_i - x) \\ e_{n,j}^*(x) &= S_{n,j,0}(x) = \frac{1}{nh\phi_x(h)} \sum_{i=1}^n c_{ij}K_i(x) & , & \quad e_{n,0}^*(x) = S_{n,0,0}(x) = \frac{1}{n\phi_x(h)} \sum_{i=1}^n K_i(x) \end{aligned}$$

where  $K_i(x) = K(h^{-1}\|x - X_i\|_{\mathcal{F}})$  and  $\phi_x(h) = P(X \in B(x, h))$  with  $B(x, h) = \{z \in \mathcal{F} \text{ such that } \|z - x\|_{\mathcal{F}} \leq h\}$ . The function  $\phi_x(h)$  is assu-

med to be a strictly positive. Then, we may write :

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

where

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

Under the approximation (2) and the fact that the error of the projection on an orthonormal basis of a Hilbert space (as for the Fourier series decomposition) is of order  $O(J^{-1})$ , we have :

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

Furthermore, consider  $T_n^*(x) = (T_{n,j}^*(x))_{j=0,\dots,J}$ ,  $e_n(x) = (e_{n,j}(x))_{j=0,\dots,J}$  and  $e_n^*(x) = (e_{n,j}^*(x))_{j=0,\dots,J}$ . We may write :

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

Consequently :

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



We deduce the following error estimate :

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

where  $e_1'$  denotes the transpose vector of the first vector of the canonical basis of  $R^J$ .

### 3.4 Uniform almost-complete convergence

Our main result is the uniform almost-complete convergence of the estimator  $\widehat{a}(\cdot)$  on a subset  $S_{\mathcal{F}}$  of  $\mathcal{F}$ . The latter verifies  $S_{\mathcal{F}} \subset \bigcup_{k=1}^{d_n} B(x_k, r_n)$  where  $x_k \in \mathcal{F}$  and  $r_n$  (respectively  $d_n$ ) is a sequence of positive real numbers.

In what follows, when there is no confusion, we denote by  $C$  or/and  $C'$  any generic positive constants.

(H1) There is a differentiable function  $\phi(\cdot)$  such that :

$$\forall x \in S_{\mathcal{F}}, 0 < C \phi(h) \leq \phi_x(h) \leq C' \phi(h) < \infty \quad \text{and} \quad \exists \eta_0 > 0, \forall \eta < \eta_0, \phi'(\eta) < C \quad \text{where } \phi' \text{ denotes the first derivative of } \phi.$$

(H2) There exists  $m \geq 2$  and  $C > 0$  such that  $\mathbb{E}(\|Y\|_{\mathcal{H}}^m | X = x) < C$  and  $\mathbb{E}(\|r(X)\|_{\mathcal{H}}^m | X = x) < C$ .

(H3) For all  $x \in S_{\mathcal{F}}$ , the operator  $\rho_x$  is continuous and satisfies  $\sup_{x \in S_{\mathcal{F}}} \|\rho_x\|_{\mathcal{H}} < C$ .

(H4) The kernel  $K$  is a positive function, differentiable with compact support in  $[0, 1]$  and satisfies :

$$\text{There exists } C > 0 \text{ such that, for all } u, v, \in \mathbb{R} : |K(u) - K(v)| \leq C |u - v|.$$

(H5) Taking  $r_n = O(\log n/n)$ , the sequence  $d_n$  verifies :

$$\frac{(\log n)^2}{n \phi(h)} < \log(d_n) < \frac{n \phi(h)}{\log n} \quad \text{and} \quad \exists \beta > 1, \sum_{n=1}^{\infty} d_n^{(1-\beta)} < \infty,$$

Notice that these conditions are very standard in this context. The covering condition on subset  $S_{\mathcal{F}}$  is linked to the topological structure of the functional space  $\mathcal{F}$ . It control Kolmogorov's entropy of the set  $S_{\mathcal{F}}$ . Such consideration has been introduced in non parametric functional statistics by [10]. The latter gives a several examples of subset  $S_{\mathcal{F}}$  and functional space  $\mathcal{F}$  for which this condition is verified. Moreover, conditions (H1), (H2), (H4) and (H5) are the same as those used in [40]. Assumption (H3) is a regularity condition used to evaluate the bias term of the estimate

**Theorem 3.4.1.** *Under hypotheses (H1)-(H5), we have :*

$$\sup_{x \in S_{\mathcal{F}}} \|\widehat{a}(x) - a(x)\|_{\mathcal{H}} = O(J^{-1}) + O(h^2) + O_{a.co.} \left( \sqrt{\frac{\log d_n}{n\phi(h)}} \right)$$

*Proof of Theorem 3.4.1* The proof is based on the following error formula :

$$(\widehat{a}(x) - a(x)) = e'_1 (S_n^{-1}(x)T_n^*(x) - S_n^{-1}(x)e_n(x) - O(J^{-1})S_n^{-1}(x)e_n^*(x))$$

Thus, Theorem 3.4.1 result will be a consequence of the following lemmas for which the proofs are given in the last section of this chapter.

**Lemma 3.4.2.** *Under conditions (H1), (H4)-(H5), we have :*

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

**Lemma 3.4.3.** *Under hypotheses (H1) and (H4), we have :*

$$\sup_{x \in S_{\mathcal{F}}} |\mathbb{E}[S_{n,k,j}(x)]| = O(1).$$

**Corollary 3.4.4.** *Under hypotheses of the Lemma 4.3.2, we have that :*

$$\sup_{x \in S_{\mathcal{F}}} |S_n(x)| = O_{a.co.}(1) \quad \text{and} \quad \sup_{x \in S_{\mathcal{F}}} |e_n^*(x)| = O_{a.co.}(1).$$

**Lemma 3.4.5.** *Under hypotheses (H1), (H3) and (H4), we have :*

$$\sup_{x \in S_{\mathcal{F}}} \|\mathbb{E} [T_{n,j}^*(x)]\|_{\mathcal{H}} = 0 \quad \text{and} \quad \sup_{x \in S_{\mathcal{F}}} \|\mathbb{E} [e_{n,j}(x)]\|_{\mathcal{H}} = O(h^2)$$

**Lemma 3.4.6.** *Under conditions (H1)-(H5), we have :*

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

and

$$\sup_{x \in S_{\mathcal{F}}} \|e_{n,j}(x) - \mathbb{E} [e_{n,j}(x)]\|_{\mathcal{H}} = O_{a.co.} \left( \sqrt{\frac{\log d_n}{n\phi(h)}} \right)$$

**Corollary 3.4.7.** *Under the hypotheses of Lemma 3.4.6, we have :*

$$\sup_{x \in S_{\mathcal{F}}} \|T_n^*(x)\|_{\mathcal{H}} = O_{a.co.} \left( \sqrt{\frac{\log d_n}{n\phi(h)}} \right)$$

and

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

### 3.5 Some concluding remarks

- *On the estimation method :* This work complements the literature on the nonparametric estimation when the variables, in the model, are of functional kind. One of the originalities of this work is that the response variable (as well as the explanatory variable) may be of infinite dimensional. The second feature of this chapter is to use locally linear estimators, while most of the literature in the field concerns kernel estimators (Nadaraya-Watson type). The difficulty with local linear estimators is that they are not written directly as one (or more) sums of independent and identically distributed variables. Notice that the

technical proofs for treating this type of estimators (in the multivariate setup) are different, then we have opted for a "direct" way of giving a decomposition (relatively complex) of our estimators as sums of independent and identically distributed variables. From this decomposition, one may think that the proofs follow very traditional paths : regularity assumption to control bias and exponential inequality (for instance, the hilbetian version of Bernstein inequality) to process dispersion terms.

- *On the convergence rate expression* Note that the convergence rate of our estimator depends on parameters related to the topological structure of the functional space of the variables and/or the method of construction of the estimator. Our results confirm the superiority of the local linear method over the kernel method. Indeed, the kernel method provides a bias of order  $O(h)$ , whereas it is of order  $O(h^2)$  for the local linear method. On the other hand, for the variance term we obtain the same finding/rate as in the multidimensional case. Only in part of our study, the variance term is expressed in terms of the concentration  $\phi(h)$  and the entropy  $d_n$  which depends, in turn, on  $S_{\mathcal{F}}$  and on the sequence  $r_n$ . These two functions are related to the topological structure of the explanatory variable. The additional term  $O(J^{-1})$ , in turn, is linked to the construction method of the estimator and controls the orthogonal projection error on the subspace generated by  $v_1, \dots, v_J$ . This additional term is not a deterioration factor because a simple choice of  $J$  i.e.,  $J = O\left(\sqrt{n/\log d_n}\right)$ , provides a standard expression of the rate of convergence :  $O\left(h^2 + \sqrt{\log d_n/n\phi(h)}\right)$ . However, in this chapter, we opted for a presentation to exploit this term through the expression of the convergence rate. This highlights the importance of the choice of  $J$ . A wise choice of this parameter will

be based on a balance between accuracy and speed in the calculation of the estimator. Indeed, too little choice of  $J$  provides quick estimator to calculate but has a slow convergence speed and vice versa.

- *Some prospects* As naturel prospect of the present work is the functional local polynomial estimation of the regression operator. This estimate increases the gain in bias term over the classical kernel method. Such a generalization can be obtained by reenforcing the regularity condition (2) in order to approximate locally  $r$  by a polynomial operator. The explicit determination of the estimator can be obtained by a straightforward modification of the present approach. In addition to this prospect, we mention more others one such as the asymptotic normality of our estimate, the conditional density estimation or the functional kNN local linear estimate, among others.

### 3.6 Proofs

For the sake for shortness we omit the *Proof of Lemma 4.3.2* which is based on the same ideas as of [40].

*Proof of Lemma 3.4.3.* We will show the result just in the case where  $k$  and  $j$  are different to 0 (the others case are identical). Notice that, the proof of this Lemma is based only on the fact that observations are identically distributed.

$$\mathbb{E}[S_{n,k,j}(x)] = \frac{1}{h^2 \phi_x(h)} \mathbb{E}[c_{1k} c_{1j} K_1(x)].$$

Then, by using the fact that the basis  $(v_j)_j$  is orthonormal and that :

$$\forall j \leq J, |c_{1j}| \leq \|x - X_1\|$$

Consequently,

$$\mathbb{E}[c_{1k} c_{1j} K_1(x)] \leq \mathbb{E}[\|x - X_1\|^2 K_1(x)] \leq Ch^2 \phi_x(h).$$

Finally, using the first part of the hypothesis (H1) permits to achieve the proof of this Lemma.  $\blacksquare$

*Proof of Lemma 3.4.5.* We will treat only the case where  $j \neq 0$ . Other cases (for instance  $j = 0$ ) are similar. The proof is also based on the identically distribution of observations.

$$\mathbb{E}[T_{n,j}^*(x)] = \frac{1}{h^2 \phi_x(h)} \mathbb{E}[c_{1k} c_{ij} K_1(x) (Y_1 - r(X_1))].$$

By conditioning by  $X_1$ , we show that :

$$\mathbb{E}[T_{n,j}^*(x)] = \frac{1}{h^2 \phi_x(h)} \mathbb{E}[c_{1k} c_{ij} K_1(x) \mathbb{E}[(Y_1 - r(X_1)) | X_1]].$$

Then

$$\mathbb{E}[T_{n,j}^*(x)] = 0$$

Concerning the term  $\mathbb{E}[e_{n,j}^*(x)]$ , we have, by the same manner, that :

$$\mathbb{E}[e_{n,j}^*(x)] = \frac{1}{h \phi_x(h)} \mathbb{E}[c_{1k} K_1(x) \rho_x(X_1 - x, X_1 - x - r(X_1))].$$

The fact that the operator  $\rho_x$  is continuous, under hypothesis (H3), we can write :

$$\|\rho_x(X_1 - x, X_1 - x - r(X_1))\|_{\mathcal{H}} \leq c \|X_1 - x\|_{\mathcal{F}}^2.$$

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

$$\forall j \leq J, \quad |c_{1j}| \leq \|x - X_1\|_{\mathcal{F}}$$

Combining these two majorations, we get :

$$\mathbb{E}[c_{1k} K_1(x) \rho_x(X_1 - x, X_1 - x - r(X_1))] \leq \mathbb{E}[\|x - X_1\|_{\mathcal{F}}^3 K_1(x)] \leq Ch^3 \phi_x(h).$$

Similarly to the Lemma 3.4.3 before, the first part of condition (H1) permits to conclude this proof.  $\blacksquare$

*Proof of Lemma 3.4.6.* Similarly to the previous Lemmas, we will give only the demonstration for  $T_{nj}^*(x)$  with  $j \neq 0$  (proofs of others cases are identical). Indeed, we use the definition of  $S_{\mathcal{F}}$  and we note  $j(x) = \arg \min_{j \in \{1, 2, \dots, d_n\}} \|x - x_k\|_{\mathcal{F}}$ , and  $\beta_i^j(x) = \langle X_i - x, v_j \rangle$ . The proof is based on the following decomposition :

$$\begin{aligned} \sup_{x \in S_{\mathcal{F}}} \|T_{n,j}^*(x) - \mathbb{E}[T_{n,j}^*(x)]\|_{\mathcal{H}} &\leq \underbrace{\sup_{x \in S_{\mathcal{F}}} \|T_{n,j}^*(x) - T_{n,j}^*(x_{j(x)})\|_{\mathcal{H}}}_{E_1} \\ &+ \underbrace{\sup_{x \in S_{\mathcal{F}}} \|T_{n,j}^*(x_{j(x)}) - \mathbb{E}[T_{n,j}^*(x_{j(x)})]\|_{\mathcal{H}}}_{E_3} \\ &+ \underbrace{\sup_{x \in S_{\mathcal{F}}} \|\mathbb{E}[T_{n,j}^*(x_{j(x)})] - \mathbb{E}[T_{n,j}^*(x)]\|_{\mathcal{H}}}_{E_2}. \end{aligned}$$

By the same arguments as those used by [40] we show that :

$$E_1 = O_{a.co.} \left( \sqrt{\frac{\log d_n}{n\phi(h)}} \right) \text{ and } E_2 = O \left( \sqrt{\frac{\log d_n}{n\phi(h)}} \right). \quad (4)$$

Finally, for the term  $E_3$ , we use the Bernstein inequality for Hilbert spaces which is based on the evaluation of :

$$\sum_{i=1}^n \mathbb{E} \|\Gamma_{i,j}\|_{\mathcal{H}}^m$$

where

$$\Gamma_{i,j} = \frac{K_i(x)\beta_i^j(x)(Y_j - r(X_j))}{nh\phi(h)} - \mathbb{E} \left[ \frac{K_i(x)\beta_i^j(x)(Y_j - r(X_j))}{nh\phi(h)} \right]$$

It is clear that for all  $j \leq m$  we have :

$$C\phi(h) \leq \mathbb{E}[K_1^k(x)] \leq C'\phi(h) \text{ and } |\beta_i^j(x)| \leq \|x - X_i\|_{\mathcal{F}} \|v_j\|_{\mathcal{F}}$$

which imply that :

$$\begin{aligned} \mathbb{E} \|h^{-1}\beta_1^j(x)(Y_1 - r(X_1))K_1(x)\|_{\mathcal{H}}^k &\leq C \mathbb{E} \left[ K_1^k(x) E \left[ \|(Y_1 - r(X_1))\|_{\mathcal{H}}^k | X_1 \right] \right] \\ &\leq C \mathbb{E}[K_1^k(x)]. \end{aligned}$$

Then, by using a Newton expansion :

$$\begin{aligned} \|\Gamma_{i,j}\|_{\mathcal{H}}^m &\leq C \frac{1}{h^m n^m \phi^m(h)} \left( \|\beta_1^j(x)(Y_1 - r(X_1))K_1(x)\|_{\mathcal{H}} + \|\mathbb{E}[\beta_1^j(x)(Y_1 - r(X_1))K_1(x)]\|_{\mathcal{H}} \right)^m \\ &\frac{1}{h^m n^m \phi^m(h)} \sum_{k=0}^m C_{k,m} \|\beta_1^j(x)(Y_1 - r(X_1))K_1(x)\|_{\mathcal{H}}^k \|\mathbb{E}[\beta_1^j(x)(Y_1 - r(X_1))K_1(x)]\|_{\mathcal{H}}^{m-k} \end{aligned}$$

where  $C_{k,m} = \frac{m!}{k!(m-k)!}$ . Finally, for all  $m \geq 1$  :

$$\sum_{i=1}^n \mathbb{E} \|\Gamma_{i,j}\|_{\mathcal{H}}^m = O(m!(n\phi(h))^{-m+1}). \quad (5)$$

Now, by applying the Bernstein inequality with respect to Hilbert spaces (cf. Theorem 2.5 in [4], page 49) we obtain :

$$\mathbb{P} \left( \left\| \sum_{i=1}^n \Gamma_{i,j} \right\|_{\mathcal{H}} > \eta_0 \sqrt{\frac{\log d_n}{n\phi(h)}} \right) \leq 2d_n^{-C\eta^2}.$$

Thus, for all  $j \leq d_n$ , the last equation leads to :

$$\mathbb{P} \left( \|T_{n,j}(x_j) - \mathbb{E}[T_{n,j}(x_j)]\|_{\mathcal{H}} > \eta \sqrt{\frac{\log d_n}{n\phi(h)}} \right) \leq 2 \exp\{-C\eta^2 \log d_n\}.$$

Since  $\log d_n > 1$ , then,

$$d_n \max_{k \in \{1, \dots, d_n\}} \mathbb{P} \left( \|T_{n,j}(x_k) - \mathbb{E}T_{n,j}(x_k)\|_{\mathcal{H}} > \eta \sqrt{\frac{\log d_n}{n\phi(h)}} \right) \leq C' d_n^{1-\beta}.$$

The second part of hypothesis (H5), leads to :

$$E_3 = O_{a.co.} \left( \sqrt{\frac{\log d_n}{n\phi(h)}} \right), \quad (6)$$

which completes the proof of this lemma. ■



# Bibliographie

- [1] Baïllo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response. *J. of Multivariate Analysis*, **100**, Pages 102–111.
- [2] Barrientos-Marin, J., Ferraty, F. and Vieu, P. (2010). Locally modelled regression and functional data. *J. of Nonparametric Statistics*, **22**, No. 5, Pages 617–632.
- [3] Berlinet, A., Elamine, A., and Mas, A. (2011). Local linear regression for functional data. *Ann. Inst. Statist. Math.*, **63**, Pages 1047–1075.
- [4] Bosq, D. (2000). *Linear processes in function spaces : Theory and applications*. Lecture Notes in Statistics, **149**, Springer.
- [5] Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference*, **147**, 1-23.
- [6] Dabo-Niang, S. and Rhomari, N. (2009). Kernel regression estimation in a Banach space. *J. Statist. Plann. Inference*, **139**, Pages 1421–1434
- [7] Demongeot, J., Laksaci, A., Madani, F. and Rachdi, M. (2013). Functional data : local linear estimation of the conditional density and its application. *Statistics*, **47**, Pages 26–44.

- [8] Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*. London, Chapman & Hall.
- [9] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis. Theory and Practice*. Springer Series in Statistics. New York.
- [10] Ferraty, F., Laksaci, A., Tadj, A. and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference*, **140**, 335-352.
- [11] Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2011). Kernel regression with functional response. *Electron. J. Statist.*, **5**, Pages 159–171.
- [12] Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2012). Estimation de la fonction de régression pour variable explicative et réponse fonctionnelles dépendantes. *C. R. Math. Acad. Sci. Paris*, **350**, Pages 717–720.
- [13] Ferraty, F., Van Keilegom, I., and Vieu, P. (2012). Regression when both response and predictor are functions. *J. of Multivariate Analysis*, **109**, Pages 10–28.
- [14] Gijbels, I., Omelka, M. and Veraverbeke, N. (2012). Multivariate and functional covariates and conditional copulas. *Electronic Journal of Statistics*, **6**, Pages 1273–1306.
- [15] Goia, A. and Vieu, P. (2015). An introduction to recent advances in high/infinite dimensional Statistics. *J. of Multivariate Analysis*, in print.
- [16] Hsing, T. and Eubank, R. (2015). Theoretical foundations of functional data analysis, with an introduction to linear operators. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester.
- [17] Kudraszowa, N.L. and Vieu, P. (2013). Uniform consistency of kNN regressors for functional variables. *Statistics & Probability Letters*, **83**, Issue 8, Pages 1863–1870.
- [18] Ramsay, J.O. and Silverman, B.W. (2002). *Applied functional data analysis. Methods and case studies*. Springer Series in Statistics. New York.

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



# Chapitre 4

## Functional local linear regression model with functional response

**Abstract.** In this chapter, we introduce a new nonparametric estimation of the regression function when both the response variable and the explanatory one are functional. Our estimate is based on local local polynomial approach. As asymptotic results, we state the rates of the uniform almost complete convergence of this estimate. The latter expressed as function of the small ball probability of the predictor and as function of the entropy of the set on which the uniformity is obtained.

**Keywords** Functional data, Kernel method, Local linear estimate, Functional response, Almost complete convergence.

### 4.1 Introduction

The functional statistics area has received an increasing interest in recent years. This great consideration is due to the interaction with other applied fields. Typically, this area concerns the modelization of variables taking values in infinite dimensional spaces, which appear in applied sciences such as

economics, soil science, epidemiology or environmental science, among others. Some key references on this topic are the monographs of Ramsay and Silverman [65], Bosq [12], Ferraty and Vieu [36]. In this context, regression model is main tool to explain how a variable of interest  $Y$  is linked with a regressor  $X$ . In this chapter we consider the problem of estimation of the regression function when both variables (response and explanatory) are functional by using local linear method.

It is well known that a local polynomial smoothing has various advantages over the kernel method, namely in the bias term moreover the kernel method, so-called local constant method, is known for being a particular case of the local polynomial method. We return to Fan and [8] and Fan and Gijbels [32], for an extensive discussion on the comparison between both these methods. Concerning the functional data case, the first result in this approach has been stated by Baïllo and Grané [4]. They obtained the  $L^2$ -convergence rate of the local linear estimate of the regression function when the explanatory variable takes values in a Hilbert space. Barrientos-Marin *et al.* [1] propose an interesting simplified version which can be used in semi-metric space. The authors established the almost complete convergence (with rate) of the proposed estimate. Recently, Chouaf and Laksaci [5] studied the spatial version of the estimator proposed by Barrientos-Marin *et al.* [1]. They obtained the almost complete consistency of this estimate when the observations are spatially dependents. All these works are only concerning the functional local modeling when the explanatory variable is functional, our interest in this chapter is to study this approach of local linear modeling when the response variable is also functional. Notice that these questions in infinite dimensional spaces are particularly interesting, at once for the fundamental problems they formulate, but also for many applications they may allow, in particular, they permit to predict an entire curve knowing another one. Noting that, the literature in regression analysis of two functional random variables is still

limited. The first consequent result on this subject has been obtained by Niang and Rhomari [18]. Under some differentiation condition they stated the convergence in  $L^p$  norm of a kernel estimator of the regression operator with the response and the explanatory variable are functionals. Recently, Ferraty and Laksaci *et al.* [40] obtained a rate of the uniform almost complete convergence of the kernel estimator of this nonparametric model in the i.i.d. case. This results has been extended to dependent case by the same authors in Ferraty and Laksaci *et al.* [42]. While the the asymptotic normality of the same estimator has been treated by Ferraty et al. [43]. The principal aim of the present work is to study this nonparametric model by employing an alternative estimator to that considered by Ferraty and Laksaci *et al.* [42]. More precisely, we construct a new estimator of the functional nonparametric regression by the local linear smoothing method and we show its almost completely convergence rate of the local linear estimate of this models uniformly over some set of the functional space.

The chapter is organized as follows. We present our model in Section 2. In Section 3 we introduce notations, Assumptions and state the main results. he proofs of the results are relegated to the last section.

## 4.2 Model

Let us introduce  $n$  pairs of random variables  $(X_i, Y_i)$  for  $i = 1, \dots, n$  that we assume drawn from the pair  $(X, Y)$  which is valued in  $\mathcal{F} \times \mathcal{H}$ , where  $\mathcal{F}$  (resp.  $\mathcal{H}$ ) is a semi-metric space equipped with a semi-metric  $d$  (resp. is a Hilbert space equipped with an inner product  $\langle \cdot, \cdot \rangle$ ). Furthermore, we assume that the lien between  $X$  and  $Y$  is modeled by the following relation

$$Y = R(X) + \epsilon, \tag{1}$$

where  $R$  is an operator from  $\mathcal{F}$  to  $\mathcal{H}$  and  $\epsilon$  is a random error variable.

The local polynomial smoothing is based on the assumption that is the functional operator is smooth enough to be locally well approximated by a polynomial. In functional statistics, there are several ways for extending this approach (see, [4] or [1] for some examples). Here we adopt the fast version proposed by [1] for which the regression function  $R(x)$  is estimated by  $\hat{a}$  solution of

$$\min_{(a,b) \in \mathbb{R}^2} \sum_{i=1}^n \|Y_i - a - b\beta(X_i, x)\|_{\mathcal{H}}^2 K(h^{-1}\delta(x, X_i)) \quad (2)$$

where  $\|\cdot\|_{\mathcal{H}}$  is the norm in the Hilbert space  $\mathcal{H}$ ,  $\beta(\cdot, \cdot)$  is a known function from  $\mathcal{F}^2$  into  $\mathbb{R}$  such that,  $\forall \xi \in \mathcal{F}$ ,  $\beta(\xi, \xi) = 0$ , with  $K$  is kernel and  $h = h_{K,n}$  is a sequence of positive real numbers and  $\delta(\cdot, \cdot)$  is a function of  $\mathcal{F} \times \mathcal{F}$  such that  $d(\cdot, \cdot) = |\delta(\cdot, \cdot)|$ . Clearly, by a simple algebra, we prove that  $(\hat{a}, \hat{b})$  solution of (2) are zero of

$$(Q'_B K Q_B) \begin{pmatrix} a \\ b \end{pmatrix} - (Q'_B K Y) = 0 \Rightarrow \begin{pmatrix} a \\ b \end{pmatrix} = (Q'_B K Y)(Q'_B K Q_B)^{-1}.$$

where  $Q'_B = \begin{pmatrix} 1 & \dots & 1 \\ B(X_1, x) & \dots & B(X_n, x) \end{pmatrix}$   
 $K = \text{diag}(K(h^{-1}\delta(x, X_1)), \dots, K(h^{-1}\delta(x, X_n)))$  and  $Y' = (Y_1, \dots, Y_n)$ .  
 Thus, we get explicitly

$$\hat{a} = (Q'_B K Y)(Q'_B K Q_B)^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and

$$\hat{b} = (Q'_B K Y)(Q'_B K Q_B)^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$



Moreover, as  $\beta(x, x) = 0$ , we can write

$$\widehat{R}(x) = \frac{\sum_{i,j=1}^n W_{ij}(x) Y_j}{\sum_{i,j=1}^n W_{ij}(x)} \quad (3)$$

where

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

with the convention  $0/0 = 0$ .

**Remark 4.2.1.** *Obviously, if  $b = 0$  then we obtain from (2) the Nadaraya-Watson estimator studied, in the functional case, in [36] and the references therein.*

### 4.3 The main result

Our main result is to establish the uniform almost complete convergence of  $\widehat{R}(x)$  on some subset  $S_{\mathcal{F}}$  of  $\mathcal{F}$ , such that :  $S_{\mathcal{F}} \subset \bigcup_{k=1}^{d_n} B(x_k, r_n)$ , where  $x_k \in \mathcal{F}$  and  $r_n$  (resp.  $d_n$ ) is a sequence of positive real numbers.

Recall that, the uniform consistency has great importance in practice. It allows us to make prediction, even if the data are not perfectly observed. On the other hand, the uniform convergence results are vital tools for some nonparametric functional data problems such as for instance data-driven bandwidth choice (cf. Benhenni and Ferraty et al. [2]). It is worth noting that, unlike to the multivariate case, the uniform consistency for functional data requires some additional tools and topological conditions. Indeed, To state our main result, we consider the following conditions.

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

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

$$\text{and } \exists \eta_0 > 0, \forall \eta < \eta_0, \phi'(\eta) < C,$$

where  $C$  and  $C'$  are strictly positive constants and where  $\phi'$  denotes the first derivative of  $\phi$ .

(U2) There exists  $C > 0$  and  $C' > 0$  such that for all real  $m \geq 1$ ,  $E(\|Y\|^m | X = x) < C$  and

$\forall (x_1, x_2) \in S_{\mathcal{F}} \times S_{\mathcal{F}}$  :

$$\|R(x_1) - R(x_2)\|_{\mathcal{H}} \leq C' d^{b_1}(x_1, x_2); \quad b_1 > 0$$

(U3) The function  $\beta(., .)$  is such that :

$$\forall x' \in \mathcal{F}, C_1 d(x, x') \leq |\beta(x, x')| \leq C_2 d(x, x'), \text{ where } C_1 > 0, C_2 > 0.$$

and

$$\forall (x_1, x_2) \in S_{\mathcal{F}} \times S_{\mathcal{F}}, \|\beta(x_1, x') - \beta(x_2, x')\| \leq C' d(x_1, x_2).$$

(U4) The kernel  $K$  is a positive, differentiable function with support  $[-1, 1]$  satisfies the following Lipschitz's condition :

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

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

$$\frac{(\log n)^2}{n \phi(h)} < \log d_n < \frac{n \phi(h)}{\log n}.$$

and

$$\exists \beta > 1, \quad \sum_{n=1}^{\infty} d_n^{(1-\beta)} < \infty.$$

(U6) The bandwidth  $h$  satisfies : there exists an integer  $n_0$ , such that

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

and

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

where  $B(x, r) = \{x' \in \mathcal{F} / d(x', x) \leq r\}$  denotes the closed-ball of center  $x$  and radius  $r$ , and  $dP(x)$  is the cumulative distribution of  $X$ .

**Theorem 4.3.1.** *Under assumptions (U1)-(U6), we have that :*

$$\sup_{x \in S_{\mathcal{F}}} \|\widehat{R}(x) - R(x)\|_{\mathcal{H}} = O(h^{b_1}) + O_{a.co.} \left( \sqrt{\frac{\log d_n}{n\phi(h)}} \right). \quad (4)$$

It Clear that, the Theorem 4.3.1's proof can be deduced directly from the following decomposition

$$\begin{aligned} \forall x \in S_{\mathcal{F}}, \widehat{R}(x) - R(x) &= \frac{1}{\widehat{R}_D(x)} \left\{ \left( \widehat{R}_N(x) - \mathbb{E}[\widehat{R}_N(x)] \right) - \left( R(x) - \mathbb{E}[\widehat{R}_N(x)] \right) \right\} \\ &+ \frac{R(x)}{\widehat{R}_D(x)} \left( 1 - \widehat{R}_D(x) \right), \end{aligned} \quad (5)$$

where

$$\widehat{R}_N(x) = \frac{1}{n(n-1)\mathbb{E}[W_{12}(x)]} \sum_{i \neq j} W_{ij}(x) Y_j$$

and

$$\widehat{R}_D(x) = \frac{1}{n(n-1)\mathbb{E}[W_{12}(x)]} \sum_{i \neq j} W_{ij}(x)$$

and of Lemmas 4.3.2, 4.3.4 and 4.3.5 below, for which the proofs are given in the Appendix.

**Lemma 4.3.2.** *Under assumptions (U1), (U3)-(H6), we obtain that :*

$$\sup_{x \in S_{\mathcal{F}}} \|\widehat{R}_D(x) - 1\| = O_{a.co.} \left( \sqrt{\frac{\log d_n}{n\phi(h)}} \right).$$

**Corollary 4.3.3.** *Under the assumptions of Lemma 4.3.2, we have that :*

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

**Lemma 4.3.4.** *Under the hypotheses (U1) and (U2), we obtain that :*

$$\sup_{x \in \mathcal{S}_{\mathcal{F}}} \left\| R(x) - \mathbb{E} \left[ \widehat{R}_N(x) \right] \right\|_{\mathcal{H}} = O(h^{b_1}).$$

**Lemma 4.3.5.** *Under the assumptions of Theorem 4.3.1, we obtain that :*

$$\sup_{x \in \mathcal{S}_{\mathcal{F}}} \left\| \widehat{R}_N(x) - \mathbb{E}[\widehat{R}_N(x)] \right\|_{\mathcal{H}} = O_{a.co.} \left( \sqrt{\frac{\log d_n}{n\phi(h)}} \right).$$

## 4.4 Appendix

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

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

**Proof of Lemma 4.3.2.** The proof of this lemma is based on the same decomposition used in Lemma 1 of [1]. Indeed,

$$\begin{aligned} \widehat{R}_D(x) = T_1 & \left[ \underbrace{\left( \frac{1}{n} \sum_{j=1}^n \frac{K_j(x)}{\phi_x(h)} \right)}_{S_2(x)} \underbrace{\left( \frac{1}{n} \sum_{i=1}^n \frac{K_i(x)\beta_i^2(x)}{h^2\phi_x(h)} \right)}_{S_4(x)} \right. \\ & \left. - \underbrace{\left( \frac{1}{n} \sum_{j=1}^n \frac{K_j(x)\beta_j(x)}{h\phi_x(h)} \right)}_{S_3(x)} \underbrace{\left( \frac{1}{n} \sum_{i=1}^n \frac{K_i(x)\beta_i(x)}{h\phi_x(h)} \right)}_{S_3(x)} \right] \end{aligned}$$

Thus, all it remains to show are the following uniform results :

$$\sup_{x \in \mathcal{S}_{\mathcal{F}}} |S_k(x) - \mathbb{E}[S_k(x)]| = O \left( \sqrt{\frac{\log d_n}{n\phi_x(h)}} \right), \quad a.co. \text{ for } k = 2, 3, 4, \quad (6)$$

and

$$\sup_{x \in S_{\mathcal{F}}} |\mathbb{E}[S_2(x)]\mathbb{E}[S_4(x)] - \mathbb{E}[S_2(x)S_4(x)] - \text{var}[S_3(x)]| = o\left(\sqrt{\frac{\log d_n}{n \phi_x(h)}}\right), \text{ a.co.}$$

and, also that, uniformly on  $x \in S_{\mathcal{F}}$  :

$$T_1 = O(1) \quad \text{and} \quad |\mathbb{E}[S_k(x)]| = O(1), \text{ for } k = 2, 3, 4.$$

Concerning the last two equations, we combine assumption (U1) to e Lemma 3 of [1] to show the claimed results. Furthermore, the proof of (6) follows the same ideas as in [38]. Indeed, by noting :  $j(x) = \arg \min_{j \in \{1, 2, \dots, d_n\}} |\delta(x, x_j)|$ , we consider the following decomposition :

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

Firstly, let us analyze the first term  $F_1$ . Since  $K$  is supported in  $[-1, 1]$ , we can write for all  $k = 2, 3, 4$  that :

$$\begin{aligned} F_1 &\leq \frac{C(k-2)}{nh^{k-2}\phi_x(h)} \sup_{x \in S_{\mathcal{F}}} \sum_{i=1}^n K_i(x) \mathbb{1}_{B(x,h)}(X_i) \\ &\quad \times \left| \beta_i^{k-2}(x) - \beta_i^{k-2}(x_{j(x)}) \mathbb{1}_{B(x_{j(x)},h)}(X_i) \right| \\ &\quad + \frac{1}{nh^{k-2}\phi_x(h)} \sup_{x \in S_{\mathcal{F}}} \sum_{i=1}^n \beta_i^{k-2}(x_{j(x)}) \mathbb{1}_{B(x_{j(x)},h)}(X_i) \\ &\quad \times \left| K_i(x) \mathbb{1}_{B(x,h)}(X_i) - K_i(x_{j(x)}) \right|. \end{aligned}$$

The Lipschitz condition on  $K$  gives

$$\begin{aligned} &\mathbb{1}_{B(x_{j(x)},h)}(X_i) \left| K_i(x) \mathbb{1}_{B(x,h)}(X_i) - K_i(x_{j(x)}) \right| \\ &\leq C\epsilon \mathbb{1}_{B(x,h) \cap B(x_{j(x)},h)}(X_i) + C \mathbb{1}_{B(x_{j(x)},h) \cap \overline{B(x,h)}}(X_i). \end{aligned}$$

While the Lipschitz condition on  $\beta$  allows to write that

$$\begin{aligned} & \mathbb{1}_{B(x,h)}(X_i) \left| \beta_i(x) - \beta_i(x_{j(x)}) \mathbb{1}_{B(x_{j(x),h})}(X_i) \right| \\ & \leq \epsilon \mathbb{1}_{B(x,h) \cap B(x_{j(x),h})}(X_i) + h \mathbb{1}_{B(x,h) \cap \overline{B(x_{j(x),h)}}}(X_i) \\ & \mathbb{1}_{B(x,h)}(X_i) \left| \beta_i^2(x) - \beta_i^2(x_{j(x)}) \mathbb{1}_{B(x_{j(x),h})}(X_i) \right| \\ & \leq \epsilon h \mathbb{1}_{B(x,h) \cap B(x_{j(x),h})}(X_i) + h^2 \mathbb{1}_{B(x,h) \cap \overline{B(x_{j(x),h)}}}(X_i) \end{aligned}$$

which implies that, for  $k = 3, 4$

$$\begin{aligned} & \mathbb{1}_{B(x,h)}(X_i) \left| \beta_i^{k-2}(x) - \beta_i^{k-2}(x_{j(x)}) \mathbb{1}_{B(x_{j(x),h})}(X_i) \right| \\ & \leq \epsilon h^{k-3} \mathbb{1}_{B(x,h) \cap B(x_{j(x),h})}(X_i) + h^{k-2} \mathbb{1}_{B(x,h) \cap \overline{B(x_{j(x),h)}}}(X_i). \end{aligned}$$

Therefore

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

where

$$\begin{aligned} F_{11}^k &= \frac{C(k-2)}{n\phi(h)} \sum_{i=1}^n \mathbb{1}_{B(x,h) \cap \overline{B(x_{j(x),h})}}(X_i), \\ F_{12} &= \frac{C\epsilon}{n\phi(h)} \sum_{i=1}^n \mathbb{1}_{B(x,h) \cap B(x_{j(x),h})}(X_i). \\ F_{13}^k &= \frac{C(k-2)\epsilon}{nh\phi(h)} \sum_{i=1}^n \mathbb{1}_{B(x,h) \cap B(x_{j(x),h})}(X_i). \\ F_{14} &= \frac{C}{n\phi(h)} \sum_{i=1}^n \mathbb{1}_{B(x_{j(x),h}) \cap \overline{B(x,h)}}(X_i). \end{aligned}$$

Thus, it suffices to evaluate the terms  $F_{11}^k$ ,  $F_{12}$ ,  $F_{13}^k$  and  $F_{14}$ . To do that we apply a standard inequality for sums of bounded random variables (cf.

Corollary A.9 in [36]) with  $Z_i$  is identified such that :

$$Z_i = \begin{cases} \frac{1}{\phi(h)} \sup_{x \in S_{\mathcal{F}}} \left[ \mathbb{I}_{B(x,h) \cap \overline{B(x_j(x),h)}}(X_i) \right] & \text{for } F_{11}^k \\ \frac{\epsilon}{h\phi(h)} \sup_{x \in S_{\mathcal{F}}} \left[ \mathbb{I}_{B(x,h) \cap B(x_j(x),h)}(X_i) \right] & \text{for } F_{12} \text{ and } F_{13}^k \\ \frac{1}{\phi(h)} \sup_{x \in S_{\mathcal{F}}} \left[ \mathbb{I}_{B(x_j(x),h) \cap \overline{B(x,h)}}(X_i) \right] & \text{for } F_{14} \end{cases}$$

Clearly, under the second part of (U1), we have for the first and the last case :

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

So that, we get :

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

In the same way, assumption (U5) allows to get, for  $F_{12}$  or  $F_{13}^k$  case

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

which implies that :

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

To achieve the study of the term  $F_1$ , it suffices to put together all the intermediate results and to use (U5) to obtain :

$$F_1 = O_{a.co.} \left( \sqrt{\frac{\log d_n}{n \phi(h)}} \right). \quad (7)$$

Concerning  $F_3$ , we use the fact that

$$F_3 \leq \mathbb{E} \left[ \sup_{x \in \mathcal{S}_{\mathcal{F}}} |S_k(x) - S_k(x_{j(x)})| \right]$$

to get

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

Now, we treat the term  $F_2$ . Indeed, it is clear that, for all  $\eta > 0$ , we have that :

$$\begin{aligned} \mathbb{P} \left( F_2^k > \eta \sqrt{\frac{\log d_n}{n\phi(h)}} \right) \\ \leq d_n \max_{j \in \{1, \dots, d_n\}} \mathbb{P} \left( |S_k(x_j) - \mathbb{E}[S_k(x_j)]| > \eta \sqrt{\frac{\log d_n}{n\phi(h)}} \right). \end{aligned}$$

We apply a Bernstein-type inequality (cf. Corollary A.8 in [36]) on

$$\Delta_{ki} = \frac{1}{nh^{k-2}\phi(h)} (K_i(x_k)\beta_i^{k-2}(x_k) - \mathbb{E}[K_i(x_k)\beta_i^{k-2}(x_k)]), \text{ for } k = 2, 3, 4,$$

for which

$$\mathbb{E}|\Delta_{ki}|^m = O(\phi(h)^{-m+1}), \text{ for } k = 2, 3, 4.$$

we get

$$\begin{aligned} \mathbb{P} \left( |S_i(x_k) - \mathbb{E}[S_i(x_k)]| > \eta \sqrt{\frac{\log d_n}{n\phi(h)}} \right) &= \mathbb{P} \left( \frac{1}{n} \left| \sum_{i=1}^n \Delta_{lki} \right| > \eta \sqrt{\frac{\log d_n}{n\phi(h)}} \right) \\ &\leq 2 \exp\{-C\eta^2 \log d_n\}. \end{aligned}$$

Thus, by choosing  $\eta$  such that  $C\eta^2 = \beta$ , we get :

$$d_n \max_{k \in \{1, \dots, d_n\}} \mathbb{P} \left( |S_i(x_k) - \mathbb{E}[S_i(x_k)]| > \eta \sqrt{\frac{\log d_n}{n\phi(h)}} \right) \leq C' d_n^{1-\beta}. \quad (8)$$



Since  $\sum_{n=1}^{\infty} d_n^{1-\beta} < \infty$ , we obtain that :

$$F_2 = O_{a.co.} \left( \sqrt{\frac{\log d_n}{n\phi(h)}} \right).$$

■

**Proof of Corollary 4.3.3.** It is clear that

$$\inf_{x \in S_{\mathcal{F}}} \widehat{R}_D(x) \leq \frac{1}{2} \Rightarrow \exists x \in S_{\mathcal{F}}, \text{ such that } 1 - \widehat{R}_D(x) \geq \frac{1}{2} \Rightarrow \sup_{x \in S_{\mathcal{F}}} |1 - \widehat{R}_D(x)| \geq \frac{1}{2}.$$

By Lemma 4.3.2, we obtain :

$$\mathbb{P} \left( \inf_{x \in S_{\mathcal{F}}} \widehat{R}_D(x) \leq \frac{1}{2} \right) \leq \mathbb{P} \left( \sup_{x \in S_{\mathcal{F}}} |1 - \widehat{R}_D(x)| > \frac{1}{2} \right).$$

Consequently :

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

■

**Proof of lemma 4.3.4.** By using the fact that the pairs  $(X_i, Y_i)$  are identically distributed, we have

$$\mathbb{E}[\widehat{R}_N(x)] = \mathbb{E} [W_{12}(x) [\mathbb{E} [-^1Y_2/X_2]]].$$

Therefore, by the assumption (U2) we get that :

$$\mathbb{1}_{B(x,h)}(X) \|R(X) - R(x)\|_{\mathcal{H}} \leq Ch^{b_1}$$

the fact that the Lipschitz's condition is uniformly on  $x$  in  $S_{\mathcal{F}}$ . completes the proof of this Lemma.

■

**Proof of Lemma 4.3.5.** The proof of this lemma follows the same steps as in Lemma 4.3.2, where  $S_i(x)$ ;  $i = 2, 3, 4$  are replaced by :

$$\begin{cases} T_2(x) &= \frac{1}{n} \sum_{j=1}^n \frac{K_j(x)Y_j}{\phi_x(h)} \\ T_3(x) &= \frac{1}{n} \sum_{j=1}^n \frac{K_j(x)\beta_j(x)Y_j}{h\phi_x(h)} \\ T_4(x) &= \frac{1}{n} \sum_{j=1}^n \frac{K_j(x)\beta_j^2(x)Y_j}{h^2\phi_x(h)} \end{cases}$$

To do that, we keep the notations used previously, namely, the definitions of  $j(x)$  and  $d_n$ . The proof is based on the the following decomposition which will be used for the three terms :

$$\begin{aligned} & \sup_{x \in S_{\mathcal{F}}} \|T_i(x) - \mathbb{E}[T_i(x)]\|_{\mathcal{H}} \\ & \leq \underbrace{\sup_{x \in S_{\mathcal{F}}} \|T_i(x) - T_i(x_{j(x)})\|_{\mathcal{H}}}_{E_1} + \underbrace{\sup_{x \in S_{\mathcal{F}}} \|T_i(x_{j(x)}) - \mathbb{E}[T_i(x_{j(x)})]\|_{\mathcal{H}}}_{E_3} \\ & \quad + \underbrace{\sup_{x \in S_{\mathcal{F}}} \|\mathbb{E}[T_i(x_{j(x)})] - \mathbb{E}[T_i(x)]\|_{\mathcal{H}}}_{E_2}. \end{aligned}$$

Similarly to the study of the term  $F_1$ , we obtain :

$$E_1 = O_{a.co.} \left( \sqrt{\frac{\log d_n}{n\phi(h)}} \right) \text{ and } E_2 = O \left( \sqrt{\frac{\log d_n}{n\phi(h)}} \right). \quad (9)$$

Finally, for the term  $E_3$ , we use the the classical Bernstein's inequality of functional space given in [12] which based on evaluation of  $\frac{1}{n^m \phi^m(h)} \sum_{i=1}^n E \|\Gamma_{ki}\|_{\mathcal{H}}^m$

$$\text{where } \Gamma_{ki} = \frac{K_i(x)\beta_i^k(x)Y_j}{nh^k\phi(h)}.$$

Note that for all  $j \leq m$  we have :  $C\phi(h) \leq \mathbb{E}[K_1^j(x)] \leq C'\phi(h)$ , which implies

that

$$\begin{aligned} \mathbb{E} \|h^{-k} \beta_1^k Y_1 K_1(x)\|_{\mathcal{H}}^j &\leq C \mathbb{E} \left[ K_1^j(x) E \left[ \|Y_1\|_{\mathcal{H}}^j | X_1 \right] \right] \\ &\leq C \mathbb{E} [K_1^j(x)]. \\ \mathbb{E} \|\Gamma_{ki}\|_{\mathcal{H}}^j &= O(j! n^j \phi(h)^{-j+1}). \end{aligned} \quad (10)$$

Now we use the Newton's binomial expansion and we get :

$$\begin{aligned} \|\Gamma_{ki}\|_{\mathcal{H}}^m &\leq C \frac{1}{n^m \phi^m(h)} (\|K_i(x) Y_i\|_{\mathcal{H}} + \|\mathbb{E}[K_i(x) Y_i]\|_{\mathcal{H}})^m \\ &= \frac{1}{n^m \phi^m(h)} \sum_{k=0}^m C_{k,m} \|Y_1 K_1(x)\|_{\mathcal{H}}^k \|\mathbb{E}[Y_1 K_1(x)]\|_{\mathcal{H}}^{m-k} \end{aligned}$$

where  $C_{k,m} = \frac{m!}{k!(m-k)!}$ . Finally, for all,  $m \geq 1$ , we have

$$\sum_{i=1}^n \mathbb{E} \|\Gamma_{ki}\|_{\mathcal{H}}^m = O(m! (n\phi(h))^{-m+1}). \quad (11)$$

We are now in position for applying the Hilbertian version of Bernstein's inequality (see, Theorem 2.5 in [12] P.49) and we get

$$\mathbb{P} \left( \left\| \sum_i^n \Gamma_{ki} \right\|_{\mathcal{H}} > \eta_0 \sqrt{\frac{\log d_n}{n\phi(h)}} \right) \leq 2d_n^{-C\eta^2}.$$

Therefore, the last equation allows to get

$$\forall j \leq d_n, \mathbb{P} \left( \|T_i(x_j) - \mathbb{E}[T_i(x_j)]\|_{\mathcal{H}} > \eta \sqrt{\frac{\log d_n}{n\phi(h)}} \right) \leq 2 \exp\{-C\eta^2 \log d_n\}.$$

Since,  $\frac{\log d_n}{n\phi(h)} > \frac{1}{n\phi(h)}$ , we can write,

$$d_n \max_{k \in \{1, \dots, d_n\}} \mathbb{P} \left( \|T_i(x_k) - \mathbb{E}T_i(x_k)\|_{\mathcal{H}} > \eta \sqrt{\frac{\log d_n}{n\phi(h)}} \right) \leq C' d_n^{1-\beta}.$$

By using the second part of condition (U5), one obtains :

$$E_3 = O_{a.co.} \left( \sqrt{\frac{\log d_n}{n\phi(h)}} \right). \quad (12)$$

The latter completes the proof of this Lemma . ■



# Bibliographie

- [1] Barrientos-Marin, J., Ferraty, F. and Vieu, P. (2010). Locally Modelled Regression and Functional Data. *J. of Nonparametric Statistics*, **22**, No. 5, Pages 617–632.
- [2] Benhenni, K., Ferraty, F., Rachdi, M. and Vieu, P. (2007). Local smoothing regression with functional data. *Computational Statistics*, **22**, No. 3, Pages 353–369.
- [3] Bosq, D. (2000). *Linear Processes in Function Spaces : Theory and applications*. Lecture Notes in Statistics, **149**, Springer.
- [4] Baïllo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response, *Journal of Multivariate Analysis*, **100**, Pages 102–111.
- [5] Chouaf, A. and Laksaci, A., (2013), On the functional local linear estimate for spatial regression. *Stat. Risk Model.* **29**, Pages 189–214.
- [6] Dabo-Niang, S. and Rhomari, N. (2009), Kernel regression estimation in a Banach space. *J. Statist. Plann. Inference* **139**, Pages 1421–1434
- [7] Demongeot, J., Laksaci, A., Madani, F. and Rachdi, M. (2013). Functional data : Local linear estimation of the conditional density and its application *Statistics*, **47** , Pages 26–44.

- [8] Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, **87**, Pages 998–1004.
- [9] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. London, Chapman & Hall.
- [10] Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *Journal of statistical planning and inference*, **140**, Pages 335–352.
- [11] Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2011). Kernel regression with functional response. *Electron. J. Stat.* **5** Pages 159–171.
- [12] Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2012). Estimation de la fonction de régression pour variable explicative et réponse fonctionnelles dépendantes. *it C. R. Math. Acad. Sci. Paris* **350** Pages 717–720.
- [13] Ferraty, F., Van Keilegom, I. and Vieu, P. Regression when both response and predictor are functions. *J. Multivariate Anal.* , **109**, Pages 10–28.
- [14] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis. Theory and Practice*. Springer Series in Statistics. New York.
- [15] Ramsay, J. O. and Silverman, B. W. (2002). *Applied functional data analysis. Methods and case studies*. Springer Series in Statistics. New York.

# Chapitre 5

## Application sur des données réelles

### 5.1 introduction

Dans ce chapitre, nous nous intéressons à l'application de notre approche sur des signaux biomédicaux. Notre objectif principal est de montrer que la régression non paramétrique bi-fonctionnelle peut être utilisée comme modèle de débruitage de ces signaux biomédicaux. Notons que, en général les procédés d'acquisition des signaux induisent des perturbations qui peuvent être gênantes pour la compréhension et le traitement de ces données. Ainsi l'élimination de bruits dans ces signaux est une étape préliminaire indispensable pour une bonne lecture des données. On trouvera dans la littérature sur le traitement du signal, plusieurs types de filtrage (cf. Hamie, 2015). Nous présentons, ici une approche alternative basée sur la technique non paramétrique qui est la plus adaptée au traitement des données fonctionnelles. Ce chapitre est présenté en trois sections, dans la première nous donnons le modèle et les données biomédicales étudiées. Les résultats de débruitage par l'estimateur à noyau et l'estimateur local linéaire sont commentés, respectivement, dans les sections 3 et 4.

## 5.2 Modèle et Données

Dans cette étude applicative, nous considérons des données d'une analyse protéomique de 112 sujets dont 64 atteints d'un cancer du côlon et 48 sujets sains (échantillon de contrôles). Rappelons que l'analyse protéomique est un outil médical qui permet de compter les protéines dans une cellule biologique en utilisant la spectrométrie de masse. Les données utilisées dans cette thèse sont disponibles sur le site web de la Ligue Contre le Cancer<sup>1</sup>.

Nous disposons des données de 5 spectres de longueur d'onde 15155 unité . Afin d'accélérer l'exécution de notre algorithme, nous avons considéré seulement des ondes de longueur 8192 unité. Nous représentons dans la figure 5.1 ces signaux.

Dans la plus part des cas, le débruitage est basé sur la décomposition du signal bruité en un signal pur et un bruit, la discrimination entre les deux s'effectue à l'aide d'une fonction de seuillage. Le choix de cette fonction est arbitraire. A notre connaissance, il n'y a pas de règle permettant un choix optimal, de cette fonction, avec une justification mathématique. Ainsi, la modélisation non paramétrique de ce problème permet de survoler la question du choix de cette fonction. En effet, il suffit d'exprimer la relation entre le signal pur et le signal bruité par le modèle de régression suivant :

$$\begin{array}{ccc}
 \text{Courbe pure} & , & \text{Courbe bruitée} \\
 \downarrow & & \downarrow \\
 Y & = & r(x) + \epsilon.
 \end{array}$$

Nous rappelons que le débruitage non paramétrique a fait l'objet de plusieurs travaux en statistique vectorielle. Citons par exemple les travaux de El Hassouni (2000) et Krounez et al. (2007). Nous renvoyons aussi à Boulanger et al.

---

1. <http://home.ocr.cancer.gov/ncifdaproteomics/ppatterns.asp>.



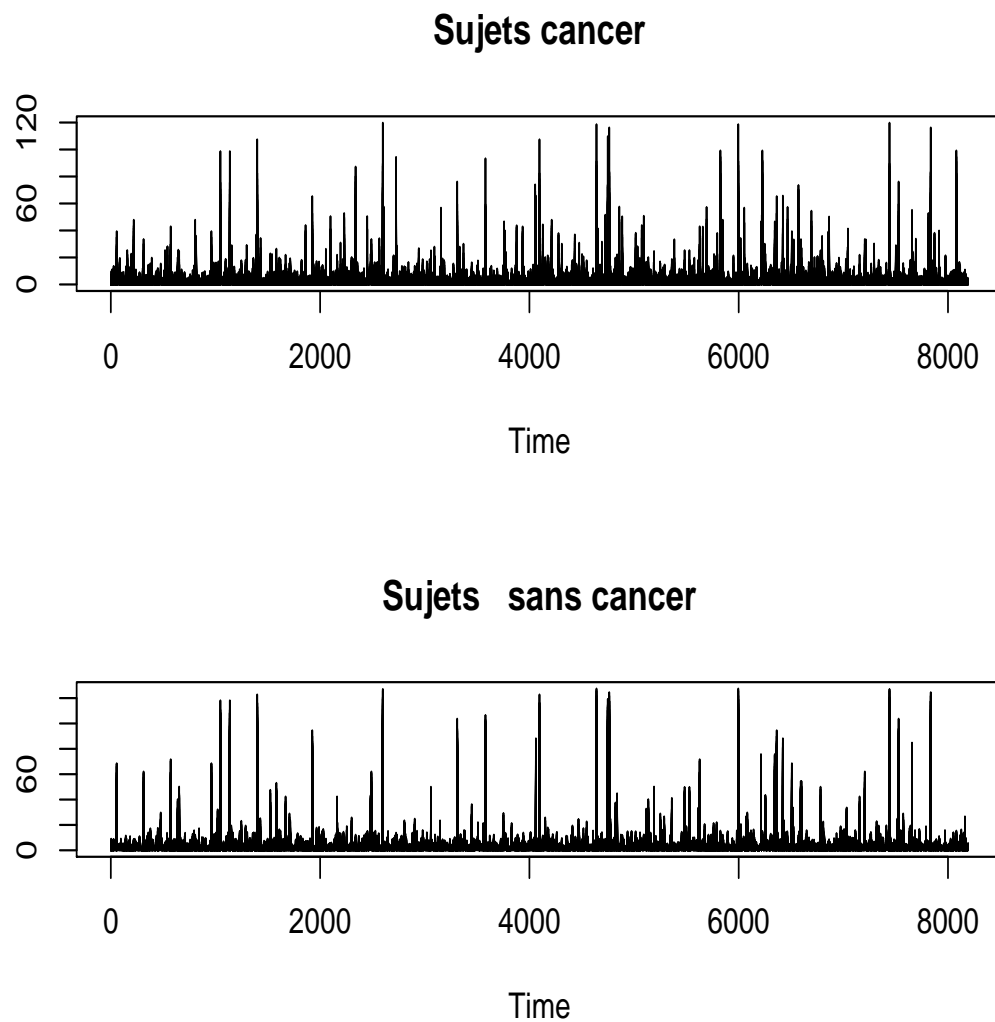


FIGURE 5.1 – Sujets cancer/ Sujets sans cancer

(2008) pour la régression non paramétrique des processus à temps continu. Dans cette partie, nous traitons le signal dans sa propre dimension qui est l'espace fonctionnel  $L_2$ . Nous considérons les deux techniques d'estimation (i) la régression classique et (ii) la régression locale linéaire.

### 5.3 Débruitage par régression locale constante

La régression classique est un cas particulier de notre approche. Elle correspond au cas particulier où  $b = 0$ . Dans le cas où les deux variables (covariable et réponse) sont fonctionnelles, ce modèle a été traité par Ferraty et al. (2013), dans lequel ils ont considéré l'estimateur suivant :

$$\hat{r}(x) = \frac{\sum_{i=1}^n K(h^{-1}d(x, X_i))Y_i}{\sum_{i=1}^n K(h^{-1}(d(x, X_i)))}.$$

L'implémentation de cet estimateur, en pratique, nécessite la détermination de certains paramètres tels que le paramètre de lissage, la métrique et le noyau  $K$ . En ce qui concerne le choix de la largeur de fenêtre  $h$ , nous avons utilisé la méthode de validation croisée sur le nombre des voisins les plus proches associés au critère suivant :

$$\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{r}^{-i}(X_i))^2$$

où

$$\hat{r}^{-i}(x) = \frac{\sum_{j \neq i}^n K(h^{-1}d(x, X_j))Y_j}{\sum_{j \neq i}^n K(h^{-1}(d(x, X_j)))}.$$

En ce qui concerne la semi-métrique, nous avons utilisé la semi métrique associée à la base de Fourier. Cette dernière est très adaptée pour l'approximation des signaux. Nous considérons aussi, un noyau quadratique et nous utilisons le code `ffunopare.knn.gcv` disponible sur le site web <https://www.math.univ-toulouse.fr>. Il est à noter que nos variables réponses  $Y_i$  ne sont pas les vraies courbes pures, mais sont aussi débruitées par le logiciel R. Le résultat de

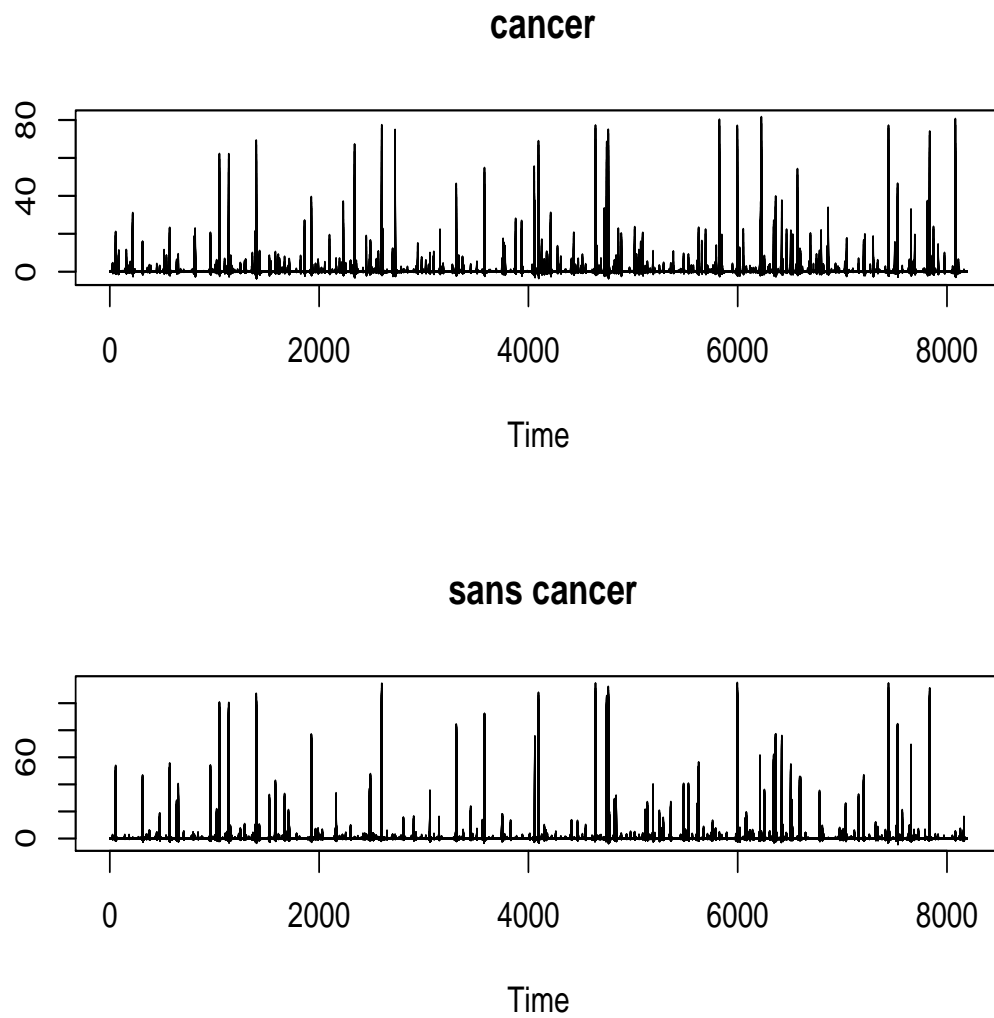


FIGURE 5.2 – cancer/ sans cancer

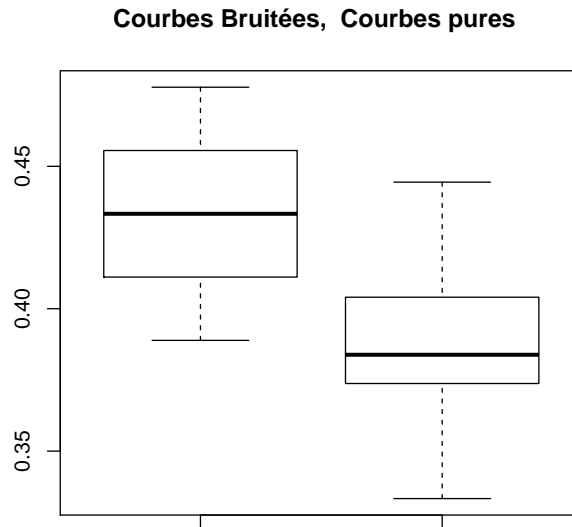


FIGURE 5.3 – comparaison de courbes

débruitage est donné par la figure 5.2. Afin de quantifier l'efficacité de notre estimation, nous allons comparer la classification de deux courbes : la courbe débruitée contre la courbe pure. A cet effet, nous utilisons le code : *funopadi.knn.lcv*, dont nous décomposons notre échantillon en deux parties, (i) échantillon de test et (ii) échantillon d'apprentissage. L'efficacité des deux approches est exprimée par le pourcentage de sujets mal classés parmi l'échantillon de test. Les résultats obtenus sont représentés par les boîtes à moustaches (cf. figure 5.3)

Il est clair que le gain dans la classification est très important sur les courbes pures contre les courbes débruitées. En effet, les sujets mal classés sur les courbes pures sont compris entre 0.33 et 0.43. Par contre dans le cas de courbes bruitées, ils varient entre 0.38 et 0.48.

## 5.4 Débruitage par régression locale linéaire

Maintenant, nous allons débruiter les signaux médicaux par la régression locale linéaire définie par :

$$\widetilde{r}(x) = (Q'_B K Q_B)^{-1} (Y' K Y) (1, 0, \dots, 0)$$

où

$$Q_B = \begin{pmatrix} 1 & c_{11} & \dots & c_{1J} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & c_{n1} & \dots & c_{nJ} \end{pmatrix},$$

avec  $K = \text{diag}(K(h^{-1}\|x - X_1\|_{\mathcal{F}}), \dots, K(h^{-1}\|x - X_n\|_{\mathcal{F}}))$ ,  $Y' = (Y_1, \dots, Y_n)$  et  $\widetilde{A}(x)' = (\widehat{a}(x), \widehat{b}_1, \dots, \widehat{b}_J)$ .

Bien entendu, pour une meilleure comparaison avec le modèle de la section précédente, nous allons traiter les deux modèles avec les mêmes conditions. Autrement dit, nous allons considérer la même méthode de choix du paramètre de lissage et nous prenons la même métrique ainsi que le même noyau. Nous mentionnons que le produit scalaire dans la matrice  $\langle \cdot, \cdot \rangle$  est obtenu en fonction de la métrique  $d$  par l'expression suivante :

$$\langle x, y \rangle = \frac{1}{2} (\|x\| + \|y\| - \|x - y\|).$$

Les résultats de débruitage, par le modèle local linéaire, sont obtenus par la figure 5.4

De même que le cas précédent, nous allons comparer la classification des courbes pures aux courbes bruitées. Les résultats de la classification sont représentés dans la figure 5.5, dont on calcule le pourcentage des courbes mal classées en utilisant le même algorithme que pour le modèle précédent. De même que pour le premier cas, les courbes pures donnent plus de visibilité que les courbes bruitées. En effet, le gain en classification est très important, car celui-ci dépasse 40% dans certaines situations. Il est à noter que les deux

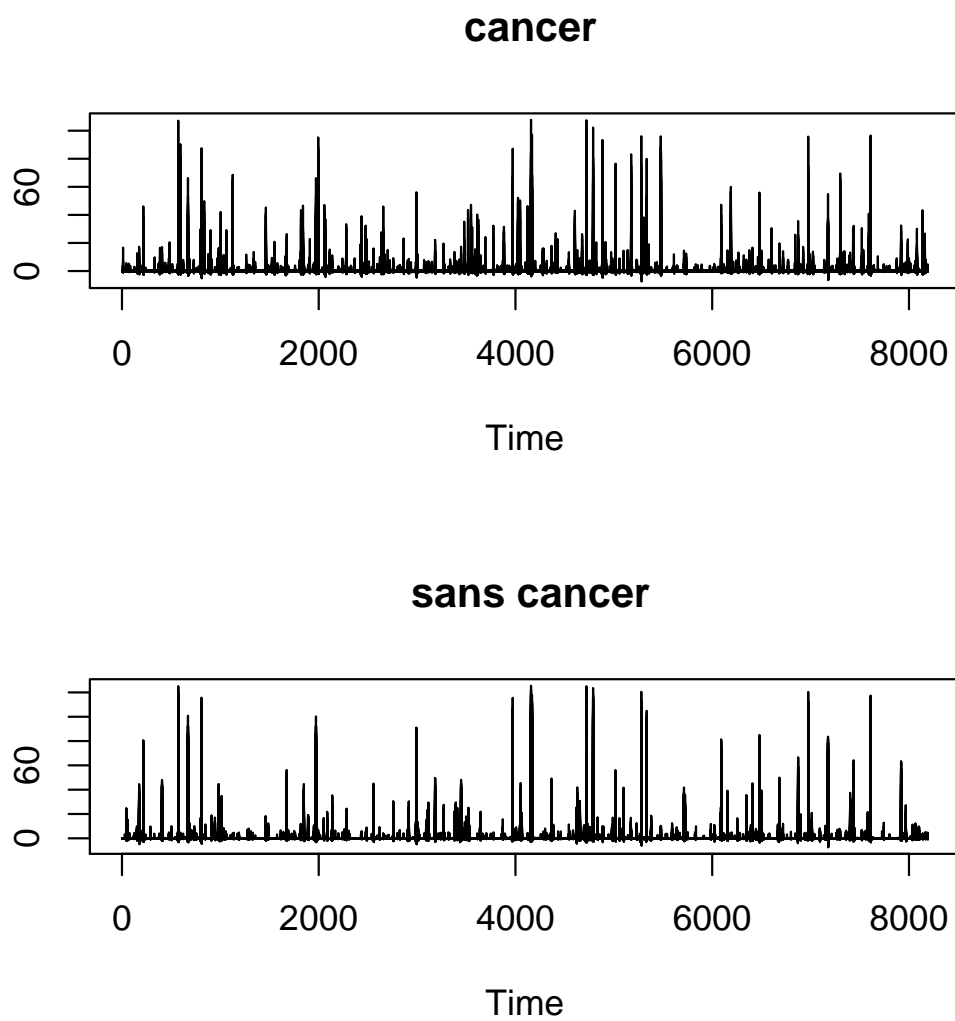


FIGURE 5.4 – cancer/ sans cancer

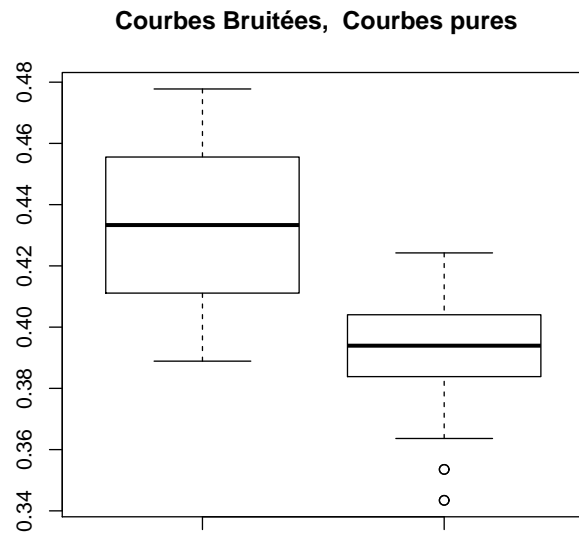


FIGURE 5.5 – comparaison de courbes

modèles sont presque équivalents, avec une légère optimalité de la régression locale linéaire. Finalement, on mentionne qu'on vu du temps d'exécution de nos programmes, les deux approches sont rapides et efficaces.



# Chapitre 6

## Conclusion et perspectives

### 6.1 Conclusion générale

Dans ce manuscrit, nous nous sommes intéressés à l'estimation locale linéaire de l'opérateur de régression dans trois situations fondamentales. Il est clair que la supériorité de cette approche sur la méthode à noyau classique est la motivation principale de ce sujet. Afin de mettre en évidence cette supériorité en NPFDA, nous avons étudié dans un premier temps, l'erreur quadratique de l'estimateur local linéaire de l'opérateur de régression introduit par Barrientos et al. (2010). Plus précisément, nous avons précisé les termes dominants dans cette erreur quadratique, à savoir le biais et la variance. Cette précision de la vitesse de convergence nous a prouvé la supériorité précédemment indiquée (notamment dans la partie biais).

La deuxième partie de cette thèse est consacrée au problème d'estimation de l'opérateur de régression quand la variable réponse est elle aussi fonctionnelle. Pour ceci, nous avons traité deux cas spécifiques. Dans le premier cas, nous avons considéré des variables hilbertiennes et nous avons adopté l'estimateur dû à Baillo et Grané (2009). Nous avons ensuite établi sa convergence presque-complète et uniforme en précisant sa vitesse de convergence.

Le résultat ainsi obtenu garde la forme usuelle de la vitesse de convergence presque-complète et uniforme. En effet, nos vitesses de convergence sont décomposées en trois termes. Tandis que les deux premiers termes contrôlent le biais et la dispersion de l'estimateur, le troisième terme contrôle la régularité de la projection de la variable hilbertienne.

Le dernier problème traité dans ce manuscrit est le cas où la variable explicative est dans un espace semi-métrique tandis que la variable réponse est hilbertienne. Dans ce cas, nous avons adopté l'estimateur de Barrientos et al. (2010). Pour ce dernier, nous avons montré la convergence uniforme presque-complète, puis nous avons donné sa vitesse de convergence. Celle-ci est exprimée en fonction de la concentration de la variable explicative et de l'entropie. Mentionnons également que le type et le cotype de l'espace fonctionnel pour la variable réponse est implicitement exploité dans la vitesse de convergence car la définition de l'estimateur nous impose de prendre une structure hilbertienne où le type et le cotype est égal à 2.

Dans la dernière partie de cette thèse nous avons considéré quelques exemples d'application sur des données simulées puis réelles. A partir de cette étude applicative nous avons pu constater que la régression locale linéaire propose des outils de prévisions alternatives à la régression classique avec une précision plus efficace.

## 6.2 Perspectives

L'importance de ce sujet peut être aussi exprimée en fonction des nombreuses perspectives qu'il engendre. C'est pour ceci que nous allons nous atteler, dans ce paragraphe, à lister quelques questions ouvertes à traiter dans le futur (à court terme ou à long terme) .

— *Normalité asymptotique.*

Dans notre thèse, nous avons établi des résultats de convergence presque-

complète. Ces derniers montrent le bon comportement asymptotique de notre estimateur. Cependant, la normalité asymptotique des estimateurs étudiés est un résultat indispensable pour plusieurs analyses statistiques alternatives telles que l'estimation par intervalle de confiance, les tests statistiques, ... Ainsi, l'obtention de ce résultat constitue une question importante à traiter dans le futur.

— *Méthode des  $k$  plus proches voisins.*

Tous nos estimateurs ont une pondération qui est issue de la méthode à noyau. Cependant, il est bien connu que ce type de pondération peut être amélioré en utilisant une pondération issue de la méthode des  $k$  plus proches voisins. Ainsi, l'estimation locale linéaire, par la méthode des  $k$  plus proches voisins, peut être considérée comme une méthode d'estimation alternative qui reste encore à développer pour nos modèles.

— *Cas de données dépendantes.*

La généralisation de nos résultats asymptotiques au cas où les données sont dépendantes est une question très importante dans la pratique. En effet, elle permet de faire de la prévision pour un processus à temps continu par l'approche locale linéaire. Plus précisément, elle permet de prévoir toute une courbe sachant qu'une autre courbe, est à disposition.

— *Méthode d'estimation par polynômes locaux.*

Il est clair que la supériorité de la méthode locale linéaire est justifiée par l'approximation linéaire localement dans un voisinage du point de conditionnement. Bien que ce gain est important, notamment pour réduire la partie biais, cette réduction peut être plus importante si l'on considère une approximation par un polynôme. Il est donc intéressant de regarder ce cas en NPFDA.

Notons, que nous sommes loin de prétendre à l'exhaustivité dans nos projets

de recherche. Surtout avec le nombre important de questions à traiter telles que le cas de données spatiales, le cas de données manquantes, le cas de données comportant ou entachée d'erreurs,...

# Bibliographie

- [1] Ardjoun, F.Z., Ait-Hennani, L. and Laksaci, A. A recursive kernel estimate of the functional modal regression under ergodic dependence condition. *J. Stat. Theory Pract.*, 10 (2016), no. 3, 475–496.
- [2] Attouch, M., Laksaci, A. and Ould-Said, E. Asymptotic distribution of robust estimator for functional nonparametric models. *Comm. Statist. Theory Methods.*, 38 (2009), no. 8-10, 1317–1335.
- [3] Attouch, M., Laksaci, A. and Ould Said, E. Asymptotic normality of a robust estimator of the regression function for functional time series data. *J. Korean Statist. Soc.*, 39 (2010), no. 4, 489–500.
- [4] Attouch, M., Gheriballah, A. and Laksaci, A. Convergence presque-complète d'un estimateur robuste de la régression non paramétrique fonctionnelle : cas spatial. *Ann. I.S.U.P.*, 56 (2012), no. 1, 3–16.
- [5] Attouch, M. and Bouabça, W. The  $k$ -nearest neighbors estimation of the conditional mode for functional data. *Rev. Roumaine Math. Pures Appl.*, 58 (2013), no. 4, 393–415.
- [6] Attouch, M. and Belabed, F.Z. The  $k$  nearest neighbors estimation of the conditional hazard function for functional data. *REVSTAT.*, 12 (2014), no. 3, 273–297.

- [7] Azzedine, N., Laksaci, A. and Ould-Said, E. On robust nonparametric regression estimation for a functional regressor. *Statist. Probab. Lett.*, 78 (2008), no. 18, 3216–3221.
- [8] Baïllo, A. and Grané, A. Local linear regression for functional predictor and scalar response. *J. Multivariate Anal.*, 100 (2009), no. 1, 102–111.
- [9] Barrientos-Marin, J., Ferraty, F. and Vieu, P. Locally modelled regression and functional data. *J. Nonparametr. Stat.*, 22, (2010), no. 5-6, 617-632.
- [10] Benhenni, K., Ferraty, F., Rachdi, M. and Vieu, P. Local smoothing regression with functional data. *Comput. Statist.*, 22 (2007), no. 3, 353–369.
- [11] Berline, A. and Servien, R. Necessary and sufficient condition for the existence of a limit distribution of the nearest-neighbour density estimator. *J. Nonparametr. Stat.*, 23 (2011), no. 3, 633–643.
- [12] Bosq, D. *Linear Processes in Function Spaces : Theory and applications*. Lecture Notes in Statistics, 149, Springer. (2000).
- [13] Burba, F., Ferraty, F. and Vieu, P.  $k$ -nearest neighbour method in functional nonparametric regression. *J. Nonparametr. Stat.*, 21 (2009), no. 4, 453–469.
- [14] Cai, Z. and Masry, E. Nonparametric estimation of additive nonlinear ARX time series : local linear fitting and projections. *Econometric Theory.*, 16 (2000), no. 4, 465–501.
- [15] Chouaf, A. and Laksaci, A. On the functional local linear estimate for spatial regression. *Stat. Risk Model.*, 29 (2012), no. 3, 189–214.
- [16] Crambes, C., Delsol, L. and Laksaci, A. Robust nonparametric estimation for functional data. *J. Nonparametr. Stat.*, 20 (2008), no. 7, 573–598.
- [17] Cuevas, A. A partial overview of the theory of statistics with functional data. *J. Statistical Planning and Inference*, 147 (2014), 1-23.

- [18] Dabo-Niang, S. and Rhomari, N. Kernel regression estimation in a Banach space. *J. Statistical Planning and Inference*, 139 (2009), 1421–1434.
- [19] Dabo-Niang, S., Rachdi, M. and Yao, A. Kernel regression estimation for spatial functional random variables. *Far East J. Theor. Stat.*, 37 (2011), no. 2, 77–113.
- [20] Demongeot, J., Ghassani, M., Rachdi, M., Ouassou, I. and Taramasco, C. Archimedean copula and contagion modeling in epidemiology. *Netw. Heterog. Media.*, 8 (2013), no. 1, 149–170.
- [21] Demongeot, J., Laksaci, A., Madani, F. and Rachdi, M. Functional data : Local linear estimation of the conditional density and its application *Statistics*, 47 (2013), 26–44.
- [22] Demongeot, J., Laksaci, A., Rachdi, M. and Rahmani, S. On the local linear modelization of the conditional distribution for functional data. *Sankhya A.*, 76 (2014), no. 2, 328–355.
- [23] Demongeot, J., Laksaci, A., Nacéri, A. and Rachdi, M. Estimation locale linéaire de la fonction de régression pour des variables hilbertiennes. *C.R. Math. Acad. Sci. Paris*, 354 (2016), 847–850.
- [24] Demongeot, J., Laksaci, A., Nacéri, A. and Rachdi, M. Local linear regression modelization when all variables are curves. *Statist. Probab. Lett.*, 121 (2017), 37–44.
- [25] Douge, L. Théorèmes limites pour des variables quasi-associées hilbertiennes. *Ann. I.S.U.P.*, 54 (2010), no. 1-2, 51–60.
- [26] Ezzahrioui, M., Ould-Said, E. Asymptotic normality of the kernel estimators of the conditional mode for functional data. *J. Nonparametr. Statist.*, 20 (2008), 3-18
- [27] Fan, J. Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, 87 (1992), no. 420, 998–1004.

- [28] Fan, J. and Gijbels, I. Variable bandwidth and local linear regression smoothers. *Ann. Statist.*, 20 (1992), no. 4, 2008–2036.
- [29] Fan, J. Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.*, 21 (1993), no. 1, 196–216.
- [30] Fan, J. and Gijbels, I. Censored regression : local linear approximations and their applications. *J. Amer. Statist. Assoc.*, 89 (1994), no. 426, 560–570.
- [31] Fan, J., Heckman, N.E. and Wand, M.P. Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *J. Amer. Statist. Assoc.*, 90 (1995), no. 429, 141–150.
- [32] Fan, J. and Gijbels, I. *Local Polynomial Modelling and its Applications*. London, Chapman & Hall. (1996).
- [33] Ferraty, F. and Vieu, P. Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. *C.R. Math. Acad. Sci. Paris.*, 330 (2000), no. 2, 139–142.
- [34] Ferraty, F., Goia, A. and Vieu, P. Functional nonparametric model for time series : a fractal approach for dimension reduction. *Test*, 11 (2002), no. 2, 317–344.
- [35] Ferraty, F., Laksaci, A. and Vieu, P. Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat. Inference Stoch. Process.*, 9 (2006), no. 1, 47–76.
- [36] Ferraty, F. and Vieu, P. *Nonparametric functional data analysis. Theory and Practice*. Springer Series in Statistics. New York. (2006).
- [37] Ferraty, F., Mas, A. and Vieu, P. Nonparametric regression on functional data : inference and practical aspects. *Aust. N. Z. J. Stat.*, 49 (2007), no. 3, 267–286.



- [38] Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statistical Planning and Inference*, 140 (2010), 335–352.
- [39] Ferraty, F., Van Keilegom, I. and Vieu, P. On the validity of the bootstrap in non-parametric functional regression. *Scand. J. Stat.*, 37 (2010), no. 2, 286–306.
- [40] Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. Kernel regression with functional response. *J. Electron. Statist.*, 5 (2011), 159–171.
- [41] Ferraty, F. and Vieu, P. *Kernel regression estimation for functional data*. The Oxford handbook of functional data analysis, 72–129, Oxford Univ. Press, Oxford, 2011.
- [42] Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2012). Estimation de la fonction de régression pour variable explicative et réponse fonctionnelles dépendantes. *C.R. Math. Acad. Sci. Paris*, 350 (2012), 717–720.
- [43] Ferraty, F., Van Keilegom, I. and Vieu, P. Regression when both response and predictor are functions. *J. Multivariate Anal.*, 109 (2012), 10–28.
- [44] Ferraty, F., Sued, M. and Vieu, P. Mean estimation with data missing at random for functional covariables. *Statistics.*, 47 (2013), no. 4, 688–706.
- [45] Gheriballah, A., Laksaci, A. and Sekkal, S. Nonparametric  $M$ -regression for functional ergodic data. *Statist. Probab. Lett.*, 83 (2013), no. 3, 902–908.
- [46] Goia, A. and Vieu, P. An introduction to recent advances in high/infinite dimensional Statistics. *J. Multivariate Analysis*, 146 (2016), 1–6.
- [47] Hsing, T. and Eubank, R. *Theoretical foundations of functional data analysis, with an introduction to linear operators*. Wiley Series in Probability and Statistics. *John Wiley & Sons, Ltd., Chichester*, (2015).

- [48] Kara-Zaitri, L., Laksaci, A., Rachdi, M. and Vieu, P. Data-driven  $k$ NN estimation in nonparametric functional data analysis, *J. Multivariate Analysis*, 153 (2016), 176-188.
- [49] Kudraszow, N. L. and Vieu, P. Uniform consistency of  $k$ NN regressors for functional variables. *Statist. Probab. Lett.*, 83 (2013), no. 8, 1863–1870.
- [50] Laïb, N. and Louani, D. Rates of strong consistencies of the regression function estimator for functional stationary ergodic data. *J. Statistical Planning and Inference*, 141 (2011), no. 1, 359–372.
- [51] Laksaci, A. Erreur quadratique de l'estimateur à noyau de la densité conditionnelle à variable explicative fonctionnelle. *C.R. Math. Acad. Sci. Paris*, 345 (2007), no. 3, 171–175.
- [52] Laksaci, A. and Hachemi, N. Note on the functional linear estimate of conditional cumulative distribution function. *JPSS J. Probab. Stat. Sci.*, 10 (2012), no. 2, 153–160.
- [53] Laksaci, A., Madani, F. and Rachdi, M. Kernel conditional density estimation when the regressor is valued in a semi-metric space. *Comm. Statist. Theory Methods.*, 42 (2013), no. 19, 3544–3570.
- [54] Laksaci, A., Nacéri, A. and Rachdi, M. *Exact Quadratic Error of the Local Linear Regression Operator Estimator for Functional Covariates*. Functional Statistic and Application, Springer, 79–90, (2015).
- [55] Lian, H. Convergence of functional  $k$ -nearest neighbor regression estimate with functional responses. *J. Electron. Statist.*, 5 (2011), 31-40.
- [56] Masry, E. Multivariate local polynomial regression for time series : uniform strong consistency and rates. *J. Time Ser. Anal.*, 17 (1996), no. 6, 571–599.

- [57] Masry, E. and Mielniczuk, J. Local linear regression estimation for time series with long-range dependence. *Stochastic Process. Appl.*, 82 (1999), no. 2, 173–193.
- [58] Masry, E. Local linear regression estimation under long-range dependence : strong consistency and rates. *IEEE Trans. Inform. Theory.*, 47 (2001), no. 7, 2863–2875.
- [59] Masry, E. Nonparametric regression estimation for dependent functional data : asymptotic normality. *Stochastic Process. Appl.*, 115 (2005), no. 1, 155–177.
- [60] Messaci, F., Nemouchi, N., Ouassou, I. and Rachdi, M. Local polynomial modelling of the conditional quantile for functional data. *Stat. Methods Appl.*, 24 (2015), no. 4, 597–622.
- [61] Naceri, A. Functional local linear regression model with functional response. *Journal of Probability and Statistical Sciences*, 27 (2015), 1–11.
- [62] Ould-Said, E. and Tatachak, A. Asymptotic properties of the kernel estimator of the conditional mode for the left truncated model. *C.R. Math. Acad. Sci. Paris*, 344 (2007), no. 10, 651–656.
- [63] Ould-Said, E. and Sadki, O. Prediction via the conditional quantile for right censorship model. *Far East J. Theor. Stat.*, 25 (2008), no. 2, 145–179.
- [64] Rachdi, M., Laksaci, A., Demongeot, J., Abdali, A. and Madani, F. Theoretical and practical aspects of the quadratic error in the local linear estimation of the conditional density for functional data. *Comput. Statist. Data Anal.*, 73 (2014), 53–68.
- [65] Ramsay, J.O. and Silverman, B.W. *Applied functional data analysis. Methods and case studies*. Springer Series in Statistics. New York. (2002).
- [66] Ruppert, D. Wand, M.P. Multivariate locally weighted least squares regression. *Ann. Statist.*, 22 (1994), no. 3, 1346–1370.

- [67] Stone, C.J. Consistent nonparametric regression. With discussion and a reply by the author. *Ann. Statist.*, 5 (1977), no. 4, 595–645.
- [68] Wang, J., Ma, W., Fan, G. and Wen, L. Local linear quantile regression with truncated and dependent data. *Statist. Probab. Lett.*, 96 (2015), 232–240.
- [69] Wang, L. Local linear estimation for regression models with locally stationary long memory errors. *J. Korean Statist. Soc.*, 45 (2016), no. 3, 381–394.
- [70] Yu, K. and Jones, M.C. Local linear quantile regression. *J. Amer. Statist. Assoc.*, 93 (1998), no. 441, 228–237.
- [71] Zhang, J. (2014). *Analysis of variance for functional data*. Monographs on Statistics and Applied Probability, 127. CRC Press, Boca Raton, FL.