

UNIVERSITÉ DE GRENOBLE

Thèse

pour obtenir le grade de
Docteur de l'Université de Grenoble

Spécialité Statistique
Arrêté ministériel : 7 août 2006

Présentée et soutenue publiquement par

Fethi Madani
le (-/-/2012)

Aspects théoriques et pratiques dans l'estimation
non paramétrique de la densité conditionnelle
pour des données fonctionnelles

JURY

Jacques Demongeot	Univ. Joseph Fourier, Grenoble	Président
Pascal Sarda	Université Le Mirail, Toulouse	Rapporteur
Elias Ould-Saïd	Univ. du Littoral Côte d'Opale, France	Rapporteur
Mustapha Rachdi	Univ. P. Mendès France, Grenoble	Directeur de thèse
Ali Laksaci	Univ. D. Liabès Sidi Bel Abbès, Algérie	Examineur
Idir Ouassou	ENSA, Marrakech, Maroc	Examineur
Sophie Lambert-Lacroix	Univ. P. Mendès France, Grenoble	Examineur

Thèse préparée au sein du laboratoire AGE Imagerie et Modélisation (AGIM) dans l'École Doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique.

Table des matières

Table des matières	3
0.1 Description et Contribution de cette thèse	9
0.2 Contexte bibliographique	10
1 Introduction aux données fonctionnelles et à l'estimation de la densité conditionnelle	15
1.1 Données fonctionnelles	15
1.2 Données fonctionnelles vs semi-métrique	18
1.2.1 Probabilités des petites boules	21
1.2.2 Champs d'application des données fonctionnelles	22
1.3 Quelques résultats sur l'estimation non-paramétrique pour des modèles fonctionnels	28
1.3.1 Notations et hypothèses	28
1.3.2 Estimation de la loi conditionnelle	29
1.3.3 Estimateur à noyau de la densité conditionnelle	30
1.3.4 Estimation du mode conditionnel	31
2 Kernel conditional density estimation when the regressor is valued in a semi-metric space	35
2.1 Introduction	36
2.2 Global and local bandwidth selection rules	38
2.3 Main Results	40
2.3.1 Assumptions	40
2.3.2 Some interpretations and examples on our hypotheses	41
2.3.3 Two theorems on global and local criteria	43
2.4 Discussion and applications	44
2.4.1 On the applicability of the method	44
2.4.2 On the finite-sample performance of the method	47
2.4.3 A real data application	51
2.5 Proofs	54
2.6 Appendix : Proofs of technical lemmas	56

Bibliography	62
3 Functional data : Local linear estimation of the conditional density and its application	67
3.1 Introduction	67
3.2 Model	68
3.3 Pointwise almost complete convergence	69
3.4 Uniform almost complete convergence	71
3.5 Application : Conditional mode estimation	73
3.6 Appendix	73
Bibliography	85
4 A fast functional locally modeled of the conditional density and mode in functional time series	87
4.1 Introduction	87
4.2 Main results	89
4.3 Concludes remarks	91
4.4 Appendix	93
Bibliography	102
5 On the quadratic error of the functional local linear estimate of the conditional density	107
5.1 Introduction	107
5.2 The model	108
5.3 Main results	109
5.4 Some comments and discussion	111
5.5 Proofs	112
Bibliography	119
6 Estimation locale linéaire des paramètres conditionnels pour des données fonctionnelles : Application sur des données simulées et réelles	121
6.1 Illustration du mode conditionnel	121
6.2 Illustration de la densité conditionnelle	126
6.3 Application sur des données réelles	127
7 Conclusion et Perspectives	133
7.1 Conclusion	133
7.2 Perspectives	134
8 Bibliographie générale	135

Résumé

Dans cette thèse, nous nous intéressons à l'estimation non paramétrique de la densité conditionnelle d'une variable réponse réelle conditionnée par une variable explicative fonctionnelle de dimension éventuellement finie.

Dans un premier temps, nous considérons l'estimation de ce modèle par la méthode à double noyaux. Nous proposons une méthode de sélection pour le choix du paramètre de lissage (global ou local) des paramètres de lissage et nous montrons son optimalité asymptotique dans le cas où les observations sont indépendantes et identiquement distribuées. Le critère adopté est issu du principe de validations croisées. Dans cette partie nous comparons également les deux types de choix (local et global).

Dans la deuxième partie, nous estimons la densité conditionnelle par la méthode des polynômes locaux. Sous certaines conditions, nous établissons des propriétés asymptotiques de cet estimateur tel la convergence presque complète et la convergence en moyenne quadratique dans le cas où les observations sont indépendantes et identiquement distribuées. Nous traitons aussi le cas où les observations sont de type α -mélangeantes, dont on montre la convergence presque complète (avec vitesse) de l'estimateur proposé. Les résultats obtenus sont également illustrés par des exemples sur des données simulées montrant l'applicabilité rapide et facile de cette méthode d'estimation dans le cadre fonctionnel.

Summary

In this thesis, we consider the problem of the nonparametric estimation of the conditional density when the response variable is real and the regressor is valued in a functional space.

In the first part, we use the double kernels method as a estimation method where we focus on the choice of the smoothing parameters. We construct data a driven method to select optimally bandwidths parameters. As main results, we study the asymptotic optimality of this selection's method in the case where observations are independent and identically distributed. Our selection rule is based on the classical cross-validation procedure and it deals with the both (global or local) choice. The finite sample performance of our approach is illustrated by some simulation results where we give a comparison between the two types of choice (local or global).

In the second part, we estimate the conditional density by the local linear method. Under some general conditions, we establish the almost complete convergence of the proposed estimator (with rate) in the both cases (i.i.d. case and the α -mixing case) . As application, we use the conditional density estimator to estimate the conditional mode estimation and we derive the same asymptotic proprieties.

Further, we study the quadratic error of this estimator by giving the asymptotic expansion of the exact expression involved in the leading in the bias and variance terms.

Liste des travaux

Publications dans des revues à comités de lecture

1. J. Demongeot, A. Laksaci, F. Madani and M. Rachdi. Local Linear estimation of the conditional density for functional data. *C. R., Math., Acad. Sci. Paris*, **348**, Issues 15-16, Pages 931-934, (2010).
2. J. Demongeot, A. Laksaci, F. Madani and M. Rachdi. Functional data : local linear estimation of the density and its application. *Statistics*, DOI : 10.1080/02331888.2011.568117 (à paraître en 2012).
3. J. Demongeot, A. Laksaci, F. Madani and M. Rachdi (2011). *A fast functional locally modeled conditional density and mode for functional time-series*. Recent Advances in Functional Data Analysis and Related Topics “Contributions to Statistics”, Physica-Verlag/Springer, 2011, 85-90, DOI : 10.1007/978-3-7908-2736-1_13
4. A. Laksaci, F. Madani and M. Rachdi. Kernel conditional density estimation when the regressor is valued in a semi-metric space. *Accepté pour publication dans : Communications Statistics-Theory and Methods*, 2012.

Communications dans des congrès

1. *Local bandwidth selection for kernel conditional density estimation when the regressor is valued in a semi-metric space*. Colloque international de Statistique des processus et Applications, CISPA 2008, Constantine : 18-19 octobre 2008.
2. *Local bandwidth selection for kernel conditional density estimation when the regressor is valued in a semi-metric space*. Journées de Statistique, Modélisation et Application JSMA'08, Alger : 22-23-24 novembre 2008.
3. *Some asymptotics for conditional parameters when the data are curves*. *International Conference on Statistics, Theory and Practice*, Sidi Bel-Abbès, 10-12 avril 2010.

Introduction générale

0.1 Description et Contribution de cette thèse

La statistique non paramétrique connaît un grand essor chez de nombreux auteurs et dans différents domaines. En effet, celle-ci possède un champ d'application très large permettant, ainsi, l'explication de certains phénomènes mal modélisés jusqu'à présent, tels que les séries chronologiques, et prédire les réalisations futures.

Il faut mentionner, par ailleurs, que les progrès atteints dans les procédés de recueil de données ont permis d'offrir la possibilité aux statisticiens de disposer de plus en plus souvent d'observations de variables dites *fonctionnelles*, c'est-à-dire de courbes. Ces données sont modélisées comme étant des réalisations d'une variable aléatoire prenant ses valeurs dans un espace abstrait de dimension éventuellement finie. Dans cette thèse, nous nous intéressons à l'estimation non paramétrique de la densité conditionnelle et les paramètres qui en découlent, comme le mode conditionnel, pour des variables aléatoires fonctionnelles.

Dans le but de présenter les travaux que nous avons réalisé durant la réalisation de cette thèse, celle-ci est organisée comme suit :

Le chapitre suivant, est un *chapitre Introductif*, qui présente une étude bibliographique des problèmes liés à l'analyse statistique des variables fonctionnelles ainsi qu'à l'estimation non paramétrique des paramètres conditionnels que ce soit dans le cadre de dimension finie ou infinie. Ensuite, dans le *chapitre 1*, nous abordons l'état de l'art des variables fonctionnelles et leurs champs d'application. De plus, afin de rendre la lecture de cette thèse simple, nous exposons les résultats obtenus, dans la littérature, concernant l'estimation de la densité et du mode conditionnels, tout en fournissant et discutant les hypothèses qui ont permis d'obtenir ces résultats.

Dans le *chapitre 2*, nous commençons par construire et étudier les propriétés asymptotiques de l'estimateur à noyau de la densité conditionnelle quand la variable explicative est à valeurs dans un espace normé. Ensuite, nous proposons deux critères (le premier *global* et le second *local*) de choix automatique du paramètre de lissage afin de rendre efficace notre

estimation. Enfin, nous établissons les résultats théoriques ainsi que pratiques d'optimalité asymptotique du paramètre sélectionné.

Une suite logique de ce chapitre veut que l'on améliore les résultats obtenus. C'est pourquoi le *chapitre 3* est consacré à l'étude d'une méthode d'estimation non paramétrique de la densité conditionnelle d'une variable scalaire Y sachant une variable fonctionnelle X i.e., une variable à valeurs dans un espace semi-métrique. Cette méthode est basée sur une estimation par polynômes locaux. Une fois la construction de notre estimateur, à l'image de ce qui se fait en dimension finie, est achevée, nous nous sommes attelé à établir sous certaines conditions, les convergences ponctuelle et uniforme presque complètes ainsi que les vitesses de convergence de cet estimateur. Nous avons utilisé, ensuite, les résultats obtenus afin de déterminer les propriétés asymptotiques de l'estimateur local linéaire du mode conditionnel.

Le *chapitre 4* quant à lui, est destiné à l'étude, sous certaines conditions de dépendance faible (mélange fort), de la convergence forte de l'estimateur du chapitre précédent, ainsi qu'à la prévision d'une série temporelle par l'estimation du mode conditionnel.

Tandis que dans le *chapitre 5*, nous avons établi les vitesses de convergence dans l'estimation en moyenne quadratique de l'estimateur étudié dans les deux chapitres précédents, le *chapitre 6* est consacré à la mise en application de ces résultats pour des données simulées puis pour des données réelles.

Enfin, dans le *chapitre 7* nous exposons des perspectives de recherche permettant d'étendre et parfois de généraliser les résultats de cette thèse.

0.2 Contexte bibliographique

L'analyse statistique pour des variables fonctionnelles a pris une ampleur considérable ces dernières années. Ce domaine de recherche en statistique connaît actuellement un grand succès auprès de la communauté des statisticiens. La preuve de cet intérêt est la publication de nombreuses publications scientifiques sur ce sujet ainsi que les nombreuses applications pratiques auxquelles ces données s'y prêtent. C'est le cas, notamment, lorsque l'on s'intéresse aux techniques d'estimation quand les données sont fonctionnelles (cf. Kneip et Gasser (1992), Ramsay et Li (1996), Rice et Silverman (1991)). Il existe, en fait, deux principales raisons à l'engouement suscité par le traitement statistique des variables fonctionnelles : (1) cela permet d'utiliser et de développer des outils théoriques performants, (2) cela offre un énorme potentiel en terme d'applications, notamment, en imagerie, en agro-alimentaire, en reconnaissance de formes, en géophysique, en économétrie, en environnement, . . . De plus, cette thématique de recherche couvre tous les domaines concernés par la communauté de statisticiens : des plus appliqués aux plus théoriques sans prédominance de l'une sur l'autre.

D'abord, signalons les efforts considérables qui ont été déployés pour la généralisation des

résultats connus et établis en dimension finie grâce à l'ouvrage de Ferraty et Vieu (2006). Celui-ci est devenu une référence en statistique non-paramétrique pour des données fonctionnelles. Notons que, l'analyse des données statistiques fait toujours intervenir le facteur dimension dans le comportement asymptotique des estimateurs établis. D'autant plus qu'il est connu que les vitesses de convergence se dégradent au fur et à mesure que la dimension augmente. Rappelons ici que les méthodes basées sur la discrétisation des données fonctionnelles ont été adoptées pour adapter les résultats de la statistique non-paramétrique au cas de données multivariées.

Vu l'avancée qu'a connu l'outil informatique dans la façon de récolter les données, d'autres alternatives sont devenues obligatoires afin de surmonter cette difficulté et d'étudier les données dans leurs propres dimensions.

D'ailleurs, le traitement des données en tant que courbes remonte aux années soixantes lorsque plusieurs études dans différentes disciplines se sont confrontées à des observations sous forme de trajectoires (cf. entre autres, Holmstrom (1961) en climatologie, Deville (1974) en démographie, Molenaar et Boomsma (1987) puis Kirkpatrick (1989) en génétique,...)

Il est bien connu qu'en statistique, le modèle de régression (paramétrique ou non-paramétrique) en dimension finie, constitue un champ de recherche et d'application très important, nous renvoyons ici aux travaux de Collomb (1981, 1985) qui dès le début des années quatre-vingt font déjà état de nombreux développements variés sur ce thème. Il convient, également, de se référer aux ouvrages de Härdle (1990), Bosq et Lecoutre (1987) et Schimek (2000) qui dressent un bilan presque exhaustif sur les diverses techniques en la matière. Ces champs de la recherche en statistique sont encore potentiellement porteurs à la fois au niveau des développements théoriques et à cause des multiples possibilités d'application.

Par ailleurs, les applications liées au modèle de régression ont une place très importante dans la prévision des séries chronologiques issues de différentes disciplines telles que la communication, les systèmes de contrôle, la climatologie ainsi que l'économétrie. Il s'agit, donc, de domaines de prévision pour lesquels les premiers résultats conséquents furent implantés par Collomb (1981) et Robinson (1983). Ce domaine de la statistique connaît des développements continus, comme en témoignent les nombreuses réalisations (cf. Gyö et al. (1989), Yoshihara (1994), Härdle et al. (1997) et Bosq (1991),...)

Commençons par signaler que, l'estimation de la loi de probabilité ou de la fonction de distribution joue un rôle important dans l'estimation d'autres paramètres fonctionnels. Les premiers travaux concernant l'estimation de la loi de probabilité des variables fonctionnelles ont été réalisés par Geffroy (1974), Gasser et al. (1998). Notons aussi que, Cadre (2001) s'est intéressé à l'étude de la médiane d'une distribution pour une variable fonctionnelle à valeurs dans un espace de Banach.

Nous faisons remarquer que les paramètres conditionnels, tels que la distribution conditionnelle, la densité conditionnelle, le mode conditionnel, le quantile conditionnel et la fonction de hasard conditionnelle, sont largement étudiés en dimension finie. A travers ces paramètres, la prévision dans les modèles non-paramétriques offre une véritable alternative à la régression non paramétrique. Il faut dire qu'en dimension finie, il existe une littérature abondante pour ces paramètres conditionnels. Roussas (1968) fut le premier à établir des propriétés asymptotiques pour l'estimateur à noyau de la distribution conditionnelle, pour des données markoviennes, pour lesquelles il a montré la convergence en probabilité. Youndjé (1993) quant à lui, il s'est intéressé à l'étude de la densité conditionnelle pour des données dépendantes ou indépendantes. On peut, notamment, citer le travail mené par Laksaci et Yousfate (2002) et dans lequel ils ont établi, pour un processus markovien stationnaire, la convergence en norme L^p de l'estimateur à noyau de la densité conditionnelle.

Vu l'intérêt que revêt l'estimation du mode et du mode conditionnel dans le domaine de la prévision, plusieurs auteurs s'en sont intéressés. Nous pouvons citer par exemple, Perzen (1962) qui a été l'un des premiers à considérer le problème de l'estimation du mode d'une densité de probabilité univariée. Il a montré que, sous certaines conditions, l'estimateur du mode obtenu en maximisant un estimateur à noyau est convergent et est asymptotiquement normal quand les données sont indépendantes et identiquement distribuées (i.i.d). Les techniques de base qu'il a développé pour cette étude ont été reprises par de nombreux auteurs dans le cas de la densité de probabilité ou de la régression. Nous n'avons mentionné ici que les principales contributions, en ayant essentiellement en vue la normalité asymptotique. Notons aussi que Nadaraya (1965) et VanRyzin (1969) ont démontré la convergence forte de l'estimateur du mode mis en place par Perzen, alors que Samanta (1973) et Konakov (1974) ont étudié des versions multivariées de cet estimateur. Les travaux d'Eddy (1980 et 1982), quant à eux, ils ont permis d'affaiblir les conditions suffisantes de normalité asymptotique qui aurait été données initialement. Par ailleurs, grâce à des conditions locales, Romano (1980), a affaibli les hypothèses précédentes. Notons aussi que Vieu (1996) a comparé deux estimateurs à noyau du mode dont le premier est défini à partir du maximum d'un estimateur de la densité de probabilité et le second à partir du zéro d'un estimateur de la dérivée de celle-ci. Ce travail a été repris par Rachdi et Sabre (2000) afin d'estimer le mode de la densité de probabilité quand les données sont entachées d'erreurs additives (les problèmes de déconvolution). Il y a aussi, entre autres, Louani (1998) qui a établi la normalité asymptotique pour la densité et ses dérivées avec application au mode.

Concernant le mode conditionnel, les propriétés de convergence et de normalité asymptotiques ont été établies par Samanta et Thaavaneswaran (1990) dans le cadre de données indépendantes et identiquement distribuées, alors que des conditions de convergence dans le cas de données ϕ -mélangeantes ont été établies par Collomb et al. (1987), dans le cas de données α -mélangeantes par Ould-Saïd (1993), dans le cas de données ergodiques par Rosa (1993) et Ould-Saïd (1997). De leur côté, Quintela et Vieu (1997) ont estimé le mode conditionnel comme étant le point annulant la dérivée d'ordre un de l'estimateur de la densité conditionnelle et ils ont établi la convergence presque complète de cet estimateur sous

la condition d' α -mélangeance. Berlinet et al. (1998), quant à eux, ils ont présenté des résultats sur la normalité asymptotique des estimateurs convergents du mode conditionnel, indépendamment de la structure de dépendance des données avec une application au cas d'un processus stationnaire α -mélangeant. Tandis que Louani et Ould-Saïd (1999) ont établi la normalité asymptotique dans le cas de données fortement mélangeantes et dans le cas de données censurées. Ould-Saïd et Cai (2005), quant à eux, ils ont établi la convergence uniforme sur un compact.

Par ailleurs, dans le cadre de données à valeurs dans un espace de dimension éventuellement finie, les travaux de Ramsay et Silverman (2002 et 2005) constituent un recueil important de méthodes statistiques, principalement du point de vue pratique, mais des développements théoriques peuvent être trouvés dans Bosq (2000) et Ferraty et Vieu (2006).

Une contribution qui s'avère importante dans la construction de l'estimateur des paramètres dans le modèle de régression linéaire est celle qui est due à Cardot et al. (1999). Elle consiste en la construction d'un estimateur pour l'opérateur de régression à partir des propriétés spectrales de l'estimateur empirique de l'opérateur de covariance de la variable explicative fonctionnelle. Ils ont établi, également, les convergences en probabilité et presque sûre de l'estimateur construit. Ce travail a été revisité dans Cuevas et al. (2002). Dans celui-ci, une étude des propriétés asymptotiques de l'estimateur de l'opérateur de régression linéaire quand la variable explicative est fonctionnelle déterministe et la réponse est fonctionnelle aléatoire a été conduite. Cardot et al. (2004a, 2004b et 2005) ont proposé et étudié des méthodes d'estimation linéaire de l'opérateur de régression par quantiles conditionnels. Une autre méthode d'estimation des quantiles conditionnels à partir de l'estimation à noyau de la fonction de répartition conditionnelle a également été proposée et étudiée par Ferraty et al. (2005), Ferraty et al. (2006), Ferraty et Vieu (2006a) et Ezzahrioui (2007). D'autres méthodes ont été proposées afin d'estimer la régression par le mode conditionnel. Celles-ci sont basées sur l'estimation de la densité conditionnelle par des estimateurs à noyau (cf. Ferraty et al. (2005), Ferraty et Vieu (2006a), Ferraty et al. (2006), Dabo-Niang et Laksaci (2006) et Ezzahrioui (2007)).

Donc, l'estimation de la densité conditionnelle en dimension éventuellement finie a connu un grand intérêt en statistique. Ce paramètre fonctionnel intervient pour l'estimation des quantiles, du mode ou de la fonction de hasard.

Signalons, qu'en dimension infinie, le mode conditionnel a connu tout récemment un intérêt croissant, malgré le peu de résultats disponibles dans la littérature. Dans ce contexte, les premiers travaux ont été réalisés par Ferraty et al. (2006). Ils ont montré, sous des conditions de régularité de la densité conditionnelle, la convergence presque complète des estimateurs à noyau de la densité conditionnelle et du mode conditionnel et ont établi leurs vitesses de convergence. Notons aussi qu'une application de leurs résultats aux données issues de l'industrie agro-alimentaire a été présentée. Dans le même contexte, Dabo-Niang et al. (2004) ont étudié un estimateur non paramétrique du mode de la densité d'une variable explicative

à valeurs dans un espace vectoriel semi-normé, de dimension éventuellement finie. Ils ont établi la convergence presque sûre avec une application de ce résultat au cas où la mesure de probabilité de la variable explicative vérifie une condition de concentration. On trouve aussi dans Dabo-Niang et Laksaci (2007) l'étude d'un estimateur à noyau du mode de la distribution d'une variable réelle Y conditionnée par une variable explicative X , à valeurs dans un espace semi-métrique. Ils ont établi la convergence en norme L^p de l'estimateur et ils ont montré que les résultats asymptotiques établis sont liés aux probabilités des petites boules de la loi de la variable explicative ainsi que la régularité de la densité conditionnelle. Notons également, qu'il y a deux autres paramètres fonctionnels qui sont d'une grande importance à savoir, le quantile et le quantile conditionnel. Ces paramètres proposent une alternative majeure dans la prévision, grâce à leur caractère robuste (cf. par exemple, les travaux de Cardot et al. (2004a, 2004b, 2005 et 2006), Ferraty et al. (2005b) et (2006)).

Pour terminer ce rapide tour d'horizon, non exhaustif, affirmons que d'un point de vue théorique, l'utilisation de variables aléatoires fonctionnelles introduit une difficulté supplémentaire puisqu'on ne peut plus se permettre de manipuler la fonction de densité de probabilité aussi facilement que dans le cas réel ou encore dans le cas vectoriel. On est donc amené à donner une écriture probabiliste qui nous conduit à des hypothèses agissant directement sur la distribution de la variable aléatoire fonctionnelle plutôt que sur la densité, comme dans le cas de dimension finie.

Chapitre 1

Introduction aux données fonctionnelles et à l'estimation de la densité conditionnelle

Dans ce chapitre, nous présentons, d'abord, quelques notions sur l'analyse des données fonctionnelles et son champ d'application, et puis, les résultats existants dans la littérature sur l'estimation de la densité conditionnelle.

1.1 Données fonctionnelles

Au cours de ces dernières années, la branche de la statistique consacrée à l'analyse des données fonctionnelles a connu un réel essor tant en termes des développements théoriques et méthodologiques que de la diversification des domaines d'application. Ceci revient aux progrès qu'a connus l'outil informatique au niveau des capacités de stockage qui permettent d'enregistrer des données de plus en plus volumineuses. Ainsi, un très grand nombre de variables peuvent être observées pour l'étude d'un même phénomène.

Une fois la réalité des variables fonctionnelles est présentée, on s'intéresse aux aspects de modélisation les concernant. Dans ce but, nous donnons quelques définitions permettant de fixer un vocabulaire. Rappelons, tout d'abord, qu'une variable aléatoire fonctionnelle est tout simplement une variable aléatoire à valeurs dans un espace de dimension éventuellement finie que nous noterons \mathcal{F} . Par exemple, cet espace \mathcal{F} peut être un espace de fonctions, d'opérateur linéaires, ... Selon la terminologie en vigueur dans la littérature, on parle aussi bien de variables aléatoires fonctionnelles que de données fonctionnelles, ce qui englobe notamment tout ce qui concerne l'analyse statistique de courbes.

Définition 1.1.1. *On appelle modèle fonctionnels, tout modèle prenant en compte au moins une variable aléatoire fonctionnelle (v.a.f).*

Définition 1.1.2. *Un modèle fonctionnel est dit paramétrique si C est indexable par un nombre fini de paramètres appartenant à \mathcal{F} , où C n'est qu'un sous-ensemble de $\mathbb{F}_{\mathcal{F}}^{\mathcal{F}'}$ ($\mathbb{F}_{\mathcal{F}}^{\mathcal{F}'}$ l'ensemble des fonctions définies sur l'espace fonctionnel \mathcal{F} et à valeurs dans l'espace \mathcal{F}'). Un modèle fonctionnel est dit non-paramétrique dans le cas contraire.*

De nombreux travaux ont été dédiés à l'étude des modèles impliquant des variables aléatoires multivariées. Ce domaine de la statistique connaît encore une activité de recherche soutenue. Cependant, les récentes innovations réalisées sur les appareils de mesure et les méthodes d'acquisition ainsi que l'utilisation de moyens informatique perfectionnés permettent souvent de récolter des données discrétisées sur des grilles de plus en plus fines, ce qui les rend fondamentalement fonctionnelles : c'est par exemple le cas en météorologie, en médecine, en imagerie satellite et dans de nombreux autres domaines d'études. C'est une des raisons pour lesquelles un nouveau champ de la statistique dédié à l'étude de données fonctionnelles, a soulevé un grand défi au début des années quatre-vingt, sous l'impulsion des travaux de Grenander (1981), Dauxois et al. (1982) et Ramsay (1982). En fait, ce domaine a été popularisé par Ramsay et Silverman (1997), puis par les différents ouvrages de Bosq (2000), Ramsay et Silverman (2002, 2005) et Ferraty et Vieu (2006). Notons que c'est un des domaines de la statistique qui est en plein essor comme en témoignent les travaux publiés et/ou cités dans des revues de premiers rangs, \dots , etc.

De plus, même si les données dont dispose le statisticien ne sont pas de nature fonctionnelle, celui-ci peut être amené à étudier des variables fonctionnelles construites à partir de son échantillon initial. Un exemple classique est celui où l'on observe plusieurs échantillons de données réelles indépendantes et où l'on est ensuite amené à comparer les densités de ces différents échantillons ou bien à considérer des modèles où elles interviennent (cf. Ramsay et Silverman, 2002). Dans le contexte particulier de l'étude des séries temporelles, l'approche introduite par Bosq (1991) fait apparaître une suite de données fonctionnelles dépendantes qui modélisent la série chronologique observée. Cette approche consiste tout d'abord à considérer le processus non pas à travers sa forme discrétisée mais comme étant un processus à temps continu puis à le découper en un échantillon de courbes successives.

Remarquons que la principale source de difficulté, que ce soit d'un point de vue théorique que pratique, provient du fait que les observations de ce type de variables sont supposées appartenir à un espace de dimension infinie.

Les tous premiers travaux dans lesquels nous retrouvons l'idée de considérer les données fonctionnelles sont relativement anciens. Rao (1958) et Tucker (1958) ont envisagé l'analyse en composantes principales et l'analyse factorielle pour des données fonctionnelles, en considérant explicitement les données fonctionnelles comme un type particulier de données. Par la suite, Ramsay (1982) a dégagé la notion de données fonctionnelles et a soulevé la question de l'adaptation des méthodes utilisées en analyse statistique de données multivariées (en dimension finie) au cadre fonctionnel.

A partir de là, les travaux portant sur la statistique des données fonctionnelles ont commencé à se multiplier pour finalement aboutir, aujourd'hui, à des ouvrages devenus des références en la matière. Par exemple, les monographies de Ramsay et Silverman (2002 et 2005), Ferraty et Vieu (2006) présentent une collection importante de méthodes statistiques spécifiques aux variables fonctionnelles dans les cadres linéaire et non linéaire. De même, Bosq (1991) a contribué au développement de méthodes statistiques permettant l'analyse de variables aléatoires fonctionnelles dépendantes (processus autorégressifs hilbertiens). Citons aussi, les travaux de Cuevas et al. (2002) qui se sont intéressés au problème de la régression linéaire d'une variable fonctionnelle sur un ensemble de données fonctionnelles déterministes "fixed functional design". D'autre part, Benhenni et al. (2010) ont considéré le problème d'estimation de l'opérateur de régression quand les données fonctionnelles sont déterministes et les erreurs sont corrélées. Cardot et al. (2005) quant à eux, ils ont proposé un estimateur non paramétrique de l'opérateur de régression quand le facteur prédictif est réel et la variable réponse est une courbe.

Par ailleurs, l'étude du modèle de régression non linéaire est beaucoup plus récente que celle du cas linéaire. Ferraty et Vieu (2000) ont établi les premiers résultats sur l'estimation non paramétrique de l'opérateur de régression non linéaire. Ces résultats ont ensuite été prolongés par Ferraty et al. (2002) en traitant le cas de données dépendantes et en établissant des convergences fortes de l'estimateur à noyau de la régression.

A leur tour, Niang et Rhomari (2003) ont étudié la convergence en norme L^p de l'estimateur de l'opérateur de régression et ont expérimenté leur résultats à la discrimination et à la classification de courbes. Rachdi et al. (2008) ont traité le problème d'estimation non paramétrique de l'opérateur de régression quand les erreurs vérifient des propriétés de longue mémoire. Ils ont établi aussi la convergence en probabilité ponctuelle puis uniforme de l'estimateur à noyau opératoire. Une autre contribution basée sur la construction d'un critère de choix automatique et optimal du paramètre de lissage pour l'estimateur de la régression quand le régresseur est de type fonctionnel a été menée par Rachdi et Vieu (2005, 2007). Tandis qu'El Methni et Rachdi (2011) ont établi l'estimation locale d'une moyenne pondérée de l'opérateur de régression pour des données fonctionnelles déterministes. Ouassou et

Rachdi (2010) ont amélioré ensuite cette estimation par l'estimateur de Stein.

Rappelons que, le fléau de la dimension rend les vitesses de convergence très faibles. Une manière de tenter de remédier à cela est de chercher une topologie qui restitue de façon pertinente les proximités entre les données. Cela peut être fait, par exemple, à l'aide d'une semi-métrique de projection basée sur les composantes principales fonctionnelles, les décompositions selon une base de Fourier, d'ondelettes, de splines, Lorsque la variable explicative est à valeurs dans un espace de Hilbert séparable, Ferraty et Vieu (2006a, Lemme 13-6) ont montré que l'on peut définir de manière générale une semi-métrique de projection qui permet de se ramener à des probabilités de petites boules de type fractal (i.e. $\exists C, \delta > 0, F_x(h) \sim C_x h^\delta$ quand $h \rightarrow 0$). On condense ainsi les données en réduisant leur dimension et on contourne ainsi le fléau de la dimension. En effet, on revient à des vitesses de convergence en puissance de n . Dans d'autres situations, on peut être confronté à des données très lisses (comme les courbes spectrométriques de masse données dans la Figure 1.2). Dans ce cas de figure, il peut être intéressant d'utiliser plutôt des semi-métriques basées sur les dérivées (cf. Ferraty et Vieu, 2006a). Ces semi-métriques peuvent également être utiles lorsque les données présentent un shift vertical artificiel (i.e non informatif vis-à-vis des réponses). Elles ont alors pour effet d'éliminer ces décalages verticaux qui nuisent à la qualité de la prédiction. Enfin, on peut envisager d'autres types de phénomènes comme, à titre d'exemple, les décalages horizontaux (cf. Dabo-Niang et al., 2006).

Face à la grande diversité des semi-métriques qu'on peut construire, on peut se poser la question sur comment choisir la semi-métrique la mieux adaptée au données. Ceci va motiver l'étude du problème de construction d'une semi-norme sur \mathcal{F} .

1.2 Données fonctionnelles vs semi-métrique

D'une façon générale, l'analyse de tout type de données nécessite la définition de la notion de distance entre celles-ci. Il est bien connu que dans un espace vectoriel de dimension finie toutes les métriques sont équivalentes. Ceci n'est plus le cas quand l'espace d'observations est de dimension infinie. C'est pourquoi le choix de la métrique (et donc de la topologie associée) est un élément crucial pour l'étude des variables aléatoires fonctionnelles.

De nombreux auteurs définissent ou étudient les variables fonctionnelles comme étant des variables aléatoires de carrés intégrables c'est-à-dire à valeurs dans $L^2(0, 1)$ (cf. notamment, Crambes et al., 2007) ou plus généralement dans un espace de Hilbert (cf. par exemple, Preda, 2007), ou de Banach (cf. Cuevas et Fraiman, 2004) ou métrique (cf. Dabo-Niang et Rhomari, 2003). Notons d'ailleurs que Bosq (2000), quant à lui, il a considéré des échantillons de variables fonctionnelles dépendantes et à valeurs dans un espace de Hilbert ou de Banach. Ces observations fonctionnelles ont été obtenues suite au découpage d'un même processus à temps continu. De plus, parmi les semi-métriques, disponibles dans la littérature, il est souvent plus intéressant de considérer des semi-métriques permettant un éventail plus large

de topologies possibles que l'on pourra choisir en fonction de la nature des données et du problème à traiter.

Signalons que, l'intérêt d'utiliser une semi-métrie plutôt qu'une métrie est que cela peut constituer une alternative aux problèmes liés à la grande dimension des données. En effet, on peut considérer une semi-métrie qui soit définie à partir d'une projection de nos données fonctionnelles sur un espace de dimension plus petite : (1) que ce soit en réalisant une analyse en composantes principales fonctionnelles de nos données (cf. Dauxois et al. (1982), Besse et Ramsay (1986), Hall et Hosseini-Nasab (2006) et Yao et Lee (2006)) ou (2) en les projetant sur une base de cardinal fini (ondelettes, splines, ...). Cela permet de réduire la dimension des données et ainsi d'augmenter la vitesse de convergence des méthodes utilisées tout en préservant la nature fonctionnelle des données. D'ailleurs, on peut choisir la base sur laquelle on projette en fonction des connaissances que l'on a de la nature de la donnée fonctionnelle. Par exemple, on pourrait choisir la base de Fourier si on suppose que la variable fonctionnelle observée est périodique. On peut se référer, pour cela, à Ramsay et Silverman (1997 et 2005) ou à Rossi et al. (2005) pour une discussion plus complète sur les différentes méthodes d'approximation par projection de données fonctionnelles. Aussi, une discussion plus approfondie de l'intérêt d'utiliser différents types de semi-métries est présentée dans le livre de Ferraty et Vieu (2006) (paragraphes 3 et 4) ainsi que dans le travail réalisé par Benhenni et al. (2007).

Pour ces différentes raisons, nous présentons ici quelque pistes (cf. Ferraty et Vieu, 2006) permettant de construire une semi-métrie. En fait, nous présentons, dans ce qui suit, seulement deux familles de semi-métries mais, naturellement, beaucoup d'autres peuvent être construites : la première est bien adaptée aux courbes dites bruitées et aux courbes irrégulières tandis que la deuxième sera plutôt employée pour le traitement de courbes tout à fait lisses (ou régulières).

Pour ce faire, nous commençons par considérer un échantillon de n courbes X_1, \dots, X_n indépendantes et identiquement distribuées de la variable aléatoire fonctionnelle $X = \{X(t), t \in [0, 1]\}$.

Notons que, l'analyse en composantes principales classique (ACP) est considérée comme étant un outil très utile pour la description et la visualisation des données dans un espace de dimension plus petite. Cette technique a été prolongée aux données fonctionnelles et plus récemment employée pour différents buts statistiques. Nous verrons que le FPCA (Functional Principal Components Analysis) est devenue un bon outil pour calculer des proximités entre les courbes dans un espace de dimension réduite. Ainsi, à partir de la semi-métrie classique L^2 , nous pouvons construire une classe paramétrique de semi-normes, que nous noterons SMPCA (Semi-Métrie basée sur l'ACP), de la manière suivante :

$$\|x\|_q^{ACP} = \sqrt{\sum_{k=1}^q \left(\int x(t)v_k(t) \right)^2 dt} \text{ pour tout } x \in \mathcal{F}$$

où v_1, \dots, v_q sont les fonctions propres orthonormales de l'opérateur de covariance :

$$\Gamma_X(s, t) = \mathbb{E}(X(t)X(s))$$

associées aux valeurs propres $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_q$.

Signalons aussi que, l'entier q n'est pas un paramètre de lissage, mais plutôt un paramètre de réglage indiquant le niveau de résolution auquel le problème est considéré.

On en déduit une famille de semi-métriques comme suit :

$$d_q^{ACP}(X_i, x) = \sqrt{\sum_{k=1}^q \left(\int (X_i(t) - x(t))v_k(t) dt \right)^2} \quad (1)$$

Notons que, l'approximation de l'intégrale dans la formule (1) peut se faire comme suit (cf Castro et al., 1986) :

$$\int_0^1 (X_i(t) - x(t))v_k(t)dt \simeq \sum_{j=1}^J w_j (X_i(t_j) - x(t_j))v_k(t_j)$$

où les poids $w_j = t_j - t_{j-1}$ et la grille (t_1, \dots, t_J) est constituée de J valeurs équidistantes dans $[0,1]$.

Si nous discrétisons deux courbes x_i et $x_{i'}$ alors, la quantité $d_q^{ACP}(x_i, x_{i'})$ sera approximée par sa version empirique :

$$d_q^{ACP}(x_i, x_{i'}) = \sqrt{\sum_{k=1}^q \left(\sum_{j=1}^J w_j (x_i(t_j) - x_{i'}(t_j))v_k(t_j) \right)^2}$$

où $\{x_i = (x_i(t_1), \dots, x_i(t_J))^t\}_{i=1, \dots, n}$ et $\{x_{i'} = (x_{i'}(t_1), \dots, x_{i'}(t_J))^t\}_{i'=1, \dots, n}$

En effet, cette famille de semi-métriques peut être utilisée seulement si les données sont équilibrées (les courbes sont observées aux mêmes points). Ceci pourrait apparaître comme un inconvénient pour l'usage d'un tel genre de semi-métriques mais, leur principal avantage est d'être utilisé même si les courbes son irrégulières. En prenant l'exemple de la prévision de la concentration maximale de l'ozone au pôle nord pendant une journée sur quatre années successives (de 2000 à 2004), étant donné la courbe de cette concentration pendant la journée précédente (cf. Figure 1.4), nous avons choisi la norme $L_{1,24}^2$ calculée, en utilisant ce genre de semi- métriques.

Une autre manière de construire une autre famille de semi-métriques est basée sur les dérivées, que nous allons noter par SMD (Semi-Métrique basée sur la Dérivée). Elle est définie

de la manière suivante :

$$d_q^{SMD}(x_i, x_{i'}) = \sqrt{\int_0^1 (x_i^{(q)}(t) - x_{i'}^{(q)}(t))^2 dt} \quad (2)$$

pour deux courbes observées x_i et $x_{i'}$, où $x^{(q)}$ désigne la dérivée d'ordre q de x . Notons, par ailleurs, que $d_0^{SMD}(x, 0)$ coïncide avec la norme classique sur l'espace L^2 de x .

De plus, on peut aussi utiliser l'approximation de chaque courbe par des B-splines (cf. De Boor (1978) ou Schumaker (1981)) et ainsi les dérivées successives seront directement calculées en différenciant plusieurs fois leurs formes analytiques. Ainsi, le calcul de l'intégrale dans (2) peut être effectué en utilisant la méthode de Gauss (cf. Lanczos, 1956). Dans la pratique, cette classe de semi-métriques sera bien adaptée et employée quand on a affaire à des courbes lisses, comme les données spectrométriques de masse (cf. Figure 1.2).

A ce stade, on pense que l'ensemble des données, lui-même, devrait être mis en avant afin de choisir la semi-métrie à employer.

En conclusion, chacune des deux familles discutées ci-dessus est adaptée à un certain genre de données : la SMPCA est prévue pour des données irrégulières, tandis que, la SMD est adaptée aux données lisses.

On peut donc affirmer, sans hésitation, que le choix de la semi-métrie permet à la fois de prendre en compte des situations plus variées et de pouvoir contourner le fléau de la dimension. Ce choix ne doit cependant pas être pris à la légère mais, doit prendre en compte, non seulement la nature des données mais aussi la nature du problème étudié.

1.2.1 Probabilités des petites boules

Le problème du fléau de la dimension est un phénomène bien connu dans le cas de modèles de régression multivariée non paramétrique. Il est bien connu que ce problème provoque une décroissance exponentielle des vitesses de convergence des estimateurs non paramétriques en fonction de la dimension (cf. Stone, 1982). Par conséquent, il est légitime de penser que les méthodes non paramétriques dans l'étude des modèles à variables fonctionnelles risquent d'avoir une vitesse de convergence très lente. Dans le cas où la variable explicative est multivariée (i.e. à valeurs dans un espace de dimension éventuellement finie (\mathcal{F}, d)), les vitesses de convergence de l'estimateur à noyau sont exprimées en fonction d'un terme de la forme h_n^d , provenant de la valeur de la probabilité que la variable explicative appartienne à la boule de centre x et de rayon h_n . Dans le cas d'une variable explicative fonctionnelle les résultats asymptotiques sont exprimés à partir de quantités plus générales appelées probabilités des petites boules et qui sont définies par :

$$F_x(h_n) := \mathbb{P}(d(X, x) \leq h_n) \text{ où } h_n \rightarrow 0$$

Au travers des différents résultats de convergence concernant l'estimateur étudié dans ce mémoire (de type Nadaraya-Watson et/ou local linéaire), on observe que la vitesse de convergence est fonction de la manière dont décroissent ces probabilités de petite boules. Il existe dans la littérature un nombre assez important de résultats probabilistes qui étudient la manière dont ces probabilités des petites boules tendent vers 0 quand d est une norme (cf. par exemple, Li et Shao (2001), Lifshits et al. (2006) et Gao et Li (2007)). On pourra également se référer au travail de Dereich (2003, Chapitre 7) qui est consacré au comportement des probabilités des petites boules dont les centres sont aléatoires. Au travers de ces travaux on peut voir, par exemple, que dans le cas de processus non-lisses tels que le mouvement brownien ou le processus d'Ornstein-Uhlenbeck, ces probabilités des petites boules sont de forme exponentielle (par rapport à h_n) et que par conséquent la vitesse de convergence de nos estimateurs est en puissance de $\ln(n)$ (cf. Ferraty et al. (2006), paragraphe 5 et Ferraty et Vieu (2006a), paragraphe 13.3.2, pour une discussion plus approfondie sur ce sujet).

Dans ce qui suit, nous allons présenter un aperçu sur l'utilité de l'analyse des données fonctionnelles dans les applications.

1.2.2 Champs d'application des données fonctionnelles

Depuis plusieurs décennies, nombreux sont les statisticiens qui ont développé des applications permettant le traitement de variables aléatoires fonctionnelles. D'une part, ce traitement permet d'utiliser ou de développer des outils théoriques performants, et d'autre part, il offre un énorme potentiel en terme d'applications (en imagerie, agro-industrie, géologie, économétrie,...). Nous exposons ci-dessous quelques exemples concrets.

Dans le domaine de la linguistique : le problème de la reconnaissance vocale est un sujet d'actualité. L'objectif est de pouvoir retranscrire phonétiquement des mots et des phrases prononcés par un individu. Les données sont des courbes correspondant à des enregistrements de phonèmes prononcés par différents individus. Des travaux ont été, également, réalisés, notamment concernant la reconnaissance vocale. On peut citer par exemple Hastie et al. (1995), Berlinet et al. (2005) ou encore Ferraty et Vieu (2003).

Etude du phénomène d'El Niño : il s'agit d'un jeu de données provenant de l'étude d'un phénomène climatologique assez important. Ce phénomène est couramment appelé El Niño. C'est un grand courant marin qui survient de manière exceptionnelle (en moyenne une à deux fois par décennie) le long des côtes péruviennes à la fin de l'hiver. Ce courant provoque des dérèglements climatiques à l'échelle de la planète. Le jeu de données est constitué de relevés de températures mensuelles de la surface océanique effectués depuis 1950 dans une zone située au large du nord du Pérou (de coordonnées $0-10^\circ$ Sud, $80-90^\circ$ Ouest) dans laquelle peut apparaître le courant marin El Niño. Ces données et leur description sont disponibles sur le site internet du centre de prévision du climat américain : <http://www.cpc.ncep.noaa.gov/data/indices/>. Il faut noter que l'évolution des températures au cours du temps est réellement un phénomène continu. Le nombre de mesures permet

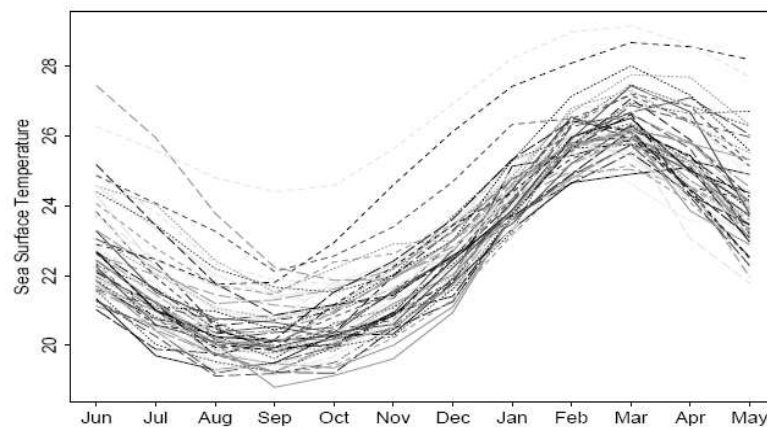


FIGURE 1.1 – Les courbes correspondant au courant d’El Niño

de prendre en considération la nature fonctionnelle des données (cf. Figure 1.1). A partir de ces données, on peut s’intéresser à la prédiction de l’évolution du phénomène à partir des données recueillies lors des années précédentes.

En industrie alimentaire : Ferraty et Vieu (2002, 2003) se sont intéressés à des données spectrométriques de masse. Ces données proviennent d’un problème de contrôle de qualité en industrie alimentaire. Ils ont étudié la teneur en graisse dans les morceaux de viande étant donné les courbes d’absorption de ces morceaux de viande (cf. pour ceci Figure 1.2). Ces données réelles ont été utilisées dans le cas où les variables sont indépendantes.

Consommation d’électricité aux USA : dans le cadre des données dépendantes, on peut considérer l’exemple d’une série chronologique qui concerne la consommation annuelle

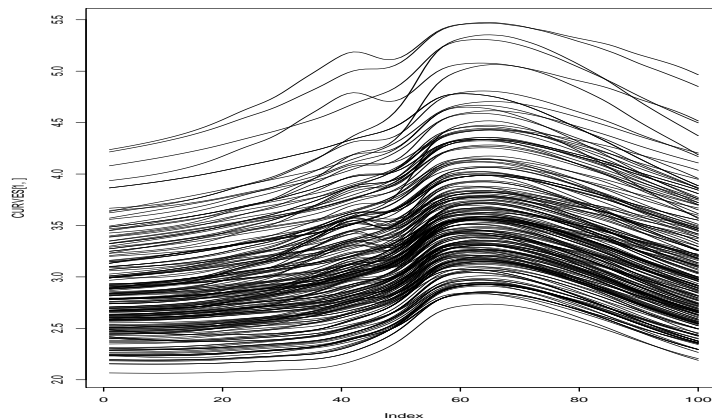


FIGURE 1.2 – Les courbes spectrométriques

d'électricité, aux USA, par des secteurs résidentiels et commerciaux de janvier 1973 jusqu'en février 2001 (338 mois). Le but de cette étude est de prévoir la consommation d'électricité de l'année suivante sachant la consommation d'électricité de toute l'année précédente. L'échantillon se compose de 28 données comme le montre la Figure 1.3. Cette série chronologique peut être regardée comme étant un ensemble de données fonctionnelles dépendantes (c'est-à-dire, une population de 28 courbes : chaque année correspond à 1 courbe).

Données de pollution : Un autre exemple de variables aléatoires fonctionnelles dépendantes portant sur l'étude de phénomènes liés à l'environnement est le problème de pollution. Il s'agit d'étudier la courbe de concentration d'ozone au Pôle Nord sur quatre années successives (de 2000 à 2004). L'objectif est de prévoir la concentration de l'ozone dans une journée étant donné la courbe de concentration de l'ozone de la veille. En procédant par un découpage journalier de la courbe de concentration annuelle de l'ozone, on obtient les courbes représentées dans Figure 1.4. Notons que plusieurs auteurs se sont intéressés aux phénomènes liés à l'environnement, on peut citer entre autres, Damon et Guillas (2002), Aneiros-Perez et al. (2004), Cardot et al. (2004, 2006), Meiring (2005).

Bref, de nombreux autres domaines d'application où l'on peut être confronté à des données de natures fonctionnelles existent et/sinon affluent. Vu l'énormité des exemples que l'on peut citer, nous sommes incapable de présenter dans cette thèse une liste exhaustive de ces applications. Sinon, nous nous contentons, dans la suite de ce paragraphe, d'un rapide tour d'horizon de ces champs d'application.

En biologie : pour l'étude des variations des courbes de croissance (cf. Rao, 1958 et Figure 1.5), et plus récemment, pour l'étude des variations de l'angle du genou durant la marche

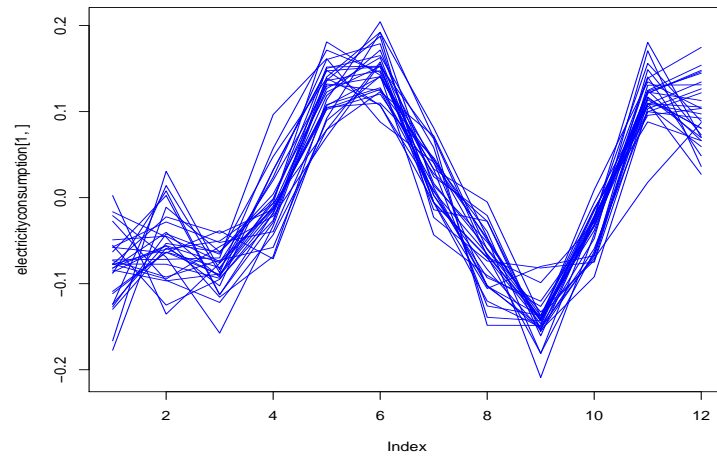


FIGURE 1.3 – Les courbes annuelles de consommation d'électricité aux USA

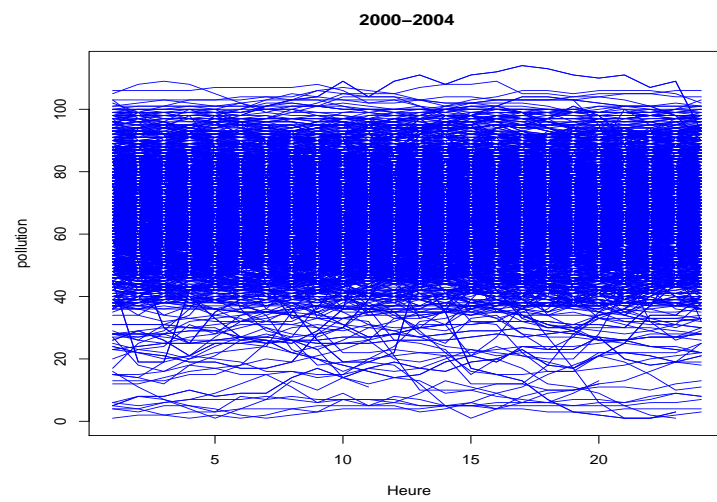


FIGURE 1.4 – Les courbes de pollution au Pôle Nord

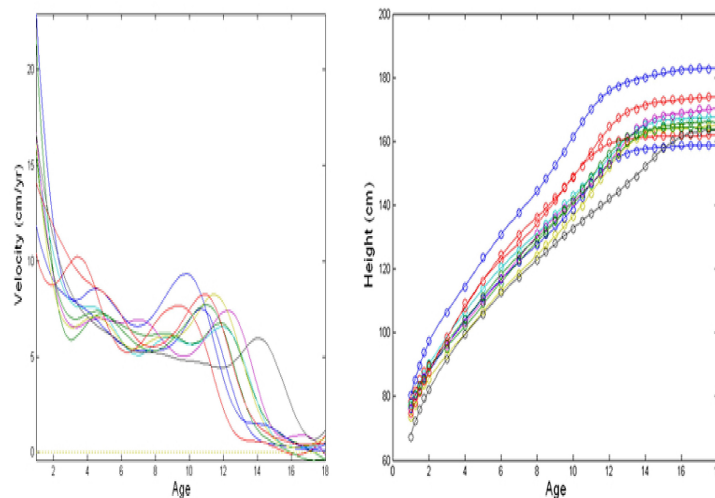


FIGURE 1.5 – Courbes de croissance

(cf. Ramsay et Silverman, 2002). Notons qu'un énorme nombre de données fonctionnelles est produit et ne demande qu'à avoir la méthodologie adéquate pour son traitement, notamment les données spectrométriques de masse (cf. pour le cancer Figure 1.6).

En biologie animal : des études sur la ponte de mouches méditerranéennes ont été effectuées et résumées par des courbes donnant, pour chaque mouche, la quantité d'oeufs pondus en fonction du temps (cf. Figure 1.7).

En économétrie : on est souvent confrontés à de nombreux phénomènes que l'on peut modéliser par des variables fonctionnelles. Parmi ces phénomènes on peut citer la volatilité des marchés financiers (cf. Müller et al., 2007), le rendement d'une entreprise (cf. Kawassaki et Ando, 2004), le commerce électronique (cf. Jank et Shmueli, 2006) ou l'intensité des transactions financières (cf. Laukaitis et Rackauskas, 2002). On peut se référer à Kneip et Utikal (2001), Benko (2006) et Benko et al. (2006) pour des références supplémentaires. Par ailleurs, nous pouvons aussi citer un exemple qui consiste à l'observation des fluctuations d'un indice boursier en fonction du temps : il s'agit typiquement d'une série temporelle qu'on découpe selon des sous-intervalles de l'espace temps (cf. Bosq, 2002).

En graphologie : l'apport des techniques de la statistique fonctionnelle a aussi trouvé une application en graphologie. Parmi les travaux réalisés sur cette problématique on peut citer, à titre d'exemple, ceux de Hastie et al. (1995) et Ramsay (2000). Ce dernier a modélisé la position du stylo (abscisses et ordonnées en fonction du temps) à l'aide d'un système d'équations différentielles de paramètres fonctionnels.

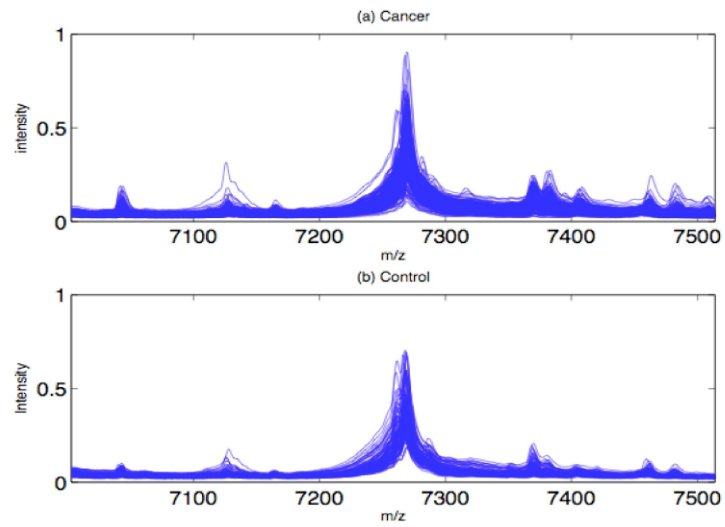


FIGURE 1.6 – Courbes spectrométriques de masse sur des cellules cancéreuses

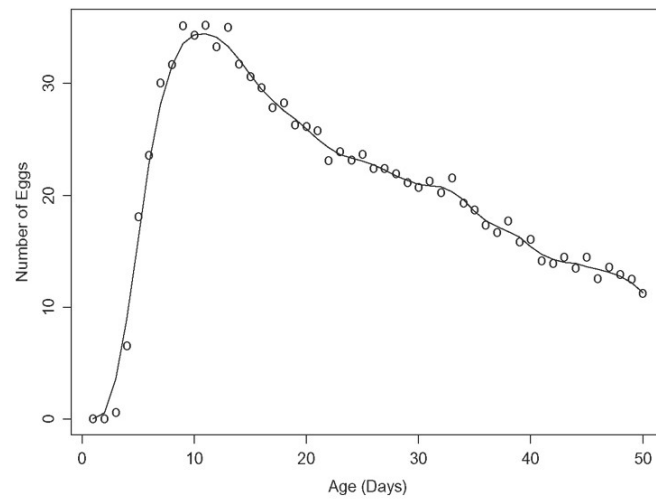


FIGURE 1.7 – Une courbe du nombre d'oeufs journaliers pondus par une mouche

Les mesures et notamment les images recueillies par satellites sont également des données dont l'étude peut être effectuée à partir des méthodologies de la statistique fonctionnelle. On peut citer, par exemple, les travaux de Vidakovic (2001) dans le domaine de la météorologie ou ceux de Dabo-Niang et al. (2004b, 2007) dans le domaine de la géophysique. Dans ces travaux, on s'intéresse à la classification des courbes recueillies par le satellite à différents endroits de l'amazone, ce qui permettrait d'identifier la nature du sol. Enfin, citons Cardot et al. (2003) et Cardot et Sarda (2006) qui ont étudié l'évolution de la végétation à partir de données satellitaires.

1.3 Quelques résultats sur l'estimation non-paramétrique pour des modèles fonctionnels

Nous rappelons, dans ce paragraphe et dans un premier temps quelques hypothèses et notations qui paraissent importantes pour la suite de ce travail de thèse. Ensuite, les résultats obtenus par Ferraty et al. (2006) et brièvement ceux obtenus par Laksaci (2005) et Ezzahrioui (2007) sur l'estimation de quelques paramètres conditionnels.

1.3.1 Notations et hypothèses

Considérons le couple de variables aléatoire (X, Y) où Y est à valeurs dans \mathbb{R} et X est à valeurs dans un espace semi-métrique (\mathcal{F}, d) qui peut être de dimension éventuellement finie. Pour $x \in \mathcal{F}$, la distribution de probabilité de Y sachant X est définie par :

$$\forall y \in \mathbb{R}, F^x(y) = \mathbb{P}(Y \leq y | X = x)$$

ou cette distribution est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R} . Notons par f^x (respectivement $f^{x(j)}$) la densité conditionnelle (respectivement sa dérivée d'ordre j) de Y sachant $X = x$. Par la suite on désignera par x le point fixe de \mathcal{F} , V_x un voisinage de x et $S_{\mathbb{R}}$ un sous-ensemble compact de \mathbb{R} . Notons aussi par : $B(x, h) = \{x' \in \mathcal{F} | d(x', x) < h\}$ la boule de centre x et de rayon h .

Voici quelques hypothèses dont nous avons besoin dans les énoncés des résultats préliminaires.

$$(H1) P(X \in B(x, h)) = \phi_x(h) > 0$$

Pour la fonction de répartition conditionnelle, celle-ci sera supposée vérifier la condition suivante :

$$(H2) \forall (y_1, y_2) \in \mathcal{S} \times \mathcal{S}, \forall (x_1, x_2) \in V_x \times V_x, |F^{x_1}(y_1) - F^{x_2}(y_2)| \leq C_x (d(x_1, x_2)^{b_1} + |y_1 - y_2|^{b_2})$$

et pour certain $j \geq 0$,

Concernant la densité conditionnelle f^x , on la supposera de classe C^j et telle que :

$$(H3) \forall (y_1, y_2) \in \mathcal{S} \times \mathcal{S}, \forall (x_1, x_2) \in V_x \times V_x, |f^{x_1(j)}(y_1) - f^{x_2(j)}(y_2)| \leq C_x (d(x_1, x_2))^{b_1} + |y_1 - y_2|^{b_2}$$

La condition de concentration (H1) joue un rôle important. Ce genre de condition est lié à la semi-métrique d . Elle quantifie et contrôle les probabilités des petites boules.

$$(H4) = \begin{cases} \forall (y_1, y_2) \in \mathbb{R}^2, |H(y_1) - H(y_2)| \leq C|y_1 - y_2| \\ \int_{\mathbb{R}} |t|^{b_2} H^{(1)}(t) dt < +\infty \end{cases}$$

(H5) Le noyau K est à support dans $(0, 1)$, tel que, $0 < C_1 < K(t) < C_2$, où C_1 et C_2 sont deux constantes strictement positives,

$$(H6) \lim_{n \rightarrow \infty} h_K = 0 \text{ et } \lim_{n \rightarrow \infty} \frac{\log n}{n \phi_x(h_K)} = 0,$$

$$(H7)- \lim_{n \rightarrow \infty} h_H = 0 \text{ et } \lim_{n \rightarrow \infty} n^\alpha h_H = \infty, \text{ pour un certain réel } \alpha > 0.$$

où H est un noyau, $h_K = h_{K,n}$ (respectivement, $h_H = h_{H,n}$) est une suite de nombres réels positifs tendant vers 0 quand n tend vers l'infini.

1.3.2 Estimation de la loi conditionnelle

Dans ce paragraphe, nous donnons un résultat de convergence de l'estimateur à noyau de la loi conditionnelle. Étant donné un élément x fixé de \mathcal{F} et soit $(X_i, Y_i)_{i=1, \dots, n}$ un échantillon de couples de variables aléatoires indépendantes à valeurs dans $\mathbb{R} \times \mathcal{F}$, l'estimateur à noyau de la loi conditionnelle $F^x(\cdot)$ est défini par :

$$\hat{F}^x(y) = \frac{\sum_{i=1}^n K\left(\frac{d(x, X_i)}{h_K}\right) H\left(\frac{y - Y_i}{h_H}\right)}{\sum_{i=1}^n K\left(\frac{d(x, X_i)}{h_K}\right)}, \forall y \in \mathbb{R}$$

Le théorème suivant donne la convergence¹ (p.co.) presque complète de l'estimateur $\hat{F}^x(y)$

1. soit $(z_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires. On dit que z_n converge presque complètement (p.co.) vers 0 si, et seulement si, $\forall \epsilon > 0, \sum_{n=1}^{\infty} \mathbb{P}(|z_n| > \epsilon) < \infty$. De plus, soit $(u_n)_{n \in \mathbb{N}^*}$ une suite de nombres réels positifs. On dit que $z_n = O(u_n)$ p.co. si, et seulement si, $\exists \epsilon > 0, \sum_{n=1}^{\infty} \mathbb{P}(|z_n| > \epsilon u_n) < \infty$: ce type de convergence implique la convergence presque sûre et la convergence en probabilité (cf. [13] pour plus de détails).

Théorème 1.3.1. (Ferraty et al. 2006). Sous les hypothèses H1-H6, on a :

$$\sup_{y \in \mathcal{S}} |\mathbb{F}_n^x(y) - F^x(y)| = O\left(h_K^{b_1}\right) + O\left(h_H^{b_2}\right) + O\left(\sqrt{\frac{\log n}{n \phi_x(h_K)}}\right), \text{ p.co.}$$

1.3.3 Estimateur à noyau de la densité conditionnelle

Dans ce paragraphe, nous présentons un estimateur à noyau de la dérivée d'ordre j de la densité conditionnelle et un résultat sur le comportement asymptotique de cet estimateur.

Cet estimateur $\hat{f}^{x(j)}$ de $f^{x(j)}$ est donné par :

$$\hat{f}^{(j)}(y|x) = \frac{h_H^{-j-1} \sum_{i=1}^n K\left(\frac{d(x, X_i)}{h_K}\right) H^{(j+1)}\left(\frac{y - Y_i}{h_H}\right)}{\sum_{i=1}^n K\left(\frac{d(x, X_i)}{h_K}\right)}, \forall y \in \mathbb{R}$$

Notons que, cet estimateur est analogue à celui introduit par Rosenblatt (1969) dans le cas où X est une variable aléatoire réelle. Il est aussi largement étudié depuis ce temps (cf. Youndjé, 1996). Afin d'établir quelques résultats de convergence, les hypothèses suivantes seront nécessaires :

$$(H8) \left\{ \begin{array}{l} \forall (y_1, y_2) \in \mathbb{R}^2, |H^{(j+1)}(y_1) - H^{(j+1)}(y_2)| \leq C|y_1 - y_2| \\ \exists \nu > 0, \forall j' \leq j + 1, \lim_{y \rightarrow \infty} |y|^{1+\nu} |H^{(j+1)}(y)| = 0 \\ H^{(j+1)} \text{ est borné} \end{array} \right.$$

$$(H9) \lim_{n \rightarrow \infty} h_K = 0 \text{ avec } \lim_{n \rightarrow \infty} \frac{\log n}{n h_H^{2j+1} \phi_x(h_K)} = 0.$$

Le théorème suivant concerne le comportement asymptotique de l'estimateur fonctionnel à noyau $\hat{f}^{x(j)}$.

Théorème 1.3.2. (Ferraty et al., 2006). Sous les hypothèses H1, H3, H4 et H6-H9, on a :

$$\sup_{y \in \mathcal{S}} |f_n^{x(j)}(y) - f^{x(j)}(y)| = O\left(h_K^{b_1}\right) + O\left(h_H^{b_2}\right) + O\left(\sqrt{\frac{\log n}{n h_H^{2j+1} \phi_x(h_K)}}\right), p.co.$$

où \mathcal{S} une sous-ensemble compact de R

1.3.4 Estimation du mode conditionnel

Cas où les données sont i.i.d.

Ce paragraphe présente un estimateur du mode conditionnel noté par $\hat{\theta}$. Notons que, l'ensemble compact \mathcal{S} est choisi de telle sorte qu'il n'y ait qu'un unique mode θ . Cet estimateur est basé sur la précédente estimation fonctionnelle de la densité conditionnelle.

Dans la suite de ce paragraphe, on utilise $\mathcal{S} = [\theta - \xi, \theta + \xi]$ comme ensemble compact. L'estimateur $\hat{\theta}$ du mode conditionnel θ est définie comme.

$$\hat{f}^x(\hat{\theta}) = \sup_{y \in \mathcal{S}} \hat{f}^x(y)$$

Notons que, l'estimateur $\hat{\theta}$ n'est pas nécessairement unique, pour assuré cette unicité et la convergence de $\hat{\theta}_n$, on suppose :

(H10) $\exists \xi > 0$, $f^x \nearrow$ dans $[\theta - \xi, \theta]$ et $f^x \searrow$ dans $[\theta, \theta + \xi]$.

(H11) f^x est j -fois continûment différentiable par rapport à y sur $[\theta - \xi, \theta + \xi]$,

et

$$(H12) \begin{cases} f^{x(l)}(\theta) = 0, si \ 1 \leq l < j \\ |f^{x(j)}(\theta)| > 0 \text{ sinon} \end{cases}$$

Signalons que ces conditions ont une grande influence sur la vitesse de convergence de l'estimateur $\hat{\theta}$ (cf. le théorème ci-dessous). De plus la convergence de cet estimateur peut être obtenue par l'hypothèse (H10) (cf. Laksaci (2005), Lemme 2.4.1).

Théorème 1.3.3. (Laksaci , 2005). Si les hypothèses du Théorème 1.3.2 et H10-H12 sont vérifiées, alors :

$$\hat{\theta} - \theta = O\left(h_K^{\frac{b_1}{j}}\right) + O\left(h_H^{\frac{b_2}{j}}\right) + O\left(\frac{\log n}{n h_H \phi_x(h_K)}\right)^{\frac{1}{2j}}, \text{ p.co.}$$

Cas où les données sont α -mélangeantes

Les résultats obtenus dans le cas des variables aléatoires fonctionnelles indépendantes et identiquement distribuées (i.i.d.) ont été prolongés au cas des variables fortement mélangeantes. Un résultat (cf. Théorème 1.3.4) s'annonce dans ce cadre grâce à des hypothèses faites dans le cas i.i.d. Ces hypothèses ont été renforcées par des conditions de concentration de la loi conjointe des couples (X, Y) et quelques hypothèses sur les coefficients de mélange.

Deux exemples d'application sont étudiés. Le premier correspond au cas i.i.d. Il concerne l'industrie agro-alimentaire (courbes spectrométriques de masse). L'autre exemple correspond au cas dépendant. Celui-ci concerne un problème de pollution (les courbes de la concentration de l'ozone sur le pôle nord) (cf. Laksaci , 2005 pour plus de détails).

Les hypothèses suivantes sont nécessaires dans l'énoncé du Théorème 1.3.4 :

$$(H13) \sup_{i \neq j} P((X_i, X_j) \in B(x, r) \times B(x, r)) = \phi_x(r) \psi_x(r) > 0,$$

(H14) Les coefficients de α -mélange de la suite (X_i, Y_i) vérifient la condition :

$$\exists a > (5 + \sqrt{17})/2, \quad \exists c > 0 \text{ tel que } \forall n, \alpha_n \leq cn^{-a}$$

,

$$(H15) \lim_{n \rightarrow \infty} h_H = 0 \text{ et } \exists \beta_1 \leq \frac{4}{(a+1)(a-2)} \text{ tel que } \lim_{n \rightarrow \infty} n^{\beta_1} h_H = \infty,$$

$$(H16) \left\{ \begin{array}{l} \lim_{n \rightarrow \infty} h_K = 0, \quad \lim_{n \rightarrow \infty} \frac{\log n}{n h_H \chi_x(h_K)} = 0. \\ \text{et} \\ \exists \beta_2 > 0, c_1 > 0, c_2 > 0, c_2 n^{(\frac{3-a}{a+1}) + \beta_2} \leq \chi_x(h_K) \leq c_1 n^{\frac{1}{1-a}}. \end{array} \right.$$

où

$\chi_x(h_K)$ désigne le maximum de la concentration entre la loi marginale et les lois conjointes de chaque couple d'observations fonctionnelles dans la boules de centre x et de rayon h_K .

Théorème 1.3.4. (*Laksaci, 2005*). *Si les hypothèses (H1), (H3)-(H5) et (H10)-(H16) sont vérifiées, alors :*

$$\hat{\theta} - \theta = O\left(h_K^{\frac{b_1}{j}}\right) + O\left(h_H^{\frac{b_2}{j}}\right) + O\left(\left(\frac{\log n}{n h_H \chi_x(h_K)}\right)^{\frac{1}{2j}}\right), p.co.$$

où b_1 et b_2 deux réels strictement positifs.

Chapitre 2

Kernel conditional density estimation when the regressor is valued in a semi-metric space

Ali Laksaci¹, Fethi Madani² and Mustapha Rachdi^{2,3}

(À paraître dans Communications Statistics- Theory and Methods, 2012)

Abstract. This paper deals with the conditional density estimation when the explanatory variable is functional. In fact, nonparametric kernel type estimator of the conditional density has been recently introduced when the regressor is valued in a semi-metric space. This estimator depends on a smoothing parameter which controls its behavior. Thus, we aim to construct and study the asymptotic properties of a data-driven criterion for choosing automatically and optimally this smoothing parameter. This criterion can be formulated in terms of a functional version of cross-validation ideas. Under mild assumptions on the unknown conditional density, it is proved that this rule is asymptotically optimal. Finally, a simulation study and an application on real data are carried out to illustrate, for finite samples, the behavior of our method. Finally, mention our results can also be considered as novel in the finite dimensional setting and several other open questions are raised in this article.

Keywords. Cross-validation, functional data, kernel estimator, nonparametric model, bandwidth selection, small balls probability

1. Université Djillali Liabès, BP. 89, Sidi Bel-Abbès 22000, Algeria. E-mail : alilak@yahoo.fr

2. Laboratoire AGIM FRE 3405 CNRS, Equipe TIMB, Université P. Mendès France (Grenoble 2), UFR SHS, BP. 47, 38040 Grenoble Cedex 09, France. E-mails : Mustapha.Rachdi@upmf-grenoble.fr and Fethi.Madani@imag.fr

3. Corresponding author

AMS Subject Classification. Primary : 62G05, Secondary : 62G07, 62G08, 62G35, 62G20.

2.1 Introduction

Conditional density estimation is a statistical technique that allows for a better understanding of the relationship between a response variable and a set of covariates, in comparison with usual regression methods. Therefore, this technique is of great importance in many scientific fields where knowledge about conditional means, obtained by regression methods, is not enough to draw valuable conclusions about the problem at hand. Moreover, conditional density functions arise in a variety of areas. One of the more useful applications involves density forecasting, where the probability density of the forecast of a time series, such as the rate of inflation, can be used to make probability statements regarding the future course of that series. However, the probability density, and its resulting interpretation, is conditional on the hypothesis that the model used to produce the forecasts is correctly specified.

Recall that, if $g(x, y)$ denotes the joint density of (X, Y) and $h(x)$ denotes the marginal density of X , then the conditional density of Y given $X = x$ is obtained by $f(x, y) = g(x, y)/h(x)$. The standard nonparametric regression does not allow the analysis of changes in modality, and standard density estimation does not allow conditioning on an explanatory variable. Notice also that conditional density estimation is, in some ways, a generalization of both nonparametric regression and standard univariate density estimation. The kernel conditional density estimation was first considered by Rosenblatt (1969) who studied the problem of estimating the density of Y given $X = x$ where X is an univariate random variable.

On the other hand, estimators of the conditional mode, the conditional distribution and the conditional median can be derived directly from estimators of $f(x, y)$. For instance in Collomb et al. (1987) it is shown how one can get an estimator of the conditional mode and how such an estimator can be used for forecasting problems (cf. to cite a few, Härdle (1990), Gannoun (1990), Youndjé (1993 and 1996) and the references therein). Moreover, It is important to mention that estimators of conditional modes are of particular interest for prediction (cf. Collomb et al. (1987) and Ferraty et al. (2005)).

Furthermore, the problem of the conditional density estimation appears to have lain free of scrutiny until it was revisited and some improved estimators were proposed (cf. Hyndman et al. (1996), and references therein for some developments). Indeed, the following modified form of Rosenblatt's estimator was considered :

$$\widehat{f}_{(a,b)}(x, y) = \frac{b^{-1} \sum_{j=1}^n K(a^{-1} \|x - X_j\|_x) K(b^{-1} \|y - Y_j\|_y)}{\sum_{j=1}^n K(a^{-1} \|x - X_j\|_x)} \quad (1)$$

where $(X_1, Y_1), \dots, (X_n, Y_n)$ is a sample of independent observations from the distribution of (X, Y) and $\|\cdot\|_x$ and $\|\cdot\|_y$ are metrics on the spaces values of X and Y , respectively.

The kernel function, $K(u)$, is assumed satisfying some specific conditions. Popular choices of $K(u)$ are defined in terms of univariate and unimodal probability density functions. Moreover, Youndjé (1993 and 1996), Hyndman et al. (1996) and others give the bias, variance, mean squared error (MSE) and convergence properties of the estimator (1) and proposed also an alternative kernel estimator with smaller MSE than the standard estimator in some commonly occurring situations. On the other, we can not continue our introduction without mentioning the work by Fan et al. (1996), who proposed an alternative conditional density estimator by generalizing Rosenblatt's estimator using local polynomial techniques. Then, Hyndman and Yao (1998) introduced two further local parametric estimators which improve on the estimators given by Fan et al. (1996). Stone (1994), meanwhile, followed a different path by using tensor products of polynomial splines to obtain conditional log density estimators. For other studies on the nonparametric estimation of the conditional density we refer also to Gannoun (1990), Youndjé (1993 and 1996), Hall et al. (1999), Härdle et al. (1991), Bashtannyk and Hyndman (2001), Gannoun et al. (2003), El Ghouch and Genton (2009) and the references therein.

In this paper, we are interested in the efficient estimation of the conditional probability density when the explanatory variables are of functional type. It should be noticed that, these questions in the infinite dimensional framework are particularly interesting, at once for the fundamental problems they formulate, but also for many applications they may allow (cf. Bosq (2000), Ramsay and Silverman (2005), Ferraty and Vieu (2006) and references therein). In fact, in this conditional context, the first results were obtained by Ferraty and Vieu (2005) and Ferraty et al. (2006). They established the almost-complete consistency, in both cases i.i.d. and strongly mixing data, of the kernel estimators of the conditional distribution function and of the conditional probability density. Moreover, they presented some applications of their results on both the conditional mode and on conditional quantiles. Among the lot of papers which are concerned with the nonparametric modelization related to the conditional distribution of a real variable given a random variable taking values in infinite dimensional spaces, we refer only to Dabo-Niang and Laksaci (2007) for the conditional mode estimation, and to Laksaci (2007) for the asymptotic expression of leading terms in the quadratic error of conditional density kernel estimators.

On the other hand, it is well known that kernel estimators have some nice asymptotic properties when the curse of dimensionality is controlled by means of suitable considerations on the small ball probabilities of the functional variable (cf. Ferraty and Vieu 2006 and references therein). However it is also well-known that, as in the standard finite dimensional framework, the smoothing parameter has to be selected suitably for insuring good practical performances (cf. Laksaci, 2007). Notice that, some papers, (cf. for instance, Youndjé et al., 1993), have treated the problem of the smoothing parameter selection in the nonparametric estimation of the conditional density, by using some techniques quite different from ours, but only in the finite dimensional setup. Furthermore, the selection of the smoothing parameter in the infinite dimensional setting is much more complicated. In particular, the so-called *scatterplot* which is a graphical tool for exploring the relationship between the explanatory variables and the scalar response is not available, and hence it becomes very hard to have some informations on the shape of the relationship between the functional variable and the

scalar response. Therefore, various areas with different (low/high) concentrations can appear in such a relationship even though it does not appear in the functional data sample (cf. for instance, the simulated curves in Section 2.4.2). It is also clear, in the infinite dimensional setup, that the concentration of the distribution of the functional explanatory variable will have an influence on the value of some appropriate bandwidth (the variance of the estimator increases when the concentration of the distribution of the functional covariates decreases which is the case when the bandwidth value's decreases (cf. conditions (17) and (14)). Moreover, in areas where the functional covariates have low concentration, the bandwidth has to be taken sufficiently large to include enough data curves, while a smaller bandwidth can be used in areas where the functional covariates have high concentration. It should, thus, be noted that Rachdi and Vieu (2007) (respectively Benhenni et al., 2007) proposed a global (respectively a local adaptive) cross-validation procedure for the regression operator estimation for functional data, which has inspired this work.

The main aim of this paper is then the construction of both global and local functional cross-validation procedures. We remark that a local bandwidth choice can significantly improve the precision of the prediction in the functional setting than the global one. In section 2, the data-driven methods are defined. The main hypotheses and results are enounced in section 3. In section 4, we propose a simulation study showing how an optimal local bandwidth choice improves the usual global selection rule for some irregular functional covariates. Finally, asymptotic theoretical support is given in section 5, and the proofs of the auxiliary results are relegated to the Appendix.

2.2 Global and local bandwidth selection rules

Let us introduce a sample of independent pairs $(X_i, Y_i)_{1 \leq i \leq n}$ identically distributed as (X, Y) which is valued in $\mathcal{F} \times \mathbb{R}$, where (\mathcal{F}, d) is a semi-metric space equipped with a semi-metric d . Assume that there exists a regular version of the conditional probability of Y given X , which is absolutely continuous with respect to the Lebesgue measure on the real line \mathbb{R} . Let $f(x, \cdot)$ denote the conditional probability density of the random variable Y given $X = x \in \mathcal{F}$, which we have to estimate. For this aim, we define the kernel estimator $\hat{f}_{(a,b)}$ of f as in (1), but by considering two different kernel functions as follows :

$$\forall x \in \mathcal{F} \text{ and } \forall y \in \mathbb{R}, \quad \hat{f}_{(a,b)}(x, y) = \frac{b^{-1} \sum_{i=1}^n K(a^{-1}d(x, X_i))H(b^{-1}(y - Y_i))}{\sum_{i=1}^n K(a^{-1}d(x, X_i))} \quad (2)$$

where K is a kernel and $a = a_{K,n}$ (respectively $b = b_{H,n}$) is a sequence of positive real numbers. Notice that the estimator (2) has been used by Roussas (1968) in the real case and by Ferraty et al. (2006) in the functional case.

The main goal of this paper is to construct and study the asymptotic behavior of a data driven method which optimally selects the smoothing parameters (a, b) . To do that, we propose to use a rule which is based on the classical leave-out-one-curve cross-validation procedure and to study its asymptotic behavior in the mean squared sense. Indeed, commonly with the

majority of the earlier works on the bandwidth selection, our rule is based on the minimization of the integrated squared error which is weighted by the probability measure, $dP_X(x)$, of the functional variable X and some nonnegative weight functions W_1 and W_2 :

$$d_1(\hat{f}_{(a,b)}, f) = \int \int \left(\hat{f}_{(a,b)}(x, y) - f(x, y) \right)^2 W_1(x)W_2(y) dP_X(x) dy \quad (3)$$

A discrete approximation of (3) is the averaged squared error given by :

$$d_2(\hat{f}_{(a,b)}, f) = \frac{1}{n} \sum_{i=1}^n \left(\hat{f}_{(a,b)}(X_i, Y_i) - f(X_i, Y_i) \right)^2 \frac{W_1(X_i)W_2(Y_i)}{f(X_i, Y_i)} \quad (4)$$

or, also, the mean integrated squared error :

$$d_3(\hat{f}_{(a,b)}, f) = \int \int \mathbb{E} \left(\hat{f}_{(a,b)}(x, y) - f(x, y) \right)^2 W_1(x)W_2(y) dP_X(x) dy \quad (5)$$

However, these loss functions depend on the conditional density f , so the smoothing parameter that minimizes these errors is not computable in practice. Thus, we must find another loss function which is asymptotically equivalent to the quadratic distances (3), (4) and (5). Following the same ideas as in Youndjé (1996) for the real case, we can write that :

$$d_1(\hat{f}_{(a,b)}, f) = A + B - 2C$$

where

$$\begin{aligned} A &= \int \int \hat{f}_{(a,b)}^2(x, y) W_1(x)W_2(y) dP_X(x) dy \\ B &= \int \int f^2(x, y) W_1(x)W_2(y) dP_X(x) dy \\ C &= \int \int \hat{f}_{(a,b)}(x, y) f(x, y) W_1(x)W_2(y) dP_X(x) dy \end{aligned}$$

Since the second term B is independent of (a, b) , the problem of minimizing d_1 is equivalent to that of minimizing $A - 2C$. A straightforward way to construct a computational procedure to select the optimal bandwidths (a, b) with respect to the error measure d_1 is to estimator the both quantities A and C . For this aim, as mentioned above, we adopt the standard leave-out-one-curve technique as in Rudemo (1982) for the probability density estimation and Rachdi and Vieu (2007) for the regression operator estimation, by considering the following criteria :

$$GCV(a, b) = \frac{1}{n} \sum_{i=1}^n W_1(X_i) \int \hat{f}_{(a,b)}^{-i^2}(X_i, y) W_2(y) dy - \frac{2}{n} \sum_{i=1}^n \hat{f}_{(a,b)}^{-i}(X_i, Y_i) W_1(X_i) W_2(Y_i) \quad (6)$$

and respectively, for a fixed $y \in \mathbb{R}$:

$$LCV_{x,y}(a, b) = \frac{1}{n} \sum_{i=1}^n W_{1,x}(X_i) \int \hat{f}_{(a,b)}^{-i^2}(X_i, z) W_{2,y}(z) dz - \frac{2}{n} \sum_{i=1}^n \hat{f}_{(a,b)}^{-i}(X_i, Y_i) W_{1,x}(X_i) W_{2,y}(Y_i) \quad (7)$$

where $W_{2,x}$ (respectively $W_{2,y}$) is some positive local weight function around x (respectively y), and for any $i = 1, \dots, n$:

$$\widehat{f}_{(a,b)}^{-i}(x, y) = \frac{b^{-1} \sum_{j \neq i}^n K(a^{-1}d(x, X_j))H(b^{-1}(y - Y_j))}{\sum_{j \neq i}^n K(a^{-1}d(x, X_j))}. \quad (8)$$

These criteria are obtained by using the fact that

$$\begin{aligned} C &= \int \int \widehat{f}_{(a,b)}(x, y) f(x, y) W_1(x) W_2(y) dP_X(x) dy \\ &= \int \int \widehat{f}_{(a,b)}(x, y) W_1(x) W_2(y) dP_{Y|X=x}(y) dP_X(x) \\ &= \int \int \widehat{f}_{(a,b)}(x, y) W_1(x) W_2(y) dP_{(X,Y)}(x, y) \\ &= \mathbb{E}_{(X,Y)} \left(\widehat{f}_{(a,b)}(X, Y) W_1(X) W_2(Y) \right) \end{aligned}$$

and

$$A = \mathbb{E}_X \left(\int \widehat{f}_{(a,b)}^2(X, y) W_1(X) W_2(y) dy \right)$$

where \mathbb{E}_Z denotes the expectation with respect to the distribution of the random variable Z .

Finally, our global (respectively, local) cross-validation procedure consists in choosing the bandwidths (a, b) which minimize $GCV(a, b)$ (respectively, $LCV_{x,y}(a, b)$) on a given set $H_n \subset \mathbb{R}^{+2}$ (respectively, $H_n(x, y) \subset \mathbb{R}^{+2}$).

2.3 Main Results

2.3.1 Assumptions

In order to deduce the asymptotic optimality of the bandwidth selected by the rule GCV (respectively, $LCV_{x,y}$), we will assume that the weight function W_1 (respectively W_2) is bounded with support in some subset S_X of \mathcal{F} (respectively on a compact subset S_Y of \mathbb{R}) and the conditional density $f(\cdot, \cdot)$ is bounded on $S_X \times S_Y$. In the sequel of this paper, when no confusion is possible, we will denote by C and C' some strictly positive generic constants and we will make the following assumptions :

The weight functions are taken, for each curve x , such that for some positive real w :

$$w = a^\nu \text{ for } 0 < \nu < 1 \text{ and } W_{1,x} \text{ is bounded and supported in } \mathcal{B}(x, w) \quad (9)$$

where $\mathcal{B}(x, h)$ denotes the closed ball with center x and radius the real $h > 0$,

$$\forall x \in S_X, 0 < C\phi(h) \leq \mathbb{P}(X \in \mathcal{B}(x, h)) \leq C'\phi(h) \quad (10)$$

where $\phi(h)$ is a positive real function such that $\lim_{h \rightarrow 0} \phi(h) = 0$.

There exist some strictly positive constants b_1, b_2 and δ , such that : $\forall (x_0, y_0) \in S_X \times S_Y$, $\forall (x_1, x_2) \in S_X \times S_X$ and $\forall (y_1, y_2) \in S_Y \times S_Y$, we have :

$$f(x_0, y_0) > \delta \quad \text{and} \quad |f(x_1, y_1) - f(x_2, y_2)| \leq C \left(d^{b_1}(x_1, x_2) + |y_1 - y_2|^{b_2} \right) \quad (11)$$

The kernel K is a bounded and Lipschitzian kernel on its support $(0, 1)$, and there exist some positive constants C and C' such that :

$$0 < C < K(t) < C' < \infty \quad (12)$$

and if $K(1) = 0$, the kernel K has to fulfill the additional condition $-\infty < C < K'(t) < C' < 0$, where K' is the first derivative of K .

The kernel H is bounded and a Lipschitzian continuous function, such that :

$$\int |t|^{b_2} H(t) dt < \infty \quad \text{and} \quad \int H^2(t) dt < \infty \quad (13)$$

The function ϕ is such that :

$$\left\{ \begin{array}{l} \exists C > 0, \exists \eta_0 > 0, \forall \eta < \eta_0, \phi'(\eta) < C \\ \text{and if } K(1) = 0, \text{ the function } \phi(\cdot) \text{ has to fulfill the additional condition :} \\ \exists C > 0, \exists \eta_0 > 0, \forall 0 < \eta < \eta_0, \int_0^\eta \phi(u) du > C \eta \phi(\eta) \end{array} \right. \quad (14)$$

For n large enough, the Kolmogorov's ϵ -entropy of S_X denoted by ψ_{S_X} (cf. for instance, Kolmogorov and Tikhomirov (1959) and Theodoros and Yannis (1997)) satisfies, for some $\gamma \in (0, 1)$:

$$\sum_{n=1}^{\infty} n^{(3\gamma+1)/2} \exp \left\{ (1-\beta) \psi_{S_X} \left(\frac{\log n}{n} \right) \right\} < \infty \text{ for some } \beta > 1 \quad (15)$$

and for all $(a, b) \in H_n$ we have :

$$\lim_{n \rightarrow +\infty} n^\gamma b = \infty \quad \text{and} \quad \phi(a) \geq C n^{-\delta} \text{ for some } \delta \in (0, 2 - 2\gamma) \quad (16)$$

2.3.2 Some interpretations and examples on our hypotheses

It is worth observing that these conditions are not very restrictive. The hypotheses (10)-(14) are very standard in the functional nonparametric setting. More precisely :

- The hypothesis (10) is a simple uniformization of the concentration property of the probability measure on the small balls. This assumption is satisfied for a large family of random functional variables. Indeed, in many examples, the small ball probability function $\mathbb{P}(X \in B(x, h))$ can be written approximatively as the product of two independent functions $g(x)$ and $\phi(h)$, as in the following examples, which can be found in Ferraty et al. (2007) :

- (i) $\mathbb{P}(X \in B(x, h)) = g(x)h^\gamma$ for some $\gamma > 0$
- (ii) $\mathbb{P}(X \in B(x, h)) = g(x)h^\gamma \exp\left(-\frac{C}{h^p}\right)$ for some $\gamma > 0$ and $p > 0$
- (iii) $\mathbb{P}(X \in B(x, h)) = \frac{g(x)}{|\log(h)|}$

Thus, condition (10) is automatically verified if the function g satisfies :

$$0 < C < \inf_{x \in S_X} g(x) \leq \sup_{x \in S_X} g(x) < C' < \infty.$$

Moreover, the function g can be specified for several well known continuous-time processes, by using the Onsager-Machlup function (cf. Ferraty et al., 2010). For instance, it is shown in Corollary 4.7.8 in Bogachev (1999, page 186), that the expression of the Onsager-Machlup function of the couple (x, z) , for the Gaussian measures on a semi-normed space $(\mathcal{F}, \|\cdot\|)$, is given by :

$$F(x, z) = \log \left(\lim_{h \rightarrow 0} \frac{\mathbb{P}(X \in B(x, h))}{\mathbb{P}(X \in B(z, h))} \right) = \frac{1}{2} \|\pi(z)\|_H^2 - \frac{1}{2} \|\pi(x)\|_H^2$$

where $\|\cdot\|_H$ is the Hilbert norm on the Cameron-Martin space of \mathcal{F} associated to a Gaussian measure, denoted by H , and $\pi(\cdot)$ is the orthogonal projection onto the orthogonal complement of the set $\{a \in H, \text{ such that } \|a\| = 0\}$. So, in this case $g(x) = \exp(-\frac{1}{2}\|\pi(x)\|_H^2)$, therefore, condition (10) is verified for subsets such as :

$$S_X = \{x \in \mathcal{F}, \text{ such that } \|\pi(x)\|_H \leq r\}.$$

- The assumption (14) summarized in the boundless of the derivative of ϕ around zero, is not very restrictive. Indeed, by some simple analytic arguments we can show that this condition is fulfilled for all the usual cases such as the exponential-type processes, the fractal-type processes, and also in the finite dimensional case. On the other hand, for a more flexibility in the choice of the kernel, we add an additional condition, specifically on the continuity of the kernel at 1. This additional condition acting on the behavior of ϕ around zero was introduced by Ferraty and Vieu (2006). It is obviously satisfied for fractal-type processes and it is always possible, if \mathcal{F} is a separable Hilbert space, to choose a semi-metric d for which this condition is fulfilled (cf. Lemma 13.6, in Ferraty and Vieu (2006), page 213)
- Assumption (15) is verified for a large class of functional spaces. We quote the following examples (cf. Ferraty et al., 2010) :
 1. The unit ball of the Cameron-Martin space associated to a Gaussian process viewed as a map in $\mathcal{C}(0, 1)$ with the spectral measure μ satisfying :

$$\int \exp(\delta|\lambda|)\mu(d\lambda) < \infty, \text{ for } \delta > 0$$

for which :

$$\psi_{S_X} \left(\frac{\log n}{n} \right) = O((\log n)^2)$$

2. The unit ball of the Cameron-Martin space associated to the standard stationary Ornstein-Uhlenbeck process viewed as a map in the Sobolev space $W_2^1(0, 1)$ with the covariance operator :

$$C(s, t) = \exp(-a|s - t|), \text{ for } a > 0$$

For this subset, we have :

$$\psi_{S_X} \left(\frac{\log n}{n} \right) = O(\log n)$$

3. The closed ball $\mathcal{B}(0, r)$ in the Sobolev space defined by the class of functions $x(t)$ on $T = [0, 2\pi)$, such that :

$$\frac{1}{2\pi} \int_0^{2\pi} x^2(t) dt + \frac{1}{2\pi} \int_0^{2\pi} x^{(m)2}(t) dt \leq r$$

where $x^{(m)}(\cdot)$ denotes the m th derivative of x . In this case :

$$\psi_{S_X} \left(\frac{\log n}{n} \right) = O(n^{1/m})$$

4. The compact subsets in the finite dimensional spaces, or in the projection semi-metric in Hilbert spaces where :

$$\psi_{S_X} \left(\frac{\log n}{n} \right) = O(\log n)$$

- Notice that, the inequality (H5bⁿ) in Ferraty et al. (2010) is not necessary here because such assumption is used to precise the convergence rate of the uniform consistency which is not necessary. In other words, the uniform consistency of the kernel estimator of the conditional density (without any precision on the convergence rate) is sufficient to show our results.
- Conditions (9) and (16) are equivalent to those used by Rachdi and Vieu (2007) and Benhenni et al. (2007) for the global and local cross-validation procedures in the operatorial regression estimation. In fact, these hypotheses are the functional versions of those used by Härdle and Marron (1985) and Youndjé (1996) in the usual real case. The condition (9) on the weight function is similar to that in Vieu (1991), and allows to give more importance to observations around the curve x .

2.3.3 Two theorems on global and local criteria

Theorem 2.3.1. *Under hypotheses (10)-(16), if the set H_n of bandwidths (a, b) is finite with :*

$$\#(H_n) = O(n^\alpha) \text{ for some } \alpha > 0, \text{ where } \# \text{ denotes the cardinality} \quad (17)$$

then, we have for $k = 1, 2, 3$, that :

$$\frac{d_k(\widehat{f}_{(a_1, b_1)}, f)}{d_k(\widehat{f}_{(a_0, b_0)}, f)} \rightarrow 1 \quad \text{almost surely (a.s.), as } n \rightarrow +\infty \quad (18)$$

where

$$(a_0, b_0) = (a_{0_{K,n}}, b_{0_{H,n}}) = \arg \inf_{(a,b) \in H_n} d_k(\widehat{f}_{(a,b)}, f)$$

$$\text{and } (a_1, b_1) = (a_{1_{K,n}}, b_{1_{H,n}}) = \arg \inf_{(a,b) \in H_n} GCV(a, b)$$

On the local framework, we suppose that (15) is verified for $S_X = \mathcal{B}(x, w)$ and we deduce the same optimality results, for the local criterion.

Theorem 2.3.2. *Under hypotheses (9)-(16), if the set $H_n(x, y)$ of bandwidths (a, b) is finite with :*

$$\#(H_n(x, y)) = O(n^{\alpha(x,y)}) \text{ for some } \alpha(x, y) > 0, \quad (19)$$

then, we have for $k = 1, 2, 3$, that :

$$\frac{d_k(\widehat{f}_{(a_1, b_1)}, f)}{d_k(\widehat{f}_{(a_0, b_0)}, f)} \rightarrow 1, \quad \text{a.s., as } n \rightarrow +\infty \quad (20)$$

where

$$(a_0, b_0) = (a_{0_{K,n}}, b_{0_{H,n}}) = \arg \inf_{(a,b) \in H_n(x,y)} d_k(\widehat{f}_{(a,b)}, f)$$

$$\text{and } (a_1, b_1) = (a_{1_{K,n}}, b_{1_{H,n}}) = \arg \inf_{(a,b) \in H_n(x,y)} LCV_{x,y}(a, b)$$

2.4 Discussion and applications

2.4.1 On the applicability of the method

It is well known that, the estimation of the conditional probability density is an important tool permitting the analysis of the input-output relation in nonparametric statistics. Such nonparametric model provides a broader range of relevant information on the covariation between two random variables. Moreover, if a conditional density estimator is available, it is easy to make the prediction via the conditional mode estimator, to derive prediction intervals or to determine the probabilities of extreme values. So, the optimality of all these statistical studies is closely linked to the construction of the optimal estimator of the conditional density. In order to emphasize the practical aspects of our study, we discuss in the rest of this section the applicability of our bandwidth selection approach on some nonparametric models, frequently used in practice, for which this question of the bandwidth selection is inherent to derive their best properties.

The conditional mode estimation : often, the prediction of the values of the response variable knowing an explained one is obtained by estimating the conditional expectation. However, the latter may not be sufficiently informative, when the conditional distribution possesses multi-modality or a highly skewed profile with heteroscedastic noise. In such cases, it would be more informative to estimator the conditional distribution itself. A pertinent predictor can be obtained by estimating the conditional mode or the conditional median. In this section, we focus on the problem of the bandwidth selection in the conditional mode estimation by the kernel method. Firstly, we recall that a natural kernel estimator $\hat{\theta}$ of the conditional mode θ is derived from the conditional density estimator as follows :

$$\hat{\theta}(a, b, x) = \arg \sup_{y \in \mathbb{R}} \hat{f}_{(a,b)}(x, y). \quad (21)$$

where $\hat{f}_{(a,b)}(x, y)$ is given in (2). Clearly, the behaviour of the conditional mode estimator depends on the choice of the two smoothing parameters a and b . In this prediction context, a naive L^2 -criterion is given by :

$$(a_{opt}, b_{opt}) = \arg \min_{a,b} \left\{ \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{\theta}^{-i}(a, b, X_i))^2 \right\}$$

where

$$\hat{\theta}^{-i}(a, b, X_i) = \arg \sup_{y \in \mathbb{R}} \hat{f}_{(a,b)}^{-i}(X_i, y)$$

with $f_{(a,b)}^{-i}(X_i, y)$ is the leave-one-out-curve estimator defined by (8). This selection method has been used by De Gooijer and Gannoun (2000) in the multivariate case and by Ferraty and Vieu (2006) in the functional setting. Although this selection method is very adequate in several practical situations but, to the best of our knowledge, their asymptotic optimality has not been addressed so far. Moreover, this selection procedure has serious problems if the estimator $\hat{\theta}$ is not unique. A reasonable way to overcome this problem is to use our bandwidth selection procedure by computing the conditional mode estimator as follows :

$$\hat{\theta}(a, b, x) = \arg \sup_{y \in \mathbb{R}} \hat{f}_{(a_1, b_1)}(x, y).$$

where (a_1, b_1) are defined in Theorem 2.3.1. Similarly to the previous criterium, the present procedure shows a great compatibility in practice (cf. Section 4.3), but their asymptotic optimality remains an open question. Furthermore, the choice of the smoothing parameter in the conditional mode estimation is one of the natural prospects of the present work.

The maximum conditional density predictive region : in many particular situations predictions's user can also be interested in the construction of a predictive interval (or region) since the latter is often more informative than a pointwise prediction. Notice that, there are several ways to determine these regions (cf. for instance De Gooijer and Gannoun, 2000). In this paragraph we focus on the maximum conditional density predictive region (MCDR) or

highest conditional density region (HCDR) introduced by Hyndman (1995). This region is defined, for any given $\alpha \in (0, 1)$, by :

$$R_\alpha = \{y : |y| < \infty, f(x, y) \geq l_\alpha(x)\}$$

where

$$l_\alpha(x) = \max \left\{ l > 0 : \int_{f(x,y) \geq l} f(x, y) dy \geq \alpha \right\}.$$

Recall that, the MCDR is of the smallest Lebesgue measure among all the predictive regions with the same coverage probability α (cf. De Gooijer and Gannoun, 2000). In the unconditional case, the estimation of the maximum density predictive region has been widely studied (cf. Samworth and Wand (2010) and the references therein). In our functional conditional context, we use the kernel estimator $\hat{f}_{(a,b)}$ of the conditional density f to give a plug-in estimator of R_α . However, as for all estimations by the kernel method, the performance of this estimation depends heavily on the choice of the bandwidth parameters (a, b) . As that has been mentioned before, many data-driven bandwidth selection have been proposed in the multivariate case. For instance, Gooijer and Gannoun (2000) compare four selection methods based on the classical leave-out-one cross-validation procedure associated to the cumulative conditional distribution, the conditional mode, the conditional mean and the conditional median. At this stage, it seems more reasonable to use a cross-validation criterion of the conditional density instead to that of the predictors (the conditional mode, the conditional mean or the conditional median). In other words, the best approximation of the MCDR can be obtained by computing :

$$\hat{R}_\alpha = \left\{ y : |y| < \infty, \hat{f}_{(a_1, b_1)}(x, y) \geq \hat{l}_\alpha(x) \right\}$$

where

$$\hat{l}_\alpha(x) = \max \left\{ l > 0 : \int_{-\infty}^{\infty} \mathbb{1}_{\{\hat{f}_{(a_1, b_1)}(x, y) \geq l\}} \hat{f}_{(a_1, b_1)}(x, y) dy \geq \alpha \right\}.$$

Similarly to the conditional mode estimation, the asymptotic optimality of this selection procedure is also an important prospect of this work. Finally, let us note that, in the real unconditional framework Samworth and Wand (2010) show the asymptotic optimality of some selection method based on the minimization of the probability of the symmetric difference between R_α and its estimator \hat{R}_α . The adaptation of these ideas in the functional conditional case is an other important prospect of this work.

The conditional expected shortfall estimation : the expected shortfall (ES) is recently considered as one of the most common risk measures in Finance. This model was introduced by Acerbi (2002) for a given level α in $(0, 1)$ by :

$$ES_\alpha = \alpha^{-1} \int_{\text{VAR}_\alpha}^{+\infty} t f_Y(t) dt$$

where $\text{VAR}_\alpha = F_Y^{-1}(\alpha)$ with f_Y (respectively, F_Y) denotes the density (respectively, the cumulative distribution) of the random variable Y representing returns on a given portfolio,

stock, bond or market index. In many situations, we have to analyze the financial risk conditionally to an exogenous variable which is continuously observed. To do that, we use the conditional expected shortfall for which the expectation is taken with respect to the conditional distribution of Y given this exogenous variable X and the VAR_α is the conditional quantiles of order α of Y given X . Accurate estimation of this conditional model depends crucially on the estimation method of the conditional density. Thus, if the kernel method is used to estimator the conditional density function, the best approximation of the conditional expected shortfall (CES) is then given by :

$$\widehat{\text{CES}}_\alpha = \alpha^{-1} \int_{\widehat{\text{VAR}}_\alpha(x)}^{+\infty} \widehat{y} f_{(a_1, b_1)}(x, y) dy$$

where $\widehat{\text{VAR}}_\alpha(x)$ is the α th conditional quantile estimator. Similarly to the previous case, the asymptotic optimality of this approximation is also another important prospect of this work.

2.4.2 On the finite-sample performance of the method

The main purpose of this Section is to show how we can implement easily and rapidly our approach on finite samples. Our first aim is to evaluate the performance of the global smoothing selection method and the second one is to compare its behavior to the local one. For these aims, we consider the following functional nonparametric model :

$$Y_i = r(X_i) + \varepsilon_i, \text{ for } i = 1, \dots, n \quad (22)$$

where the ε_i 's are generated independently according to a $\mathcal{N}(0, 1)$ distribution. The sampled functional explanatory variables X_i for $i = 1, \dots, n$, which is assumed to be independent of ε_i for $i = 1, \dots, n$, is generated according to the following expressions :

$$X_i(t) = a_i \sin(4(b_i - t)) + b_i + \eta_{i,t}, \quad \forall t \in [0, 1] \text{ and } i = 1, 2, \dots, n$$

where b_i (respectively, $\eta_{i,t}$) is $\mathcal{N}(0, 3)$, (respectively, $\mathcal{N}(0, 0.5)$), while the n random variables a_i 's are generated according to a $\mathcal{N}(4, 3)$. All the curves X_i 's are discretized on the same grid generated from 100 equispaced measurements in $(0, 1)$ (cf. Figure 6.1).

On the other hand, the scalar response Y_i , defined by (22), is computed by considering the following operator :

$$r(x) = \int_0^1 \frac{dt}{1 + |x(t)|}$$

Recall that, with this model definition, the conditional density of Y given $X = x$ can be explicitly defined by :

$$f(x, y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y - r(x))^2\right)$$

In order to check the efficiency of this global selection method over finite samples, we compare the averaged squared error (d_2) of the global cross-validation bandwidth (a^{GCV}, b^{GCV}) to

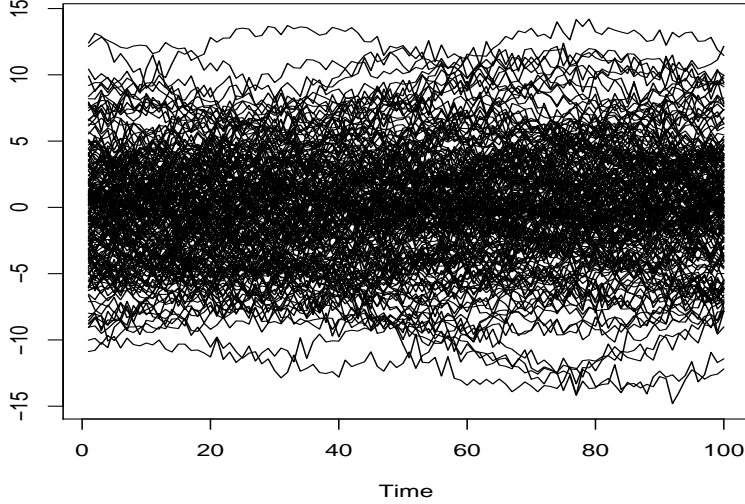


FIGURE 2.1 – A sample of 200 irregular curves

that given by the bandwidth (a^{Gd_2}, b^{Gd_2}) which minimizes the global averaged squared error. Noting that the use of d_2 as a criterion of accuracy, is motivated by the fact that the averaged squared error is easier to deal with from a computational point of view and it is asymptotically equivalent to d_1 and d_3 (cf. Proof of Lemma 2.5.3).

For practical purposes, we select the parameters (a^{GCV}, b^{GCV}) and (a^{Gd_2}, b^{Gd_2}) over a finite set H_n defined by (a_q, b_q) where a_q (respectively, b_q) is the quantile of order q of the vector of all distances between the curves (respectively, between the response variable). Concerning the weight functions, W_1 and W_2 , we recall that these functions are introduced to reduce the boundary effects through their supports, but the role of their expressions are not very determinant in practice as pointed out by Härdle and Marron (1985). In our simulation study, we take

$$W_1(t) = \begin{cases} 1 & \text{if } \min_{i=1, \dots, n} d(t, X_i) < a \\ 0 & \text{otherwise} \end{cases}$$

$$W_2(z) = \begin{cases} 1 & \text{if } z \in [\min\{Y_i, i = 1, \dots, n\} \times 0.9, \max\{Y_i, i = 1, \dots, n\} \times 1.1] \\ 0 & \text{otherwise} \end{cases}$$

The kernel K (respectively, H) is chosen to be quadratic on $(0, 1)$ (respectively, on $(-1, 1)$). Another important point permitting to ensure a good behavior of the method, is the

n	50	100	150	200	250
$d_2(\widehat{f}_{(a^{GCV}, b^{GCV})}, f)$	0.1183	0.0850	0.0774	0.0395	0.0262
$d_2(\widehat{f}_{(a^{Gd_2}, b^{Gd_2})}, f)$	0.0593	0.04284	0.0392	0.03333	0.02361
$\frac{d_2(\widehat{f}_{(a^{Gd_2}, b^{Gd_2})}, f)}{d_2(\widehat{f}_{(a^{GCV}, b^{GCV})}, f)}$	1.9949	1.9859	1.9744	1.1851	1.1096

TABLE 2.1 – Averaged squared errors according to various values of n .

use of a semi-metric that is well adapted to the kind of data that we have to deal with. Here, we use a semi-metric based on the q first eigenfunctions of the empirical covariance operator associated with the q greatest eigenvalues (cf. Ferraty and Vieu (2006) for more discussions). This choice is motivated by the shape of the curves X (cf. Figure 6.1). Similarly to Benhenni et al. (2007), we take $q = 3$. The results of the four selected sample sizes, $n \in \{50, 100, 200, 250\}$, are gathered in Table 2.1.

Then, it can be seen that Table 2.1 shows the good behavior of our functional procedure, in this sense that, the d_2 errors computed by using the smoothing parameters given by our bandwidth selectors is very close to the minimum of the true/real errors. This fact is illustrated in the last row by the ratio $d_2(\widehat{f}_{(a^{GCV}, b^{GCV})}, f)/d_2(\widehat{f}_{(a^{Gd_2}, b^{Gd_2})}, f)$, where we observe that our results admit sufficiently good performances even though for a small sample size $n = 50$. It should be noted that, our automatic selection method computes faster the results also for large sample sizes $n \in \{100, 200, 250\}$.

Now, we compare this global procedure to the local one. For this purpose, we consider a sample of 150 observations of the couple (X, Y) , randomly splitted into two subsets (100 observations as the learning sample and 50 ones are used as the test sample) and we construct the kernel estimator of the conditional density for each curve in the test sample on the 100 equispaced point y_j for $j = 1, \dots, 100$ in the interval

$[\min\{Y_i, i = 1, \dots, n\} \times 0.9, \max\{Y_i, i = 1, \dots, n\} \times 1.1]$ by using the both smoothing parameters selectors (local and global). The local bandwidths (a^{LCV}, b^{LCV}) is selected over $H_n(x, y)$ the set of $(a(x), b(y))$ such that, for $a(x)$ (respectively, for $b(x)$) the ball centered at x (respectively, the interval centered at y) with radius $a(x)$ (respectively, with radius $b(y)$) contains exactly k neighbors of x (respectively, of y). Further, we use the following local weight functions :

$$W_{1,x}(t) = \begin{cases} 1 & \text{if } d(t, x) < a(x) \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad W_{2,y}(z) = \begin{cases} 1 & \text{if } |z - y| < b(y) \\ 0 & \text{otherwise} \end{cases}$$

We choose the global bandwidths over the same set H_n defined above and we use the same

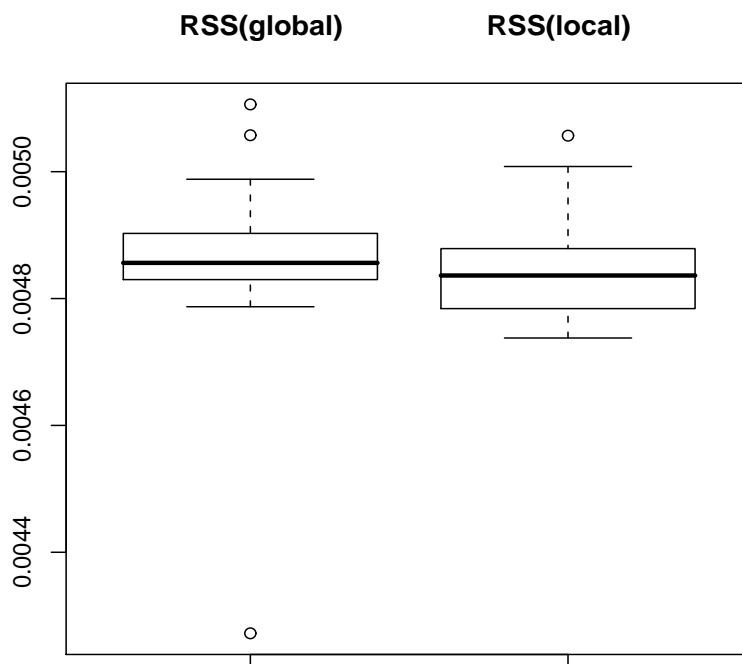


FIGURE 2.2 – Comparison of RSS errors

functions W_1 and W_2 . Moreover, we consider the same semi-metric and the same kernels K and H as in the first illustration. Then, we examine the accuracy of our conditional density estimates as a collection of slices by using the generalized sum of squared residuals (RSS) defined by :

$$RSS(local) = \frac{1}{500} \sum_{i=1}^{50} \sum_{j=1}^{100} \left[f(x_i, y_j) - \hat{f}_{(a^{LCV}, b^{LCV})}(x_i, y_j) \right]^2$$

and

$$RSS(global) = \frac{1}{500} \sum_{i=1}^{50} \sum_{j=1}^{100} \left[f(x_i, y_j) - \hat{f}_{(a^{GCV}, b^{GCV})}(x_i, y_j) \right]^2$$

We have carried out several tests (exactly 25 tests) by changing observations between the learning and the test samples. In Figure 2 we plot the box-plot of the given RSS errors in both cases. It appears clearly that, the local bandwidth choice outperforms better than the global selection method.

2.4.3 A real data application

In this Section, we test the efficiency of our procedure in the conditional mode estimation as discussed in Section 4.1. More precisely, we compare our local bandwidth selection procedure to the local one used by Ferraty and Vieu (2006). For this purpose, we consider the prevision problem when we are interested by the logarithm of the total precipitations given the monthly maximum temperatures via the conditional mode estimation using the both selection methods.

Notice that, the data used for this study, come from the US National Climatic data center and are available on the Web⁴, and are collected in 98 climatic stations in USA from 2004 until 2010. According to the notations of the previous section, the functional predictor X_i is the monthly maximum temperatures in the i th climatic station from 2004 until 2010 (in tenths of degrees F) and Y_i is the logarithm of the total precipitations (in hundredths of inches) in the same station and the same period. The functional predictors X_i for $i = 1, \dots, 98$ are plotted in Figure 6.3.

The practical utilization of our selection method in the conditional mode estimation has been discussed in the previous section. For this study, we keep the same arguments used in the previous section. In other words, we keep the same weight functions $W_{1,x}$, $W_{2,y}$, the same subsets $H_n(x, y)$, the same kernels K and H and according to the shape of our curves, we use the semi-metric based on the functional principal components analysis (cf. also Besse et al., 1997) with $q = 8$. We split our data into two subsets. The first sample, of size $n = 80$ corresponds to the learning sample which will be used, as a sample, to compute our conditional mode estimators at the 18 remaining curves (considered as the test sample). Then, the conditional density estimation based on the LCV, LCDE method say, may be presented as follows :

$$\forall \text{ curve } x_i \text{ in the test sample, } \hat{\theta}(a^{LCV}, b^{LCV}, x_i) = \arg \sup_{y \in \mathbb{R}} \hat{f}_{(a^{LCV}, b^{LCV})}(x_i, y)$$

While for the Ferraty and Vieu's method (F-V-Method, say) we use the R-routine named *funopare.mode.lcv*⁵. Recall that, the parameters (a, b) in this R-routine are locally chosen over the same type of set $H_n(x, y)$ as follows :

$$\forall \text{ curve } x_i \text{ in the test sample, } \hat{\theta}(a^{FV}, b^{FV}, x_i) = \arg \sup_{y \in \mathbb{R}} \hat{f}_{(a^{FV_i^*}, b^{FV_i^*})}(x_i^*, y).$$

where $x_i^* = \arg \min_{x_j \in \text{learning sample}} d(x_i, x_j)$ and

$$(a^{FV_i^*}, b^{FV_i^*}) = \arg \min_{(a,b) \in H_n(x_i^*, y)} |Y_i^* - \hat{\theta}(a, b, x_i^*)|.$$

4. in the ftp address : "<ftp://ftp.ncdc.noaa.gov/pub/data/ushcn/v2/monthly>"

5. available at the website www.lsp.ups-tlse.fr/staph/npfda

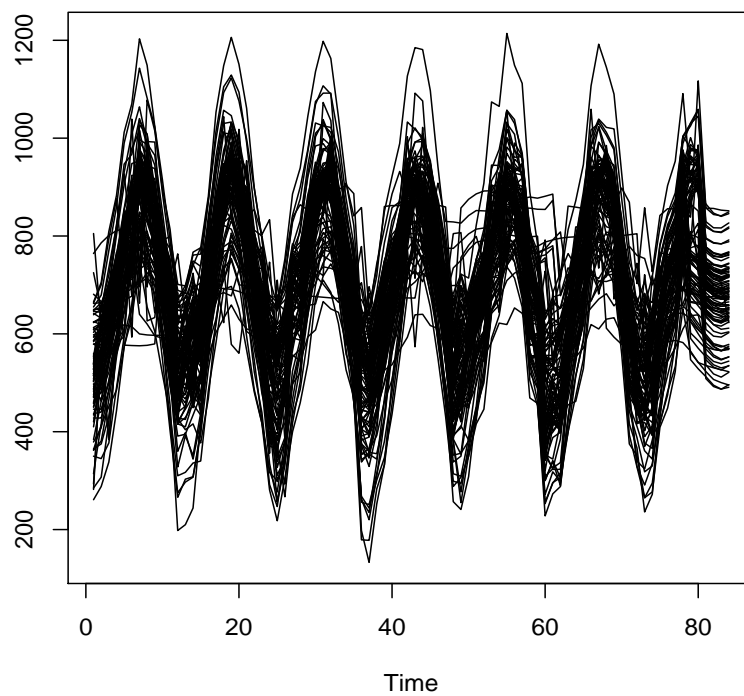


FIGURE 2.3 – Monthly maximum temperatures in 98 climatic stations in USA

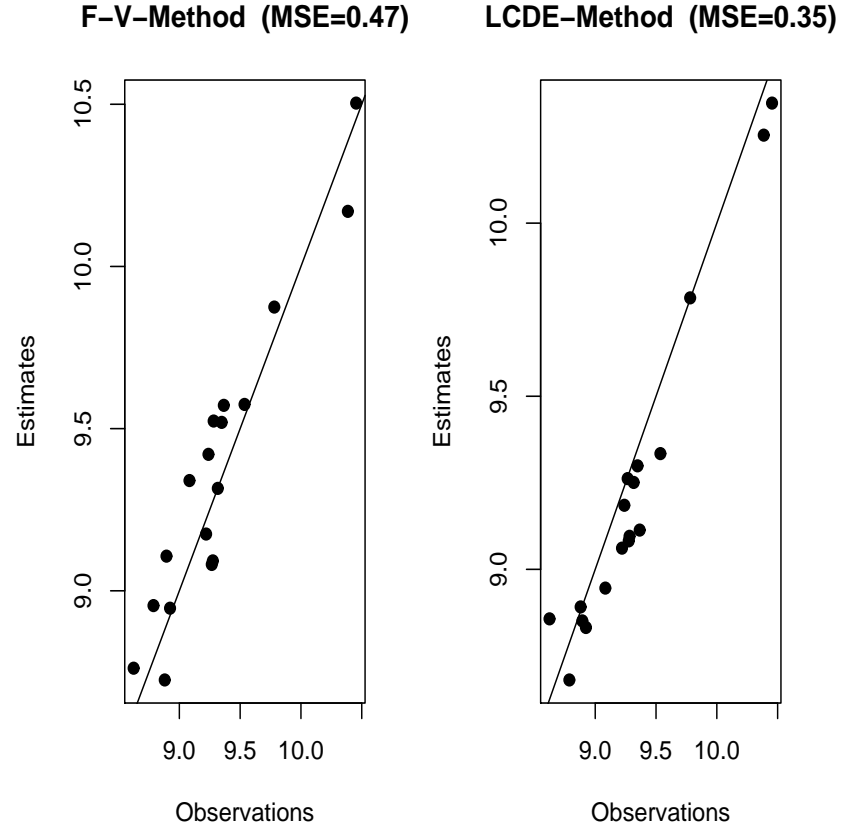


FIGURE 2.4 – Comparison of the prediction results between the F-V-Method and the LCDE-Method

The performance of the both selection rules, in terms of prediction, is evaluated by computing the mean squared prediction errors (MSE), defined by the following quantities :

$$\text{MSE}(\text{LCDE}) = \frac{\sum_{i \in \mathcal{I}_1} \left(Y_i - \hat{\theta}(a^{\text{LCV}}, b^{\text{LCV}}, X_i) \right)^2}{18}$$

and

$$\text{MSE}(\text{FV}) = \frac{\sum_{i \in \mathcal{I}_1} \left(Y_i - \hat{\theta}(a^{\text{FV}}, b^{\text{FV}}, X_i) \right)^2}{18}$$

with \mathcal{I}_1 is the indexes set's of the sample test.

Clearly, the comparison of both scatterplots (cf. Figure 6.4) indicates that the LCDE Method gives better results than those given by the F-V-Method. This is confirmed by relative mean squared prediction errors $\text{MSE}(\text{LCDE}) = 0.35$ and $\text{MSE}(\text{FV}) = 0.47$. It is worth noting that

the superiority of our local procedure to the local one used in Ferraty and Vieu (2006) can be justified by the fact that, for each curve x_i in the sample test, our local procedure minimizes over all k -nearest neighbors smoothing parameter, while in Ferraty and Vieu (2006) consider, only, the k -nearest neighbors of the nearest curve of x_i^* in the learning sample.

2.5 Proofs

Recall that, C denotes a generic constant. As specified above, the technical parts which are similar to the finite dimensional or regression operator estimation for functional data techniques are omitted. Thus, we encourage readers who are interested in these proofs to keep at hand the standard finite dimensional literature (that is, the paper by Härdle and Marron, 1985) or those relative to the infinite dimensional framework (that is, the papers by Rachdi and Vieu (2007), and Benhenni et al. (2007)).

Proof of theorem 2.3.1. We consider the following decomposition, for all $(x, y) \in S_X \times S_Y$, and for all $(a, b) \in H_n$:

$$\begin{aligned} \left(\widehat{f}_{(a,b)}(x, y) - f(x, y)\right)^2 &= \left(\widehat{f}_N(x, y) - f(x, y)\widehat{f}_D(x)\right)^2 \\ &\quad + 2\left(1 - \widehat{f}_D(x)\right)\widehat{f}_D(x)\left(\widehat{f}_{(a,b)}(x, y) - f(x, y)\right)^2 \\ &\quad + \left(1 - \widehat{f}_D(x)\right)^2\left(\widehat{f}_{(a,b)}(x, y) - f(x, y)\right)^2 \end{aligned}$$

where

$$\widehat{f}_D(x) = \frac{1}{n\mathbb{E}[K(a^{-1}d(x, X))]} \sum_{i=1}^n K(a^{-1}d(x, X_i)) \quad (23)$$

and

$$\begin{aligned} \widehat{f}_N(x, y) &= \frac{1}{nb\mathbb{E}[K(a^{-1}d(x, X))]} \sum_{i=1}^n K(a^{-1}d(x, X_i))H(b^{-1}(y - Y_i)) \\ &= \widehat{f}_D(x)\widehat{f}_{(a,b)}(x, y) \end{aligned} \quad (24)$$

It follows from the uniform consistency⁶ of $\widehat{f}_D(x)$ to $\mathbb{E}[\widehat{f}_D(x)] = 1$ (cf. Ferraty et al., 2010), that :

$$\forall k = 1, 2, 3 \quad d_k(\widehat{f}_{(a,b)}(x, y), f(x, y)) = d_k(\widehat{f}_N(x, y), f(x, y)\widehat{f}_D(x)) + o_{a.s.}(d_k(\widehat{f}_{(a,b)}(x, y), f(x, y)))$$

6. In Ferraty et al. (2010), the main aim is to state the rate of the uniform almost-complete convergence of the functional component. Such result can be easily extended here (without precision of the convergence rate) to $\sup_{a \in H_n}$ by using the second part of assumption (15)

It suffices, then, to show the claimed result for d_k^* defined by :

$$d_k^*(\widehat{f}_{(a,b)}, f) = d_k(\widehat{f}_N(x, y), f(x, y)\widehat{f}_D(x)) \quad \text{for all } k = 1, 2, 3.$$

Notice that by following the same steps as those used for proving Lemmas 5, 6 and 7 in Rachdi and Vieu (2007), and by noting :

$$K^*(x, X_i) = \frac{1}{b\mathbb{E}[K(a^{-1}d(x, X))]} \left(K(a^{-1}d(x, X_i))H(b^{-1}(y - Y_i) - bf(x, y)K(a^{-1}d(x, X_i))) \right)$$

we obtain that :

$$\sup_{h \in H_n} \left(\frac{|d_k^*(\widehat{f}_{(a,b)}, f) - d_l^*(\widehat{f}_{(a,b)}, f)|}{d_3^*(\widehat{f}_{(a,b)}, f)} \right) \rightarrow 0, \quad \text{a.s., for all } k \neq l$$

On the other hand, we introduce the error measure d_5 defined by :

$$d_5(\widehat{f}_{(a,b)}, f) = \frac{1}{n} \sum_{i=1}^n \left(\widehat{f}_{(a,b)}^{-i}(X_i, Y_i) - f(X_i, Y_i) \right)^2 \frac{W_1(X_i)W_2(Y_i)}{f(X_i, Y_i)}$$

where

$$\widehat{f}_{(a,b)}^{-i}(X_i, Y_i) = \frac{\widehat{f}_N^{-i}(X_i, Y_i)}{\widehat{f}_D^{-i}(X_i)}$$

with

$$\widehat{f}_N^{-i}(x, y) = \frac{1}{nb\mathbb{E}[K(a^{-1}d(x, X))]} \sum_{j \neq i}^n K(a^{-1}d(x, X_j))H(b^{-1}(y - Y_j))$$

and

$$\widehat{f}_D^{-i}(x) = \frac{1}{n\mathbb{E}[K(a^{-1}d(x, X))]} \sum_{j \neq i}^n K(a^{-1}d(x, X_j))$$

Remark that :

$$GCV(a, b) - d_5(\widehat{f}_{(a,b)}, f) = CT(a, b) + T$$

where

$$CT(a, b) = \frac{1}{n} \sum_{i=1}^n W_1(X_i) \int \widehat{f}_{(a,b)}^{-i^2}(X_i, y)W_2(y) dy - \frac{1}{n} \sum_{i=1}^n \widehat{f}_{(a,b)}^{-i^2}(X_i, Y_i) \frac{W_1(X_i)W_2(Y_i)}{f(X_i, Y_i)}$$

and

$$T = \frac{1}{n} \sum_{i=1}^n f(X_i, Y_i)W_1(X_i)W_2(Y_i)$$

Thus, the proof of this theorem is complete if we can prove that d_5 is asymptotically equivalent to d_3 and that :

$$\sup_{h \in H_n} \left| \frac{CT(a, b)}{d_3(\widehat{f}_{(a,b)}, f)} \right| \rightarrow 0, \quad \text{a.s., as } n \rightarrow +\infty$$

That is why the demonstration of Theorem 2.3.1 is achieved by the following Lemmas 2.5.1, 2.5.2 and 2.5.3, for which the proofs are given in the Appendix (cf. Section 6). ■

Lemma 2.5.1. *Under hypotheses (10), (11), (12) and (14), we have that :*

$$d_3(\widehat{f}_{(a,b)}, f) \geq C \frac{1}{nb\phi(a)}$$

Lemma 2.5.2. *Under hypotheses (10), (11), (12), (14) and (17), we obtain that :*

$$\sup_{(a,b) \in H_n} |nb\phi(a)CT(a, b)| \rightarrow 0, \quad \text{a.s., as } n \rightarrow +\infty$$

Lemma 2.5.3. *Under hypotheses (10), (11), (12), (14) and (17), we have that :*

$$\sup_{(a,b) \in H_n} \left| \frac{d_5(\widehat{f}_{(a,b)}, f) - d_3^*(\widehat{f}_{(a,b)}, f)}{d_3^*(\widehat{f}_{(a,b)}, f)} \right| \rightarrow 0, \quad \text{a.s., as } n \rightarrow +\infty$$

Proof of Theorem 2.3.2. The main ideas of the proof are essentially contained in Rachdi and Vieu (2007) and, in the same way, the proof of Theorem 2.3.1 above, but the computations here are more complicated since we have the additional problems of dealing with non constant weight functions and functional data. The reader should have the above mentioned papers at hand in order to get all the details of this proof. To make things easier, from hypothesis (9), the weight function is bounded and has a compact support with nonempty interior (the closure of $\mathcal{B}(x, w)$). Thus, hypothesis (5) in Rachdi and Vieu (2007) is satisfied and the proof of this theorem is shown using the same steps as in Theorem 2.3.1's proof. It is therefore omitted here. We just mention that the first step of the proof consists in showing the result over a finite subset of H_n , and in the second step, the result is extended to the continuous set of bandwidths H_n by using the Hölder continuity property of the functions K and f (cf. Vieu (1991) and Youndjé et al. (1993)). ■

2.6 Appendix : Proofs of technical lemmas

In what follows, we will denote, for all $i = 1, \dots, n$, by :

$$K_i(x) = K(a^{-1}d(x, X_i)) \quad \text{and} \quad H_i(y) = H(b^{-1}(y - Y_i))$$

Proof of Lemma 2.5.1. It is clear that :

$$d_3(\widehat{f}_{(a,b)}, f) \geq \int \int \text{Var} \left[\widehat{f}_{(a,b)}(x, y) \right] W_1(x) W_2(y) dP_X(x) dy$$

Therefore, it is sufficient to evaluate the variance term $\text{Var} \left[\widehat{f}_{(a,b)}(x, y) \right]$. To do that, by using a similar computational techniques as in Laksaci (2007) and by taking into account the fact that $\mathbb{E} \left[\widehat{f}_D(x) \right] = 1$, we get that :

$$\begin{aligned} \text{Var} \left[\widehat{f}_{(a,b)}(x, y) \right] &= \text{Var} \left[\widehat{f}_N(x, y) \right] - 2\mathbb{E}(\widehat{f}_N(x, y)) \text{Cov}(\widehat{f}_N(x, y), \widehat{f}_D(x)) \\ &\quad + \left(\mathbb{E}(\widehat{f}_N(x, y)) \right)^2 \text{Var}(\widehat{f}_D(x)) + o \left(\frac{1}{nb\phi(a)} \right) \end{aligned} \quad (25)$$

According to the definitions (24) and (23) of the estimators \widehat{f}_N and \widehat{f}_D , we obtain that :

$$\begin{aligned} \text{Var} \left(\widehat{f}_N(x, y) \right) &= \frac{1}{n(b\mathbb{E} [K(a^{-1}d(x, X))])^2} \text{Var} (K_1(x)H_1(y)) \\ \text{Cov}(\widehat{f}_N(x, y), \widehat{f}_D(x)) &= \frac{1}{nb(\mathbb{E} [K(a^{-1}d(x, X))])^2} \text{Cov}(K_1(x)H_1(y), K_1(x)) \end{aligned}$$

and

$$\text{Var} \left(\widehat{f}_D(x) \right) = \frac{1}{n(\mathbb{E} [K(a^{-1}d(x, X))])^2} \text{Var} (K_1(x))$$

Moreover, under (12) and after some simple calculations, we can show, for all $i, j = 1, 2$, that :

$$\mathbb{E} \left[K_1^i(x)H_1^j(y) \right] = b\mathbb{E} \left[K_1^i(x)f(X_1, y) \right] \int H^j(t)dt + o(b\mathbb{E}[K_1(x)]) \quad (26)$$

and

$$0 < C\phi(a) \leq \mathbb{E} [K_1^i(x)] \leq C'\phi(a) \quad (27)$$

By comparing asymptotically the three quantities in (25), we can see that the first term is leading. Hence, from (26) and (27) we obtain that :

$$\text{Var} \left[\widehat{f}_{(a,b)}(x, y) \right] = \frac{1}{nb(\mathbb{E} [K(a^{-1}d(x, X))])^2} \mathbb{E} [K_1^2(x)f(X_1, y)] \int H^2(t)dt + o \left(\frac{1}{nb\phi(a)} \right)$$

Furthermore, by using the fact that the conditional density does not vanish on a neighbor of $S_X \times S_Y$, we obtain that :

$$d_3(\widehat{f}_{(a,b)}, f) \geq C \frac{1}{nb\phi(a)}$$

which completes the proof of this lemma. \blacksquare

Proof of Lemma 2.5.2. From the definition of $CT(a, b)$, we have for all $(a, b) \in H_n$:

$$\begin{aligned} |CT(a, b)| &= \left| \frac{1}{n} \sum_{i=1}^n \int \widehat{f}_{(a,b)}^{-i^2}(X_i, y)W_2(y)W_1(X_i)dy - \frac{2}{n} \sum_{i=1}^n \widehat{f}_{(a,b)}^{-i^2}(X_i, Y_i) \frac{W_1(X_i)W_2(Y_i)}{f(X_i, Y_i)} \right| \\ &= \left| \sum_{i=1}^n \frac{1}{n\widehat{f}_D^{-i^2}(x)} \left[\int \widehat{f}_N^{-i^2}(X_i, y)W_2(y)W_1(X_i)dy - 2\widehat{f}_N^{-i^2}(X_i, Y_i) \frac{W_1(X_i)W_2(Y_i)}{f(X_i, Y_i)} \right] \right| \end{aligned}$$

This combined with the uniform consistency of $\widehat{f}_D^{-i}(x)$ to 1 (cf. Ferraty et al., 2008), is enough to prove that, as $n \rightarrow +\infty$:

$$\sup_{(a,b) \in H_n} b\phi(a) \left| \sum_{i=1}^n \left[\int \widehat{f}_N^{-i^2}(X_i, y) W_2(y) W_1(X_i) dy - 2\widehat{f}_N^{-i^2}(X_i, Y_i) \frac{W_1(X_i) W_2(Y_i)}{f(X_i, Y_i)} \right] \right| \rightarrow 0, \text{ a.s.}$$

To do this, one can use similar arguments as in the real case (cf. Youndjé, 1996). Indeed, let for all $1 \leq i, j, k \leq n$:

$$\begin{aligned} a_{ij} &= K_j^2(X_i) W_1(X_i) \\ b_{ij} &= \int H_j^2(y) W_2(y) dy - H_j^2(Y_i) \frac{W_2(Y_i)}{f(X_i, Y_i)} \\ U_{ij} &= a_{ij} b_{ij} \\ c_{ijk} &= K_j(X_i) K_k(X_i) W_1(X_i) \\ d_{ijk} &= \int H_j(y) H_k(y) W_2(y) dy - H_j(Y_i) H_k(Y_i) \frac{W_2(Y_i)}{f(X_i, Y_i)} \\ V_{ijk} &= c_{ijk} d_{ijk} \end{aligned}$$

Now, we have to examine the following limits :

$$\sup_{(a,b) \in H_n} \left| \frac{b\phi(a)}{(n-1)^2} \sum_{i \neq j} U_{ij} \right| \rightarrow 0, \text{ a.s.}$$

and

$$\sup_{(a,b) \in H_n} \left| \frac{b\phi(a)}{(n-1)^2} \sum_{i \neq j \neq k \neq i} V_{ijk} \right| \rightarrow 0, \text{ a.s.}$$

The proof of the above two limits follows by adopting the same steps as in the proof of Lemma 3 in Rachdi and Vieu (2007). Notice that, by the Borel-Cantelli Lemma, it is enough to show that there is a $\xi, \xi' > 0$ such that for all $p \in \mathbb{N}^*$ there are constants C, C' so that :

$$\mathbb{E} \left(n^{-2} b\phi(a) \sum_{i \neq j} U_{ij} \right)^{2p} \leq C n^{-\xi p} \quad (28)$$

and

$$\mathbb{E} \left(n^{-2} b\phi(a) \sum_{i \neq j \neq k \neq i} V_{ijk} \right)^{2p} \leq C' n^{-\xi' p} \quad (29)$$

To prove (28) we have that :

$$\mathbb{E} \left(n^{-2} b\phi(a) \sum_{i \neq j} U_{ij} \right)^{2p} = (n^{-2} b\phi(a))^{2p} \sum_{i_1 \neq j_1} \cdots \sum_{i_{2p} \neq j_{2p}} \mathbb{E} (U_{i_1 j_1} \cdots U_{i_{2p} j_{2p}})$$

To compute these quantities, we have to show that :

$$\mathbb{E} (U_{i_1 j_1} \cdots U_{i_{2p} j_{2p}}) = 0, \text{ if } m > 2p \quad (30)$$

where m denotes the cardinality of the set $\{i_1, j_1, \dots, i_{2p}, j_{2p}\}$.

On the other hand, observe that, for all i, j :

$$\begin{aligned} \mathbb{E}[b_{ij}|X_1, \dots, X_n] &= \int \int b^{-2} H^2(b^{-1}(y-z)) f(X_j, z) W_2(y) dy dz \\ &\quad - \int \int b^{-2} H^2(b^{-1}(y-z)) \frac{f(X_i, y) f(X_j, z)}{f(X_i, y)} W_2(y) dy dz \\ &= 0 \end{aligned}$$

and thus, we may write :

$$\mathbb{E}[U_{ij}|X_1, \dots, X_n] = a_{ij} \mathbb{E}[b_{ij}|X_1, \dots, X_n] = 0$$

and

$$\mathbb{E}[U_{ij}|X_i] = \mathbb{E}[\mathbb{E}[U_{ij}|X_1, \dots, X_n]|X_i] = 0$$

So, if $m > 2p$ there exists an $a \in \{1, \dots, 2p\}$ such that i_a (or j_a) appears only once in $\{i_1, j_1, \dots, i_{2p}, j_{2p}\}$, therefore, it suffices to compute the expectation by conditioning with respect to X_{i_a} (or X_{j_a}) to show (30).

On the other hand, for $2 \leq m \leq 2p$, we deduce from (10) and (12) that :

$$|\mathbb{E} (U_{i_1 j_1} \cdots U_{i_{2p} j_{2p}})| \leq \frac{C}{b^{4p} \phi(a)^{4p}} \mathbb{E} \left[\prod_{i,j=1}^m K^{\tau_{ij}} (a^{-1} d(X_i, X_j)) \prod_{i=1}^m W_1^{\beta_i}(X_i) \right]$$

where, $\beta_i = 0, 1$, and $\sum_{ij} \tau_{ij} = 2p$ and for all $i = 1, \dots, m$, there is an j such that $\tau_{ij} \neq 0$, and $\beta_i = 1$ or $\beta_j = 1$.

Similar arguments as those invoked for proving Lemma 9 in Rachdi and Vieu (2007) can be used, which allow us to obtain that :

$$\text{for all } m = 2, \dots, 2p, |\mathbb{E} (U_{i_1 j_1} \cdots U_{i_{2p} j_{2p}})| \leq \frac{C}{b^{4p} (\phi(a))^{4p}} \phi^{\frac{m}{2}}(a)$$

Therefore, by (14), we have :

$$\begin{aligned} \mathbb{E} \left(n^{-2} b \phi(a) \sum_{i \neq j} U_{ij} \right)^{2p} &\leq C \sum_{m=2}^{2p} n^{m-4p} b^{-2p} (\phi(a))^{\frac{m}{2}-2p} \\ &= C \frac{1}{(nb\phi(a))^{2p}} (\phi(a))^p \sum_{m=2}^{2p} \left(n \sqrt{(\phi(a))} \right)^{m-2p} \\ &\leq C \frac{1}{(nb\phi(a))^{2p}} (\phi(a))^p \\ &\leq C \frac{1}{(n^{1-\gamma})^{2p} (\phi(a))^p}. \end{aligned} \quad (31)$$

It suffices now to combine (31) together with the assumption (16), to get (28).

Concerning (29), we use analogous arguments as for showing (28). By denoting m' the cardinality of the set $\{i_1, j_1, k_1, \dots, i_{2p}, j_{2p}, k_{2p}\}$ and because $\mathbb{E}[d_{ijk}|X_1, \dots, X_n] = 0$ we deduce that, if $m' > 3p$, we have that :

$$\mathbb{E} (V_{i_1 j_1 k_1} \cdots V_{i_{2p} j_{2p} k_{2p}}) = 0$$

Moreover, when $3 \leq m' \leq 3p$, a straightforward modification of the proof of Lemma 5.4.2 in Younjé (1996), gives :

$$\mathbb{E} [V_{i_1 j_1 k_1} \cdots V_{i_{2p} j_{2p} k_{2p}}] \leq \frac{C}{b^{4p} \phi(a)^{4p}} (b\phi(a))^{m'/2}$$

This leads directly to write, under (14), that :

$$\begin{aligned} \mathbb{E} \left(n^{-2} b\phi(a) \sum_{i \neq j \neq k \neq i} V_{ijk} \right)^{2p} &\leq C \sum_{m'=3}^{3p} n^{m'} b^{-4p} (b\phi(a))^{m'/2} (\phi(a))^{-4p} (n^{-2} b\phi(a))^{2p} \\ &= C \frac{1}{((n(b\phi(a))^{1/2})^p)^{2p}} \sum_{m'=3}^{3p} (n(b\phi(a))^{1/2})^{m'-3p} \\ &\leq C \frac{1}{(n^{(1-\gamma/2)p} \phi(a)^{p/2})^{2p}} \end{aligned}$$

Once again, we use (16) to complete the proof of this lemma. ■

Proof of Lemma 2.5.3. By using the same arguments as those used at the beginning of the proof of the Theorem 2.3.1, and by introducing :

$$d_5^*(\widehat{f}_{(a,b)}, f) = \frac{1}{n} \sum_{i=1}^n \left(\widehat{f}_N^{-i}(X_i, Y_i) - f(X_i, Y_i) \widehat{f}_D^{-i}(X_i) \right)^2 \frac{W_1(X_i) W_2(Y_i)}{f(X_i, Y_i)}$$

we can write

$$d_5(\widehat{f}_{(a,b)}, f) = d_5^*(\widehat{f}_{(a,b)}, f) + o(d_5(\widehat{f}_{(a,b)}, f))$$

Therefore, because of the asymptotic equivalence between d_3 and d_2^* the proof of this lemma will be completed as soon as we show that :

$$\sup_{h \in H_n} \left| \frac{d_5^*(\widehat{f}_{(a,b)}, f) - d_2^*(\widehat{f}_{(a,b)}, f)}{d_3(\widehat{f}_{(a,b)}, f)} \right| \rightarrow 0, \text{ a.s.}$$

To do that, we consider the following decomposition :

$$\begin{aligned} (\widehat{f}_N^{-i} - f \widehat{f}_D^{-i})^2 &= (\widehat{f}_N - f \widehat{f}_D)^2 + 2 (\widehat{f}_N^{-i} - \widehat{f}_N + f(\widehat{f}_D - \widehat{f}_D^{-i})) (\widehat{f}_N - f \widehat{f}_D) \\ &+ (\widehat{f}_N^{-i} - \widehat{f}_N + f(\widehat{f}_D - \widehat{f}_D^{-i}))^2 \end{aligned}$$

Furthermore, observe that, for all $i = 1, \dots, n$, we have that :

$$\begin{aligned} \widehat{f}_N^{-i}(x, y) - \widehat{f}_N(x, y) &= \frac{1}{n-1} \widehat{f}_N(x, y) \\ &\quad - \frac{1}{(n-1)b\mathbb{E}[K(a^{-1}d(x, X))]} K\left(\frac{d(x, X_i)}{a}\right) H\left(\frac{y - Y_i}{b}\right) \end{aligned}$$

and

$$\widehat{f}_D^{-i}(x) - \widehat{f}_D(x) = \frac{1}{n-1} \widehat{f}_D(x) - \frac{1}{(n-1)\mathbb{E}[K(a^{-1}d(x, X))]} K\left(\frac{d(x, X_i)}{a}\right)$$

Since K is a bounded function, we deduce from the uniform consistency of \widehat{f}_D and \widehat{f}_N that

$$|\widehat{f}_N^{-i}(x, y) - \widehat{f}_N(x, y)| \leq \frac{C}{(n-1)b\mathbb{E}[K(a^{-1}d(x, X))]} \leq \frac{C}{(n-1)b\phi(a)}, \text{ a.s.}$$

and

$$|\widehat{f}_D^{-i}(x) - \widehat{f}_D(x)| \leq \frac{C}{(n-1)b\mathbb{E}[K(a^{-1}d(x, X))]} \leq \frac{C}{(n-1)b\phi(a)}, \text{ a.s.}$$

Hence, we get :

$$\begin{aligned} \left| d_5^*(\widehat{f}_{(a,b)}, f) - d_2^*(\widehat{f}_{(a,b)}, f) \right| &\leq \frac{C}{(n-1)b\phi(a)} \sup_{(x,y) \in S_X \times S_Y} \left| \widehat{f}_N(x, y) - f(x, y) \widehat{f}_D \right| \\ &\quad + \frac{C}{(n-1)^2 b^2 (\phi(a))^2} \end{aligned}$$

By combining this last result with Lemma 2.5.1 we obtain the claimed result. ■

Bibliographie

- [1] Acerbi, C. (2002). Spectral measures of risk : a coherent representation of subjective risk aversion. *J. Bank. Financ.*, **26**, Pages 1505-1518.
- [2] Bashtannyk, D.M. and Hyndman, R.J. (2001). Bandwidth selection for kernel conditional density estimation. *Comput. Statist. Data Anal.*, **36**, Pages 279-298.
- [3] Benhenni, K., Ferraty, F., Rachdi, M. and Vieu, P. (2007). Local smoothing regression with functional data. *Comput. Statist.*, **22**, No. 3, Pages 353–369.
- [4] Besse, P., Cardot, H. and Ferraty, F. (1997). Simultaneous nonparametric regressions of unbalanced longitudinal data. *Comput. Statist. Data Anal.*, **24**, No. 3, Pages 255–270.
- [5] Bogachev, V.I. (1999). *Gaussian measures*. Math surveys and monographs, **62**, Amer. Math. Soc.
- [6] Bosq, D. (2000). *Linear Processes in Function Spaces : Theory and applications*. Lecture Notes in Statistics, **149**, Springer.
- [7] Collomb, G., Härdle, W. and Hassani, S. (1987). A note on prediction via estimation of conditional mode function. *J. of Statist. Plan. and Inf.*, **15**, Pages 227–236.
- [8] Dabo-Niang, S. and Laksaci, A. (2007). Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle. *Pub. Inst. Stat. Univ. Paris*, **3**, Pages 27–42.
- [9] De Gooijer, J. and Gannoun, A. (2000) Nonparametric conditional predictive regions for time series *Comput. Statist. Data Anal.*, **33**, Pages 259–257.
- [10] El Ghouch, A. and Genton, M. (2009). Local polynomial quantile regression with parametric features. *J. Amer. Statist. Assoc.*, **104**, No. 488, Pages 1416–1429.
- [11] Fan, J., Yao, Q. and Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, **83**, Pages 189–206.
- [12] Ferraty, F., Mas, A. and Vieu, P. (2007). Advances in nonparametric regression for functional variables. *Aust. and New Zeal. J. of Statist.*, **49**, Pages 1-20.
- [13] Ferraty, F., Tadj, A., Laksaci, A. and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. of Statist. Plan. and Inf.*, **140**, Pages 335–352 .
- [14] Ferraty, F., Laksaci, A. and Vieu, P. (2005). Functional times series prediction via conditional mode. *C. R. Acad. Sci. Maths. Paris*, **340**, Pages 389–392.

- [15] Ferraty, F., Laksaci, A. and Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statist. Inf. for Stoch. Proc.*, **9**, Pages 47–76.
- [16] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis. Theory and Practice*. Springer-Verlag.
- [17] Gannoun, A. (1990). Estimation non paramétrique de la médiane conditionnelle. *Pub. Inst. Stat. Univ. Paris*, **35**, No. 1, Pages 11-22.
- [18] Gannoun, A., Saracco, J. and Yu, K. (2003). Nonparametric prediction by conditional median and quantiles. *J. of Statist. Plan. and Inf.*, **117**, No. 2, Pages 207–223.
- [19] Hall, P., Wolk, R.C. and Yao, Q. (1999). Methods for estimating a conditional distribution function. *J. Amer. Statist. Assoc.*, **94**, Pages 154–163.
- [20] Härdle, W. (1991). *Smoothing Techniques with Implementation in S*. Springer, New York.
- [21] Härdle, W., Janssen, P. and Serfling, R. (1991). Strong consistency rates for estimators of conditional functionals. *Ann. Statist.*, **16**, No. 4, Pages 1428–1449.
- [22] Härdle, W. and Marron, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.*, **13**, No. 4, Pages 1465–1481.
- [23] Hyndman, R.J. (1995). Highest-density forecast regions for non-linear and non-normal time series models. *J. Forecast.*, **14**, Pages 431–441.
- [24] Hyndman, R.J., Bashtannyk, D.M. and Grunwald, G.K. (1996). Estimating and visualizing conditional densities. *J. Comput. Graph. Statist.*, **5**, Pages 315–336.
- [25] Hyndman, R.J. and Yao, Q. (1998). Nonparametric estimation and symmetry tests for conditional density functions. *Working paper 17/98, Department of Econometrics and Business Statistics, Monash University*.
- [26] Kolmogorov, A.N. and Tikhomirov, V.M. (1959). ϵ -entropy and ϵ -capacity. *Uspekhi Mat. Nauk.*, **14**, Pages 3–86. (Eng. Transl. Amer. Math. Soc. Transl. Ser., **2**, Pages 277–364, (1964)).
- [27] Laksaci, A. (2007). Convergence en moyenne quadratique de l'estimateur à noyau de la densité conditionnelle avec variable explicative fonctionnelle. *Pub. Inst. Stat. Univ. Paris*, **3**, Pages 69–80.
- [28] Ouassou, I. and Rachdi, M. (2009). Stein type estimation of the regression operator for functional data. *Advances and Applications in Statistical Sciences*, **1**, No. 2, Pages 233-250.
- [29] Rachdi, M. and Vieu, P. (2007). Nonparametric regression for functional data : automatic smoothing parameter selection. *J. of Statist. Plan. and Inf.*, **137**, Pages 2784–2801.
- [30] Ramsay, J.O. and Silverman, B.W. (2005). *Functional data analysis*. Second Edition, Springer-Verlag. New York.
- [31] Rosenblatt, M. (1969). Conditional probability density and regression estimators. In *Multivariate Analysis II*, Ed. P.R. Krishnaiah. Academic Press, New York and London.

- [32] Roussas, G.G. (1968). On some properties of nonparametric estimates of probability density functions. *Bull. Soc. Math. Greece (N.S.)*, **9**, Pages 29–43.
- [33] Samworth, R. J. and Wand, M. P. (2010). Asymptotics and optimal bandwidth selection for highest density region estimation. *Ann. Statist.*, **38**, Pages 1767–1792.
- [34] Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.*, **9**, Pages 65–78.
- [35] Stone, C.J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.*, **22**, No. 1, Pages 118–184.
- [36] Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. Chapman & Hall, London.
- [37] Theodoros, N. and Yannis, G.Y. (1997). Rates of convergence of estimate, Kolmogorov entropy and the dimensionality reduction principle in regression. *Ann. Statist.*, **25**, No. 6, Pages 2493–2511.
- [38] Vieu, P. (1991). Nonparametric regression : optimal local bandwidth choice. *J. R. Stat. Soc.*, **53**, No. 2, Pages 453–464.
- [39] Youndjé, E. (1993). *Estimation non paramétrique de la densité conditionnelle par la méthode du noyau*. PhD Thesis from the Rouen University (in French).
- [40] Youndjé, E. (1996). Propriétés de convergence de l'estimateur à noyau de la densité conditionnelle. *Rev. Roumaine Math. Pures Appl.*, **41**, Pages 535–566.
- [41] Youndjé, E., Sarda, P. and Vieu, P. (1993). Kernel estimator of conditional density bandwidth selection for dependent data. *C. R. Acad. Sci. Math. Paris*, **316**, No. 9, Pages 935–938.

Chapitre 3

Functional data : Local linear estimation of the conditional density and its application

Jacques Demongeot¹, Ali Laksaci²
Fethi Madani¹ and Mustapha Rachdi^{3,4}

C. R., Math., Acad. Sci. Paris, **348**, *Issues 15-16, Pages 931-934, (2010).*
Statistics, DOI : 10.1080/02331888.2011.568117 (à paraître en 2012)

Abstract. In this paper, we introduce a new nonparametric estimation of the conditional density of a scalar response variable given a random variable taking values in a semi-metric space. Under some general conditions, we establish the pointwise and uniform almost complete consistencies with rates of this estimator. Moreover, as an application, we use the obtained results to derive some asymptotic properties for the local linear estimator of the conditional mode.

Keywords. Functional data, Local linear estimator, Conditional density, Conditional mode, Nonparametric model, Small balls probability

AMS Subject Classification. Primary : 62G05, Secondary : 62G07, 62G08, 62G35, 62G20.

3.1 Introduction

This paper deals with local polynomial modeling of the conditional density function when the explanatory variable is of functional type. It is well known that a local polynomial smoothing has various advantages over the kernel method, namely this method has superior bias properties to the previous one (cf. [4] and [6] for an extensive discussion on the comparison

1. Laboratoire AGIM FRE 3405 CNRS, Equipe TIMB, Faculté de Médecine de Grenoble, Université J. Fourier, 38700 La Tronche, France. E-mail : Jacques.Demongeot@imag.fr

2. Université Djillali Liabès, BP. 89, Sidi Bel-Abbès 22000, Algeria. E-mail : alilak@yahoo.fr

3. Laboratoire AGIM FRE 3405 CNRS, Equipe TIMB, Université P. Mendès France (Grenoble 2), UFR SHS, BP. 47, 38040 Grenoble Cedex 09, France. E-mails : Mustapha.Rachdi@upmf-grenoble.fr and Fethi.Madani@imag.fr

4. Corresponding author

between both these methods). 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 (cf. [5, 11, 29, 30]). Moreover, the kernel method is known for being a particular case of the local polynomial method.

Except the fact that the conditional density plays an important role in nonparametric prediction, there are several tools in nonparametric statistic, such as the conditional mode, the conditional median or the conditional quantiles, which are based on the preliminary estimator of the functional “parameter” proposed in this paper. In the nonparametric functional statistics, the first results about the almost-complete consistency were obtained in [10], for conditional density/distribution functions estimation when the data are independent and identically distributed. The strong mixing case has been studied by [16]. On the other hand, the convergence in L^p -norm mode of the kernel estimator of the conditional mode was stated in [8], and some asymptotics of the conditional quantile and mode estimators in [10]. While in [12] the asymptotic expansion of the exact expression involved in the leading terms of the quadratic error of the kernel estimators of the conditional density is established, in [9] the uniform almost complete convergence of some nonparametric conditional models is showed. For some more recent advances in the nonparametric statistics for functional data we refer to [3, 5, 25] and the references therein.

In this work, we introduce a new nonparametric estimation of the conditional density for functional data. Our estimator is based on the local linear approach. Notice that the local linear estimator of the conditional density has been widely studied, when the explicative variable lies in a finite dimensional space, and there are many references on this topic (cf. for instance [12, 19]). For a general treatment/study of local polynomial estimation, we refer to [13, 24]. Thus, in this paper, we are concerned in proving, under some general conditions, the almost complete convergence with rates of the constructed estimator. More precisely, in Section 3, we show the pointwise consistency. The uniform version of this asymptotic result is given in Section 4. The interest of the uniform consistency comes mainly from the fact that the pointwise performance of all estimators is not sufficient to quantify its efficiency, but, some stability is needed, in the sense that this performance should be uniform over a neighborhood. Notice that, in the functional statistics, the uniform convergence is not a direct extension of the previous pointwise results, but, it requires some additional tools and conditions. In Section 5, we will emphasize the consequence of the previous results to the estimation of the conditional mode.

3.2 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 $\mathcal{F} \times \mathbb{R}$, where \mathcal{F} is a semi-metric space equipped with a semi-metric d .

Furthermore, we assume that there exists a regular version of the conditional probability of Y given X , which is absolutely continuous with respect to Lebesgue measure on \mathbb{R} and has bounded density, denoted by f^x . Local polynomial smoothing is based on the assumption that functional parameter is smooth enough to be locally well approximated by a polynomial. In functional statistics, there are several ways for extending the local linear ideas (cf. [1, 2, 5]).

Here we adopt the fast functional locally modeling, that is, we estimate the conditional density f^x by \hat{a} which is obtained by minimizing the following quantity :

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

where $\beta(\cdot, \cdot)$ is a known function from \mathcal{F}^2 into \mathbb{R} such that, $\forall \xi \in \mathcal{F}$, $\beta(\xi, \xi) = 0$, with K and H are kernels and $h_K = h_{K,n}$ (resp. $h_H = h_{H,n}$) is chosen as 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 get explicitly the following definition of \hat{f}^x :

$$\hat{f}^x(y) = \frac{\sum_{i,j=1}^n W_{ij}(x) H(h_H^{-1}(y - Y_i))}{h_H \sum_{i,j=1}^n W_{ij}(x)} \quad (2)$$

where

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

with the convention $0/0 = 0$.

Remark 3.2.1.

- Obviously, if $b = 0$ then we obtain from (1) the Nadaraya-Watson estimator studied, in the functional case, in [9, 10, 28] and the references therein.
- The minimization of (1) may be achieved by a “wiggly” \hat{b} that forces \hat{f}^x to adapt to all the data points in a neighborhood of x (see Chapter 15 in [29] for a similar reasoning in the context of linear regression). In [6] is expressed the same idea stating that optimizing in b is an infinite-dimensional problem.

3.3 Pointwise almost complete convergence

In what follows x denotes a fixed point in \mathcal{F} , N_x denotes a fixed neighborhood of x , $S_{\mathbb{R}}$ will be a fixed compact subset of \mathbb{R} , and $\phi_x(r_1, r_2) = \mathbb{P}(r_2 \leq \delta(X, x) \leq r_1)$.

Notice that our nonparametric model will be quite general in the sense that we will just need the following assumptions :

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

(H2) The conditional density f^x is such that : there exist $b_1 > 0$, $b_2 > 0$, $\forall (y_1, y_2) \in S_{\mathbb{R}}^2$ and $\forall (x_1, x_2) \in N_x \times N_x$

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

where C_x is a positive constant depending on x .

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

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

(H4) K is a positive, differentiable function with support $[-1, 1]$.

(H5) H is a positive, bounded, Lipschitzian continuous function, such that :

$$\int |t|^{b_2} H(t) dt < \infty \text{ and } \int H^2(t) dt < \infty.$$

(H6) The bandwidth h_K satisfies : there exists an integer n_0 , such that

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

and

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

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 .

(H7) The bandwidth h_H satisfies :

$$\lim_{n \rightarrow \infty} n^\gamma h_H = \infty \text{ for some } \gamma > 0 \text{ and } \lim_{n \rightarrow \infty} \frac{\ln n}{nh_H \phi_x(h_K)} = 0.$$

Observe that these conditions are very standard in this context. The conditions (H1), (H3) and (H6) are the same as those used in [1]. Assumptions (H2) is a regularity condition which characterizes the functional space of our model and is needed to evaluate the bias term in the asymptotic results of this paper. The hypotheses (H5) and (H7) are technical conditions and are, also, similar to those considered in [10].

The following theorem gives the almost-complete convergence⁵ (a.co.) of \widehat{f}^x .

Theorem 3.3.1. *Under assumptions (H1), (H2), (H3), (H4), (H5), (H6) and (H7), we have that :*

$$\sup_{y \in S_{\mathbb{R}}} |\widehat{f}^x(y) - f^x(y)| = O \left(h_K^{b_1} + h_H^{b_2} \right) + O \left(\sqrt{\frac{\ln n}{nh_H \phi_x(h_K)}} \right), \text{ a.co.}$$

Remark that, the proof of Theorem 4.2.1 is a direct consequence of the decomposition :

$$\begin{aligned} \forall y \in S_{\mathbb{R}}, \widehat{f}^x(y) - f^x(y) &= \frac{1}{\widehat{f}_D^x} \left\{ \left(\widehat{f}_N^x(y) - \mathbb{E}[\widehat{f}_N^x(y)] \right) - \left(f^x(y) - \mathbb{E}[\widehat{f}_N^x(y)] \right) \right\} \\ &+ \frac{f^x(y)}{\widehat{f}_D^x} \left(1 - \widehat{f}_D^x \right), \end{aligned} \quad (3)$$

where

$$\widehat{f}_N^x(y) = \frac{1}{n(n-1)h_H \mathbb{E}[W_{12}(x)]} \sum_{i \neq j} W_{ij}(x) H(h_H^{-1}(y - Y_i)).$$

and

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

and of Lemmas 5.3.1, 5.3.2 and 5.3.3 below, for which the proofs are given in the Appendix.

5. Let $(z_n)_{n \in \mathbb{N}}$ be a sequence of real r.v.'s; we say that z_n converges almost completely (a.co.) to zero if, and only if, $\forall \epsilon > 0, \sum_{n=1}^{\infty} \mathbb{P}(|z_n| > \epsilon) < \infty$. Moreover, let $(u_n)_{n \in \mathbb{N}^*}$ be a sequence of positive real numbers; we say that $z_n = O(u_n)$ a.co. if, and only if, $\exists \epsilon > 0, \sum_{n=1}^{\infty} \mathbb{P}(|z_n| > \epsilon u_n) < \infty$: This kind of convergence implies both almost sure convergence and convergence in probability (cf. [13] for details).

Lemma 3.3.1. (cf. [1])

Under assumptions (H1), (H3), (H4) and (H6), we have that :

$$1 - \widehat{f}_D^x = O\left(\sqrt{\frac{\ln n}{n \phi_x(h_K)}}\right), \text{ a.co.}$$

and

$$\exists \delta > 0, \text{ such that } \sum_{i=1}^{\infty} \mathbb{P}\left(\widehat{f}_D^x < \delta\right) < \infty.$$

Lemma 3.3.2. Under assumptions (H1), (H2) and (H4), we obtain :

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

Lemma 3.3.3. Under assumptions of Theorem 4.2.1, we get :

$$\sup_{y \in S_{\mathbb{R}}} \left| \widehat{f}_N^x(y) - \mathbb{E}[\widehat{f}_N^x(y)] \right| = O\left(\sqrt{\frac{\ln n}{n h_H \phi_x(h_K)}}\right), \text{ a.co.}$$

3.4 Uniform almost complete convergence

This Section is devoted to the uniform version of Theorem 4.2.1. More precisely, our purpose is to establish the uniform almost complete convergence of \widehat{f}^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.

In practice, the uniform consistency has great importance because it allows us to make prediction, even if the data are not perfectly observed. Moreover, the uniform convergence results are indispensable tools for some nonparametric functional data problems such as for instance data-driven bandwidth choice (cf. [4, 28]), or bootstrapping (cf. [16]). It is worth noting that, in the multivariate case, the uniform consistency is a standard extension of the pointwise one, however, in our functional case, some additional tools and topological conditions are required. Thus, in addition to the conditions introduced in the previous section, we need the following ones.

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

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

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

(U2) The conditional density f^x satisfies, for some strictly positive constant C , that :

$$\begin{aligned} \forall (y_1, y_2) \in S_{\mathbb{R}} \times S_{\mathbb{R}}, \forall (x_1, x_2) \in S_{\mathcal{F}} \times S_{\mathcal{F}} : \\ |f^{x_1}(y_1) - f^{x_2}(y_2)| \leq C \left(d^{b_1}(x_1, x_2) + |y_1 - y_2|^{b_2} \right). \end{aligned}$$

(U3) The function $\beta(\cdot, \cdot)$ satisfies (H3) and, for some strictly positive constant C' , the following Lipschitz's condition :

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

(U4) The kernel K satisfies (H4) and, for some strictly positive constant C , the following Lipschitz's condition :

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

(U5) For some $\gamma \in (0, 1)$, $\lim_{n \rightarrow +\infty} n^\gamma h_H = \infty$, and for $r_n = O\left(\frac{\ln n}{n}\right)$ the sequence d_n satisfies :

$$\frac{(\ln n)^2}{n^{1-\gamma} \phi(h_K)} < \ln d_n < \frac{n^{1-\gamma} \phi(h_K)}{\ln n},$$

and

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

Notice that conditions (U1) and (U2) are, respectively, the uniform versions of (H1) and (H2). Indeed, conditions (U1) and (U5) are linked with the topological structure of the functional variable. Therefore, as in the pointwise case, the choice of the topological structure, controlled here by means of the function $\delta(\cdot, \cdot)$, plays a crucial role. So, a right choice of this function improves the convergence rate of the estimator. More precisely, we will see thereafter, that a good semi-metric is that increases the concentration of the probability measure of the functional variable X as well as minimizes d_n . It should be noticed that, both conditions (U1) and (U2) are verified for several continuous time-processes (cf. for instance [9] for some examples).

Theorem 3.4.1. *Under assumptions (U1), (U2), (U3), (U4), (U5), (H5) and (H6), we have that :*

$$\sup_{x \in S_{\mathcal{F}}} \sup_{y \in S_{\mathbb{R}}} |\hat{f}^x(y) - f^x(y)| = O(h_K^{b_1}) + O(h_H^{b_2}) + O_{a.co.} \left(\sqrt{\frac{\ln d_n}{n^{1-\gamma} \phi(h_K)}} \right). \quad (4)$$

It is clear that, as for Theorem 4.2.1, the Theorem 3.4.1's proof can be deduced directly from the decomposition (1) and from the following intermediate results which correspond to the uniform versions of Lemmas 5.3.1, 5.3.2, 5.3.3 and for which the proofs are, also, given in the Appendix.

Lemma 3.4.1. *Under assumptions (U1), (U3), (U4), (U5) and (H6), we obtain that :*

$$\sup_{x \in S_{\mathcal{F}}} |\hat{f}_D^x - 1| = O_{a.co.} \left(\sqrt{\frac{\ln d_n}{n \phi(h_K)}} \right).$$

Corollary 3.4.1. *Under the assumptions of Lemma 3.4.1, we have that :*

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

Lemma 3.4.2. *Under the hypotheses (U1), (U2) and (H5), we obtain that :*

$$\sup_{x \in S_{\mathcal{F}}} \sup_{y \in S_{\mathbb{R}}} |f^x(y) - \mathbb{E} [\hat{f}_N^x(y)]| = O(h_K^{b_1}) + O(h_H^{b_2}).$$

Lemma 3.4.3. *Under the assumptions of Theorem 3.4.1, we obtain that :*

$$\sup_{x \in S_{\mathcal{F}}} \sup_{y \in S_{\mathbb{R}}} |\widehat{f}_N^x(y) - \mathbb{E}[\widehat{f}_N^x(y)]| = O_{a.co.} \left(\sqrt{\frac{\ln d_n}{n^{1-\gamma} \phi(h_K)}} \right).$$

3.5 Application : Conditional mode estimation

Let us now study the almost complete convergence of the kernel estimator of the conditional mode of Y given $X = x$, denoted by $\theta(x)$, uniformly on a fixed ‘‘compact’’ subset $S_{\mathcal{F}}$ of \mathcal{F} . For this aim, we assume that $\theta(x)$ satisfies, on $S_{\mathcal{F}}$, the following uniform uniqueness property (cf. [27, 34] for the univariate case and [26] for the multivariate case).

(U6) $\forall \epsilon_0 > 0, \exists \eta > 0, \forall r : S \rightarrow S_{\mathbb{R}}$, we have that :

$$\sup_{x \in S_{\mathcal{F}}} |\theta(x) - r(x)| \geq \epsilon_0 \Rightarrow \sup_{x \in S_{\mathcal{F}}} |f^x(r(x)) - f^x(\theta(x))| \geq \eta.$$

Moreover, we suppose, also, that there exists some integer $j > 1$ such that $\forall x \in S_{\mathcal{F}}$, the function f^x is j -times continuously differentiable on *interior*($S_{\mathbb{R}}$) with respect to y , and that :

(U7)

$$\left\{ \begin{array}{l} f^{x(l)}(\theta(x)) = 0, \text{ if } 1 \leq l < j \\ \text{and } f^{x(j)}(\cdot) \text{ is uniformly continuous on } S_{\mathbb{R}} \\ \text{such that } |f^{x(j)}(\theta(x))| > C > 0 \end{array} \right.$$

where $f^{x(j)}$ denotes the j th order derivative of the conditional density f^x .

We estimate the conditional mode $\theta(x)$ by the random variable $\widehat{\theta}(x)$ defined by :

$$\widehat{\theta}(x) = \arg \sup_{y \in S_{\mathbb{R}}} \widehat{f}^x(y).$$

Thus, from Theorem 3.4.1, we derive the following Corollary.

Corollaire 3.5.1. *Under the hypotheses of Theorem 3.4.1, and if the conditional density f^x satisfies assumptions (H9) and (H10), then we get :*

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

3.6 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 $i = 1, \dots, n$:

$$K_i(x) = K(h^{-1}\delta(x, X_i)), \beta_i(x) = \beta(X_i, x) \text{ and } H_i(y) = H(h_H^{-1}(y - Y_i)).$$

Proof of lemma 5.3.2. Since the pairs (X_i, Y_i) are identically distributed, then from assumption (H4) we obtain :

$$\forall y \in S_{\mathbb{R}}, \mathbb{E}[\widehat{f}_N^x(y)] = \mathbb{E} [W_{12}(x) [\mathbb{E} [h_H^{-1} H_1(y)/X]]].$$

By the classical change of variables $t = (y - z)/h_H$, we obtain :

$$h_H^{-1} \mathbb{E} [H_1(y)/X] = \int_{\mathbb{R}} H(t) f^X(y - h_H t) dt,$$

therefore

$$|\mathbb{E}[H_1(y)/X] - f^x(y)| \leq \int_{\mathbb{R}} H(t) |f^X(y - h_H t) - f^x(y)| dt.$$

Thus, by the assumption (H2) we get that :

$$\forall y \in S_{\mathbb{R}}, \mathbf{1}_{B(x, h_K)}(X) |\mathbb{E}(H_1(y)/X) - f^x(y)| \leq \int_{\mathbb{R}} H(t) (h_K^{b_1} + |t|^{b_2} h_H^{b_2}) dt.$$

Since H is a probability density function, then the claimed result in this lemma is a direct consequence of (H5). \blacksquare

Proof of lemma 5.3.3. The compactness property of $S_{\mathbb{R}}$, allows us to write that : there exists a sequence of real numbers $(t_k)_{k=1, \dots, s_n}$, such that $S_{\mathbb{R}} \subset \bigcup_{k=1}^{s_n} (t_k - l_n, t_k + l_n)$ with $l_n = n^{-\frac{3\gamma}{2} - \frac{1}{2}}$ and $s_n = O(l_n^{-1})$.

Let $t_y = \arg \min_{t \in \{t_1, \dots, t_{s_n}\}} |y - t|$ and consider the following decomposition :

$$\begin{aligned} \sup_{y \in S_{\mathbb{R}}} \left| \widehat{f}_N^x(y) - \mathbb{E}[\widehat{f}_N^x(y)] \right| &\leq \underbrace{\sup_{y \in S_{\mathbb{R}}} \left| \widehat{f}_N^x(y) - \widehat{f}_N^x(t_y) \right|}_{A_1} + \underbrace{\sup_{y \in S_{\mathbb{R}}} \left| \widehat{f}_N^x(t_y) - \mathbb{E}[\widehat{f}_N^x(t_y)] \right|}_{A_2} + \\ &+ \underbrace{\sup_{y \in S_{\mathbb{R}}} \left| \mathbb{E}[\widehat{f}_N^x(t_y)] - \mathbb{E}[\widehat{f}_N^x(y)] \right|}_{A_3}. \end{aligned} \quad (5)$$

Firstly, for the terms A_1 and A_3 , we use the Lipschitz's condition on the kernel H to show that :

$$\begin{aligned} \sup_{y \in S_{\mathbb{R}}} \left| \widehat{f}_N^x(y) - \widehat{f}_N^x(t_y) \right| &\leq \sup_{y \in S_{\mathbb{R}}} \frac{1}{n(n-1) h_H \mathbb{E}[W_{12}(x)]} \sum_{i \neq j} |H_i(y) - H_i(t_y)| W_{ij}(x), \\ &\leq \sup_{y \in S} \frac{C|y - t_y|}{h_H} \left(\frac{1}{n(n-1) h_H \mathbb{E}[W_{12}(x)]} \sum_{i \neq j} W_{ij}(x) \right), \\ &\leq C \frac{l_n}{h_H^2} \widehat{f}_D^x. \end{aligned} \quad (6)$$

The almost complete consistency of \widehat{f}_D^x (cf. Lemma 5.3.1) and (14) permit to write that :

$$\sup_{y \in S_{\mathbb{R}}} \left| \widehat{f}_N^x(y) - \widehat{f}_N^x(t_y) \right| \leq C \frac{l_n}{h_H^2}$$

Since $l_n = n^{-\frac{3\gamma}{2} - \frac{1}{2}}$, then

$$\sup_{y \in S_{\mathbb{R}}} \left| \widehat{f}_N^x(y) - \widehat{f}_N^x(t_y) \right| = o \left(\sqrt{\frac{\ln n}{n h_H \phi_x(h_K)}} \right). \quad (7)$$

Indeed, the term A_3 may be considered as a direct consequence of the following known inequality :

$$\sup_{y \in S_{\mathbb{R}}} \left| \mathbb{E}[\widehat{f}_N^x(y)] - \mathbb{E}[\widehat{f}_N^x(t_y)] \right| \leq \mathbb{E} \left[\sup_{y \in S_{\mathbb{R}}} \left| \widehat{f}_N^x(y) - \widehat{f}_N^x(t_y) \right| \right]. \quad (8)$$

Secondly, about the term A_2 , we can write for all $\eta > 0$:

$$\begin{aligned} & \mathbb{P} \left(\sup_{y \in S_{\mathbb{R}}} \left| \widehat{f}_N^x(t_y) - \mathbb{E}[\widehat{f}_N^x(t_y)] \right| > \eta \sqrt{\frac{\ln n}{n h_H \phi_x(h_K)}} \right) \\ &= \mathbb{P} \left(\max_{t_y \in \{t_1, \dots, t_{s_n}\}} \left| \widehat{f}_N^x(t_y) - \mathbb{E}[\widehat{f}_N^x(t_y)] \right| > \eta \sqrt{\frac{\ln n}{n h_H \phi_x(h_K)}} \right) \\ &\leq s_n \max_{t_y \in \{t_1, \dots, t_{s_n}\}} \mathbb{P} \left(\left| \widehat{f}_N^x(t_y) - \mathbb{E}[\widehat{f}_N^x(t_y)] \right| > \eta \sqrt{\frac{\ln n}{n h_H \phi_x(h_K)}} \right). \end{aligned}$$

all it remains to compute is the following quantity :

$$\mathbb{P} \left(\left| \widehat{f}_N^x(t_y) - \mathbb{E}[\widehat{f}_N^x(t_y)] \right| > \eta \sqrt{\frac{\ln n}{n h_H \phi_x(h_K)}} \right), \text{ for all } t_y \in \{t_1, \dots, t_{s_n}\}.$$

This later quantity's value is given by a straightforward adaptation of the proof of Lemma 2 in [1]. To do that, we consider the following decomposition :

$$\begin{aligned} \widehat{f}_N^x(t_y) &= \underbrace{\frac{n^2 h_K^2 \phi_x^2(h_K)}{n(n-1) \mathbb{E}[W_{12}]} }_{T_1} \left[\underbrace{\left(\frac{1}{n} \sum_{j=1}^n \frac{K_j(x) H_j(t_y)}{h_H \phi_x(h_K)} \right)}_{T_2} \underbrace{\left(\frac{1}{n} \sum_{i=1}^n \frac{K_i(x) \beta_i^2(x)}{h_K^2 \phi_x(h_K)} \right)}_{T_3} \right. \\ &\quad \left. - \underbrace{\left(\frac{1}{n} \sum_{j=1}^n \frac{K_j(x) \beta_j(x) H_j(t_y)}{h_H h_K \phi_x(h_K)} \right)}_{T_4} \underbrace{\left(\frac{1}{n} \sum_{i=1}^n \frac{K_i(x) \beta_i(x)}{h_K \phi_x(h_K)} \right)}_{T_5} \right]. \end{aligned}$$

It follows that :

$$\widehat{f}_N^x(t_y) - \mathbb{E}[\widehat{f}_N^x(t_y)] = T_1((T_2 T_3 - \mathbb{E}[T_2 T_3]) - (T_4 T_5 - \mathbb{E}[T_4 T_5])).$$

Moreover, observe that :

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

and in the same way :

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

So, the claimed result will be obtained as soon as the following assertions have been checked :

$$\sum_n s_n \mathbb{P} \left\{ |T_i - \mathbb{E}[T_i]| > \eta \sqrt{\frac{\ln n}{n h_H \phi_x(h_K)}} \right\} < \infty, \quad \text{for } i = 2, 3, 4, 5, \quad (9)$$

$$T_1 = O(1) \quad \text{and} \quad \mathbb{E}[T_i] = O(1) \quad \text{for } i = 2, 3, 4, 5, \quad (10)$$

and almost completely :

$$|\mathbb{E}[T_2]\mathbb{E}[T_3] - \mathbb{E}[T_2T_3] - \mathbb{E}[T_4]\mathbb{E}[T_5] + \mathbb{E}[T_4T_5]| = o \left(\sqrt{\frac{\ln n}{n h_H \phi_x(h_K)}} \right). \quad (11)$$

Proof of (17) : For this aim, we use the Bernstein's exponential inequality for which the main point is to evaluate asymptotically the m th order moment of :

$$Z_i^{l,k} = \frac{1}{h_K^l h_H^k \phi_x(h_K)} \left(K_i(x) H_i^k(t_y) \beta_i^l(x) - \mathbb{E} \left[K_i(x) H_i^k(t_y) \beta_i^l(x) \right] \right)$$

for $l = 0, 1, 2$, and $k = 0, 1$.

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

$$\begin{aligned} & \mathbb{E} \left| \left(K_i(x) H_i^k(t_y) \beta_i^l(x) - \mathbb{E} \left[K_i(x) H_i^k(t_y) \beta_i^l(x) \right] \right)^m \right| \\ &= \mathbb{E} \left| \sum_{d=0}^m C_m^d \left(K_i(x) H_i^k(t_y) \beta_i^l(x) \right)^d \left(\mathbb{E} \left[K_i(x) H_i^k(t_y) \beta_i^l(x) \right] \right)^{m-d} (-1)^{m-d} \right| \\ &\leq \sum_{d=0}^m C_m^d \left(\mathbb{E} \left| K_i(x) H_i^k(t_y) \beta_i^l(x) \right|^d \right) \left| \mathbb{E} \left[K_i(x) H_i^k(t_y) \beta_i^l(x) \right] \right|^{m-d} \\ &\leq \sum_{d=0}^m C_m^d \mathbb{E} \left| K_1^\delta(x) \beta_1^{dl}(x) \mathbb{E}[H_1^{dk}(t_y)|X_1] \right| \left| \mathbb{E} \left[K_1(x) \beta_1^l(x) \mathbb{E}[H_1^k(t_y)|X_1] \right] \right|^{m-d} \end{aligned}$$

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

Using the same arguments as those invoked in the proof of Lemma 5.3.2, and replacing H by H^d , we show that, for all $d \leq m$:

$$\mathbb{E} \left[H_1^\delta(t_y)/X \right] = h_H \int_{\mathbb{R}} H^\delta(t) f^X(t_y - h_H t) dt.$$

So, conditions (H2) and (H5) allow us to write that :

$$\mathbb{E} \left[H_1^{dk}(t_y)/X \right] = O(h_H^k), \quad \text{for all } d \leq m \text{ and } k = 0, 1.$$

Moreover, it is shown in [1] that :

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

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

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

Thus, to achieve this proof, it suffices to use the classical Bernstein's inequality (see Corollary A8 in [11], page 234), first with $a_n = (h_H \phi_x(h_K))^{-1/2}$ to treat the terms T_2 and T_4 , and second with $a_n = (\phi_x(h_K))^{-1/2}$ for the terms T_3 and T_5 . In conclusion, we obtain for all $\eta > 0$:

$$\begin{aligned} \mathbb{P} \left\{ |T_2 - \mathbb{E}[T_2]| > \eta \sqrt{\frac{\ln n}{n h_H \phi_x(h_K)}} \right\} &\leq C' n^{-C\eta^2}, \\ \mathbb{P} \left\{ |T_4 - \mathbb{E}[T_4]| > \eta \sqrt{\frac{\ln n}{n h_H \phi_x(h_K)}} \right\} &\leq C' n^{-C\eta^2} \end{aligned}$$

and for $i = 3, 5$

$$\mathbb{P} \left\{ |T_i - \mathbb{E}[T_i]| > \eta \sqrt{\frac{\ln n}{n h_H \phi_x(h_K)}} \right\} \leq \mathbb{P} \left\{ |T_i - \mathbb{E}[T_i]| > \eta \sqrt{\frac{\ln n}{n \phi_x(h_K)}} \right\} \leq C' n^{-C\eta^2}.$$

Therefore, an appropriate choice of η permits to deduce that :

$$s_n \mathbb{P} \left\{ |T_i - \mathbb{E}[T_i]| > \eta \sqrt{\frac{\ln n}{n h_H \phi_x(h_K)}} \right\} \leq C' n^{-1-\gamma}, \text{ for } i = 2, 3, 4, 5.$$

Proofs of (18) and (25). Notice that, the first part of (18) has been treated in [1]. We now proceed in proving the second part of (18) and (25). For this aim, since the pairs (X_i, Y_i) , $i = 1, \dots, n$ are identically distributed, we obtain that :

$$\left\{ \begin{array}{l} \mathbb{E}[T_2] = \frac{\mathbb{E}[K_1(x)H_1(t_y)]}{h_H \phi_x(h_K)}, \quad \mathbb{E}[T_3] = \frac{\mathbb{E}[K_1(x)\beta_1^2(x)]}{h_K \phi_x(h_K)}, \\ \mathbb{E}[T_4] = \frac{\mathbb{E}[K_1(x)H_1(t_y)\beta_1(x)]}{h_K h_H \phi_x(h_K)}, \quad \mathbb{E}[T_5] = \frac{\mathbb{E}[K_1(x)\beta_1(x)]}{h_K \phi_x(h_K)}, \\ \text{and} \\ \mathbb{E}[T_2]\mathbb{E}[T_3] - \mathbb{E}[T_2T_3] - \mathbb{E}[T_4]\mathbb{E}[T_5] + \mathbb{E}[T_4T_5] \\ = \left(1 - \frac{n(n-1)}{n^2}\right) h_K^{-2} \phi_x(h_K)^{-2} \mathbb{E}[K_1(x)\beta_1^2(x)] \mathbb{E}[K_1(x)H_1(t_y)] \end{array} \right.$$

Thus, for both equations (18) and (25), we have to evaluate :

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

As previously, we condition on X_1 to show that, for all $l = 0, 1, 2$, and $k = 0, 1$, we have :

$$\mathbb{E} \left[K_i(x) H_i^k(t_y) \beta_i^l(x) \right] = O(h_H^k \mathbb{E} \left[K_i(x) \beta_i^l(x) \right])$$

and by Lemma 3 in [1], we obtain that :

$$\mathbb{E} \left[K_i(x) H_i^k(t_y) \beta_i^l(x) \right] = O(h_H^k h_K^l \phi_x(h_K)). \quad (12)$$

Equality (12) leads directly to :

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

$$|\mathbb{E}[T_2]\mathbb{E}[T_3] - \mathbb{E}[T_2T_3] - \mathbb{E}[T_4]\mathbb{E}[T_5] + \mathbb{E}[T_4T_5]| = O(h_H n^{-1})$$

which implies that :

$$\mathbb{E}[T_2]\mathbb{E}[T_3] - \mathbb{E}[T_2T_3] - \mathbb{E}[T_4]\mathbb{E}[T_5] + \mathbb{E}[T_4T_5] = O\left(\sqrt{\frac{\ln n}{n h_H \phi_x(h_K)}}\right).$$

Finally, the Lemma 5.3.3 is a direct consequence of the assertions (7), (8), (17), (18) and (25). \blacksquare

Proof of Lemma 3.4.1. The proof of this lemma is based on the same decomposition's kind as used to prove Lemma 5.3.3. Indeed,

$$\widehat{f}_D^x = T_1 \left[\underbrace{\left(\frac{1}{n} \sum_{j=1}^n \frac{K_j(x)}{\phi_x(h_K)} \right)}_{S_2(x)} \underbrace{\left(\frac{1}{n} \sum_{i=1}^n \frac{K_i(x)\beta_i^2(x)}{h_K^2 \phi_x(h_K)} \right)}_{S_4(x)} - \underbrace{\left(\frac{1}{n} \sum_{j=1}^n \frac{K_j(x)\beta_j(x)}{h_K \phi_x(h_K)} \right)}_{S_3(x)} \underbrace{\left(\frac{1}{n} \sum_{i=1}^n \frac{K_i(x)\beta_i(x)}{h_K \phi_x(h_K)} \right)}_{S_3(x)} \right]$$

and in the same fashion, all it remains to show are the following uniform convergences :

$$\sup_{x \in S_{\mathcal{F}}} |S_k(x) - \mathbb{E}[S_k(x)]| = O\left(\sqrt{\frac{\ln d_n}{n \phi_x(h_K)}}\right), \quad a.co. \text{ for } k = 2, 3, 4, \quad (13)$$

and $\sup_{x \in S_{\mathcal{F}}} |\mathbb{E}[S_2(x)]\mathbb{E}[S_4(x)] - \mathbb{E}[S_2(x)S_4(x)] - \text{var}[S_3(x)]| = o\left(\sqrt{\frac{\ln d_n}{n \phi_x(h_K)}}\right)$, *a.co.* and, also that, uniformly on $x \in S_{\mathcal{F}}$: $T_1 = O(1)$ and $|\mathbb{E}[S_k(x)]| = O(1)$, for $k = 2, 3, 4$. Clearly, the last two equations are direct consequences of the assumption (U1) and of the Lemma 3 in [1]. While the proof of (13) follows the same ideas as in [9]. Indeed, by noting : $j(x) = \arg \min_{j \in \{1, 2, \dots, d_n\}} |\delta(x, x_j)|$, we consider the following decomposition :

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

We have, then, to evaluate each term F_j^k , $j = 1, 2, 3$. Since F_1^k and F_3^k have almost the same treatment, we will consider the following two items :

- *Treatment of the terms F_1^k and F_3^k .* Firstly, let us analyze the first term F_1^k for $k = 2, 3, 4$. Since K is supported in $[-1, 1]$, we can write for all $k = 2, 3, 4$ that :

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

The Lipschitz condition on K allows us directly to write

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

In a similar way, the Lipschitz condition on β gives

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

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

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

Thus,

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

where

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

Now, to evaluate these terms F_{11}^k , F_{12} , F_{13}^k and F_{14} , we apply a standard inequality for sums of bounded random variables (cf. Corollary A.9 in [11]) with Z_i is identified such that :

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

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

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

So that, we get :

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

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

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

which implies that :

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

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

$$F_1^k = O_{a.co.} \left(\sqrt{\frac{\ln d_n}{n \phi(h_K)}} \right). \quad (14)$$

Furthermore, since :

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

we have also :

$$F_3^k = O\left(\sqrt{\frac{\ln d_n}{n \phi(h_K)}}\right).$$

– *Treatment of the term F_2^k .* For all $\eta > 0$, we have that :

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

Set,

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

By using similar proof as for showing Lemma 2 of [1], we get for all $j = 1, \dots, d_n$ and $i = 1, \dots, n$, that :

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

So, one can apply a Bernstein-type inequality (cf. Corollary A.8 in [11]) to obtain directly :

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

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

$$d_n \max_{k \in \{1, \dots, d_n\}} \mathbb{P} \left(|S_i(x_k) - \mathbb{E}[S_i(x_k)]| > \eta \sqrt{\frac{\ln d_n}{n\phi(h_K)}} \right) \leq C' d_n^{1-\beta}. \quad (15)$$

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

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

■

Proof of Corollary 3.4.1. Clearly, we have that :

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

According to the Lemma 3.4.1, we obtain :

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

Consequently :

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

■

Proof of Lemma 3.4.2. It suffices to combine the proofs of the previous lemmas, and assuming the Lipschitz's condition uniformly on (x, y) in $S_{\mathcal{F}} \times S_{\mathbb{R}}$. ■

Proof of Lemma 3.4.3. The proof of this lemma follows the steps as for proving Lemma 3.4.1, where $S_2(x)$ and $S_4(x)$ are replaced by :

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

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

$$\begin{aligned} |T_i^x(y) - \mathbb{E}[T_i^x(y)]| &\leq \underbrace{\sup_{x \in \mathcal{S}_{\mathcal{F}}} \sup_{y \in \mathcal{S}_{\mathbb{R}}} |T_i^x(y) - T_i^{x_{j(x)}}(y)|}_{E_1} \\ &+ \underbrace{\sup_{x \in \mathcal{S}_{\mathcal{F}}} \sup_{y \in \mathcal{S}_{\mathbb{R}}} |T_i^{x_{j(x)}}(y) - T_i^{x_{j(x)}}(t_y)|}_{E_2} \\ &+ \underbrace{\sup_{x \in \mathcal{S}_{\mathcal{F}}} \sup_{y \in \mathcal{S}_{\mathbb{R}}} |T_i^{x_{j(x)}}(t_y) - \mathbb{E}[T_i^{x_{j(x)}}(t_y)]|}_{E_3} \\ &+ \underbrace{\sup_{x \in \mathcal{S}_{\mathcal{F}}} \sup_{y \in \mathcal{S}_{\mathbb{R}}} |\mathbb{E}[T_i^{x_{j(x)}}(t_y)] - \mathbb{E}[T_i^{x_{j(x)}}(y)]|}_{E_4} \\ &+ \underbrace{\sup_{x \in \mathcal{S}_{\mathcal{F}}} \sup_{y \in \mathcal{S}_{\mathbb{R}}} |\mathbb{E}[T_i^{x_{j(x)}}(y)] - \mathbb{E}[T_i^x(y)]|}_{E_5}. \end{aligned}$$

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

$$E_1 = O_{a.co.} \left(\sqrt{\frac{\ln d_n}{n^{1-\gamma} \phi(h_K)}} \right) \text{ and } E_5 = O \left(\sqrt{\frac{\ln d_n}{n^{1-\gamma} \phi(h_K)}} \right). \quad (16)$$

Concerning the term E_2 , by using the Lipschitz's condition on the kernel H , one can write :

$$|T_i^{x_{j(x)}}(y) - T_i^{x_{j(x)}}(t_y)| \leq C \frac{1}{n h_K^l h_H \phi(h_K)} \sum_{i'=1}^n K_{i'}(x_{j(x)}) \beta_{i'}^l(x_{j(x)}) |H_{i'}(y) - H_{i'}(t_y)| \leq \frac{l_n}{h_H^2} S_i(x_{j(x)}),$$

where

$S_i(\cdot)$ for $i = 2, 3, 4$, are defined and treated in Lemma 3.4.1's proof. Thus, by using the facts that : $\lim_{n \rightarrow +\infty} n^\gamma h_H = \infty$ and $l_n = n^{-\frac{3}{2}\gamma - \frac{1}{2}}$, we obtain :

$$E_2 = O_{a.co.} \left(\sqrt{\frac{\ln d_n}{n^{1-\gamma} \phi(h_K)}} \right) \text{ and } E_4 = O \left(\sqrt{\frac{\ln d_n}{n^{1-\gamma} \phi(h_K)}} \right). \quad (17)$$

Finally, for the term E_3 , we use the same arguments as for proving Lemma 3.4.1, to show that for all $\eta > 0$,

$$\mathbb{P} \left(E_3 > \eta \sqrt{\frac{\ln d_n}{n h_H \phi(h_K)}} \right)$$

$$\begin{aligned}
&= \mathbb{P} \left(\max_{j \in \{1, 2, \dots, s_n\}} \max_{k \in \{1, \dots, d_n\}} |T_i^{x_k}(t_j) - \mathbb{E}[T_i^{x_k}(t_j)]| > \eta \sqrt{\frac{\ln d_n}{n h_H \phi(h_K)}} \right) \\
&\leq s_n d_n \max_{j \in \{1, 2, \dots, s_n\}} \max_{k \in \{1, \dots, d_n\}} \mathbb{P} \left(|T_i^{x_k}(t_j) - \mathbb{E}[T_i^{x_k}(t_j)]| > \eta \sqrt{\frac{\ln d_n}{n h_H \phi(h_K)}} \right).
\end{aligned}$$

This last probability can be treated by using the classical Bernstein's inequality, with $a_n = (h_H \phi_x(h_K))^{-1/2}$. Recall that, the choice of a_n is motivated by moment of order m of $Z_i^{l,k}$ computed in Lemma 5.3.3's proof. That allows, finally, to :

$$\forall j \leq s_n, \mathbb{P} \left(|T_i^{x_k}(t_j) - \mathbb{E}[T_i^{x_k}(t_j)]| > \eta \sqrt{\frac{\ln d_n}{n h_H \phi(h_K)}} \right) \leq 2 \exp\{-C\eta^2 \ln d_n\}.$$

Therefore, since $s_n = O(l_n^{-1}) = O(n^{\frac{3}{2}\gamma + \frac{1}{2}})$, and by choosing $C\eta^2 = \beta$ one has :

$$s_n d_n \max_{j \in \{1, 2, \dots, s_n\}} \max_{k \in \{1, \dots, d_n\}} \mathbb{P} \left(|T_i^{x_k}(t_j) - \mathbb{E}[T_i^{x_k}(t_j)]| > \eta \sqrt{\frac{\ln d_n}{n h_H \phi(h_K)}} \right) \leq C' s_n d_n^{1-C\eta^2}.$$

By using the fact that $\lim_{n \rightarrow +\infty} n^\gamma h_H = \infty$ and the second part of condition (U5), one obtains :

$$E_3 = O_{a.co.} \left(\sqrt{\frac{\ln d_n}{n^{1-\gamma} \phi(h_K)}} \right). \quad (18)$$

Thus, Lemma 3.4.3's result can be easily deduced from (16), (17) and (18). \blacksquare

Proof of Corollary 4.3.4. By a simple manipulation, we show that :

$$|f^x(\widehat{\theta}(x)) - f^x(\theta(x))| \leq 2 \sup_{y \in S_{\mathbb{R}}} |\widehat{f}^x(y) - f^x(y)|. \quad (19)$$

We use the following Taylor expansion of the function f^x :

$$f^x(\widehat{\theta}(x)) = f^x(\theta(x)) + \frac{1}{j!} f^{x(j)}(\theta'(x)) (\widehat{\theta}(x) - \theta(x))^j,$$

for some $\theta'(x)$ between $\theta(x)$ and $\widehat{\theta}(x)$. It is clear that, from conditions (U6), (29) and Theorem 3.4.1, we obtain that :

$$\sup_{x \in S_{\mathcal{F}}} |\widehat{\theta}(x) - \theta(x)| \rightarrow 0, \text{ a.co.}$$

Next, under condition (U7), we obtain that :

$$\sup_{x \in S_{\mathcal{F}}} |f^{x(j)}(\theta'(x)) - f^{x(j)}(\theta(x))| \rightarrow 0, \text{ a.co.}$$

Consequently, we can get $\tau > 0$ such that :

$$\sum_{n=1}^{\infty} \mathbb{P} \left(\inf_{x \in S_{\mathcal{F}}} f^{x(j)}(\theta'(x)) < \tau \right) < \infty,$$

and we have :

$$\sup_{x \in S_{\mathcal{F}}} |\widehat{\theta}(x) - \theta(x)|^j \leq C \sup_{x \in S_{\mathcal{F}}} \sup_{y \in S_{\mathbb{R}}} |\widehat{f}^x(y) - f^x(y)|, \text{ a.co.}$$

So, the claimed result is a direct consequence of this last inequality together with Theorem 3.4.1's result. \blacksquare

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

Bibliographie

- [1] Barrientos-Marin, J. (2007). Some Practical Problems of Recent Nonparametric Procedures : Testing, Estimation, and Application. PhD thesis from the Alicante University (Spain).
- [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] Benhenni, K., Ferraty, F., Rachdi, M. and Vieu, P. (2007). Local smoothing regression with functional data. *Computational Statistics*, **22**, No. 3, Pages 353–369.
- [4] Bosq, D. (2000). *Linear Processes in Function Spaces : Theory and applications*. Lecture Notes in Statistics, **149**, Springer.
- [5] Baïllo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response, *Journal of Multivariate Analysis*, **100**, Pages 102–111.
- [6] Cai, T.-T. and Hall, P. (2006). Prediction in functional linear regression, *Annals of Statistics*, **34**, Pages 2159–2179.
- [7] Chu, C.-K. and Marron, J.-S. (1991). Choosing a kernel regression estimator. With comments and a rejoinder by the authors. *Statist. Sci.*, **6**, Pages 404–436.
- [8] Ezzahrioui, M. and Ould-Saïd, E. (2008). Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *J. Nonparametr. Stat.*, **20**, Pages 3–18.
- [9] Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, **87**, Pages 998–1004.
- [10] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. London, Chapman & Hall.
- [11] Fan, J. and Yim, T.-H. (2004). A cross-validation method for estimating conditional densities. *Biometrika*, **91**, Pages 819–834.
- [12] 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.
- [13] Ferraty, F., Laksaci, A. and Vieu, P. (2005). Functional times series prediction via conditional mode. *C. R., Math., Acad. Sci. Paris*, **340**, Pages 389–392.
- [14] Ferraty, F., Laksaci, A. and Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat. Inference Stoch. Process.*, **9**, Pages 47–76.
- [15] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis. Theory and Practice*. Springer Series in Statistics. New York.

- [16] Ferraty, F., Van Keilegom, I. and Vieu, P. (2008). On the validity of the bootstrap in nonparametric functional regression. *Scandinavian J. of Statist.*, **37**, No. 2, Pages 286–306.
- [17] Hyndman, R. and Yao, Q. (2002). Nonparametric estimation and symmetry tests for conditional density functions. *J. Nonparametr. Stat.*, **14**, Pages 259–278.
- [18] Müller, H.-G. and Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Stat.*, **33**, No. 2, Pages 774–805.
- [19] Ould-Saïd, E. and Cai, Z. (2005). Strong uniform consistency of nonparametric estimation of the censored conditional mode function. *J. Nonparametr. Stat.*, **17**, Pages 797–806.
- [20] Rachdi, M. and Vieu, P. (2007). Nonparametric regression for functional data : automatic smoothing parameter selection. *Journal of Statistical Planning and Inference*, **137**, Pages 2784–2801.
- [21] Ramsay, J.-O. and Silverman, B.-W. (1997). *Functional data analysis*. Springer Series in Statistics. New York.
- [22] Ramsay, J. O. and Silverman, B. W. (2002). *Applied functional data analysis. Methods and case studies*. Springer Series in Statistics. New York.
- [23] Sarda, P. and Vieu, P. (2000). *Kernel Regression*. Pages 43–70, Wiley, New York.
- [24] Vieu, P. (1996). A note on density mode estimation. *Statist. Probab. Lett.*, **26**, Pages 297–307.
- [25] Youndjé, E. (1993). Estimation non paramétrique de la densité conditionnelle par la méthode du noyau. PhD thesis (in French), Rouen University.

Chapitre 4

A fast functional locally modeled of the conditional density and mode in functional time series

Jacques Demongeot¹, Ali Laksaci², Fethi Madani^{1,3,*} and Mustapha Rachdi⁴

Recent Advances in Functional Data Analysis and Related Topics “Contributions to Statistics”, Physica-Verlag/Springer, 2011, 85-90, DOI : 10.1007/978-3-7908-2736-1_13

Abstract. In this paper, we study the asymptotic behavior of the nonparametric local linear estimation of the conditional density of a scalar response variable given a random variable taking values in a semi-metric space. Under some general conditions on the mixing property of the data, we establish the pointwise almost complete convergence, with rates, of this estimator. This approach can be applied in time series analysis to the prediction problem via the conditional mode estimation.

Keywords and Phrases. Functional data, Local linear estimator, Conditional density, Conditional mode, Nonparametric model, Small balls probability, Mixing data

AMS Subject Classification. Primary : 62G05, Secondary : 62G07, 62G08, 62G35, 62G20.

4.1 Introduction

Let (X_i, Y_i) for $i = 1, \dots, n$ be n pairs of random variables that we assume drawn from the pair (X, Y) which is valued in $\mathcal{F} \times \mathbb{R}$, where \mathcal{F} is a semi-metric space equipped with

1. Laboratoire TIMC-IMAG, UMR CNRS 5525, Equipe AGIM³, Faculté de Médecine de Grenoble, Université J. Fourier, 38700, La Tronche, France. E-mail : Jacques.Demongeot@imag.fr

2. Université Djillali Liabès, BP. 89, Sidi Bel-Abbès 22000, Algeria. E-mail : alilak@yahoo.fr

3. Corresponding author : fethi_madani@yahoo.fr and Fethi.Madani@imag.fr

4. Laboratoire TIMC-IMAG, UMR CNRS 5525, Equipe AGIM³, Université de Grenoble 2, UFR SHS, BP. 47, 38040 Grenoble Cedex 09, France. E-mail : Mustapha.Rachdi@upmf-grenoble.fr

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

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

where $\beta(\cdot, \cdot)$ is a known function from \mathcal{F}^2 into \mathbb{R} such that, $\forall \xi \in \mathcal{F}$, $\beta(\xi, \xi) = 0$, with K and H are kernels and $h_K = h_{K,n}$ (resp. $h_H = h_{H,n}$) is a sequence of positive real numbers and $\delta(\cdot, \cdot)$ is a function of $\mathcal{F} \times \mathcal{F}$. Clearly, by a simple algebra, we get explicitly the following definition of \hat{f}^x :

$$\hat{f}^x(y) = \frac{\sum_{i,j=1}^n W_{ij}(x) H(h_H^{-1}(y - Y_i))}{h_H \sum_{i,j=1}^n W_{ij}(x)}$$

where

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

with the convention $0/0 = 0$.

Notice that the local linear estimator of the conditional density has been widely studied, when the explicative variable lies in a finite dimensional space (cf. for instance [12, 13, 19, 24]). Such estimation method has various advantages over the kernel method, namely, this method has superior bias properties than the kernel ones (cf. [4, 6] for an extensive discussion on the comparison between the both methods). Moreover, it is known that the kernel method can be viewed as a particular case of this method.

Nowadays, the progress of informatics tools permits the recuperation of increasingly bulky data. These large data sets are available essentially by real time monitoring, and computers can manage such databases. The object of statistical study can then be curves (consecutive discrete recordings are aggregated and viewed as sampled values of a random curve) not numbers or vectors. Functional data analysis (FDA) (cf. [5], [11] and [30]) can help to analyze such high-dimensional data sets. In this area, the conditional density is an important model to study the association between functional covariates scalar responses. In particular, there are several prediction tools in nonparametric statistic, such the conditional mode, the conditional median or the conditional quantiles, are based on the preliminary estimate of this functional parameter. Since, the contribution of [10] this subject has been widely studied in functional statistics (see, [8, 10, 9, 12] for some results on the conditional density estimation and its mode in the functional framework. All these works give some asymptotic results on the Nadaraya-Watson type estimator of this model. As direct extension of this method we focus, here, the estimation by local polynomial method. Such estimation procedure has recently investigated for the functional data. For example, we quote the realizations of [1, 2, 2, 5, 21, 22] which are concerned with the local linear type regression operator estimation for independent and identically distributed functional data.

In this work, we introduce the local linear nonparametric estimation of the conditional density and its mode for strongly mixed functional data. To the best of our knowledge the local linear nonparametric estimation in functional time series has not been addressed so far. The current work is the first contribution in this topic. As asymptotic result we establish, under some general conditions, the almost complete convergence rate of our estimate. The interest of this work comes mainly from the fact that the main fields of application of functional statistical methods relate to the analysis of continuous-time stochastic processes. Our study, for instance, can be applied to predict future values of some process by cutting the whole past of this process into continuous paths.

The organization of the remainder of the paper is as follows : The following Section is dedicated to fixing notations, hypotheses and the presentation of the main results. Section 3 is concentrated on some conclusions, discussions and applications of our study. The proofs of the auxiliary results are relegated to the Appendix.

4.2 Main results

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

Definition 4.2.1. *Let $\{Z_i, i = 1, 2, \dots\}$ be a strictly stationary sequence of random variables. Given a positive integer n , set*

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

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

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

Throughout the paper, x denotes a fixed point in \mathcal{F} , N_x denotes a fixed neighborhood of x , $f^{x(j)}$ denotes the j th order derivative of the conditional density f^x . S will be a fixed compact subset of \mathbb{R} , $B(x, r) = \{x' \in \mathcal{F} / |\delta(x', x)| \leq r\}$ and $\phi_x(r_1, r_2) = \mathbb{P}(r_2 \leq \delta(X, x) \leq r_1)$.

Notice that, our nonparametric model will be quite general in the sense that we will just need the following assumptions :

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

(H2) The conditional density f^x is such that : there exist $b_1 > 0$, $b_2 > 0$, $\forall (y_1, y_2) \in S^2$ and $\forall (x_1, x_2) \in N_x \times N_x$

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

where C_x is a positive constant depending on x .

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

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

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

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

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

(H6) K is a positive, differentiable function with support $[-1, 1]$.

(H7) H is a positive, bounded, Lipschitzian continuous function, such that :

$$\int |t|^{b_2} H(t) dt < \infty \text{ and } \int H^2(t) dt < \infty.$$

(H8) The bandwidth h_K satisfies : there exists an integer n_0 , such that

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

and

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

where $dP(x)$ is the cumulative distribution.

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

$$(H10) \begin{cases} \lim_{n \rightarrow \infty} h_K = 0, \lim_{n \rightarrow \infty} \frac{\chi_x^{(1/2)}(h_K) \log n}{n h_H \phi_x^2(h_K)} = 0, \\ \text{and } \exists \eta_0 > \frac{3\beta_1 + 1}{a + 1}, C n^{\frac{(3-a)}{(a+1)} + \eta_0} \leq h_H \chi_x^{1/2}(h_K) \\ \text{where } \chi_x(h) = \max(\phi_x^2(h), \varphi_x(h)) \end{cases}$$

The following theorem gives the almost-complete convergence⁵ (a.co.) of \hat{f}^x .

Theorem 4.2.1. *Under assumptions (H1)-(H10), we have :*

$$\sup_{y \in S} |\hat{f}^x(y) - f^x(y)| = O(h_K^{b_1} + h_H^{b_2}) + O \left(\sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n h_H \phi_x^2(h_K)}} \right), \text{ a.co.}$$

The proof of Theorem 4.2.1 is a direct consequence of the decomposition :

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

where

$$\hat{f}_N^x(y) = \frac{1}{n(n-1)h_H \mathbb{E}[W_{12}(x)]} \sum_{i \neq j} W_{ij}(x) H(h_H^{-1}(y - Y_i)).$$

5. Let $(z_n)_{n \in \mathbb{N}}$ be a sequence of real r.v.'s; we say that z_n converges almost completely (a.co.) to zero if, and only if, $\forall \epsilon > 0, \sum_{n=1}^{\infty} \mathbb{P}(|z_n| > \epsilon) < \infty$. Moreover, let $(u_n)_{n \in \mathbb{N}^*}$ be a sequence of positive real numbers; we say that $z_n = O(u_n)$ a.co. if, and only if, $\exists \epsilon > 0, \sum_{n=1}^{\infty} \mathbb{P}(|z_n| > \epsilon u_n) < \infty$: This kind of convergence implies both almost sure convergence and convergence in probability (cf. Sarda and Vieu (2000) for details).

and

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

and of Lemmas 5.3.1, 5.3.2 and 5.3.3 below, for which the proofs are given in the Appendix.

Lemma 4.2.1. *Under assumptions (H1), (H3), (H4), (H6), (H8) and (H10), we have that :*

$$1 - \widehat{f}_D^x = O\left(\sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n \phi_x^2(h_K)}}\right), \text{ a.co.}$$

and

$$\exists \delta > 0, \text{ such that } \sum_{i=1}^{\infty} \mathbb{P}\left(\widehat{f}_D^x < \delta\right) < \infty.$$

Lemma 4.2.2. *Under assumptions (H1), (H2) and (H7), we obtain :*

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

Lemma 4.2.3. *Under assumptions of Theorem 4.2.1, we get :*

$$\sup_{y \in S} \left| \widehat{f}_N^x(y) - \mathbb{E}[\widehat{f}_N^x(y)] \right| = O\left(\sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n h_H \phi_x^2(h_K)}}\right), \text{ a.co.}$$

4.3 Concludes remarks

- *On the assumptions* The hypotheses used in this work are not unduly restrictive and they are rather classical in the setting of nonparametric functional statistics. Indeed, the conditions (H1), (H3), (H6) and (H8) are the same as those used by [?]. Specifically (H1) is needed to deal with the functional nonparametric of our model by controlling the concentration properties of the probability measure of the variable X . The latter, is quantified, here with respect the bi-functional operator δ which can be related to the topological structure on the functional space \mathcal{F} by taking $d = |\delta|$. While (H3) is a mild regularity condition permits to control the shape of locating function β . Such condition is verified, for instance, if we take $\delta = \beta$. However, as pointed out in [?], this consideration of $\delta = \beta$ is not very adequate in practice, because these bi-functional operators do not plays similar role. We return to [?] for more discussions on these conditions and some examples of β and δ . As usually in nonparametric problems, the infinite dimension of the model is controlled by mean of a smoothness condition (H2). This condition is needed to evaluate the bias component of the rates of convergence. The first part of (H4) is a standard choice for the mixing coefficient in time series. While the second part of this condition measure the local dependence of the observations. Let us note that this last has been exploited in the expression of the convergence rate. Assumptions (H7), (H9) and (H7) are standard technical conditions in nonparametric estimation. They are imposed for the sake of simplicity and brevity of the proofs.
- *Some particular cases*
 - *The Nadaraya-Watson case* In a first attempt we will look at what happen if $b = 0$. It is clear that, in this particular case, the conditions (H3) and (H8) are not necessary to get our result and Theorem 4.2.1 can be reformulated in the following way.

Corollary 4.3.1. *Under assumptions (H1), (H2), (H4)-(H7) and (H9)-(H10), we have :*

$$\sup_{y \in S} |\widehat{f}_{NW}^x(y) - f^x(y)| = O\left(h_K^{b_1} + h_H^{b_2}\right) + O\left(\sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n h_H \phi_x^2(h_K)}}\right), \text{ a.co.}$$

where $\widehat{f}_{NW}^x(y)$

- *The multivariate case* In the vectorial case, when $\mathcal{F} = \mathbb{R}^p$, $p \geq 1$ and if the probability density of the random variable X (resp. the jointly density of (X_i, X_j)) denoted by f (resp. by $f_{i,j}$), is of \mathcal{C}^1 class, then $\phi_x(h) = O(h^p)$, and $\varphi_x(h) = O(h^{2p})$ which implies that $\chi_x(h) = O(h^{2p})$. Then our Theorem leads straitforwardly to the next Corollary,

Corollary 4.3.2. *Under assumptions (H2)-,(H3)and (H5)-(H10), we have :*

$$\sup_{y \in S} |\widehat{f}^x(y) - f^x(y)| = O\left(h_K^{b_1} + h_H^{b_2}\right) + O\left(\sqrt{\frac{\log n}{n h_H h_K^p}}\right), \text{ a.co.}$$

where

We point out that, in the special case when $\mathcal{F} = \mathbb{R}$ our estimate is identified to the estimator of [12] by taking $\beta(x, X) = X - x$ and $\delta(x, X) = x - X$.

- *The independent case* In this situation, the conditions (H4), (H5) and the last part of (H10) are automatically verified and $\chi_x(h) = \varphi_x(h) = \phi_x^2(h)$. So, we obtain the following result

Corollary 4.3.3. *Under assumptions (H1)-(H3) and (H6)-(H9), we have :*

$$\sup_{y \in S} |\widehat{f}^x(y) - f^x(y)| = O\left(h_K^{b_1} + h_H^{b_2}\right) + O\left(\sqrt{\frac{\log n}{n h_H \phi_x(h_K)}}\right), \text{ a.co.}$$

- *Application to functional times series prediction* The most important application of our study when the observations are dependent and of functional nature is the prediction of future values of some continuous time process by using the conditional mode $\theta(x) = \arg \sup_{y \in S} f^x(y)$ as prediction tool. The latter is estimated by the random variable $\widehat{\theta}(x)$ which is such that :

$$\widehat{\theta}(x) = \arg \sup_{y \in S} \widehat{f}^x(y).$$

In practice, we proceed as follows : let $(Z_t)_{t \in [0, b]}$ be a continuous time real valued random process. From Z_t we may construct N functional random variables $(X_i)_{i=1, \dots, N}$ defined by :

$$\forall t \in [0, b], \quad X_i(t) = Z_{N^{-1}((i-1)b+t)},$$

and a real characteristic $Y_i = G(X_{i+1})$. So, we can predict the characteristic Y_N by the conditional mode estimate $\widehat{Y} = \widehat{\theta}(X_N)$ given by using the $N - 1$ pairs of r.v $(X_i, Y_i)_{i=1, \dots, N-1}$. Such prediction is motivated by the following consistency result :

Corollary 4.3.4. *Under the hypotheses of Theorem 4.2.1, and if the function f^x is j -times continuously differentiable on (S) with respect to y , and that :*

$$\left\{ \begin{array}{l} f^{x(l)}(\theta(x)) = 0, \text{ if } 1 \leq l < j \\ \text{and } f^{x(j)}(\cdot) \text{ is uniformly continuous on } S \\ \text{such that } |f^{x(j)}(\theta(x))| > C > 0, \end{array} \right. \quad (2)$$

then we get :

$$|\widehat{\theta}(x) - \theta(x)|^j = O(h_K^{b_1}) + O(h_H^{b_2}) + O_{a.co.} O \left(\sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n h_H \phi_x^2(h_K)}} \right).$$

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 $i = 1, \dots, n$:

$$K_i(x) = K(h^{-1}\delta(x, X_i)), \beta_i(x) = \beta(X_i, x) \text{ and } H_i(y) = H(h_H^{-1}(y - Y_i)).$$

Proof of lemma 5.3.1. The proof is based on the same decomposition used by Barrientos *et al.* (2007)

$$\widehat{f}_D^x = \underbrace{\frac{n^2 h_K^2 \phi_x^2(h_K)}{n(n-1)\mathbb{E}[W_{12}]}}_{T_1} \left[\underbrace{\left(\frac{1}{n} \sum_{j=1}^n \frac{K_j(x)}{\phi_x(h_K)} \right)}_{T_2} \underbrace{\left(\frac{1}{n} \sum_{i=1}^n \frac{K_i(x) \beta_i^2(x)}{h_K^2 \phi_x(h_K)} \right)}_{T_4} - \underbrace{\left(\frac{1}{n} \sum_{j=1}^n \frac{K_j(x) \beta_j(x)}{h_K \phi_x(h_K)} \right)^2}_{T_3} \right].$$

Thus, we can write

$$\widehat{f}_D^x - \mathbb{E}[\widehat{f}_D^x] = T_1 \left((T_2 T_4 - \mathbb{E}[T_2 T_4]) - (T_3^2 - \mathbb{E}[T_3^2]) \right).$$

Moreover, we have

$$\begin{aligned} T_2 T_4 - \mathbb{E}[T_2 T_4] &= (T_2 - \mathbb{E}[T_2]) (T_4 - \mathbb{E}[T_4]) + (T_4 - \mathbb{E}[T_4]) \mathbb{E}[T_2] \\ &\quad + (T_2 - \mathbb{E}[T_2]) \mathbb{E}[T_4] + \mathbb{E}[T_2] \mathbb{E}[T_4] - \mathbb{E}[T_2 T_4] \end{aligned}$$

and in the same way :

$$\begin{aligned} T_3^2 - \mathbb{E}[T_3^2] &= (T_3 - \mathbb{E}[T_3])^2 + 2 (T_3 - \mathbb{E}[T_3]) \mathbb{E}[T_3] \\ &\quad + \mathbb{E}^2[T_3] - \mathbb{E}[T_3^2]. \end{aligned}$$

So, the claimed result can be derived from the following assertions

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

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

$$\text{Cov}(T_2, T_4) = o\left(\sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n \phi_x^2(h_K)}}\right) \quad (5)$$

$$\text{and } \text{Var}[T_3] = o\left(\sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n \phi_x^2(h_K)}}\right). \quad (6)$$

It is shown that in Barrientos *et al.* (2007) that

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

Theses lasts results are not affected by the dependence structure of the data. It suffices to prove (3), (5) and (6) to finish the proof of the claimed result. For (3), we set

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

Then, it can be seen that

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

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

$$\begin{aligned} \frac{1}{h_K^k} K_i(x) \beta_i^k(x) &\leq \frac{1}{h_K^k} K_i(x) |\delta(X_i, x)|^k \\ &\leq \frac{1}{h_K^k} K(h^{-1} \delta(x, X_i)) |\delta(X_i, x)|^k \mathbb{1}_{]-1, 1[}(h^{-1} \delta(x, X_i)) \\ &\leq K(h^{-1} \delta(x, X_i)) \leq C \end{aligned} \quad (7)$$

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

$$\begin{aligned} \mathbb{P} \{ |T_{k+2} - \mathbb{E}[T_{k+2}]| > \varepsilon \} &= \mathbb{P} \left\{ \left| \frac{1}{n \phi_x(h_K)} \sum_{i=1}^n \Delta_i^k \right| > \varepsilon \right\} \\ &\leq \mathbb{P} \left\{ \left| \sum_{i=1}^n \Delta_i^k \right| > \varepsilon n \phi_x(h_K) \right\} \\ &\leq C(A_1 + A_2) \end{aligned} \quad (8)$$

where

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

and

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

with

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

Next, we evaluate the asymptotic behavior of S_n^{2*} . For this we use the technique of Masry (1986). We define the sets

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

and

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

where $m_n \rightarrow \infty$, as $n \rightarrow \infty$. Let $J_{1,n}$ and $J_{2,n}$ be the sum of covariance over S_1 and S_2 respectively. Because of (H1), (H4) and (7) we can write :

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

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

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

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

$$|J_{2,n}| = \left| \sum_{(i,j) \in E_2} Cov[\Delta_i^k, \Delta_j^k] \right| \leq C \frac{nm_n^{-a+1}}{a-1}. \quad (9)$$

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

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

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

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

Finally, as $a > 2$

$$S_n^2 = O(n\chi_x^{1/2}(h_K)). \quad (10)$$

Now, we apply (8) to $\varepsilon = \lambda \frac{\sqrt{S_n^2 \log n}}{n\phi_x(h_K)}$ and $r = C(\log n)^2$. It follows that

$$A_2 \leq Cn^{1-(a+1)/2} \chi_x(h_K)^{-(a+1)/4} (\log n)^{(3a-1)/2}.$$

Next, using the left side of (H10) we obtain

$$A_2 \leq Cn^{-1-\eta(a+1)/2} (\log n)^{(3a-1)/2}.$$

So, it exists some real $\nu > 0$ such that

$$A_2 \leq Cn^{-1-\nu}. \quad (11)$$

By means of (10), we show that

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

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

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

Thus, for λ large enough :

$$\exists \nu' > 0, \quad A_1 \leq Cn^{-\lambda^2/2} \leq Cn^{-1-\nu'}. \quad (12)$$

Hence

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

Finally, by following similar arguments used to prove (10) we get

$$\begin{aligned} Cov(T_2, T_4) &= O\left(\frac{\chi_x^{(1/2)}(h_K)}{n \phi_x^2(h_K)}\right) \\ \text{and } Var[T_3] &= o\left(\frac{\chi_x^{(1/2)}(h_K)}{n \phi_x^2(h_K)}\right). \end{aligned}$$

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

Proof of lemma 5.3.2. The bias term is standard is not affected by the dependence condition of (X_i, Y_i) . So, by the equiprobability of the couples (X_i, Y_i) we have

$$\forall y \in S, \mathbb{E}[\widehat{f}_N^x(y)] = \frac{1}{\mathbb{E}[W_{12}]} \mathbb{E} [W_{12}(x) [\mathbb{E} [h_H^{-1} H_1(y)/X]]].$$

By the classical change of variables $t = (y - z)/h_H$, we obtain :

$$h_H^{-1} \mathbb{E} [H_1(y)/X] = \int_{\mathbb{R}} H(t) f^X(y - h_H t) dt,$$

therefore

$$|\mathbb{E}[H_1(y)/X] - f^x(y)| \leq \int_{\mathbb{R}} H(t) |f^X(y - h_H t) - f^x(y)| dt.$$

Thus, by the assumption (H2) we get that :

$$\forall y \in S, \mathbb{1}_{B(x, h_K)}(X) |\mathbb{E}(H_1(y)/X) - f^x(y)| \leq \int_{\mathbb{R}} H(t) (h_K^{b_1} + |t|^{b_2} h_H^{b_2}) dt.$$

Hence,

$$\forall y \in S, |\mathbb{E}[\widehat{f}_N^x(y)] - f^x(y)| \leq \frac{1}{\mathbb{E}[W_{12}]} \mathbb{E} [W_{12}(x) |\mathbb{E} [h_H^{-1} H_1(y)/X] - f^x(y)|] \leq C(h_K^{b_1} + h_H^{b_2}).$$

Proof of lemma 5.3.3.

Using the compactness of S , we can write that $S \subset \bigcup_{k=1}^{r_n} S_k$ where $S_k = (t_k - l_n, t_k + l_n)$. Taking $t_y = \arg \min_{t \in \{t_1, \dots, t_{z_n}\}} |y - t|$, we have

$$\begin{aligned}
\sup_{y \in S} \left| \widehat{f}_N^x(y) - \mathbb{E}[\widehat{f}_N^x(y)] \right| &\leq \underbrace{\sup_{y \in S} \left| \widehat{f}_N^x(y) - \widehat{f}_N^x(t_y) \right|}_{A_1} + \underbrace{\sup_{y \in S} \left| \widehat{f}_N^x(t_y) - \mathbb{E}[\widehat{f}_N^x(t_y)] \right|}_{A_2} + \\
&\quad + \underbrace{\sup_{y \in S} \left| \mathbb{E}[\widehat{f}_N^x(t_y)] - \mathbb{E}[\widehat{f}_N^x(y)] \right|}_{A_3}. \tag{13}
\end{aligned}$$

- Concerning (A_1) :

$$\begin{aligned}
\sup_{y \in S} \left| \widehat{f}_N^x(y) - \widehat{f}_N^x(t_y) \right| &\leq \sup_{y \in S} \frac{1}{n(n-1)h_H \mathbb{E}[W_{12}(x)]} \sum_{i \neq j} |H_i(y) - H_i(t_y)| W_{ij}(x), \\
&\leq \sup_{y \in S} \frac{C|y - t_y|}{h_H} \left(\frac{1}{n h_H \mathbb{E}[W_{12}(x)]} \sum_{i \neq j} W_{ij}(x) \right), \\
&\leq C \frac{l_n}{h_H^2} \widehat{f}_D^x \\
&\leq C \frac{l_n}{h_H^2}. \tag{14}
\end{aligned}$$

The second inequality is obtained by considering a Lipschitz argument whereas the last one comes from the almost comply consistency of \widehat{f}_D^x (see, Lemma 5.3.1). Take now

$$l_n = n^{-\frac{3}{2}\beta_1 - \frac{1}{2}}, \tag{15}$$

and note that, because of (H9), we have

$$l_n/h_H^2 = o\left(\sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n h_H \phi_x^2(h_K)}} \right).$$

Thus, for n large enough, we can write

$$\sup_{y \in S_{\mathbb{R}}} \left| \widehat{f}_N^x(y) - \widehat{f}_N^x(t_y) \right| = o_{a.co.} \left(\sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n h_H \phi_x^2(h_K)}} \right). \tag{16}$$

- Concerning (A_2) : we can write for all $\eta > 0$:

$$\begin{aligned}
&\mathbb{P} \left(\sup_{y \in S} \left| \widehat{f}_N^x(t_y) - \mathbb{E}[\widehat{f}_N^x(t_y)] \right| > \eta \sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n h_H \phi_x^2(h_K)}} \right) \\
&= \mathbb{P} \left(\max_{t_y \in \{t_1, \dots, t_{z_n}\}} \left| \widehat{f}_N^x(t_y) - \mathbb{E}[\widehat{f}_N^x(t_y)] \right| > \eta \sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n h_H \phi_x^2(h_K)}} \right) \\
&\leq z_n \max_{t_y \in \{t_1, \dots, t_{z_n}\}} \mathbb{P} \left(\left| \widehat{f}_N^x(t_y) - \mathbb{E}[\widehat{f}_N^x(t_y)] \right| > \eta \sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n h_H \phi_x^2(h_K)}} \right).
\end{aligned}$$

all it remains to compute is the following quantity :

$$\mathbb{P} \left(\left| \widehat{f}_N^x(t_y) - \mathbb{E}[\widehat{f}_N^x(t_y)] \right| > \eta \sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n h_H \phi_x^2(h_K)}} \right), \text{ for all } t_y \in \{t_1, \dots, t_{z_n}\}.$$

The later is given by a straightforward adaptation of the proof of Lemma (5.3.1). To do that, we consider the following decomposition :

$$\begin{aligned} \widehat{f}_N^x(t_y) = & \underbrace{\frac{n^2 h_K^2 \phi_x^2(h_K)}{n(n-1) \mathbb{E}[W_{12}]} }_{S_1} \left[\underbrace{\left(\frac{1}{n} \sum_{j=1}^n \frac{K_j(x) H_j(t_y)}{h_H \phi_x(h_K)} \right)}_{S_2} \underbrace{\left(\frac{1}{n} \sum_{i=1}^n \frac{K_i(x) \beta_i^2(x)}{h_K^2 \phi_x(h_K)} \right)}_{S_3} \right. \\ & \left. - \underbrace{\left(\frac{1}{n} \sum_{j=1}^n \frac{K_j(x) \beta_j(x) H_j(t_y)}{h_H h_K \phi_x(h_K)} \right)}_{S_4} \underbrace{\left(\frac{1}{n} \sum_{i=1}^n \frac{K_i(x) \beta_i(x)}{h_K \phi_x(h_K)} \right)}_{S_5} \right] \end{aligned}$$

which implies that

$$\widehat{f}_N^x(t_y) - \mathbb{E}[\widehat{f}_N^x(t_y)] = S_1((S_2 S_3 - \mathbb{E}[S_2 S_3]) - (S_4 S_5 - \mathbb{E}[S_4 S_5])).$$

Clearly

$$\begin{aligned} S_2 S_3 - \mathbb{E}[S_2 S_3] &= (S_2 - \mathbb{E}[S_2]) (S_3 - \mathbb{E}[S_3]) + (S_3 - \mathbb{E}[S_3]) \mathbb{E}[S_2] \\ &\quad + (S_2 - \mathbb{E}[S_2]) \mathbb{E}[S_3] + \mathbb{E}[S_2] \mathbb{E}[S_3] - \mathbb{E}[S_2 S_3] \end{aligned}$$

and

$$\begin{aligned} S_4 S_5 - \mathbb{E}[S_4 S_5] &= (S_4 - \mathbb{E}[S_4]) (S_5 - \mathbb{E}[S_5]) + (S_5 - \mathbb{E}[S_5]) \mathbb{E}[S_4] \\ &\quad + (S_4 - \mathbb{E}[S_4]) \mathbb{E}[S_5] + \mathbb{E}[S_4] \mathbb{E}[S_5] - \mathbb{E}[S_4 S_5] \end{aligned}$$

So, our claimed result is direct consequences of the following assertions

$$\sum_n z_n \mathbb{P} \left\{ |S_k - \mathbb{E}[S_k]| > \eta \sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n h_H \phi_x(h_K)}} \right\} < \infty, \quad \text{for } k = 2, 3, 4, 5, \quad (17)$$

$$S_1 = O(1) \quad \text{and} \quad \mathbb{E}[S_k] = O(1) \quad \text{for } k = 2, 3, 4, 5, \quad (18)$$

$$Cov(S_2, S_3) = o \left(\sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n \phi_x^2(h_K)}} \right) \quad (19)$$

$$\text{and } Cov(S_4, S_5) = o \left(\sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n \phi_x^2(h_K)}} \right). \quad (20)$$

Observe that the case $k = 3, 5$ has been already obtained in Lemma (5.3.1). Thus, we focus only the case $k = 2, 4$.

Firstly, for (17), we use the same arguments as those invoked in the proof of Lemma 5.3.1 and we compute asymptotically

$$s_n'^{2*} = \sum_{i \neq j=1}^n |Cov(\Lambda_i^k, \Lambda_j^k)|,$$

where

$$\Lambda_i^k = \frac{1}{h_K^k} \left(K_i(x) H_i(t_y) \beta_i^k(x) - \mathbb{E} \left[K_i(x) H_i(t_y) \beta_i^k(x) \right] \right) \text{ for } k = 0, 1.$$

We split the sum into two sets defined by

$$S'_1 = \{(i, j) \text{ such that } 0 \leq |i - j| \leq m'_n\}$$

and

$$S'_2 = \{(i, j) \text{ such that } m'_n + 1 \leq |i - j| \leq n - 1\}$$

where $m'_n \rightarrow \infty$. Then

$$s_n'^{2*} \leq \sum_{S'_1} Cov(\Lambda_i^k, \Lambda_j^k) + \sum_{S'_2} Cov(\Lambda_i^k, \Lambda_j^k).$$

On one hand, by (H1) and (H4)-(H7), we have :

$$\begin{aligned} Cov(\Lambda_i^k, \Lambda_j^k) &= \mathbb{E}[\Lambda_i^k \Lambda_j^k] - \mathbb{E}[\Lambda_i^k] \mathbb{E}[\Lambda_j^k] \\ &= \mathbb{E}[\Delta_i^k \Delta_j^k \mathbb{E}[H_i(t_y) H_j(t_y) | X_i, X_j]] - \mathbb{E}^2[\Delta_1^k \mathbb{E}[H_1(t_y) | X]] \\ &\leq C \left(h_H^2 \mathbb{E}[\Delta_i^k \Delta_j^k] + h_H^2 \mathbb{E}^2[\Delta_1^k] \right) \\ &\leq C h_H^2 (\varphi_x(h_K) + \phi_x^2(h_K)) \\ &\leq C h_H^2 \chi_x(h_K). \end{aligned}$$

On the other hand this covariance can be controlled by means of the Davydov-Rio's inequality for bounded mixing processes

$$\forall i \neq j \quad Cov(\Lambda_i^k, \Lambda_j^k) \leq C \alpha(|i - j|).$$

Then, by using similar arguments as those invoked in the proof of Lemma (5.3.1)

$$s_n'^{2*} \leq C(nm_n h_H^2 \chi_x(h_K) + nm_n^{1-a}).$$

We can take

$$m'_n = \left(\frac{1}{h_H^2 \chi_x(h_K)} \right)^{1/a},$$

and we conclude as for the proof of Lemma 5.3.1 that :

$$s_n'^{2*} = O(n(h_H^2 \chi_x(h_K))^{(a-1)/a}). \quad (21)$$

The computation of the variance term can be done by following the same ideas as in bias term given in Lemma (5.3.2) and is based in the fact that

$$\mathbb{E}[H_1^2(t_y)|X] = O(h_H) \quad \text{and} \quad \mathbb{E}[H_1(t_y)|X] = O(h_H).$$

Therefore,

$$\text{Var}[\Lambda_i^k] = O(h_H \chi_x^{(1/2)}(h_K)).$$

This last result, together with (21) show that

$$s_n'^2 = \sum_{i,j=1}^n |\text{Cov}(\Lambda_i^k, \Lambda_j^k)| = O(nh_H \chi_x^{1/2}(h_K)). \quad (22)$$

Once again, similar arguments as those invoked for proving Lemma 5.3.1 can be used, and we obtain successively, for all $\epsilon > 0$, and $k = 2, 4$:

$$\sum_n z_n \mathbb{P} \{ |S_k - \mathbb{E}[S_k]| > \epsilon \} \leq C l_n^{-1} (A'_1 + A'_2),$$

where

$$A'_1 = 4 \left(1 + \frac{\epsilon^2}{r s_n'^2} \right)^{-r/2} \quad \text{and} \quad A'_2 = 4c n r^{-1} \left(\frac{r}{\epsilon} \right)^{a+1}.$$

To state the desired result we take $\epsilon = C_1 \eta \frac{\sqrt{s_n'^2 \log n}}{n h_H \phi_x(h_K)}$ and $r = c \log n^2$ and we use the fact that $s_n'^2 = O(nh_H \chi_x^{1/2}(h_K))$. We note that A'_1 and A'_2 are exactly the same as the terms A_1 and A_2 appearing in the proof of Lemma 5.3.1. We also note that, the choice of l_n made in (15) and our conditions on h_K (resp. on χ_x) are immediately insuring the existence, for η large enough of some $\nu > 0$ such that :

$$l_n^{-1} (A'_1 + A'_2) \leq C n^{-1-\nu}.$$

Finally, we arrive at

$$\sum_n z_n \mathbb{P} \left\{ |S_i - \mathbb{E}[S_i]| > \eta \sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n h_H \phi_x^2(h_K)}} \right\} < \infty. \quad (23)$$

Let us now prove the result (18) and (25). The proof of this last follows exactly along the same line as the proof of (22).

$$\text{Cov}(S_2, S_3) = O\left(\frac{\chi_x^{(1/2)}(h_K)}{n \phi_x^2(h_K)}\right) = o\left(\sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n h_H \phi_x^2(h_K)}}\right) \quad (24)$$

$$\text{and } \text{Cov}(S_4, S_5) = O\left(\frac{\chi_x^{(1/2)}(h_K)}{n \phi_x^2(h_K)}\right) = o\left(\sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n h_H \phi_x^2(h_K)}}\right). \quad (25)$$

While the proof of (4) is based on the fact that

$$\mathbb{E}[S_2] = \frac{\mathbb{E}[K_1(x)H_1(t_y)]}{h_H \phi_x(h_K)}, \quad \text{and} \quad \mathbb{E}[S_4] = \frac{\mathbb{E}[K_1(x)H_1(t_y)\beta_1(x)]}{h_K h_H \phi_x(h_K)},$$

So it remains to evaluate the quantities :

$$\mathbb{E} \left[K_i(x) H_i(t_y) \beta_i^k(x) \right], \text{ for } k = 0, 1.$$

By conditioning on X_1 and by applying Lemma 3 in Barrientos (2007), we obtain that :

$$\mathbb{E} \left[K_i(x) H_i(t_y) \beta_i^k(x) \right] = O(h_H h_K^k \phi_x(h_K)). \quad (26)$$

Leading

$$\mathbb{E}[S_k] = O(1), \text{ for } k = 2, 4.$$

Finally, we arrive at :

$$\sup_{y \in \mathcal{S}_{\mathbb{R}}} \left| \widehat{f}_N^x(y) - \mathbb{E}[\widehat{f}_N^x(t_y)] \right| = O_{a.co.} \left(\sqrt{\frac{\chi_x^{(1/2)}(h_K) \log n}{n h_H \phi_x^2(h_K)}} \right). \quad (27)$$

- Concerning (T_3) : because of (H8) and (H9) we have :

$$\sup_{y \in \mathcal{S}} \left| \mathbb{E}[\widehat{f}_N^x(y)] - \mathbb{E}[\widehat{f}_N^x(t_y)] \right| \leq C \frac{l_n}{h_H^2}.$$

Using analogous arguments as for T_1 , we can show for n large enough that

$$\mathbb{P} \left(\sup_{y \in \mathcal{S}} \left| \mathbb{E}[\widehat{f}_N^x(y)] - \mathbb{E}[\widehat{f}_N^x(t_y)] \right| > \frac{\eta}{3} \sqrt{\frac{\log n}{n h_H \chi_x(h_K)}} \right) = 0. \quad (28)$$

Now, our lemma can be easily deduced from (16), (27) and (28). ■

Proof of Corollary 4.3.4. By a simple manipulation, we show that :

$$|f^x(\widehat{\theta}(x)) - f^x(\theta(x))| \leq 2 \sup_{y \in \mathcal{S}_{\mathbb{R}}} |\widehat{f}^x(y) - f^x(y)|. \quad (29)$$

We use the following Taylor expansion of the function f^x :

$$f^x(\widehat{\theta}(x)) = f^x(\theta(x)) + \frac{1}{j!} f^{x(j)}(\theta'(x)) (\widehat{\theta}(x) - \theta(x))^j,$$

for some $\theta'(x)$ between $\theta(x)$ and $\widehat{\theta}(x)$. It is clear that, from conditions (U6), (29) and Theorem 3.4.1, we obtain that :

$$|\widehat{\theta}(x) - \theta(x)| \rightarrow 0, \text{ a.co.}$$

Next, under condition (U7), we obtain that :

$$|f^{x(j)}(\theta'(x)) - f^{x(j)}(\theta(x))| \rightarrow 0, \text{ a.co.}$$

Consequently, we can get $\tau > 0$ such that :

$$\sum_{n=1}^{\infty} \mathbb{P} \left(f^{x(j)}(\theta'(x)) < \tau \right) < \infty,$$

and we have :

$$|\widehat{\theta}(x) - \theta(x)|^j \leq C \sup_{y \in \mathcal{S}} |\widehat{f}^x(y) - f^x(y)|, \text{ a.co.}$$

So, the claimed result is a direct consequence of this last inequality together with Theorem 3.4.1's result. ■

Bibliographie

- [1] Barrientos-Marin, J. (2007). Some Practical Problems of Recent Nonparametric Procedures : Testing, Estimation, and Application. PhD thesis (in French) from the Paul Sabatier's University (Toulouse).
- [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] Benhenni, K., Griche-Hedli, S., Rachdi, M. (2010). Estimation of the regression operator from functional fixed-design with correlated errors. *Journal of Multivariate Analysis*, **101**, Pages 476–490.
- [4] Benhenni, K., Ferraty, F., Rachdi, M. and Vieu, P. (2007). Local smoothing regression with functional data. *Computational Statistics*, **22**, No. 3, Pages 353–369.
- [5] Bosq, D. (2000). *Linear Processes in Function Spaces : Theory and applications*. Lecture Notes in Statistics, **149**, Springer.
- [6] Baillo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response, *Journal of Multivariate Analysis*, **100**, Pages 102–111.
- [7] Chu, C.-K. and Marron, J.-S. (1991). Choosing a kernel regression estimator. With comments and a rejoinder by the authors. *Statist. Sci.*, **6** (1991), Pages 404–436.
- [8] Dabo-Niang, S. and Laksaci, A. (2007). Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle. *Pub. Inst. Stat. Univ. Paris*, **3**, Pages 27–42.
- [9] El Methni, M. and Rachdi, M. (2010). Local weighted average estimation of the regression operator for functional data. *Commun. Stat., Theory and Methods*, to appear.
- [10] Ezzahrioui, M. and Ould-Saïd, E. (2008). Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *J. Nonparametr. Stat.*, **20**, Pages 3–18.
- [11] Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, **87**, Pages 998–1004.
- [12] Fan, J. and Yim, T.-H. (2004). A cross-validation method for estimating conditional densities. *Biometrika*, **91**, Pages 819–834.
- [13] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. London, Chapman & Hall.
- [14] Ferraty, F., Goia, A. and Vieu, P. (2002). Functional nonparametric model for time series : a fractal approach to dimension reduction. *TEST*, **11**, No. 2, Pages 317–344.
- [15] Ferraty, F., Laksaci, A., Tadj, A. and Vieu, Ph. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *Journal of statistical planning and inference*, **140**, Pages 335–352.

- [16] Ferraty, F., Laksaci, A. and Vieu, P. (2005). Functional times series prediction via conditional mode. *C. R., Math., Acad. Sci. Paris*, **340**, Pages 389–392.
- [17] Ferraty, F., Laksaci, A. and Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat. Inference Stoch. Process.*, **9**, Pages 47–76.
- [18] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis. Theory and Practice*. Springer Series in Statistics. New York.
- [19] Hyndman, R. and Yao, Q. (2002). Nonparametric estimation and symmetry tests for conditional density functions. *J. Nonparametr. Stat.*, **14**, Pages 259–278.
- [20] Laksaci, A. (2007). Convergence en moyenne quadratique de l'estimateur à noyau de la densité conditionnelle avec variable explicative fonctionnelle. *Pub. Inst. Stat. Univ. Paris*, **3**, Pages 69–80.
- [21] Laksaci, A., Madani, F. and Rachdi, M. (2012). Kernel conditional density estimation when the regressor is valued in a semi-metric space *Communications in Statistics Theory and Methods*. (to appear).
- [22] Demongeot, J., Laksaci, A., Madani, F. and Rachdi, M. (2011). Functional data : Local linear estimation of the conditional density and its application *Statistics, Volume 00, Pages 00-00, DOI : 10.1080/02331888.2011.568117*.
- [23] Demongeot, J., Laksaci, A., Madani, F. and Rachdi, M. (2010) Local linear estimation of the conditional density for functional data. *C. R., Math., Acad. Sci. Paris*, **348**, Pages 931-934.
- [24] Müller, H.-G. and Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Stat.*, **33**, No. 2, Pages 774-805.
- [25] Ouassou, I. and Rachdi, M. (2010). Stein type estimation of the regression operator for functional data. *Advances and Applications in Statistical Sciences*, **1**, No 2, Pages 233-250.
- [26] Ould-Saïd, E. and Cai Z. (2005). Strong uniform consistency of nonparametric estimation of the censored conditional mode function. *J. Nonparametr. Stat.*, **17**, Pages 797-806.
- [27] Rachdi, M. and Sabre, R. (2000). Consistent estimates of the mode of the probability density function in nonparametric deconvolution problems. *Statist. Probab. Lett.*, **47**, Pages 105–114.
- [28] Rachdi, M. and Vieu, P. (2007). Nonparametric regression for functional data : automatic smoothing parameter selection. *Journal of Statistical Planning and Inference*, **137**, Pages 2784–2801.
- [29] Ramsay, J.-O. and Silverman, B.-W. (1997). *Functional data analysis*. Springer Series in Statistics. New York.
- [30] Ramsay, J.-O. and Silverman, B.-W. (2002). *Applied functional data analysis. Methods and case studies*. Springer Series in Statistics. New York.
- [31] Rio, E. (1990). *Exponential inequalities ...* Springer-Verlag, New York.
- [32] Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants*. Springer, ESAIM, Collection Mathématiques et Applications
- [33] Sarda, P. and Vieu, P. (2000). *Kernel Regression*. Pages 43–70, Wiley, New York.

- [34] Vieu, P. (1996). A note on density mode estimation. *Statist. Probab. Lett.*, **26**, Pages 297–307.
- [35] Youndjé, E. (1993). Estimation non paramétrique de la densité conditionnelle par la méthode du noyau. PhD thesis (in French) from the Rouen's University.

Chapitre 5

On the quadratic error of the functional local linear estimate of the conditional density

5.1 Introduction

The observation of functional variables has become usual due, for instance, to the development of measuring instruments that allow one to observe variables at finer and finer resolutions. Then, as technology progresses, we are able to handle larger and larger datasets. At the same time, monitoring devices such as electronic equipment and sensors (for registering images, temperature, etc.) have become more and more sophisticated. This high-tech revolution offers the opportunity to observe phenomena in an increasingly accurate way by producing statistical units sampled over a finer and finer grid, with the measurement points so close that the data can be considered as observations varying over a continuum. Such continuous (or functional) data may occur in biomechanics (e.g. human movements), chemometrics (e.g. spectrometric curves), econometrics (e.g. the stock market index), geophysics (e.g. spatio-temporal events such as El Niño or time series of satellite images), or medicine (electro-cardiograms/electro-encephalograms). It is well known that standard multivariate statistical analyses fail with functional data. However, the great potential for applications has encouraged new methodologies able to extract relevant information from functional datasets. This Handbook aims to present a state of the art exploration of this high-tech field, by gathering together most of major advances in this area. The main statistical topics (classification, inference, factor-based analysis, regression modelling, resampling methods, time series, random processes) are covered in the setting of functional data. The twin challenges of the subject are the practical issues of implementing new methodologies and the theoretical techniques needed to expand the mathematical foundations and toolboxes. This chapter and the following, therefore, mixes practical, methodological and theoretical aspects of the subject, sometimes within the same chapter (cf. Chapter 2 before). As a consequence, these results should appeal to a wide audience of engineers, practitioners and graduate students, as well as academic researchers, not only in statistics and probability but also in numerous related application areas.

It seems, then, natural to assume that the data are actually observations from a random variable taking values in a functional space. In this chapter, we are interested in the local polynomial modeling of the conditional density function when the explanatory variable is of functional type. Such study is motivated by the fact that the local polynomial smoothing has various advantages over the kernel method, namely this method has superior bias properties to the previous one (see, for example Chu and Marron (1991) and Fan (1992) for an extensive discussion on the comparison between both these methods). Moreover, as noticed by Fan and Yao (2003) that the conditional density provides a very informative summary of response variable that allows us to examine the overall shape of the conditional distribution. In the nonparametric functional statistics, the first results about the conditional distribution were obtained in Ferraty et al. (2006). In this last, it is established the almost complete convergence of the kernel estimator of the conditional density and its derivatives. the quadratic error of this estimate has been studied by Laksaci (2007). The latter gave the asymptotic expansion of the exact expression involved in the leading terms of the quadratic error of the considered estimate. Recently, Ferraty et al. (2010) stated the uniform almost complete convergence of the kernel estimate of some nonparametric conditional models, in particular, of the conditional density model.

Since the open question (“How can the local polynomial ideas be adapted to infinite dimensional settings?”) stated by Ferraty and Vieu (2006), the local linear smoothing in the functional data setting, have been considered by many authors. We cite, for instance Barrientos et al. (2010), Baillo and Grané (2009), El Methni and Rachdi (2011) which are concerned with the local linear type regression operator estimation for independent and identically distributed functional data. While, the first contribution on the local polynomial modeling of the conditional density function when the explanatory variable is functional were considered by Demongeot et al. (2010). The authors established the almost complete consistency (in pointwise and uniform) of a fast functional local linear estimate of the conditional density when the explanatory variable is functional and the observations are i.i.d. Their study is extended to dependent case by Demongeot et al. (2011).

In this chapter, we give the convergence rate in mean square of of a fast functional local linear estimate considered by Demongeot et al. (2010). The expression of this convergence rate shows the superiority of this method with respect to the kernel method, namely in the bias terms. It should be noted that the accuracy of our asymptotic results leads to interesting perspectives from a practical point of view, in particular, minimizing mean squared errors can govern automatic bandwidth selection procedures.

We present our model in Section 5.2. In Section 5.3 we give some notations, hypotheses and the presentation of the main results. Section 5.4 is devoted to some discussions and comments on the result. The proofs of the results are relegated to the last section of this chapter.

5.2 The 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 $\mathcal{F} \times \mathbb{R}$, where \mathcal{F} is a semi-metric space equipped with a semi-metric d .

Furthermore, we assume that there exists a regular version of the conditional probability of Y given X , which is absolutely continuous with respect to the Lebesgue measure on \mathbb{R} and has two-times continuously differentiable density, denoted by f^x . Local polynomial smoothing is based on the assumption that functional parameter is smooth enough to be locally well approximated by a polynomial. In functional statistics, there are several ways for extending the local linear ideas (cf. Barrientos et al.(2010), Baillo. and Grané (2009)). Here we adopt the fast functional locally modeling, that is, we estimate the conditional density f^x by \hat{a} which is obtained by minimizing the following quantity :

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

where $\beta(\cdot, \cdot)$ (resp. $\delta(\cdot, \cdot)$) is a known operator from \mathcal{F}^2 into \mathbb{R} such that, $\forall \xi \in \mathcal{F}$, $\beta(\xi, \xi) = 0$ (resp. $\delta(\xi, \xi) = 0$), with K and H are kernels and $h_K = h_{K,n}$ (resp. $h_H = h_{H,n}$) is chosen as a sequence of positive real numbers. Clearly, by a simple algebra, we get explicitly the following definition of \hat{f}^x :

$$\hat{f}^x(y) = \frac{\sum_{i,j=1}^n W_{ij}(x) H(h_H^{-1}(y - Y_i))}{h_H \sum_{i,j=1}^n W_{ij}(x)} \quad (2)$$

where

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

with the convention $0/0 = 0$.

5.3 Main results

In the remainder of the chapter, we set :

$$\begin{aligned} \phi_x(r_1, r_2) &= \mathbb{P}(r_2 \leq \delta(x, X) \leq r_1) \\ \psi_l(\cdot, y) &= \frac{\partial^l f^x(y)}{\partial y^l} \end{aligned}$$

and

$$\Psi_l(s) = E[\psi_l(X) - \psi_l(x) | \delta(x, X) = s], \text{ for some } l \in \{0, 2\}$$

We will 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 :

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

(H2) For $l \in \{0, 2\}$, the quantities $\Psi'_l(0)$ exist, where Ψ'_l denotes the first derivative of Ψ_l .

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

$$\left\{ \begin{array}{l} \forall z \in \mathcal{F}, C_1 |\delta(x, z)| \leq |\beta(x, z)| \leq C_2 |\delta(x, z)|, \text{ where } C_1 > 0, C_2 > 0, \\ \sup_{u \in B(x, r)} |\beta(u, x) - \delta(x, u)| = o(r) \\ \text{and } h_K \int_{B(x, h_K)} \beta(u, x) dP(u) = o\left(\int_{B(x, h_K)} \beta^2(u, x) dP(u)\right) \end{array} \right.$$

where $B(x, r) = \{z \in \mathcal{F} / |\delta(x, z)| \leq r\}$ and $dP(x)$ is the probability distribution of X .
(H4) K is a positive, differentiable function supported within $[-1, 1]$. Its derivative K' satisfies $K'(t) < 0$, for $-1 \leq t < 1$, and $K(1) > 0$.

(H5) H is a positive function, integrable, bounded, symmetric and such that $\int H(t) dt = 1$ et $\int t^2 H(t) dt < \infty$,

(H6) The bandwidths h_K and h_H satisfy : $\lim_{n \rightarrow \infty} h_K = 0$, $\lim_{n \rightarrow \infty} h_H = 0$ et $\lim_{n \rightarrow \infty} n h_H \phi(h_K) = \infty$.

Notice that, (H1) and (H2) are a simple adaptation of the conditions H_1 and H_3 in Ferraty et al. (2007) when we replace the metric by some bi-functional δ . The second part of the condition (H3) is unrestrictive and is verified, for instance, if $\delta(\cdot, \cdot) = \beta(\cdot, \cdot)$ or, moreover, if

$$\lim_{\delta(x, u) \rightarrow 0} \left| \frac{\beta(u, x)}{\delta(x, u)} - 1 \right| = 0$$

because, $\forall u \in B(x, r)$, we have

$$\frac{|\beta(u, x) - \delta(x, u)|}{r} \leq \left| \frac{\beta(u, x)}{\delta(x, u)} - 1 \right|.$$

The rest of (H3) has been introduced and commented by Barrientos et al. (2010) by giving a several examples of bi-functional operators δ and β satisfying this condition. Conditions (H4)-(H6) are standards and classically used in the context of quadratic errors in functional statistic.

Theorem 5.3.1. *Under assumptions (H1)-(H6), we have that :*

$$\begin{aligned} E \left[\hat{f}^x(y) - f^x(y) \right]^2 &= B_H^2(x, y) h_H^4 + B_K^2(x, y) h_K^2 \\ &+ \frac{V_{HK}(x, y)}{n h_H \phi_x(h_K)} + o(h_H^4) + o(h_K^2) + o\left(\frac{1}{n h_H \phi_x(h_K)}\right) \end{aligned} \quad (3)$$

with

$$\begin{aligned} B_H(x, y) &= \frac{1}{2} \frac{\partial^2 f_Y^X(x, y)}{\partial y^2} \int t^2 H(t) dt \\ B_K(x, y) &= \Psi'_0(0) \frac{\left(K(1) - \int_{-1}^1 (u^3 K(u))' \chi_x(u) du \right)}{\left(K(1) - \int_{-1}^1 (u^2 K(u))' \chi_x(u) du \right)} \end{aligned}$$

and

$$V_{HK}(x, y) = (f_Y^X(x, y)) \left(\int H^2(t) dt \right) \left[\frac{\left(K^2(1) - \int_{-1}^1 (u^4 K^2(u))' \chi_x(u) du \right)}{\left(K(1) - \int_{-1}^1 (u^2 K(u))' \chi_x(u) du \right)^2} \right]$$

If we set that :

$$\widehat{f}_N(x, y) = \frac{1}{n(n-1)EW_{12}} \sum_{i \neq j, 1}^n W_{ji} H_i(y) \text{ and } \widehat{f}_D(x) = \frac{1}{n(n-1)EW_{12}} \sum_{i \neq j, 1}^n W_{ij}$$

where $K_i = K(h_K^{-1} \delta(x, X_i))$, $H_i(y) = H(h_H^{-1}(y - Y_i))$, for all $i = 1, \dots, n$, then we obtain the following lemmas which will be useful for Theorem 5.3.1's proof.

Lemma 5.3.1. *Under the hypotheses of Theorem 5.3.1, we have :*

$$E \left[\widehat{f}_N(x, y) \right] - f_Y^X(x, y) = B_H(x, y) h_H^2 + B_K(x, y) h_K + o(h_H^2) + o(h_K)$$

Lemma 5.3.2. *Under the hypotheses of Theorem 5.3.1, we have :*

$$\text{Var} \left[\widehat{f}_N(x, y) \right] = \frac{V_{HK}(x, y)}{nh_H \phi_x(h_K)} + o\left(\frac{1}{nh_H \phi_x(h_K)}\right)$$

Lemma 5.3.3. *Under the hypotheses of Theorem 5.3.1, we have :*

$$\text{Cov}(\widehat{f}_N(x, y), \widehat{f}_D(x)) = O\left(\frac{1}{n\phi_x(h_K)}\right)$$

Lemma 5.3.4. *Under the hypotheses of Theorem 5.3.1, we have :*

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

5.4 Some comments and discussion

1. Remarks on the model :

In the present work, the functional space of our model is characterized by the regularity condition (H2). Of course this condition is closely related to the existence of the differentiability of the operators $\frac{\partial^l f_Y^X(\cdot, y)}{\partial y^l}$ and $f_Y^X(\cdot, y)$ (cf. Ferraty et al. (2007) for more discussions on the link between the existence of the derivative of ψ_l and Ψ_l). It should be noted that, this condition is used in order to keep the usual form of the quadratic error (cf. Vieu, 1991). However, if we replace (H2) by a Lipschitz condition as :

$$\forall (y_1, y_2) \in N_y \times N_y \quad \forall (x_1, x_2) \in N_x \times N_x,$$

$$\|f_Y^X(x_1, y_1) - f_Y^X(x_2, y_2)\| \leq C (|\delta(x_1, x_2)|^2 + |y_1 - y_2|^2)$$

which is less restrictive than the condition (H2), we obtain a result as follows

$$\left[\widehat{f}_Y^X(x, y) - f_Y^X(x, y) \right]^2 = O(h_H^4 + h_K^2) + O\left(\frac{1}{nh_H \phi_x(h_K)}\right).$$

But, such expression of the convergence rate is inexact and can not be used to determine the smoothing parameter. In other words, this condition of the differentiability is a good compromise for obtaining an expression asymptotically exact of the convergence rate.

5.5 Proofs

Proof of Theorem 5.3.1. Since

$$E \left[\widehat{f}_Y^X(x, y) - f_Y^X(x, y) \right]^2 = \left[E \left(\widehat{f}_Y^X(x, y) \right) - f_Y^X(x, y) \right]^2 + \text{Var} \left[\widehat{f}_Y^X(x, y) \right]$$

the proof of this Theorem is based on the separate calculate separately of the two parts : bias and variance terms. Then we have to distinguish two stages in this proof. Firstly, for the bias term, we recall that, for all $z \neq 0, p \in \mathbb{N}^*$, we can write :

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

By using this decomposition for $z = \widehat{f}_D(x)$ and $p = 1$ we show that

$$\begin{aligned} \widehat{f}_Y^X(x, y) - f_Y^X(x, y) &= \left(\widehat{f}_N(x, y) - f_Y^X(x, y) \right) \\ &\quad - \left(\widehat{f}_N(x, y) - E\widehat{f}_N(x, y) \right) \left(\widehat{f}_D(x) - 1 \right) \\ &\quad - (E\widehat{f}_N(x, y)) \left(\widehat{f}_D(x) - 1 \right) \\ &\quad + \left(\widehat{f}_D(x) - 1 \right)^2 \widehat{f}_Y^X(x, y) \end{aligned} \tag{4}$$

which imply that :

$$\begin{aligned} E \left[\widehat{f}_Y^X(x, y) \right] - f_Y^X(x, y) &= \left(E\widehat{f}_N(x, y) - f_Y^X(x, y) \right) - \text{Cov}(\widehat{f}_N(x, y), \widehat{f}_D(x)) \\ &\quad + E \left(\widehat{f}_D(x) - E\widehat{f}_D(x) \right)^2 \widehat{f}_Y^X(x, y) \end{aligned}$$

As the kernel H is bounded, we can find a constant $C > 0$ such that $\widehat{f}_Y^X(x, y) \leq Ch_H^{-1}$. Hence,

$$\begin{aligned} E \left[\widehat{f}_Y^X(x, y) \right] - f_Y^X(x, y) &= \left(E\widehat{f}_N(x, y) - f_Y^X(x, y) \right) - \text{Cov}(\widehat{f}_N(x, y), \widehat{f}_D(x)) \\ &\quad + \text{Var} \left(\widehat{f}_D(x) \right) O(h_H^{-1}). \end{aligned}$$

Secondly, concerning the variance term, we use similar ideas as those used by Sarda and Vieu (2000), and Bosq et Lecoutre (1987) to deduce that :

$$\begin{aligned} \text{Var} \left[\widehat{f}_Y^X(x, y) \right] &= \text{Var} \left[\widehat{f}_N(x, y) \right] \\ &\quad - 2(E\widehat{f}_N(x, y))\text{Cov}(\widehat{f}_N(x, y), \widehat{f}_D(x)) \\ &\quad + (E\widehat{f}_N(x, y))^2 \text{Var}(\widehat{f}_D(x)) + o \left(\frac{1}{nh_H \phi(h_K)} \right) \end{aligned}$$

Finally, the proof of Theorem 5.3.1 becomes a direct consequence of Lemmas 5.3.1, 5.3.2, 5.3.3 and 5.3.4.

■

Proof of Lemma 5.3.1 We start by writing :

$$E \left[\widehat{f}_N(x, y) \right] = E \left[\frac{1}{n(n-1)E[W_{12}]} \sum_{j \neq i, 1}^n W_{ij} H_i \right] = \frac{E[W_{12}H_1]}{E[W_{12}]} = \frac{1}{E[W_{12}]} E [W_{12}E[H_1/X]] \quad (5)$$

To evaluate the quantity $E[H_1/X]$, we use the usual change of variable $t = h_H^{-1}(y-z)$. Thus,

$$E[H_1/X] = \frac{1}{h_H} \int H \left(\frac{y-z}{h_H} \right) f_Y^X(X, z) dz = \int H(t) f_Y^X(X, y - h_H t) dt.$$

As $f_Y^X(X, \cdot)$ is of class C^2 in y , then, we can use the Taylor development of order two as follows

$$f_Y^X(X, y - h_H t) = f_Y^X(X, y) - h_H t \frac{\partial f_Y^X(X, y)}{\partial y} + \frac{h_H^2 t^2}{2} \frac{\partial^2 f_Y^X(X, y)}{\partial y^2} + o(h_H^2).$$

Moreover, under (H5), we get :

$$E[H_1/X] = f_Y^X(X, y) + \frac{h_H^2}{2} \frac{\partial^2 f_Y^X(X, y)}{\partial y^2} \int t^2 H(t) dt + o(h_H^2).$$

which can be re-written as

$$E[H_1/X] = \psi_0(X, y) + \frac{h_H^2}{2} \left(\int t^2 H(t) dt \right) \psi_2(X, y) + o(h_H^2).$$

It follows, from (5), that :

$$E \left[\widehat{f}_N(x, y) \right] = \frac{1}{E[W_{12}]} (E [W_{12}\psi_0(X, y)] + E [W_{12}\psi_2(X, y)] + o(h_H^2))$$

Now, by the same arguments as those used by in Barrientos et al. (2010) for the regression function, we show that :

$$\begin{aligned} E [W_{12}\psi_l(X, y)] &= \psi_l(x, y)E[W_{12}] + E [W_{12}(\psi_l(X, y) - \psi_l(x, y))] \\ &= \psi_l(x, y)E[W_{12}] + E [W_{12}E [\psi_l(X, y) - \psi_l(x, y)|\delta(x, X)]] \\ &= \psi_l(x, y)E[W_{12}] + E [W_{12}(\Psi_l(\delta(x, X)))] . \end{aligned}$$

Moreover, since $\Psi_l(0) = 0$, for $l \in \{0, 2\}$, we have that :

$$E [W_{12} (\Psi_l(\delta(x, X)))] = \Psi'_l(0)E [\delta(x, X)W_{12}] + o(E [\delta(x, X)W_{12}]).$$

Hence,

$$\begin{aligned} E [\widehat{f}_N(x, y)] &= f_Y^X(x, y) + \frac{h_H^2}{2} \frac{\partial^2 f_Y^X(X, y)}{\partial y^2} \int t^2 H(t) dt + o\left(h_H^2 \frac{\mathbb{E}[\delta(x, X)W_{12}]}{E[W_{12}]}\right) \\ &\quad + \Psi'_0(0) \frac{\mathbb{E}[\delta(x, X)W_{12}]}{E[W_{12}]} + o\left(\frac{E[\delta(x, X)W_{12}]}{E[W_{12}]}\right) \end{aligned}$$

It is clear that,

$$\begin{aligned} E [\delta(x, X)W_{12}] &= E [K_1 \beta_1^2 \delta_1] EK_1 - E[K_1 \beta_1]E[K_1 \delta_1 \beta_1]E [W_{12}] \\ &= E[K_1 \beta_1^2]EK_1 - (E[K_1 \beta_1])^2. \end{aligned}$$

Observe that, under (H4), for all $a > 0$

$$E[K_1^a \beta_1] \leq C \int_{B(x, h_K)} \beta(u, x) dP(u)$$

Using, the last part of (H3) we get :

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

So, we can see that :

$$E[K_1^a \beta_1] = o(h_K^2 \phi_x(h_K)) \quad (6)$$

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

$$E[K_1^a \beta_1^b] = E[K_1^a \delta^b(x, X)] + E\left[K_1(\beta^b(X, x) - \delta^b(x, X))\right]$$

and, the second part of (H3) implies that,

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

whereas the first part gives,

$$\mathbb{1}_{B(x, h_K)} |\beta(X, x)| \leq \mathbb{1}_{B(x, h_K)} |\delta(x, X)| \leq b.$$

Thus,

$$\begin{aligned} E\left[K_1^a(\beta^b(X, x) - \delta^b(x, X))\right] &\leq b \sup_{u \in B(x, h_K)} |\beta(u, x) - \delta(x, u)| E[K_1^a |\delta|^b(x, X)] \\ &\leq b \sup_{u \in B(x, h_K)} |\beta(u, x) - \delta(x, u)| h_K^b E[K_1^a] \\ &\leq b \sup_{u \in B(x, h_K)} |\beta(u, x) - \delta(x, u)| h_K^b \phi_x(h_K) \end{aligned}$$

then,

$$E[K_1^a \beta_1^b] = E[K_1^a \delta^b(x, X)] + o(h_K^b \phi_x(h_K))$$

Concerning the first term, we write :

$$\begin{aligned}
h_K^{-b} E[K_1^a \delta^b] &= \int v^b K^a(v) dP^{h_K^{-1} \delta(x, X)}(v) \\
&= \int_{-1}^1 \left[K^a(1) - \int_v^1 ((u^b K^a(u))') du \right] dP^{h_K^{-1} \delta(x, X)}(v) \\
&= \left(K(1) \phi_x(h_K) - \int_{-1}^1 (u^b K^a(u))' \phi_x(-h_K, uh_K) du \right) \\
&= \phi_x(h_K) \left(K(1) - \int_{-1}^1 (u^b K^a(u))' \frac{\phi_x(-h_K, uh_K)}{\phi_x(h_K)} du \right)
\end{aligned}$$

Finally, under (H1), we obtain :

$$E[K_1^a \beta_1^b] = h^b \phi_x(h_K) \left(K(1) - \int_{-1}^1 (u^b K^a(u))' \chi_x(u) du \right) + o(h_K^b \phi_x(h_K)) \quad (7)$$

It follows that :

$$E[W_{12}] = h_K^2 \phi_x^2(h_K) \left(K(1) - \int_{-1}^1 (u^2 K(u))' \chi_x(u) du \right)$$

$$\left(K(1) - \int_{-1}^1 K'(u) \chi_x(u) du \right) + o(h_K^2 \phi_x^2(h_K))$$

and

$$E[\delta(x, X) W_{12}] = h_K^3 \phi_x^2(h_K) \left(K(1) - \int_{-1}^1 (u^3 K(u))' \chi_x(u) du \right)$$

$$\left(K(1) - \int_{-1}^1 K'(u) \chi_x(u) du \right) + o(h_K^3 \phi_x^2(h_K)).$$

Consequently

$$E[\widehat{f}_N(x, y)] = f_Y^X(x, y) + \frac{h_H^2}{2} \frac{\partial^2 f_Y^X(x, y)}{\partial y^2} \int t^2 H(t) dt + o(h_H^2)$$

$$+ h_K \Psi_0'(0) \frac{(K(1) - \int_{-1}^1 (u^3 K(u))' \chi_x(u) du)}{(K(1) - \int_{-1}^1 (u^2 K(u))' \chi_x(u) du)} + o(h_K^2)$$

Proof of Lemma 5.3.2 It is clear that

$$Var(\widehat{f}_N(x, y)) = \frac{1}{(nh_H(n-1)E[W_{12}])^2} Var\left(\sum_{i \neq j=1}^n W_{ij} H_i\right) \quad (8)$$

$$\begin{aligned}
&= \frac{1}{(n(n-1)h_H(EW_{12}))^2} (n(n-1)E[W_{12}^2 H_1^2] + n(n-1)E[W_{12}W_{21}H_1H_2] \\
&\quad n(n-1)(n-2)E[W_{12}W_{13}H_1^2] + n(n-1)(n-2)E[W_{12}W_{23}H_1H_2] \quad (9)
\end{aligned}$$

$$n(n-1)(n-2)E[W_{12}W_{31}H_1H_3] + n(n-1)(n-2)E[W_{12}W_{32}H_1H_3] \quad (10)$$

$$-n(n-1)(4n-6)(E[W_{12}H_1])^2 \quad (11)$$

Observe that the previous lemma gives $\frac{E[W_{12}H_1]}{E[W_{12}]} = O(1)$. Furthermore, by a simple manipulation and by using (6) and (7) we arrive at :

$$\left\{ \begin{array}{l} E[W_{12}^2 H_1^2] = O(h_K^4 h_H \phi_x^2(h_K)), \\ E[W_{12}W_{13}H_1^2] = E[\beta_1^4 K_1^2 H_1^2](E[K_1])^2 + o(h_K^4 h_H \phi_x^3(h_K)) \\ E[W_{12}W_{21}H_1H_2] = O(h_K^4 h_H^2 \phi_x^2(h_K)), \\ E[W_{12}W_{23}H_1H_2] = O(h_K^4 h_H^2 \phi_x^3(h_K)), \\ \quad \quad \quad \text{hfill} \\ E[W_{12}W_{31}H_1H_3] = O(h_K^4 h_H^2 \phi_x^3(h_K)) \\ E[W_{12}W_{32}H_1H_3] = O(h_K^4 h_H^2 \phi_x^3(h_K)) \end{array} \right.$$

Therefore, the second quantity is the leading term in $Var(\widehat{f}_N(x, y))$. This term can be evaluated, by the same arguments used in the pervious proof. Indeed :

$$\begin{aligned} E[\beta_1^4 K_1^2 H_1^2] &= E[\beta_1^4 K_1^2 E(H_1^2/X)] \\ &= E\left[\beta_1^4 K_1^{22} \left(\frac{y-z}{h_H}\right) f_Y^X(X, z) dz\right] \\ &= h_H E[\beta_1^4 K_1^{22}(t) f_Y^X(X, y - h_H t) dt]. \end{aligned}$$

From the first order Taylor's expansion, we have

$$f_Y^X(X, y - h_H t) = f_Y^X(X, y) + O(h_H) = f_Y^X(X, y) + o(1).$$

Next,

$$E[\beta_1^4 K_1^2 H_1^2] = h_H \int H^2(t) dt E[\beta_1^4 K_1^2 f_Y^X(X, y)] + o(h_H E[\beta_1^4 K_1^2]).$$

Once again we follow the same steps as in the previous Lemma to write

$$E[\beta_1^4 K_1^2 f_Y^X(X, y)] = f_Y^X(x, y) E[\beta_1^4 K_1^2] + o(E[\beta_1^4 K_1^2])$$

which implies that

$$E[\beta_1^4 K_1^2 H_1^2] = h_H f_Y^X(x, y) \int H^2(t) dt (E[\beta_1^4 K_1^2]) + o(h_H E[\beta_1^4 K_1^2]).$$

Combining the last equation with (11) we obtain

$$\begin{aligned} \text{Var}(\widehat{f}_N(x, y)) &= \frac{f_Y^X(x, y)}{nh_H\phi_x(h_K)} \left(\int H^2(t) dt \right) \left[\frac{\phi_x(h_K)E[\beta_1^4 K_1^2]}{(E[\beta_1^2 K_1])^2} \right] + o\left(\frac{1}{nh_H\phi_x(h_K)}\right). \quad (12) \\ \text{Var}(\widehat{f}_N(x, y)) &= \frac{f_Y^X(x, y)}{nh_H\phi_x(h_K)} \left(\int H^2(t) dt \right) \left[\frac{(K^2(1) - \int_{-1}^1 (u^4 K^2(u))' \chi_x(u) du)}{(K(1) - \int_{-1}^1 (u^2 K(u))' \chi_x(u) du)^2} \right] \end{aligned}$$

So, by (7) we obtain :

$$+o\left(\frac{1}{nh_H\phi_x(h_K)}\right).$$

■

Proof of Lemma 5.3.3 The proof of this lemma is very similar to the proof of the Lemma 5.3.2. It suffices to write :

$$\begin{aligned} \text{Cov}(\widehat{f}_N(x, y), \widehat{f}_D(x)) &= \frac{1}{h_H(n(n-1)E[W_{12}])^2} \text{Cov}(\sum_{i \neq j}^n W_{ij} H_i, \sum_{i' \neq j' = 1}^n W_{i'j'}) \\ &= \frac{1}{h_H(n(n-1)E[W_{12}])^2} (n(n-1)E[W_{12}^2 H_1] \\ &\quad + n(n-1)E[W_{12}W_{21}H_1]n(n-1)(n-2)E[W_{12}W_{13}H_1] \\ &\quad + n(n-1)(n-2)E[W_{12}W_{23}H_1]n(n-1)(n-2)E[W_{12}W_{31}H_1] \\ &\quad + n(n-1)(n-2)E[W_{12}W_{32}H_1] \\ &\quad - n(n-1)(4n-6)(E[W_{12}H_1]E[W_{12}]) \end{aligned}$$

By a simple algebra, we deduce that :

$$\left\{ \begin{array}{l} E[W_{12}^2 H_1] = O(h_K^4 h_H \phi_x^2(h_K)), \\ E[W_{12}W_{13}H_1] = O(h_K^4 h_H \phi_x^2(h_K)), \\ E[W_{12}W_{21}H_1] = O(h_K^4 h_H \phi_x^2(h_K)), \quad E[W_{12}W_{23}H_1] = O(h_K^4 h_H \phi_x^2(h_K)), \\ E[W_{12}W_{31}H_1] = O(h_K^4 h_H \phi_x^2(h_K)), \\ E[W_{12}W_{32}H_1] = O(h_K^4 h_H \phi_x^2(h_K)) \end{array} \right.$$

Furthermore, as $E[W_{12}] = O(h_K^2 \phi_x(h_K))$, we have :

$$\text{Cov}(\widehat{f}_N(x, y), \widehat{f}_D(x)) = O\left(\frac{1}{n\phi_x(h_K)}\right). \quad (13)$$

■

Proof of Lemma 5.3.4 Similarly to Lemma 5.3.2, we write :

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

Because of :

$$\left\{ \begin{array}{l} E[W_{12}^2] = O(h_K^4 \phi_x^2(h_K)), \\ E[W_{12}W_{13}] = O(h_K^4 \phi_x^2(h_K)), \\ E[W_{12}W_{21}] = O(h_K^4 \phi_x^2(h_K)), \\ E[W_{12}W_{23}] = O(h_K^4 \phi_x^2(h_K)), \\ E[W_{12}W_{31}] = O(h_K^4 \phi_x^2(h_K)), \\ E[W_{12}W_{32}] = O(h_K^4 \phi_x^2(h_K)) \end{array} \right.$$

we have that :

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

which completes the proof.

■

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] Baïllo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response, *Journal of Multivariate Analysis*, **100**, Pages 102–111.
- [3] Bosq, D., Lecoutre, J. P., (1987), *Théorie de l'estimation fonctionnelle.*, Ed. Economica.
- [4] Chu, C.-K. and Marron, J.-S. (1991). Choosing a kernel regression estimator. With comments and a rejoinder by the authors. *Statist. Sci.*, **6**, Pages 404–436.
- [5] El Methni, M. and Rachdi, M. (2010). Local weighted average estimation of the regression operator for functional data. *Commun. Stat., Theory and Methods*, Volume 00, Pages 00–00.
- [6] Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, **87**, Pages 998–1004.
- [7] Fan, J., Yao, Q. and Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, **83**, Pages 189–206.
- [8] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series : Nonparametric and Parametric Methods*, Springer-Verlag, New York.
- [9] 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.
- [10] Ferraty, F., Laksaci, A. and Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat. Inference Stoch. Process.*, **9**, Pages 47–76.
- [11] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis. Theory and Practice*. Springer Series in Statistics. New York.
- [12] Laksaci, A. (2007). Convergence en moyenne quadratique de l'estimateur à noyau de la densité conditionnelle avec variable explicative fonctionnelle. *Pub. Inst. Stat. Univ. Paris*, **3**, Pages 69–80.
- [13] Sarda, P., Vieu, P. (2000), *Kernel regression.*, Schimek, Michael G. (ed.), Smoothing and regression. Approaches, computation and application. Chichester : Wiley. Wiley Series in Probability and Statistics. 43-70.
- [14] Vieu, P. (1991), Quadratic errors for nonparametric estimates under dependence., *J. Multivariate Anal.*, **39**, 324-347.

Chapitre 6

Estimation locale linéaire des paramètres conditionnels pour des données fonctionnelles : Application sur des données simulées et réelles

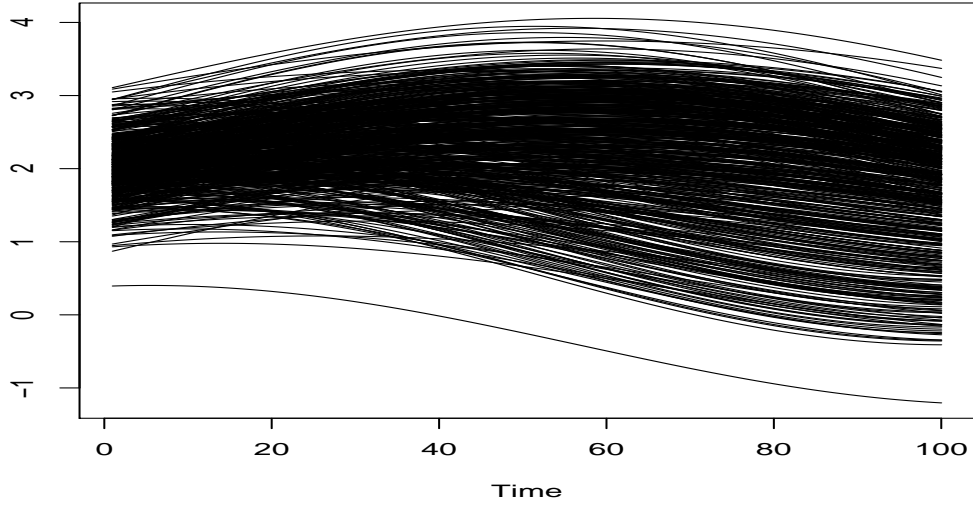
La densité conditionnelle est un outil fondamental pour décrire la relation entre deux variables aléatoires. Dans ce chapitre, nous allons déterminer ce lien par la méthode d'estimation non paramétrique locale linéaire. L'objectif principal est de montrer à l'aide de données simulées puis réelles l'applicabilité de cette méthode dans le cadre fonctionnel. Dans un premier temps, nous illustrons le mode conditionnel comme outil de prévision très lié à l'estimation de la densité conditionnelle. Ensuite, nous proposons une implémentation facile et rapide de l'estimateur de la densité conditionnelle proposé dans le chapitre 2. Enfin, une application sur des données réelles prouvera la supériorité de la méthode d'estimation par polynômes locaux sur la méthode à noyau. Nous abordons également, dans cette étude, de nombreuses questions d'ordre pratique telles que le choix du paramètre de lissage et des deux opérateurs β et δ (cf. les chapitres précédentes). C'est pourquoi, nous énumérerons différentes méthodes permettant de faire une sélection optimale de ces éléments.

6.1 Illustration du mode conditionnel

Nous récoltons des observations fonctionnelles générées à l'aide du processus suivant :

$$X_i(t) = \cos((1 - W_i)t) + \sin(W_it) + b_i, \quad \forall t \in [0, \pi] \text{ et } i = 1, 2, \dots, n$$

où W_i (respectivement, b_i) est distribué selon la loi uniforme $U(0, 1)$ (respectivement, $normal\mathcal{N}(1, 0.4)$). Nous supposons que ces courbes sont observées sur une grille formée d'une discrétisation en 100 points de l'intervalle $(0, \pi)$. Ces variables fonctionnelles sont représentées dans la Figure 6.1.

FIGURE 6.1 – Les courbes X_i

Concernant les variables réponse Y_i , nous les récoltons selon le modèle de régression suivant :

$$Y = r(X) + \varepsilon \text{ où } r(x) = 4 \exp\left(\frac{1}{1 + \int_0^\pi |x(t)|^2 dt}\right)$$

et ε est une variable aléatoire suivant la loi normale $\mathcal{N}(0, 0.3)$. L'objectif de cette illustration est de montrer l'utilité de la densité conditionnelle dans un contexte de prévision. A cette fin, nous partageons notre échantillon en deux paquets d'observations : un échantillon d'apprentissage $(X_i, Y_i)_{i=1, \dots, 500}$ et un échantillon de test $(X_i, Y_i)_{i=501, \dots, 550}$. Pour ce dernier sous-échantillon (échantillon de test), nous supposons que les valeurs réponses sont inconnues et nous allons les approximer par $\theta(X_i)$, où :

$$\theta(x) = \arg_y \max f^x(y)$$

que nous estimons par :

$$\hat{\theta}(x) = \arg_y \max \hat{f}_{(h_K, h_H)}(x, y)$$

où

$$\hat{f}_{(h_K, h_H)}(x, y) = \frac{\sum_{i,j=1}^n W_{ij}(x) H(h_H^{-1}(y - Y_i))}{h_H \sum_{i,j=1}^n W_{ij}(x)}$$

avec

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

en considérant la convention : $0/0 = 0$.

Dans le but de vérifier l'efficacité de ce modèle dans cette analyse prévisionnelle, on compare les quantités $(Y_i)_{i=501, \dots, 550}$ aux quantités $(\hat{\theta}(X_i))_{i=501, \dots, 550}$. Comme nous l'avions mentionné ci-dessus, la détermination pratique de cet estimateur dépend du choix des paramètres h_K , h_H , β et δ . Concernant le choix des paramètres de lissage h_K et h_H , rappelons que celui-ci s'appuie sur le critère de sélection qui a été adopté par Ferraty et Vieu (2006) dans le cadre

de la méthode d'estimation à noyau. En effet, les paramètres h_K et h_H s'obtiennent en minimisant le critère suivant :

$$\text{pour chaque courbe } X_i \text{ dans l'échantillon test, } \text{err}(h_K, h_H) = |Y_{i^*} - \theta(X_{i^*})| \quad (1)$$

où i^* désigne l'indice de la courbe la plus proche de X_i parmi toutes les courbes de l'échantillon d'apprentissage. Plus précisément, ces paramètres de lissage sont sélectionnés en minimisant le critère (1), sur l'ensemble des plus proches voisins. Notons que cette méthode de sélection est très compatible avec nos jeux des données. Cependant, à notre connaissance, il n'y a pas eu d'étude théorique qui permet de prouver l'optimalité de cette méthode, même lorsqu'on utilise la méthode à noyau. D'autre part, le choix des opérateurs δ et β joue un rôle essentiel. Concrètement, nos résultats théoriques offrent, sous certaines conditions, une grande souplesse dans le choix de ces deux opérateurs (cf. les chapitres 2-4). Dans notre cas, nous nous sommes concentrés sur le cas où β et δ sont issus de la métrique de l'indice fonctionnel. Plus précisément, nous considérons les opérateurs :

$$\beta(x, x') = \langle \theta_1, x - x' \rangle$$

Il est clair, qu'avec ce choix, on peut contrôler l'approximation des courbes avec signe (ou courbes signées). Il est à noter que lorsque les courbes sont très lisses (i.e., régulières), on peut remplacer le produit scalaire $\langle x, x' \rangle$ par $\langle x^{(q)}, x'^{(q)} \rangle$ où la valeur de l'entier naturel q est choisie en fonction du degré de lissage des courbes. Ainsi, l'opérateur β sera choisi en fonction des paramètres θ_1 et q . Rappelons que les idées dans Ait-Saidi et al. (2007), peuvent être adaptées afin de trouver une méthode de sélection pratique pour θ_1 . Cependant, cette adaptation, dans le cas de l'estimation de la densité conditionnelle, exige des outils et des résultats préliminaires supplémentaires (cf. la discussion dans Attaoui et al. (2010)).

Dans cette illustration, nous sélectionnons l'indice fonctionnel θ_1 sur l'ensemble des vecteurs propres de l'opérateur de covariance empirique :

$$\frac{1}{100} \sum_{i=1}^{100} (X_i - \bar{X})^t (X_i - \bar{X}).$$

Nous renvoyons à Barrientos et al. (2010) pour plus de discussions sur l'importance et la motivation de ce choix. Finalement, l'opérateur β est choisi en fonction des deux paramètres :

- q : l'ordre de la dérivée des données fonctionnelles.
- q_1 : l'ordre de la valeur propre associée au vecteur propre fonctionnel θ_1

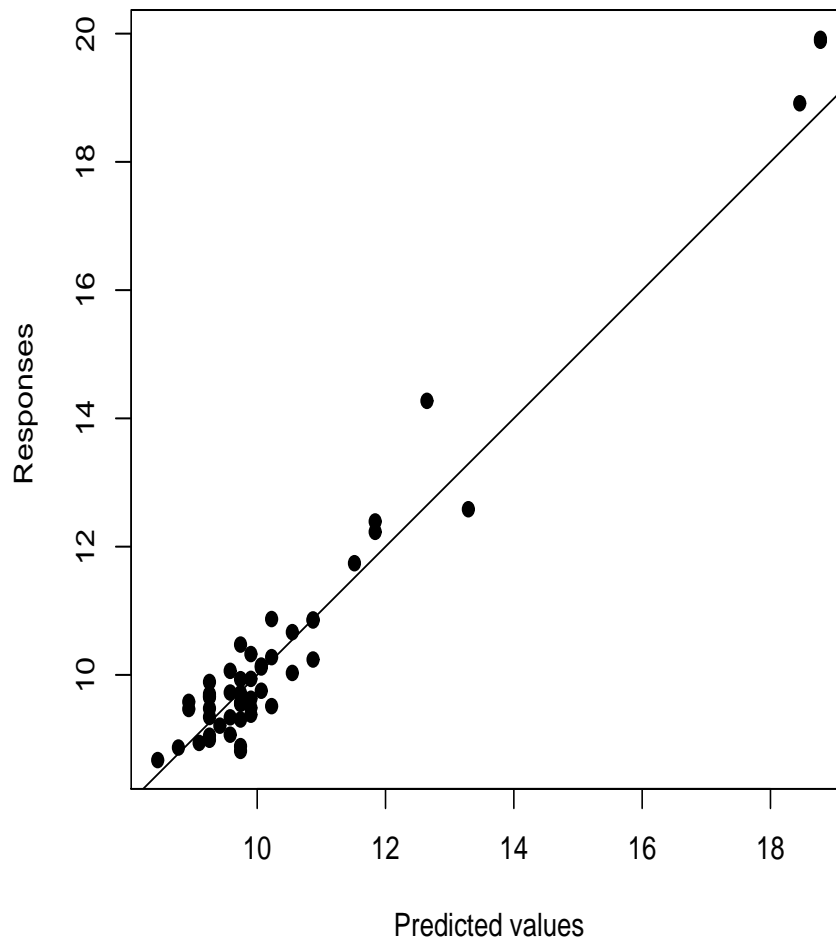
De même pour l'opérateur δ , nous pouvons prendre aussi :

$$\delta(x, x') = \langle \theta_2, x - x' \rangle$$

et nous considérons la même procédure que pour β . Toutefois, nous pouvons identifier δ à la métrique d i.e.,

$$\delta(x, x') = d(x, x') = \sqrt{\langle x - x', x - x' \rangle}$$

Afin de mettre en évidence ces nombreuses possibilités de choix des paramètres intervenant dans nos estimations, nous avons effectué nos simulations en considérant deux cas et plusieurs valeurs de q , q_1 et q_2 (l'ordre de la valeur propre associée au vecteur propre fonctionnel θ_2). A ce stade, nous remarquons que les résultats sont légèrement meilleurs pour $q = 0$, $q_1 = 1$ et $q_2 = 2$ (cf. Figure 6.2).

FIGURE 6.2 – Les résultats pour $\delta(.,.) = d(.,.)$ et $q_1 = 1$

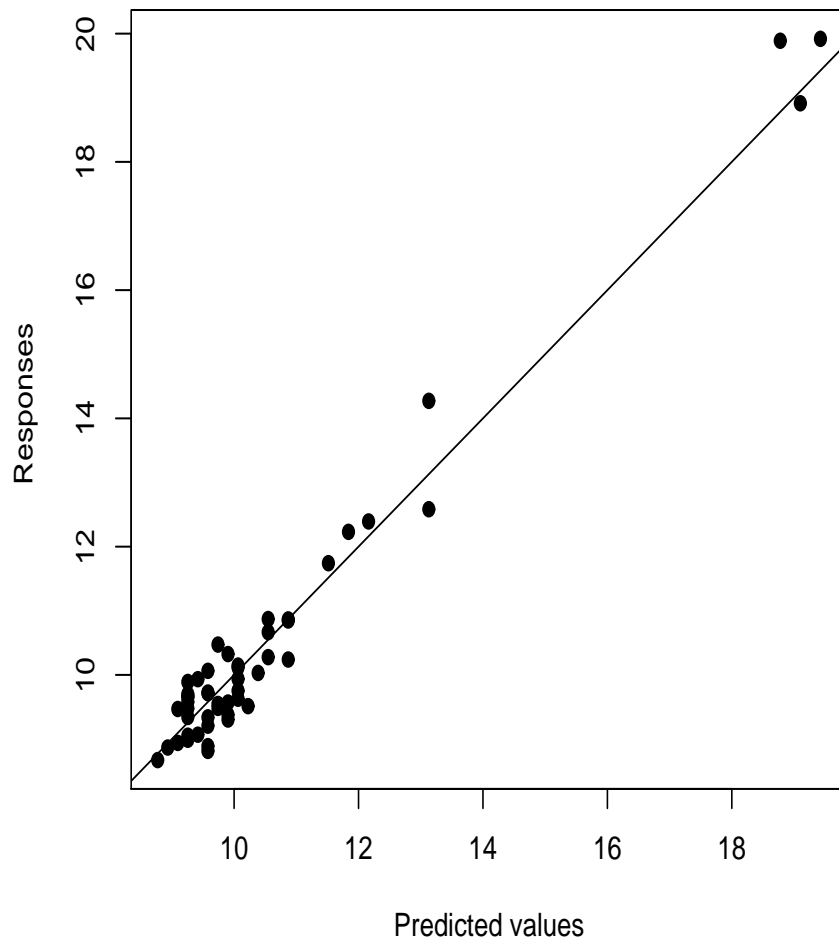


FIGURE 6.3 – Les résultats pour $\delta(x, x') = \langle \theta_2, x - x' \rangle$ où $(q_1, q_2) = (2, 1)$

En conclusion, nous pouvons affirmer que l'estimation du mode conditionnel, en utilisant la méthode par polynômes locaux, est très efficace comme modèle de prévision quand les données sont de type fonctionnel. Ceci est d'autant vrai, même si l'on ne dispose pas de résultats théoriques permettant de déterminer automatiquement les différents paramètres intervenant dans l'estimateur. De plus, on constate que ce modèle est facile à manipuler et que nos programmes garantissent une rapidité certaine dans l'obtention des résultats. D'un autre côté, on constate que le deuxième choix de δ est légèrement meilleur que le premier choix i.e., quand δ est identifié à la métrique. En effet, nous obtenons :

$$\text{MSE} = \frac{1}{50} \sum_{i=501}^{550} (Y_i - \hat{\theta}(X_i))^2 = 0.27$$

pour le premier cas et 0.19 pour le deuxième.

6.2 Illustration de la densité conditionnelle

Dans ce paragraphe, nous gardons le même jeu de données que dans la section précédente et nous comparons l'estimateur :

$$\hat{f}_{(h_K, h_H)}(x, y) = \frac{\sum_{i,j=1}^n W_{ij}(x) H(h_H^{-1}(y - Y_i))}{h_H \sum_{i,j=1}^n W_{ij}(x)}$$

à la vraie densité conditionnelle qui est donnée par :

$$f^x(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y - r(x))^2\right)$$

Cette illustration est motivée par le fait que la densité conditionnelle pourrait être utilisée à d'autres finalités (le test de multi-modalité des données, l'estimation de la fonction de hasard,...) et pas seulement comme étant une étape préliminaire à l'estimation du mode conditionnel. Il est donc très intéressant de montrer l'applicabilité de cette méthode sans avoir comme objet fondamental l'obtention du mode conditionnel. Le défi principal est de trouver des critères de choix différents de ceux proposés dans le cas précédent. Pour cela, nous proposons d'utiliser des idées similaires à celles vues dans le deuxième chapitre pour la méthode du noyau. Autrement dit, le critère de choix naturel des paramètres h_K et h_H est basé sur la minimisation des erreurs :

$$\begin{aligned} d_1(\hat{f}_{(h_K, h_H)}, f) &= \int \int \left(\hat{f}_{(h_K, h_H)}(x, y) - f(x, y) \right)^2 W_1(x) W_2(y) dP_X(x) dy \\ d_2(\hat{f}_{(h_K, h_H)}, f) &= \frac{1}{n} \sum_{i=1}^n \left(\hat{f}_{(h_K, h_H)}(X_i, Y_i) - f(X_i, Y_i) \right)^2 \frac{W_1(X_i) W_2(Y_i)}{f(X_i, Y_i)} \end{aligned}$$

où

$$d_3(\hat{f}_{(h_K, h_H)}, f) = \int \int \mathbb{E} \left(\hat{f}_{(h_K, h_H)}(x, y) - f(x, y) \right)^2 W_1(x) W_2(y) dP_X(x) dy$$

Cependant, pour la détermination pratique de nos paramètres, nous considérons le critère de validation croisée suivant :

$$\frac{1}{n} \sum_{i=1}^n W_1(X_i) \int \widehat{f}_{(h_K, h_H)}^{-i^2}(X_i, y) W_2(y) dy - \frac{2}{n} \sum_{i=1}^n \widehat{f}_{(h_K, h_H)}^{-i}(X_i, Y_i) W_1(X_i) W_2(Y_i)$$

où

$$\widehat{f}_{(h_K, h_H)}^{-k}(X_k, y) = \frac{\sum_{j \neq k, j=1}^n W_{ij}(X_k) H(h_H^{-1}(y - Y_j))}{h_H \sum_{i,j=1}^n W_{ij}(X_k)}$$

Il s'agit, en fait, d'une adaptation de l'étude effectuée dans le chapitre 2 sur l'estimateur à noyau de la densité conditionnelle. Avec cette technique, nous obtenons des résultats de simulation satisfaisants mais l'optimalité asymptotique de la méthode reste à prouver (il s'agit donc d'une question ouverte). Dans ce qui suit, nous gardons les mêmes opérateurs β et δ , et nous supposons que $W_1 = W_2 \equiv 1$. Enfin, nous utilisons les 119 observations (X_i, Y_i) pour calculer l'estimateur de $f(y|X_{120})$ quand y appartient à l'intervalle $[0.9 \times \min_{i=1, \dots, 119} (Y_i), 1.1 \times \max_{i=1, \dots, 119} (Y_i)]$. Les résultats de nos investigations figurent dans la Figure 6.4.

Afin de mettre en évidence le rôle, crucial, des paramètres de lissage dans cette estimation, nous perturbons ce choix en considérant deux couples de valeurs arbitraires : (1) un couple de valeurs plus petites que les paramètres optimaux $(h_K, h_H) = (0.29, 1.40)$ et (2) l'autre couple comporte des valeurs plus grandes que les valeurs fournies par notre critère, c'est-à-dire, nous considérons le couple $(h_K, h_H) = (0.66, 3.40)$. Nous remarquons que le couple de paramètres optimaux relativement à notre critère fournit des résultats d'estimation nettement meilleurs, car on aboutit à une erreur quadratique moyenne $MSE = 0.002$, comme erreur d'estimation, alors que pour le cas (1) nous obtenons $MSE = 0.01$ et pour le cas (2) l'erreur $MSE = 0.006$.

6.3 Application sur des données réelles

Dans cette Section, nous utilisons un jeu de données réelles afin d'atteindre deux buts. La première finalité est de donner la bonne mise en application de notre technique sur le plan pratique, et la seconde est de fournir une comparaison de l'estimation du mode conditionnel par la méthode du noyau à celle obtenue par la méthode d'estimation par polynômes locaux. Pour ce faire, nous considérons les courbes spectrométriques de masse de 197 morceaux de viande et nous fixons comme objectif la prévision du taux de matière grasse, Y , dans un morceau de viande connaissant sa courbe spectrométrique de masse, X . Ces courbes X sont présentées dans la Figure 6.5.

Plus précisément, dans ce qui suit, nous comparons la prévision via le mode conditionnel, en utilisant les deux quantités $\widehat{\theta}_{FLM}$ (estimation du mode par la méthode d'estimation fonctionnelle localement linéaire) et $\widehat{\theta}_{KM}$ (estimation du mode par la méthode d'estimation à noyau). Pour ceci, nous partageons les 197 observations, disponibles sur le site internet du groupe de travail STAPH de l'Université Paul Sabatier (Toulouse 3)¹, comme suit : nous

1. <http://www.math.univ-toulouse.fr/staph/npfda/>

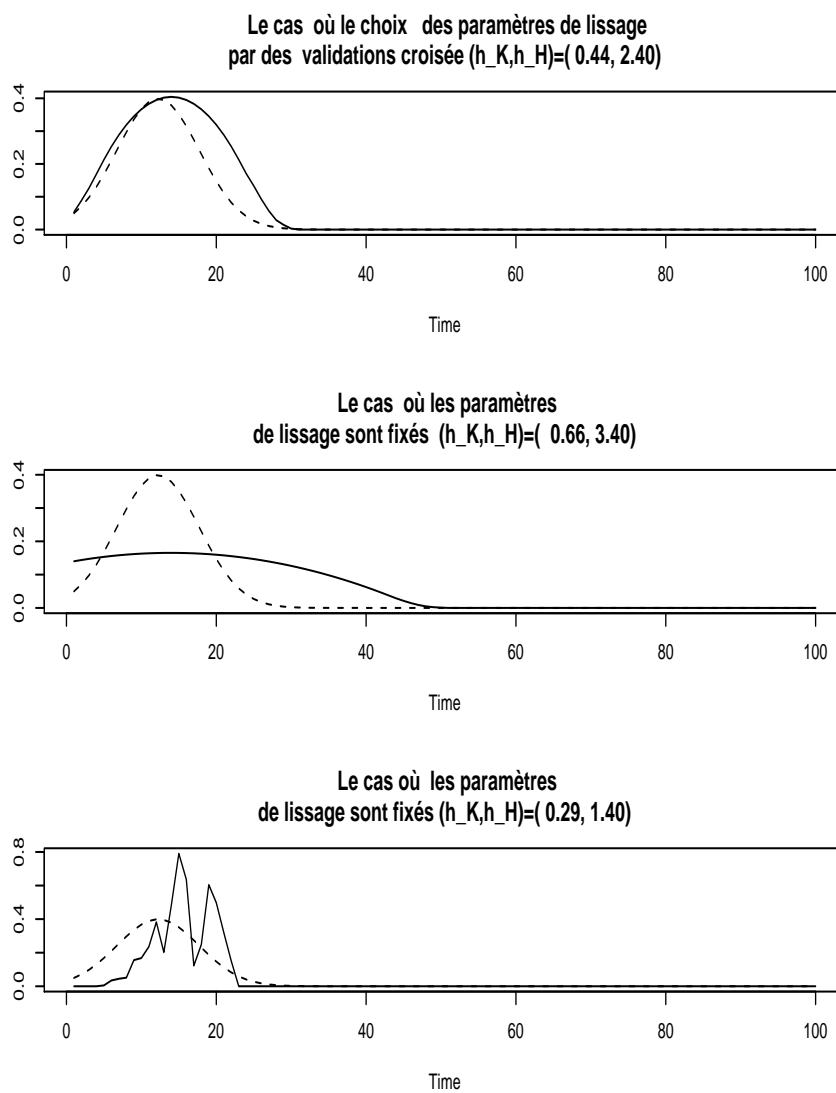


FIGURE 6.4 – La densité conditionnelle f^x est en pointillés et son estimateur $\hat{f}_{(h_K, h_H)}$ est en ligne continue

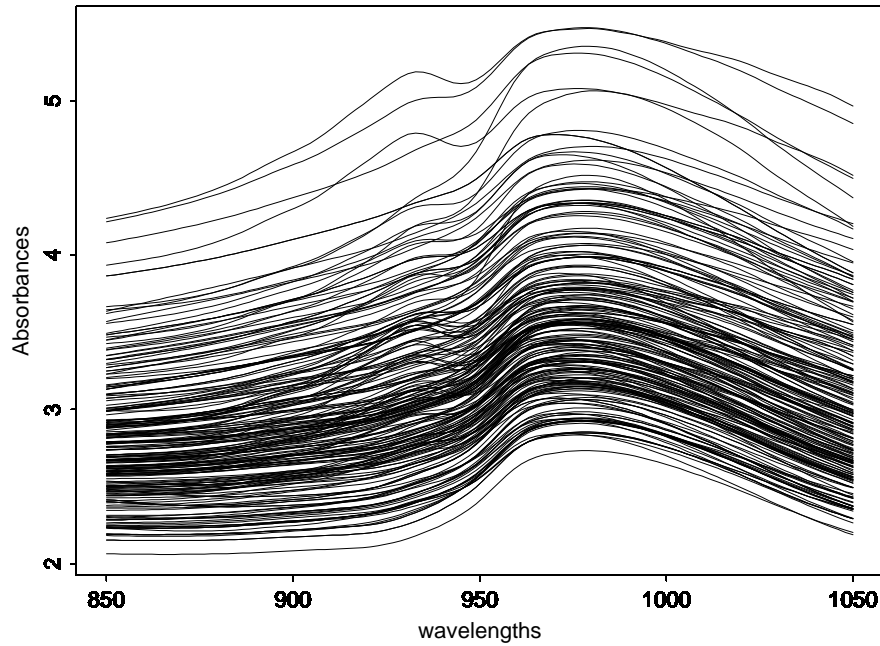


FIGURE 6.5 – Données spectrométriques de masse

considérons les 170 premières observations comme échantillon d'apprentissage et les 17 dernières comme échantillon du test. Pour calculer $\hat{\theta}_{FLM}$, nous utilisons la même procédure que, dans la première section, pour le choix des paramètres (h_K, h_H) ainsi que pour l'opérateur β . Concernant l'opérateur δ , nous prenons :

$$\delta(x, x') = d(x, x') = \sqrt{\int (x^{(q)} - x'^{(q)})^2}$$

Il est à rappeler que cette métrique pour $q = 2$ peut être considérée comme la plus adaptée pour ce type de données. Aussi, nous constatons, ici, que l'opérateur β associé aux mêmes valeurs (que précédemment) i.e., $q = 2$ et $q_1 = 69$, donne des résultats de prévision optimales relativement à l'erreur :

$$\text{MSE}(\text{FLM}) = \frac{1}{17} \sum_{i=171}^{197} (Y_i - \hat{\theta}_{FLM}(X_i))^2$$

Par ailleurs, pour calculer $\hat{\theta}_{KM}$, nous utilisons le programme dû à Ferraty et Vieu (2006) téléchargeable sur le site <http://www.math.univ-toulouse.fr/staph/npfda/> STAPH pour la même métrique d (cf. note de bas de page pour l'adresse Internet). En considérant un noyau quadratique, nous obtenons les résultats confinés dans la Figure 6.6.

A partir des résultats obtenus, nous pouvons conclure clairement que, la méthode d'estimation par polynômes locaux est nettement meilleure et plus performante que la méthode d'estimation à noyau : en fait, on obtient $\text{MSE}(\text{FLM}) = 3.84$ contre $\text{MSE}(\text{KM}) = 5.42$. De

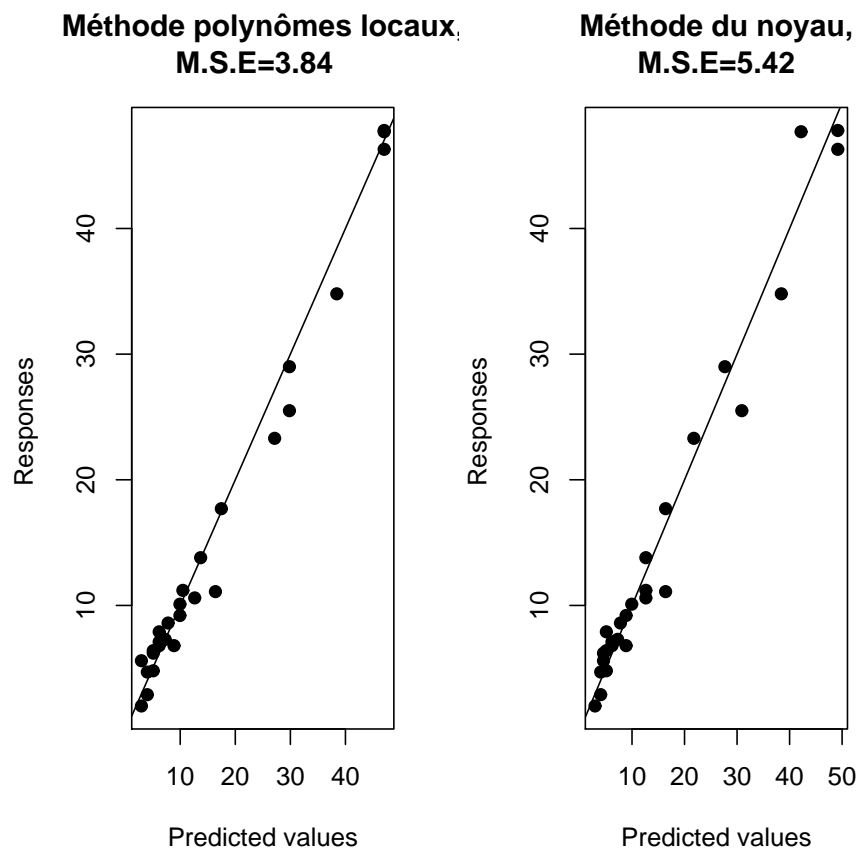


FIGURE 6.6 – Les résultats de prévision pour les deux méthodes d'estimation

plus, nos résultats sont aussi comparables à d'autres outils de prévision tels que la régression et la médiane pour lesquelles l'erreur de prévision est de 3.5 pour la régression et de 3.44 pour la médiane conditionnelle (cf. Ferraty and Vieu, 2006).

Chapitre 7

Conclusion et Perspectives

7.1 Conclusion

Dans cette thèse, nous avons abordé une étude globale de l'estimation de la densité conditionnelle quand la variable explicative fonctionnelle. Deux méthodes d'estimation ont été considérées, la première concerne la méthode de noyau tandis que la deuxième concerne la méthode d'estimation par polynômes locaux.

Pour la première méthode nous avons étudié une question cruciale dans l'estimation non paramétrique par la méthode à noyau. Il s'agit du problème de choix du paramètre de lissage. A ce sujet, nous avons proposé une méthode de sélection automatique pour les deux paramètres de lissages. L'optimalité asymptotique de notre méthode est obtenue sous des conditions standards en statistique fonctionnelle. En pratique, cette méthode est efficace, très facile à implémenter, et elle s'exécute rapidement. De plus, notre critère de choix est utilisable pour d'autres modèles non paramétriques liés à la densité conditionnelle. Ainsi, on peut dire que notre contribution donne une réponse pertinente à la question de Ferraty et Vieu (2006) et elle ouvre aussi des perspectives sur de nombreuses questions de recherche (cf. le chapitre 2).

Dans la deuxième partie, nous avons considéré une autre approche pour l'estimation de la densité conditionnelle quand les données sont fonctionnelles. L'estimateur proposé est une généralisation au cas fonctionnel de l'estimateur par local linéaire introduit par Fan et Gijbels (1996). Comme résultats asymptotiques nous avons établi la vitesse de convergence presque complète (ponctuelle et uniforme) et nous avons donné l'expression asymptotiquement exacte de l'erreur quadratique de cet estimateur. Les expressions des vitesses de convergence obtenues ont la même forme que l'estimateur à noyau, dont les deux structures fonctionnelles sont bien exploitées. Plus précisément, la dimensionnalité du modèle dans la partie biais, alors que la dimensionnalité de l'espace fonctionnel de la variable explicative a été explicité dans la partie dispersion. De même que dans le cas précédent, ces résultats asymptotiques sont obtenus sous des conditions très classiques en statistique non-paramétrique fonctionnelle et que les estimateurs sont facile à utiliser en pratique. De plus l'importance de cette deuxième partie est aussi exprimé par le nombre important de perspectives de recherche qu'il offre. Dans la section suivante, nous listons quelques exemples de ces perspectives.

7.2 Perspectives

Pour conclure les travaux de cette thèse, nous exposons dans ce qui suit, quelques développements futurs possibles en vue d'améliorer et d'étendre nos résultats.

Sur le choix des paramètres de lissage pour l'estimation local linéaire

Une question naturelle pour l'utilisation pratique de cet estimateur, dans la partie simulation, nous avons considéré le critère de validation croisés suivant :

$$\frac{1}{n} \sum_{i=1}^n W_1(X_i) \int \widehat{f}_{(h_K, h_H)}^{-i^2}(X_i, y) W_2(y) dy - \frac{2}{n} \sum_{i=1}^n \widehat{f}_{(h_K, h_H)}^{-i}(X_i, Y_i) W_1(X_i) W_2(Y_i)$$

où

$$\widehat{f}_{(h_K, h_H)}^{-k}(X_k, y) = \frac{\sum_{k \neq i, j=1}^n W_{ij}(X_k) H(h_H^{-1}(y - Y_i))}{h_H \sum_{i, j=1}^n W_{ij}(X_k)}$$

L'étude de l'optimalité asymptotique de cette méthode n'a pas été conduite, mais elle constitue une perspective de recherche à court terme.

Sur la normalité asymptotique de l'estimateur locale linéaire

Notons que la normalité asymptotique est importante en pratique et nous permet de déterminer l'intervalle de confiance où de faire des tests statistiques. Dans le futur proche, nous essaierons d'étendre les résultats de Ezzahrioui et Ould Saï (2008) sur la méthode à noyau (cas fonctionnel) pour notre estimateur local linéaire.

L'estimateur locale linéaire dans le cadre spatial

La généralisation de nos résultats au cadre spatial est une suite logique à suivre. Nous pensons que cette généralisation peut être facilement atteinte par l'adaptation des résultats de Laksaci (2010) et Rachdi et Niang et Yao (2011).

D'autres perspectives pourraient être traitées à long terme telles que, le cas où les variables sont entachées par des erreurs d'observations, le processus d'erreur est à longue mémoire,...

Par ailleurs, un autre problème mérite d'être soulevé. Il concerne la généralisation de nos résultats sur la méthode par polynômes locaux d'ordre supérieur strictement à 1.

Chapitre 8

Bibliographie générale

Acerbi, C. (2002). Spectral measures of risk : a coherent representation of subjective risk aversion. *J. Bank. Financ.*, **26**, Pages 1505-1518.

Aneiros-Pérez, G. et Vieu, P. (2006) Semi-functional partial linear regression. *Statistics & Probability Letters*. **76** (11) 1102-1110.

Antoniadis, A. et Sapatinas, T. (2007) Estimation and inference in functional mixed-effect models *Comput. Statist. & Data Anal.* **51** (10) 4793-4813.

Araki, Y., Konishi, S. et Imoto, S. Functional discriminant analysis for microarray gene expression data via radial basis function networks. (English summary) COMPSTAT 2004|Proceedings in Computational Statistics, Physica, Heidelberg 613-620.

Barrientos-Marin, J. (2007). Some Practical Problems of Recent Nonparametric Procedures : Testing, Estimation, and Application. PhD thesis from the Alicante University (Spain).

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.

Baïllo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response, text *Journal of Multivariate Analysis*, **100**, Pages 102–111.

Bashtannyk, D.M. and Hyndman, R.J. (2001). Bandwidth selection for kernel conditional density estimation. *Comput. Statist. Data Anal.*, **36**, Pages 279-298.

Benhenni, K., Ferraty, F. et Rachdi, M. (2007) Local smoothing regression with functional data. *Computational Statistics*, **22**, (3) 353-369.

Benhenni, K., Hedli-Griche, S., Rachdi, M. (2010) Estimation of the regression operator from functional fixed-design with correlated errors. *Journal of Multivariate Analysis*, **101**, Issue 2, 476-490.

Benhenni, K., Hedli-Griche, S., Rachdi, M. et Vieu, P. (2008) Consistency of the regression estimator with functional data under long memory conditions. *Statistics & Probability Letters*, **78**, Issue 8, 1043-1049.

- Benko, M. (2006) Functional Data Analysis with Applications in Finance Mémoire de Thèse. université de Berlin.
- Benko, M., Hardle, W. et Kneip, A. (2006) Common Functional Principal Components SFB 649 Discussion Papers SFB649DP2006-010, Humboldt University, Berlin, Germany.
- Bogachev, V.I. (1999). *Gaussian measures*. Math surveys and monographs, **62**, Amer. Math. Soc.
- Bosq, D. (1991), Modelization, nonparametric estimation and prediction for continuous time processes. In Nonparametric functional estimation and related topics (Spetses, 1990), 509-529, *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.* **19**, Kluwer Acad. Publ., Dordrecht.
- Bosq, D. (2000) *Linear Processes in Function Spaces : Theory and Applications* Lecture Notes in Statistics, **149**, Springer-Verlag, New York.
- Bosq, D. et Delecroix. V. (1985) Nonparametric prediction of a Hilbert space valued random variable. *Stochastic Processes Appl.* **19**, 271-280.
- Bosq, D. et Lecoutre. J.P. (1987) *Théorie de l'estimation fonctionnelle*. Economica, Paris.
- Berlinet, A., Gannoun, A. et Matzner-Lober, E. (1998). Asymptotic normality of convergent estimators of the conditional mode. *CANADIAN J. Statistics.* **26** , 2, 365-380.
- Berlinet, A., Gannoun, A. et Matzner-Lober, E. (2001) Asymptotic normality of convergent estimates of the conditional quantiles. *Statistics.* **35**, 139-169.
- Berlinet, A., Biau, G. and Rouvière, L. (2005) Functional supervised classification with wavelets. *Annales de l'ISUP*, **2**, 61-80.
- Besse, P., Cardot, H. and Ferraty, F. (1997). Simultaneous nonparametric regressions of unbalanced longitudinal data. *Comput. Statist. Data Anal.*, **24**, No. 3, Pages 255–270.
- Besse, P. et Ramsay, J.O. (1986) Principal components analysis of sampled functions. *Psychometrika* **51** (2) 285-311.
- Cadre. B. (2001) Convergent estimators for the L_1 -median of Banach valued random variable. *Statistics.* **35**, 509-521.
- Cai, T.-T. and Hall, P. (2006). Prediction in functional linear regression, *Annals of Statistics*, **34**, Pages 2159–2179.
- Cardot, H., Ferraty, F. et Sarda, P. (1999) Functional linear model. *Statist & Probability Letters.* **45**, (1) 11-22.
- Cardot, H., Ferraty, F. and Sarda, P. (2003) Spline Estimators for the Functional Linear Model *Statistica Sinica* **13** (3) 571-591.
- Cardot, H., Crambes, C. and Sarda, P. (2004) Spline estimation of conditional quantities for functional covariates. *C. R. Acad. Sci., Paris.* **339**, (2) 141-144.

- Cardot, H., Crambes, C. and Sarda, P. (2004a) Conditional Quantiles with Functional Covariates : an Application to Ozone Pollution Forecasting. *Contributed paper in Compstat Prague 2004 Proceedings* 769-776.
- Cardot, H. et Sarda, P. (2005). Quantile regression when the covariates are functions. *J. Nonparametr. Stat.* **17** (7) 841-856.
- Cardot, H. et Sarda, P. (2005a) Estimation in generalized linear models for functional data via penalized likelihood. *J. Multivariate Anal.* **92** (1) 24-41.
- Cardot, H., Crambes, C. and Sarda, P. (2006) *Ozone pollution forecasting using conditional mean and conditional quantiles with functional covariates*. Statistical methods for biostatistics and related fields, Härdle, Mori and Vieu (Eds.), Springer.
- Cardot, H. et Sarda, P. (2006) *Linear regression models for functional data*. The art of semiparametrics 49-66, Contrib. Statist., Physica-Verlag/Springer, Heidelberg.
- Cardot, H., Crambes, C. and Sarda, P. (2006a) *Ozone pollution forecasting using conditional mean and conditional quantiles with functional covariates*. Statistical methods for biostatistics and related fields, Härdle, Mori and Vieu (Eds.), Springer.
- Cardot, H., Mas, A. et Sarda, P. (2007) CLT in functional linear regression models. (English summary) *Probab. Theory Related Fields* **138** (3-4) 325-361.
- Castro, P.E., Lawton, W.H., et Sylvestre, E.A. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics*, 28 :329-337.
- Chu, C.-K. and Marron, J.-S. (1991). Choosing a kernel regression estimator. With comments and a rejoinder by the authors. *Statist. Sci.*, **6**, Pages 404–436.
- Crambes, C., Kneip, A. and Sarda, P. (2007) Smoothing splines estimators for functional linear regression *Ann. Statist.* **37**, (1) 35-72.
- Collomb, G. (1981) Estimation non-paramétrique de la régression : *Revue bibliographique Inter. Statist. Review.* **49**, 75-93.
- Collomb, G. (1985) Nonparametric regression : an up to date bibliography *Statistics*. **2**, 297-307.
- Collomb, G., Härdle, W., and Hassani, S (1987) A note on prediction via conditional mode estimation. *J. Statist. Plann. and Inf.* **15**, 227-236.
- Cuevas, A., Febrero, M. and Fraiman, R. (2002) Linear function regression : the case of fixed design and functional response *Canad. J. of Statistics* **30** 285-300.
- Cuevas, A. et Fraiman, R. (2004) *On the bootstrap methodology for functional data*. (English summary) COMPSTAT 2004|Proceedings in Computational Statistics 127-135 Physica, Heidelberg.
- Dabo-Niang, S. (2002) Estimation de la densité dans un espace de dimension infinie : Application aux diffusions. *C. R. Acad. Sci., Paris.* **334**, (3) 213-216.

- Dabo-Niang, S. and Rhomari, N. (2003) Estimation non paramétrique de la régression avec variable explicative dans un espace métrique. (French. English, French summary) [Kernel regression estimation when the regressor takes values in metric space] *C. R. Math. Acad. Sci. Paris* **336** (1) 75-80.
- Dabo-Niang, S., Ferraty, F. et Vieu, P. (2004a) Estimation du mode dans un espace vectoriel semi-normé. (French) [Mode estimation in a semi-normed vector space] *C. R. Math. Acad. Sci. Paris* **339** 9 659-662.
- Dabo-Niang, S. (2004b) Kernel density estimator in an infinite-dimensional space with a rate of convergence in the case of diffusion process. *Appl. Math. Lett.* **17** (4) 381-386.
- Dabo-Niang, S., Ferraty, F. et Vieu, P. (2006) Mode estimation for functional random variable and its application for curves classification. *Far East J. Theor. Stat.* **18** (1) 93-119.
- Dabo-Niang, S. and Laksaci, A. (2007) Estimation non paramétrique de mode conditionnel pour variable explicative fonctionnelle. (French) [Nonparametric estimation of the conditional mode when the regressor is functional] *C. R. Math. Acad. Sci. Paris* **344** (1) 49-52.
- Damon, J. et Guillas, S. (2002) The inclusion of exogenous variables in functional autoregressive ozone forecasting *Environmetrics* **13** 759-774.
- Dauxois, J., Pousse, A. and Romain, Y. (1982) Asymptotic theory for the principal component analysis of a vector random function : some applications to statistical inference *J. Multivariate Anal.* **12** (1) 136-154.
- De Boor, C. (1978) *A practical guide to Splines*. Springer, New-York.
- De Gooijer, J. and Gannoun, A. (2000) Nonparametric conditional predictive regions for time series *Comput. Statist. Data Anal.*, **33**, Pages 259–257.
- Delsol, L. (2007) Régression sur variables fonctionnelle : Estimation, Tests de structure et Application. Thèse de Doctorat, Université de Toulouse.
- Demongeot, J., Laksaci, A., Madani, F. and Rachdi, M. (2010) Local linear estimation of the conditional density for functional data. *C. R., Math., Acad. Sci. Paris*, **348**, Pages 931-934.
- Demongeot, J., Laksaci, A., Madani, F. and Rachdi, M. Functional data : Local linear estimation of the conditional density and its application *Statistics, Volume 00, Pages 00-00*, DOI : 10.1080/02331888.2011.568117.(à paraître en 2012)
- Demongeot, J., Laksaci, A., Madani, F. and Rachdi, M. (2011). *A fast functional locally modeled conditional density and mode for functional time-series*. Recent Advances in Functional Data Analysis and Related Topics “Contributions to Statistics”, Physica-Verlag/Springer, 2011, 85-90, DOI : 10.1007/978-3-7908-2736-1_13
- Dereich, S. (2003) High resolution coding of stochastic processes and small ball probabilities. PhD Thesis.
- Deville, J. C. (1974) Méthodes statistiques et numériques de l'analyse harmonique. *Ann. Insee*, **15**, 3-101.

- Eddy, W. F. (1982) The asymptotic distribution of kernel estimators of the mode. *Z. W. Giebete*, **59**, 2279-290.
- El Ghouch, A. and Genton, M. (2009). Local polynomial quantile regression with parametric features. *J. Amer. Statist. Assoc.*, **104**, No. 488, Pages 1416–1429.
- Ezzahrioui, M. (2007) Pr evision dans le smod eles conditionnels en dimension infinie. Th ese de Doctorat, Universit e du Littoral C ote d’Opale.
- Ezzahrioui, M. and Ould-Sa id, E. (2008). Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *J. Nonparametr. Stat.*, **20**, Pages 3–18.
- Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, **87**, Pages 998–1004.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. London, Chapman & Hall.
- Fan, J., Yao, Q. and Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, **83**, Pages 189–206.
- Fan, J. and Yao, Q. (2003). *Nolinear Time Series : Nonparametric and Parametric Methods*, Springer-Verlag, New York.
- Fan, J. and Yim, T.-H. (2004). A cross-validation method for estimating conditional densities. *Biometrika*, **91**, Pages 819–834.
- Ferraty, F., Goia, A. et Vieu, P. (2002) R egression non-param etrique pour des variables al eatoires fonctionnelles m elangeantes. (French) [Nonparametric regression for mixing functional random variables] *C. R. Math. Acad. Sci. Paris* **334** (3) 217-220.
- Ferraty F., Goia A. and Vieu P. (2002b) Functional nonparametric model for time series : a fractal approach for dimension reduction. *Test*, **11**, (2) 317-344
- Ferraty, F., Mas, A. and Vieu, P. (2007) Advances in nonparametric regression for fonctionnal data. Aust. and New Zeal. *J. of Statist* **49** 1-20.
- Ferraty, F., Rabhi, A. et Vieu, P. (2008) Estimation non-param etrique de la fonction de hasard avec variable explicative fonctionnelle. *Rom. J. Pure & Applied Math.* **53** (1) 1-18.
- Ferraty, F., Tadj, A., Laksaci, A. and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. of Statist. Plan. and Inf.*, **140**, Pages 335–352 .
- Ferraty, F. and Vieu, P. (2000) Dimension fractale et estimation de la r egression dans des espaces vectoriels semi-norm es *C. R. Math. Acad. Sci. Paris*, **330**, 403-406.
- Ferraty, F. et Vieu, P. (2002) The functional nonparametric model and application to spectrometric data. *Comput. Statist.* **17** (4) 545-564.

- Ferraty, F. and Vieu, P. (2003) Curves discrimination : a nonparametric functional approach. *Comput. Statist. and Data Anal.* **44**, 161-173.
- Ferraty, F. and Vieu, P. (2004) Nonparametric models for functional data, with application in regression times series prediction and curves discrimination. *J. Nonparametric. Statist.* **16**, 111-127.
- Ferraty, F., Laksaci, A. et Vieu, P. (2005) Functional time series prediction via conditional mode estimation. *C. R. Math. Acad. Sci. Paris* **340** (5) 389-392.
- Ferraty, F. and Vieu, P. (2006) *Nonparametric functional data analysis. Theory and Practice.* Springer Verlag.
- Ferraty, F., Laksaci, A. and Vieu, P. (2006) Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat. Inference Stoch. Process* **9** (1) 47-76.
- Ferraty F. and Vieu P. (2006a) *Nonparametric modelling for functional data.* Springer-Verlag, New York.
- Ferraty, F., Rabhi, A. et Vieu, P. (2008) Estimation non-paramétrique de la fonction de hasard avec variable explicative fonctionnelle. *Rom. J. Pure Applied Math.* **53**, 1-18.
- Ferraty, F., Van Keilegom, I. and Vieu, P. (2008). On the validity of the bootstrap in nonparametric functional regression. *Scandinavian J. of Statist.*, **37**, No. 2, Pages 286–306.
- Gannoun, A. (1990). Estimation non paramétrique de la médiane conditionnelle. *Pub. Inst. Stat. Univ. Paris*, **35**, No. 1, Pages 11-22.
- Gannoun, A., Saracco, J. and Yu, K. (2003). Nonparametric prediction by conditional median and quantiles. *J. of Statist. Plan. and Inf.*, **117**, No. 2, Pages 207–223.
- Gao, F. et Li, W.V. (2007) Small ball probabilities for the Slepian Gaussian fields. *Trans. Amer. Math. Soc.* **359** (3) 1339-1350 (electronic).
- Gasser, T., Hall, P. et Presnell, B. (1998) Nonparametric estimation of the mode of a distribution of random curves. *J. R. Statomptes. Soc. Ser. B Stat. Methodol.* **60** (4) 681-691.
- Geffroy, J. (1974) Sur l'estimation d'une densité dans un espace métrique. *C. R. Aca. Sci., Paris*, **278**, 1449-1452.
- Grenander, U. (1981) Abstract inference. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York. ix+526 pp.
- Györfi, L., Härdle, W., Sarda, P. and Vieu, P. (1989) Nonparametric curve estimation for time series. *Lecture Notes in Statistics.* **60**, Springer-Verlag.
- Hall, P. et Hosseini-Nasab, M. (2006) On properties of functional principal components analysis *J.R. Stat. Soc.* **B** 68 (1) 109-126.
- Hall, P., Poskitt, P. and Presnell, D. (2001) A functional data-analytic approach to signal discrimination. *Technometrics* **43** 1-9.

Hall, P., Wolk, R.C. and Yao, Q. (1999). Methods for estimating a conditional distribution function. *J. Amer. Statist. Assoc.*, **94**, Pages 154–163.

Härdle, W. (1990) Applied nonparametric regression. *Cambridge Univ. Press*, Cambridge, UK.

Härdle, W. (1991). *Smoothing Techniques with Implementation in S*. Springer, New York.

Härdle, W., Jenssen, P. and Serfling, R. (1991). Strong consistency rates for estimators of conditional functionals. *Ann. Statist.*, **16**, No. 4, Pages 1428–1449.

Härdle, W., Lütkepohl, H. and Chen, R. (1997) A review of nonparametric time series analysis. *Inter. Statist. Rev.* (**65**), 73-85.

Härdle, W. and Marron, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.*, **13**, No. 4, Pages 1465–1481.

Hassani, S., Sarda, P., et Vieu, P. (1995) Approche non paramétrique en théorie de fiabilité : revue bibliographique. *Rev. Statist. Appl.* (**35**), 27-41.

Hastie, T., Buja, A. et Tibshirani, R. (1995) Penalized discriminant analysis. *Ann. Statist.* **13** 435-475.

Hedli-Griche, S. (2008) Estimation de l'opérateur de régression pour des données fonctionnelles et des erreurs corrélées. PhD Thesis.

Holmstrom, I. (1961) On method for parametric representation of the state of the atmosphere. *Tellus.* **15** 127-149.

Hyndman, R.J. (1995). Highest-density forecast regions for non-linear and non-normal time series models. *J. Forecast.*, **14**, Pages 431–441.

Hyndman, R.J., Bashtannyk, D.M. and Grunwald, G.K. (1996). Estimating and visualizing conditional densities. *J. Comput. Graph. Statist.*, **5**, Pages 315–336.

Hyndman, R.J. and Yao, Q. (1998). Nonparametric estimation and symmetry tests for conditional density functions. *Working paper 17/98, Department of Econometrics and Business Statistics, Monash University.*

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

Jank W. et Shmueli, G. (2006) Functional Data Analysis in Electronic Commerce *Research Statistical Science* **21** (2) 155-166.

Kawasaki, Y. et Ando, T. (2004) *Functional data analysis of the dynamics of yield curves*. COMPSTAT 2004|Proceedings in Computational Statistics 1309-1316, Physica, Heidelberg.

Kirkpatrick, M. and Heckman, N. (1989) A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. *J. Math. Biol.* **27**, (4) 429-450.

- Kneip, A. et Gasser, T. (1992) Statistical tools to analyze data representing a sample of curves. *Ann. Statist.* **20** (3) 1266-1305.
- Kneip, A. et Utikal, K.J. (2001) Inference for density families using functional principal component analysis. With comments and a rejoinder by the authors. *J. Amer. Statist. Assoc.* **96** (454) 519-542.
- Kolmogorov, A.N. and Tikhomirov, V.M. (1959). ϵ -entropy and ϵ -capacity. *Uspekhi Mat. Nauk.*, **14**, Pages 3–86. (Eng. Transl. Amer. Math. Soc. Transl. Ser., **2**, Pages 277–364, (1964)).
- Konakov, V.D. (1974) On the asymptotic of the mode of multidimensional distributions. *Theory of probab. Appl.*, **19**, 794-799.
- Laksaci, A. (2005). Contribution aux modèles non paramétriques conditionnels pour variables explicatives fonctionnelles. Thèse de Doctorat, Université de Toulouse.
- Laksaci, A. (2007). Convergence en moyenne quadratique de l'estimateur à noyau de la densité conditionnelle avec variable explicative fonctionnelle. *Pub. Inst. Stat. Univ. Paris*, **3**, Pages 69–80.
- Laksaci, A. et Yousfate, A. (2002) Estimation fonctionnelle de la densité de l'opérateur de transition d'un processus de Markov à temps discret. *C. R. Acad. Sci., Paris* **334**, (11) 1035-1038.
- Laksaci, A., Madani, F. and Rachdi, M. (2012). Kernel conditional density estimation when the regressor is valued in a semi-metric space *Communications in Statistics Theory and Methods*. (to appear).
- Lanczos, C. (1956) *Applied analysis*. Prentice Hall.
- Lecoutre, J.P., et Ould-Saïd, E. (1995) Hazard rate estimation for mixing and censored processes. *J. Nonparametric Statist.* (5), 83-89.
- Laukaitis, A. et Rackauskas, A. (2002) Functional data analysis of payment systems Nonlinear Analysis : Modelling and Control **7** (2) 53-68.
- Li, W.V. and Shao, Q.M. (2001) *Gaussian processes : inequalities, small ball probabilities and applications*. In :C.R. Rao and D. Shanbhag (eds.) Stochastic processes, Theory and Methods. Handbook of Statitics, 19, North-Holland, Amsterdam.
- Lifshits, M.A., Linde, W. et Shi, Z. (2006) Small deviations of Riemann- Liouville processes in Lq-spaces with respect to fractal measures. *Proc. London Math. Soc.* (3) **92** (1) 224-250.
- Louani, D. (1998) On the asymptotique normality of the function and its derivatives under censoring. *Comm. Statist., Theory and methods.* **27**, 2909-2924.
- Louani, D. et Ould-Said, E. (1999) Asymptotique normality of kernel estimators of the conditional mode under strong mixing hypothesis. *J. Nonparametric Statist.* **11**, (4) 413-442.

- Meiring, W. (2005) *Mid-Latitude Stratospheric Ozone Variability : a Functional Data Analysis Study of Evidence of the Quasi-Biennial Oscillation, Time Trends and Solar Cycle in Ozoneonde Observations Technical report 403* Statistic and Applied Probability, University of California, Santa Barbara.
- Molenaar, P. et Boosma, D. (1987) The genetic analysis of repeated measures : the karhunen-loeve expansion. *Behavior Genetics*. **17**, 229-242.
- Müller, H.-G. and Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Stat.*, **33**, No. 2, Pages 774–805.
- Müller, H.G., Sen, R. et Stadtmüller, U. (2007) Functional data Analysis for Volatility Process.soumis
- Nadaraya, E.A. (1965) On estimation of density functions and regression curves. *Theory Prob. Appl.*, **10**. 186-190.
- Ouassou, I. and Rachdi, M. (2009). Stein type estimation of the regression operator for functional data. *Advances and Applications in Statistical Sciences*, **1**, No. 2, Pages 233-250.
- Ould-Saïd, E. (1993) Estimation nonparamétrique du mode conditionnelle. Application à la prévision. *C. R. Acad. Sci., Paris*. **I36**, 943-947.
- Ould-Saïd, E. (1997) A note on ergodic processes prediction via estimation of the conditional mode function. *Scand. J. Statist.* **24**, 231-239.
- Ould-Saïd, E. and Cai, Z. (2005) Strong uniform consistency of nonparametric estimation of the censored conditional mode function. *J. Nonparametric Statist.* **17**, 797-806.
- Perzen, E. (1962) On estimation of a probability density function and mode. *Ann. Math. Stat.* **33**, 1065-1076.
- Preda, C. (2007) Regression models for functional data by reproducing kernel Hilbert spaces methods. *J. Statist. Plann. Inference* **137** (3) 829-840.
- Quentela Del Rio, A., Vieu, P., (1997). A nonparametric conditional mode estimate. *Non-parametric J. Statist.*, **8**, 253-266.
- Rachdi, M., El Methni, M. (2011) Local weighted average estimation of the regression operator for functional data. *Comm. Stat., Theory and Methods*, (Sous presse).
- Rachdi, M., Sabre, R. (2000) Consistent estimates of the mode of the probability density function in nonparametric deconvolution problems, *Statistics & Probability Letters*. **47** 1051-114.
- Rachdi, M. et Vieu, P. (2005) Sélection automatique du paramètre de lissage pour l'estimation non-paramétrique de la régression pour des données fonctionnelles. *C. R., Math., Acad. Sci. Paris* **341** (6) 365-368.
- Rachdi, M. et Vieu, P. (2007). Nonparametric regression for functional data : automatic smoothing parameter selection. *J. Statist. Plann. and Inf.*, **137**, 2784-2801, (2007).

- Ramsay, J.O. (1982) When the data are functions. *Psychometrika* **47** (4) 379-396.
- Ramsay, J.O. (2000a) Differential equation models for statistical functions. *Canad. J. Statist.* **28** (2) 225-240.
- Ramsay, J.O. (2000b) Functional components of variation in handwriting. *Journal of the American Statistical Association* **95** 9-15.
- Ramsay, J.O., Bock, R. et Gasser, T. (1995) Comparison of height acceleration curves in the Fels, Zurich, and Berkeley growth data, *Annals of Human Biology* **22** 413-426.
- Ramsay, J. and Li, X. (1996) Curve registration. *J. R. Statist. Soc. B*, **60**, 351-363.
- Ramsay, J.O., Munhall, K.G., Gracco V.L. and Ostry D.J. (1996) Functional data analysis of lip motion. *J Acoust Soc Am* **99** 3178-3727.
- Ramsay, J. and Silverman, B. (1997) *Functional Data Analysis* Springer-Verlag, New York.
- Ramsay, J. and Silverman, B. (2002) *Applied functional data analysis : Methods and case studies* Spinger-Verlag, New York.
- Ramsay, J. and Silverman, B. (2005) *Functional Data Analysis* (Second Edition) Spinger-Verlag, New York.
- Rao, C. R. (1958) Some statistical methods for comparison of growth curves. *Biometrics* **14** 1-17.
- Rice, J. and Silverman, B. (1991) Estimating the mean and the covariance structure non-parametrically when the data are curves. *J. R. Statist. Soc. B*, **53**, 233-243.
- Rio, E. (1990). *Exponential inequalities ...* Springer-Verlag, New York.
- Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants*. Springer, ESAIM, Collection Mathématiques et Applications
- Robinson, P. (1983) Nonparametric estimators for time series. *J. of Time Ser. Anal.* **4**, 185-207.
- Romano, J. P. (1988) On weak convergence and optimality of kernel density estimates of the mode. *Ann. Statist.* **16**, **916**, 629-647.
- Rosenblatt, M. (1969). Conditional probability density and regression estimators. In *Multivariate Analysis II*, Ed. P.R. Krishnaiah. Academic Press, New York and London.
- Rossi, F., Delannay, N., Conan-Guez, B. et Verleysen, M. (2005c) Representation of Functional Data in Neural Networks *Neurocomputing* **64** 183-210
- Roussas, G. (1968) On some properties of nonparametric estimates of probability density functions. *Bull. Soc. Math. Grèce(N.S.)*, **9**, (1), 29-43.
- Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.*, **9**, Pages 65–78.

- Samanta, M. (1973) Nonparametric estimation of the mode of a multivariate density. *South African Statist. J.* **7**, 109-117.
- Samanta, M. (1989) Non-parametric estimation of conditional quantiles. *Statistics & Probability Letters.* **7**, 407-412.
- Samanta, M. and Thavaneswaran, A. (1990) Non-parametric estimation of conditional mode. *Comm. statist. Theory and Meth.* **16**, 4515-4524.
- Samworth, R. J. and Wand, M. P. (2010). Asymptotics and optimal bandwidth selection for highest density region estimation. *Ann. Statist.*, **38** , Pages 1767–1792.
- Sarda, P. and Vieu, P. (2000). *Kernel Regression*. Pages 43–70, Wiley, New York.
- Schimek, M. (2000) *Smoothing and regression : Approches, computation and application*. Ed. M.G. Schimek, Wiley Series in Probability and Statistics.
- Schumaker, M. (1981) *SSpline functions : basic theory*. Wiley.
- Stone, C.J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *Ann. Stat.*, **12** , 1285-1297.
- Stone, C.J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.*, **22**, No. 1, Pages 118–184.
- Stute, W. (1985) Conditional Empirical processes. *Ann. Statist.* **14**, 638-647.
- Theodoros, N. and Yannis, G.Y. (1997). Rates of convergence of estimate, Kolmogorov entropy and the dimensionality reduction principle in regression. *Ann. Statist.*, **25**, No. 6, Pages 2493–2511.
- Tucker, L.R. (1958) Determination of parameters of functional equations by factor analysis. *Psychometrika* **23** 19-23.
- Valderrama, M.J., Ocaña, F.A. et Aguilera, A.M. (2002) Forecasting PC-ARIMA models for functional data. COMPSTAT (Berlin) 25-36.
- Van Ryzin, J. (1969). On strong consistency of density estimation. *The annals of mathematical Statistics.* **40**, **5**, 1765-1772.
- Vidakovic, B. (2001) Wavelet-Based Functional Data Analysis : Theory, Applications and Ramifications Proceedings PSFVIP-3 F3399.
- Vieu, P. (1991), Quadratic errors for nonparametric estimates under dependence., *J. Multivariate Anal.*, **39**, 324-347.
- Vieu, P. (1996) A note on density mode function. *Statistics & Probability Letters.* **26**, (4) 297-307.
- Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. Chapman & Hall, London.

Wertz, W. (1981) Nonparametric estimators in abstract and homogeneous spaces. Lecture note in Mathematics. University of Technology ; Wien.

Yao, F. and Lee, T.C.M. (2006) Penalised spline models for functional principal component analysis *J.R. Stat. Soc. B* **68** (1) 3-25.

Yoshihara, K. I. (1994) *Weakly dependent stochastic sequences and their applications*. vol IV : Curve estimation based on weakly dependent data. Sanseido, Tokyo.

Youndjé, E. (1993). *Estimation non paramétrique de la densité conditionnelle par la méthode du noyau*. PhD Thesis from the Rouen University (in French).

Youndjé, E. (1996). Propriétés de convergence de l'estimateur à noyau de la densité conditionnelle. *Rev. Roumaine Math. Pures Appl.*, **41**, Pages 535–566.

Youndjé, E., Sarda, P. and Vieu, P. (1993). Kernel estimator of conditional density bandwidth selection for dependent data. *C. R. Acad. Sci. Math. Paris*, **316**, No. 9, Pages 935–938.