Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Discipline ou spécialité :

Statistique

Présentée et soutenue par :

Amel TADJ

le : mercredi 20 avril 2011

Titre :

Sur les modèles non paramétriques conditionnels en statistique fonctionnelle

---

Ecole doctorale :
Mathématiques Informatique Télécommunications (MITT)

Unité de recherche :
C.N.R.S. - U.M.R. 5219

Directeur(s) de Thèse :

Frédéric Ferraty et Philippe Vieu	Univ. Toulouse
Ali Laksaci	Univ. Sidi Bel Abbès

Rapporteurs :

Hervé Cardot	Univ. Bourgogne
Stéphane Girard	INRIA Rhône-Alpes. Grenoble

Autre(s) membre(s) du jury

Mohamed Attouch	Univ. Sidi Bel Abbès	Examinateur
Ali Gannoun	Univ. Montpellier	Examinateur
Aldo Goia	Univ. Piemonte Orientale	Examinateur

Dédicace

À mes parents.

Remerciements

En premier lieu, je tiens à remercier chaleureusement Frédéric Ferraty, Ali Laksaci et Philippe Vieu, pour leur disponibilité, leur dynamisme et leur gentillesse. Travailler avec eux fut un grand plaisir qui, je l'espère, pourra se poursuivre. Il ont su me guider avec un enthousiasme constant et communicatif, et m'encourager pendant ces quatres années.
Leurs grandes qualités scientifiques et humaines ont été indispensables à l'élaboration de cette thèse. Pour tout cela, je ne les remercierai jamais assez.

Ma gratitude va à Hervé Cardot et Stéphane Girard pour avoir accepté de rapporter sur mes travaux. Leur lecture attentive et leurs remarques judicieuses ont été précieuses.

Je suis également reconnaissante à Mohamed Attouch, Ali Gannoun et Aldo Goia pour leur participation à mon jury. J'apprécie l'intérêt que tous ont porté à mon travail. C'est pour moi un grand honneur d'avoir un tel jury.

Je remercie les membres du groupe STAPH qui m'ont accueillie comme une collègue à part entière et m'ont permis de passer quatre années très agréables.

Enfin, je souhaite de tout cœur remercier ma famille : mes parents, mes sœurs Karima et Ilhem et mes frères Mohamed et Zakaria, ainsi que mes beaux frères Aziz et Nacer pour leur amour et leur soutien sans faille. Je remercie vivement mes nièces Zineb, Yousra, Lamia, Malak et mon neveu Farès. Je les remercie tous de m'avoir supportée et encouragée pendant les moments de doute. Je n'aurai jamais pu faire cette thèse sans eux.

Je tiens enfin à remercier mon oncle Abderrahmane Bendeddouche qui m'a apporté son soutien pendant toute la réalisation de ce travail. Je lui dédie ce mémoire de thèse.

Un grand merci au laboratoire de mathématiques de l'université de Sidi Bel Abbès, ainsi qu'à son directeur Mouffak Benchohra pour tout le soutien moral et logistique.

Je remercie l'école doctorale EDMITT, le service des relations internationales et le service de scolarité pour leur disponibilité et leur efficacité.

A tous mes amis de l'université qui m'ont permis de m'évader de mes travaux et de décompresser : Nour Eddine Amroun, Mustapha Mechab, Fethallah Tebboune et Nasreddine Bouchenak.

Je souhaite enfin remercier mes copines Salima Mimouni, Malika Belarbi et Safia Sekkal pour leur conviviabilité et leur bonne humeur.

Table des matières

Chapitre 1

Introduction

1.1 Résumé

La problématique abordée dans cette thèse est l'estimation non paramétrique des modèles conditionnels à variable explicative fonctionnelle en traitant deux cas : le cas où la variable réponse est réelle et le cas d'une variable réponse fonctionnelle. On établit la convergence uniforme presque complète d'estimateurs non paramétriques pour certains modèles conditionnels.

Dans un premier temps, nous considérons une suite d'observations i.i.d. et nous construisons des estimateurs par la méthode du noyau pour la fonction de régression généralisée, la fonction de répartition conditionnelle, la densité conditionnelle, la fonction de hasard conditionnelle et le mode conditionnel. Nous étudions la convergence uniforme presque complète de ces estimateurs en précisant leurs vitesses. A titre illustratif, nous donnons des exemples d'applications sur des données simulées.

Dans un second temps, on généralise nos résultats au cas d'une variable réponse fonctionnelle (appartenant à un espace de Banach) et on estime la régression classique. Cette généralisation a été étudiée dans les deux cas : les observations i.i.d. ainsi que le cas dépendant. Dans ce dernier, nous avons fixé comme objectif la convergence presque complète ponctuelle lorsque les observations sont β-mélangeantes.

Nos résultats asymptotiques exploitent bien la structure topologique de l'espace fonctionnel de nos observations et le caractère fonctionnel de nos modèles. En effet, toutes nos vitesses de convergence sont quantifiées en fonction de la concentration de la mesure de probabilité de la variable fonctionnelle, de l'entropie de Kolmogorov et du degré de régularité des modèles. Notons également que dans le cas où la variable réponse est aussi fonctionnelle, nos vitesses de convergence contiennent un terme additionnel qui dépend du type de l'espace de Banach de la variable réponse.

1.2 Summary

In this thesis, we consider the problem of the nonparametric estimation in the conditional models when the regressor takes its values in infinite dimension space. More precisely, we treated two cases when the response variable is real and functional. One establishes almost complete uniform convergence of nonparametric estimators for certain conditional models.

Firstly, we consider a sequence of i.i.d. observations. In this context, we build kernel estimators of the conditional cumulative distribution, the conditional density, the conditional hazard function and the conditional mode. We give the uniform consistency rate of these estimators. We illustrate our results by giving an application on simulated samples.

Secondly, we generalize our results when the response variable is in a Banach space. We estimate the regression function. In this context, we treat both cases : i.i.d and dependent observations. In the last case, we consider that the observations are β-mixing and we establishes almost complete pointwise convergence.

Our asymptotic results exploit the topological structure of functional space for the observations. Let us note that all the rates of convergence are based on an hypothesis of concentration of the measure of probability

of the functional variable on the small balls and also on the Kolmogorov's entropy which measures the number of the balls necessary to cover some set. Moreover, when the response variable is functional the rate of convergence contains a new term which depends on type of Banach space.

1.3 Liste des travaux

Publications

1. F. Ferraty, A. Laksaci, A. Tadj, P. Vieu. Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference.* 140, 335-352.
2. F. Ferraty, A. Laksaci, A. Tadj, P. Vieu. Kernel regression with functional response (en révision favorable à Electronic Journal of Statistics).
3. F. Ferraty, A. Laksaci, A. Tadj, P. Vieu. Estimation de la régression pour variable explicative et réponse fonctionnelles dépendantes (Soumis pour publication).

Communications

1. F. Ferraty, A. Laksaci, A. Tadj, P. Vieu. Rate of uniform consistency for nonparametric estimates with functional variables. 5èmes Journées de Statistique Fonctionnelle et Opérationnelle (STAPH), Dijon, les 18 et 19 Juin 2009.
2. F. Ferraty, A. Laksaci, A. Tadj, P. Vieu. Modèle de régression non paramétrique fonctionnel à variable réponse Banachique. Journées Internationales de Statistique Théorique et Appliquée, Sidi Bel Abbés, les 10, 11 et 12 Avril 2010.

1.4 Les modèles non paramétriques conditionnels comme outils d'analyse prévisionnelle

Etudier les liens entre deux variables aléatoires est une question très importante en statistique. D'un point de vue historique, ce problème a été abordé pour la première fois dans un contexte géométrique par Galileo Galilei en 1632. L'idée principale de ce dernier est d'ajuster un nuage de point par une droite permettant d'interpréter la relation entre des données contaminées. Une formulation mathématique pour ce problème, connue dans la littérature sous le nom de régression linéaire, a été donnée par Legendre et Gauss indépendamment, en (1805) et (1809) et est basée sur le principe des moindres carrés. En statistique, ce problème peut être modélisé de la manière suivante : supposons qu'on dispose de deux variables aléatoires dépendantes X et Y, la prévision de Y sachant X se fait à travers X par une application r. Autrement dit, on cherche une fonction r telle que $r(X)$ soit une bonne approximation de Y selon un critère donné. Ainsi, le problème devient la minimisation de la fonction de risque suivante :

$$err(r) = \mathbb{E}l(Y - r(X)), \tag{1.1}$$

où l est une fonction de perte. Si cette fonction de perte est convexe et admet un minimum unique, alors on peut prendre la quantité

$$\widehat{r}(.) = \arg\min_{c\in\mathbb{R}} \mathbb{E}(l(Y - c)|X = x)$$

comme approximation de Y sachant $X = x$. En conséquence, on peut dire qu'en statistique non paramétrique, les outils naturels pour faire la prévision sont les modèles liés à la distribution conditionnelle de ces variables aléatoires. Parmi ces modèles les plus sollicités citons la régression classique, la régression modale (mode conditionnel), les quantiles de régressions (la médiane conditionnelle)…. Dans la suite de ce paragraphe, nous rappelons les définitions de ces outils, en explicitant la fonction de risque associée à chaque modèle et en mettant l'accent sur la diversité et l'étendue de la littérature disponible sur ces modèles.

1.4.1 La régression classique

Ayant observé X, la valeur moyenne de Y est l'approximation la plus utilisée comme prédicteur de Y. La fonction de risque associée à ce prédicteur est

$$l(z) = z^2.$$

Autrement dit, la régression classique est l'unique solution de

$$\widehat{r}(x) = \mathbb{E}(Y|X = x).$$

Cette estimation est équivalente à l'estimation par la méthode de maximum vraisemblance classique lorsque Y est X sont liées par la relation suivante :

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

où ϵ est de loi normale de variance finie σ^2 et indépendante de X. L'optimalité de cet outil de prévision dans le cas où le bruit est gaussien l'a rendu très populaire.

Dans notre contexte non paramétrique, les premiers résultats ont été obtenus par Tukey (1961). Tandis que l'estimation par la méthode du noyau a été utilisée pour la première fois en 1964 séparément par Nadaraya et Watson. Cette méthode d'estimation a connu un développement continu. En effet, Devroye (1978) a établi la convergence uniforme presque sûre de cet estimateur. Le taux de convergence optimal pour la régression non paramétrique a été donné par Stone (1980, 1982). Collomb (1981, 1983, 1984, 1985) apporte une contribution déterminante sur ce modèle. Ces travaux se sont focalisés sur l'utilisation de la régression dans la prévision de séries chronologiques. Les premiers résultats asymptotiques sur l'estimation non paramétrique de la fonction de régression sur les processus α-mélangeants ont été élaborés par Györfi *et al.* (1989). Dans ce cadre α-mélangeant, Vieu (1991) a donné les termes asymptotiquement exacts de l'erreur quadratique de l'estimateur à noyau de la fonction de régression. Nous renvoyons à Bosq et Lecoutre (1987), Schimek (2000), Sarda et Vieu (2000) pour un large éventail de références.

Plus récemment, le modèle de régression non paramétrique a été considéré pour la prévision spatiale (voir de Lu et Chen (2004), Biau et Cadre (2004), Carbon *et al.* (2007), Li et Tran (2009)).

1.4.2 Les quantiles de régressions

L'inconvénient de la régression classique est que l'estimation de la fonction de régression est sensible aux valeurs aberrantes et peut se montrer inappropriée dans certains cas, comme lorsque la distribution est multimodale ou fortement asymétrique. Ce manque de robustesse peut être résolu par la prévision en utilisant les quantiles conditionnels. Cet outil de prévision est obtenu en introduisant la fonction de risque

$$l(z) = |z| + 2p(z-1)z.$$

Ainsi le prédicteur est la solution du problème d'optimisation

$$\hat{r}(x) = \min_{c \in \mathbb{R}} \mathbb{E}(l(Y-c)|X=x) \tag{1.2}$$

La robustesse de cette méthode provient de la convexité et la bornitude de la fonction de risque. Ces deux propriétés sont indispensables pour la robustesse d'un modèle, la première sert à assurer l'unicité de la solution du problème d'optimisation (1.2), tandis que la deuxième est utilisée pour réduire l'influence des observations aberrantes sur la prévision.

L'utilisation des quantiles de régression en prévision a été étudiée par plusieurs auteurs. A titre d'exemple, Stone (1977) est le premier qui a estimé le quantile conditionnel. Il a établi la convergence en probabilité d'un estimateur basé sur l'estimation empirique de la fonction de répartition conditionnelle. La normalité asymptotique et la convergence uniforme de l'estimateur à noyau de quantiles conditionnels ont été obtenues par Samanta (1989) dans le cas i.i.d.. En 1991, Roussas a traité le cas où les observations sont issues d'un processus de Markov. Il a établi la convergence presque sûre d'un estimateur à noyau pour ce paramètre. Berlinet *et al.* (1998) donnent un théorème général sur la normalité asymptotique des estimateurs des quantiles conditionnels, indépendamment de la corrélation entre les observations. Zhou et Liang (2003) ont utilisé l'approche ci-dessus (1.2) pour estimer la médiane conditionnelle. Ils ont montré la normalité asymptotique de l'estimateur construit lorsque les observations sont α mélangeantes. Une généralisation de cet estimateur a été proposée par Gannoun et *al.* (2003). Ces derniers ont établi la convergence presque complète et la normalité asymptotique de ces estimateurs. Récemment, Li et Lin (2007) ont rétabli les résultats de Zhou et Liang (2003) pour le cas de variables associées. Nous renvoyons à Lemdani et *al.* (2009) pour l'estimation des quantiles conditionnels sur des données tronquées.

1.4.3 La régression modale

Un autre prédicteur alternatif à la régression classique est le mode conditionnel. Par définition, le mode conditionnel est la valeur qui maximise la densité conditionnelle. On suppose qu'il existe un compact S où le mode est unique noté θ. Ce prédicteur est obtenu, en considérant :

$$\theta(x) = \arg\max_{y \in S} f_{Y|X=x}(y),$$

où $f_{Y|X=x}$ est la densité conditionnelle de Y sachant $X = x$. En utilisant le critère (1.1), nous pouvons remarquer que le mode conditionnel est obtenu en considérant la fonction de risque

$$l(z) = \begin{cases} 0 & \text{si } z = 0, \\ 1 & \text{sinon .} \end{cases}$$

Un premier estimateur du mode conditionnel fut étudié par Collomb et *al.* (1987). Ces derniers ont établi la convergence uniforme de cet estimateur. Dans cet article, Collomb a donné un exemple d'un processus ϕ-mélangeant où le mode conditionnel prévoit mieux que la régression. Samanta et Thavaneswaran (1990) ont étudié la normalité asymptotique. Motivé par sa supériorité sur la régression en prévision, le mode conditionnel a fait l'objet de plusieurs travaux : Quintela-Del-Rio et Vieu (1997) sur le processus α-mélangeant, Ould-Saïd (1997) pour le cas ergodique, Louani et Ould-Saïd (1999) pour la normalité asymptotique dans le cas des observations fortement mélangeantes. Loannides et Matzner-Løber (2004) proposent un estimateur pour le mode conditionnel lorsque les variables sont entachées d'erreurs. Nous renvoyons a Ould-Saïd et Cai (2005) pour le cas de données censurées.

1.5 Modèles non paramétriques conditionnels et variables fonctionnelles : état de l'art et contributions de la thèse

La statistique fonctionnelle occupe désormais une place importante dans la recherche en statistique. Il s'agit de la modélisation statistique des données qui sont des courbes supposées observées sur toutes leurs trajectoires. Ceci est pratiquement possible en raison de la précision des appareils de mesures modernes et de l'importante capacité de stockage qu'offrent les systèmes informatiques actuels. Il est facile d'obtenir une discrétisation très fine d'objets mathématiques tels que des courbes, surfaces,..... Ce type de variables se retrouve dans de nombreux domaines, comme la météorologie, la chimie quantitative, la biométrie, l'économétrie ou l'imagerie médicale. Parmi les ouvrages de référence en la matière, on peut citer les monographies de Ramsay et Silverman (1997) pour les aspects appliqués, Bosq (2000) pour les aspects théoriques, Ferraty et Vieu (2006) pour une étude non paramétrique et Ferraty et Romain (2011) pour des développements récents. Dans le même contexte, nous renvoyons à Manteiga et Vieu (2007) ainsi que Ferraty (2010). L'objectif de ce paragraphe est de faire une étude bibliographique sur les modèles non paramétriques conditionnels considérés dans cette thèse, permettant de comparer nos résultats obtenus avec ceux qui existent déjà. Cependant, vu l'étendue de la littérature disponible dans ce domaine, nous ne pouvons pas faire un exposé exhaustif. Ainsi, nous allons restreindre notre étude bibliographique aux modèles non paramétriques.

1.5.1 Sur le modèle de régression

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

Tout au long de ce travail de thèse, nous avons considéré une variable explicative fonctionnelle. Notre apport dans ce domaine concerne en premier temps la convergence uniforme de l'estimateur à noyau de l'opérateur de régression lorsqu'on considère une variable réponse scalaire. Dans la deuxième et la troisième partie de la thèse, nous étudions le cas où la variable réponse est aussi fonctionnelle. Comme résultats asymptotiques, nous obtenons en premier temps la convergence uniforme d'un estimateur à noyau dans le cas où les observations sont indépendantes et identiquement distribuées et ensuite, nous établissons la convergence presque complète lorsque les observations sont β-mélangeantes.

1.5.2 Sur la fonction de répartition conditionnelle

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

La contribution de la thèse sur ce modèle est l'étude de la convergence uniforme sur les deux arguments réel et fonctionnel de l'estimateur de la fonction de répartition conditionnelle. La vitesse de convergence de cet estimateur est précisée. Les résultats obtenus sont détaillés dans le chapitre 2 de cette thèse. Ce sont les premiers résultats uniformes disponibles dans la littérature portant sur l'estimation de la fonction de répartition conditionnellement à une variable fonctionnelle.

1.5.3 Sur la densité conditionnelle et ses applications

L'estimation de la fonction de densité conditionnelle et ses dérivées, en statistique fonctionnelle, a été introduite par Ferraty *et al.* (2006). Ces auteurs ont obtenu la convergence presque complète dans le cas i.i.d. Depuis cet article, une littérature abondante s'est développée sur l'estimation de la densité conditionnelle et ses dérivées, notamment afin de l'utiliser pour estimer le mode conditionnel. En effet, en considérant des observations α-mélangeantes, Ferraty *et al.* (2005) ont établi la convergence presque complète d'un estimateur à noyau du mode conditionnel défini par la variable aléatoire maximisant la densité conditionnelle. Alternativement, Ezzahrioui et Ould-Saïd (2005, 2006) ont estimé le mode conditionnel par le point qui annule la dérivée de l'estimateur à noyau de la densité conditionnelle. Ces derniers se sont concentrés sur la normalité asymptotique de l'estimateur proposé dans les deux cas (i.i.d. et mixing). La précision des termes dominants de l'erreur quadratique de l'estimateur à noyau de la densité conditionnelle a été obtenue par Laksaci (2007). Nous renvoyons à Laksaci *et al.* (2010) pour la question du choix du paramètre de lissage dans l'estimation de la densité conditionnelle à variable explicative fonctionnelle.

Dans cette thèse, nous proposons d'étudier la convergence uniforme d'un estimateur à noyau pour la densité d'une variable réelle conditionnée par une variable aléatoire fonctionnelle. Nous donnons avec précision la vitesse de convergence de cet estimateur et dérivons des propriétés asymptotiques de l'estimateur du mode conditionnel. Le lecteur trouvera l'ensemble de ces résultats au chapitre 2.

1.5.4 Sur la fonction de hasard conditionnelle

La littérature sur l'estimation de la fonction de hasard conditionnelle est relativement restreinte en statistique fonctionnelle. L'article de Ferraty *et al.* (2008) est un travail précurseur sur le sujet. Dans cette publication les auteurs ont établi la convergence presque complète d'un estimateur à noyau de la fonction de hasard conditionnelle, lorsque les observations sont indépendantes et identiquement distribuées. Dans le même contexte, Ezzahrioui (2007) a étudié la normalité asymptotique. Le cas α-mélangeant a été traité par Quintela-Del-Rio (2010). Ce dernier a établi la convergence presque complète et la normalité asymptotique de l'estimateur proposé par Ferraty *et al.* (2008). L'auteur a illustré ces résultats asymptotiques par une application sur des données sismiques. On pourra regarder également le récent article de Laksaci et Mechab (2010) sur l'estimation de la fonction de hasard conditionnelle pour des données fonctionnelles spatialement dépendantes.

Dans cette thèse, nous étudions l'estimation de la fonction de hasard d'une réponse réelle conditionnellement à une variable fonctionnelle. Nous obtenons la vitesse de convergence presque complète uniforme sur les deux arguments fonctionnel et réel. Ces premiers résultats uniformes sont détaillés dans le chapitre 2.

1.6 Présentation du cadre de travail et des résultats

1.6.1 Présentation de la problématique de la thèse

La première thématique de cette thèse porte essentiellement sur l'étude de la convergence uniforme en statistique non paramétrique fonctionnelle. Rappelons que l'une des principales motivations de l'engouement de la statistique non paramétrique fonctionnelle est la solution qu'elle offre pour le problème du fléau de la dimension. Ce phénomène bien connu en statistique non paramétrique concerne la dégradation considérable de la qualité de l'estimation lorsque la dimension augmente. La convergence uniforme en statistique non paramétrique fonctionnelle engendre un autre problème de dimensionnalité. En effet, d'une manière générale le traitement de la convergence uniforme sur un ensemble donné est lié au nombre des boules qui recouvrent cet ensemble. En dimension finie pour un ensemble compact, ce nombre est de l'ordre de r^d où r est le rayon des boules, d est la dimension de l'espace. Du point de vue probabiliste, cette relation est justifiée par le fait que la probabilité de l'ensemble est majorée par le nombre de boules multipliés par r^d qui est la mesure de Lebesgue d'une boule de rayon r. Donc, on peut dire qu'il y a une relation entre le nombre des boules, la dimension de l'espace et la mesure de probabilité utilisée. Ainsi, il est naturel de s'interroger sur la vitesse de convergence uniforme des estimateurs lorsque la dimension est infinie. Bien entendu, ce nombre dépend de la structure topologique de l'espace de variable fonctionnelle considérée mais les questions les plus importantes sont :

1. Peut-on trouver un compromis entre le rayon de la boule et le nombre des boules pour assurer la convergence uniforme des estimateurs construits ?
2. Peut-on optimiser la vitesse de convergence en fonction de la structure topologique considérée ?

L'étude effectuée dans la première partie de cette thèse constitue une réponse à cette question et la notion d'entropie jouera un rôle clé dans notre démarche.

La deuxième problématique abordée est le cas où la variable réponse est fonctionnelle. Un tel sujet a été l'objet de plusieurs études dans le cas fonctionnel paramétrique : nous renvoyons le lecteur à Bosq (1989, 1990, 1991), Besse et Cardot (1996), Antoniadis et Sapatinas (2003) ou encore Laukaitis (2007). Dans la deuxième partie de cette thèse, nous donnons plusieurs résultats dans le cadre non paramétrique lorsque la variable réponse est fonctionnelle. Il s'agit, à notre connaissance, de la première étude générale dans ce contexte.

1.6.2 Le plan de la thèse

Ce premier chapitre est consacré à la présentation des notations et des résultats asymptotiques obtenus. Puis, cette thèse se divise en trois parties. La première partie s'intéresse uniquement à une variable réponse réelle et au cas d'observations i.i.d.. Dans ce contexte, nous étudions la convergence uniforme des estimateurs à noyau de l'opérateur de régression, de la fonction de répartition conditionnelle ainsi que de la densité conditionnelle. Puis, nous dérivons des résultats portant sur l'estimation de la fonction de hasard conditionnelle ou bien celle du mode conditionnel. Dans la deuxième partie, nous nous intéressons à la situation où la variable réponse est Banachique (i.e. à valeurs dans un espace de Banach) et toujours au cas d'observations i.i.d.. Le chapitre 4 propose un rappel d'outils probabilistes pour variables aléatoires Banachiques. Les résultats de convergence uniforme de l'estimateur à noyau de l'opérateur de régression sont détaillés dans le chapitre 5. Enfin, une troisième partie propose un premier résultat de convergence ponctuelle de l'estimateur à noyau de l'opérateur de régression lorsqu'on considère la réponse fonctionnelle et des observations dépendantes.

1.6.3 Présentation des modèles et ses estimateurs

Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathcal{F} \times \mathbb{R}$, où $\mathcal{F}$ est un espace semi-métrique. On note d la semi-métrique de $\mathcal{F}$. Pour tout $x \in \mathcal{F}$, on définit la fonction de régression généralisée par

$$m_\varphi(x) = \mathbb{E}\left[\varphi(Y) \mid X = x\right],$$

où φ est une fonction mesurable connue (voir Ferraty et Vieu 2006). Notons que cette fonction peut regrouper plusieurs modèles non paramétriques tels la régression classique pour $\varphi = \text{ Id }$ (Ferraty et Vieu (2002)), ou la fonction de répartition conditionnelle si $\varphi(t) = \mathbb{I}_{]-\infty,y]}(t)$ (pour $y \in \mathbb{R}$). La fonction de répartition conditionnelle de Y sachant $X = x$, notée F^x est donnée par

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

On suppose que cette distribution conditionnelle de Y sachant X est absolument continue par rapport à la mesure de Lebesgue dont on note f^x sa densité. On désigne par $f^{x(j)}$ la dérivée d'ordre j de cette densité conditionnelle.

Soit $(X_i, Y_i)_{i=1,\cdots,n}$ des couples ayant la même loi que (X, Y). On considère un estimateur pour la fonction de régression généralisée, noté $\widehat{m_\varphi}$, défini par

$$\widehat{m_\varphi}(x) = \frac{\sum_{i=1}^n K(h_K^{-1} d(x, X_i))\varphi(Y_i)}{\sum_{i=1}^n K(h_K^{-1} d(x, X_i))}, \ \forall x \in \mathcal{F},$$

où K est un noyau et h_K est une suite de réels positifs. Ainsi, on construit un estimateur pour la fonction de répartition conditionnelle, noté $\widehat{F^x}$ (voir, Ferraty *et al.* (2006)), défini par

$$\widehat{F^x}(y) = \sum_{i=1}^n W_{ni}(x)\mathbb{I}_{\{Y_i \leq y\}}, \qquad \forall y \in \mathbb{R}, \ \forall x \in \mathcal{F},$$

où

$$W_{ni}(x) = \frac{K(h_K^{-1} d(x, X_i))}{\sum_{i=1}^n K(h_K^{-1} d(x, X_i))}.$$

De même, on introduit un estimateur pour la densité conditionnelle, noté $\widehat{f^x}$, défini par

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

où H est un noyau et h_H est une suite de réels positifs.

Pour l'estimation de la fonction de hasard conditionnelle, on note par h^x la fonction de hasard conditionnelle et $\widehat{h^x}$ son estimateur à noyau, donnée respectivement par

$$h^x(y) = \frac{f^x(y)}{1 - F^x(y)}, \ \forall y, \ F^x(y) < 1, \ \forall x \in \mathcal{F},$$

$$\widehat{h^x}(y) = \frac{\widehat{f^x}(y)}{1 - \widehat{F^x}(y)}.$$

On étudie aussi le mode conditionnel. Un estimateur de θ est la variable aléatoire $\hat{\theta}$ solution du problème d'optimisation suivant :

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

1.6.4 Présentation des résultats obtenus

Réponse réelle et cas i.i.d.

Nous supposons pour commencer que l'échantillon que nous étudions est constitué de variables indépendantes et nous établissons la convergence uniforme de notre estimateur, en donnant l'expression explicite des termes de vitesse de convergence.

Théorème 1 *Soit $S_{\mathcal{F}} \subset \mathcal{F}$, on a*

$$\sup_{x \in S_{\mathcal{F}}} |\widehat{m_\varphi}(x) - m_\varphi(x)| = O(h_K^b) + O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h_K)}}\right),$$

où $\phi(h_K)$ est la concentration de la mesure de probabilité de la variable fonctionnelle X dans la boule de centre x et de rayon h_K et $\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)$ est la fonction d'entropie.

Théorème 2

$$\sup_{x \in S_{\mathcal{F}}} \sup_{y \in S_{\mathbb{R}}} |\widehat{F}^x(y) - F^x(y)| = O(h_K^{b_1}) + O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h_K)}}\right).$$

Théorème 3

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

Théorème 4

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

Corollaire 1

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

La démonstration de ces résultats et le détail des conditions imposées seront donnés dans l'article Ferraty *et al.* (2010) qui est présenté dans le chapitre 2 de cette thèse.

Réponse fonctionnelle et cas i.i.d.

Dans cette partie, nous généralisons les résultats obtenus précédemment à une variable réponse appartenant à un espace de Banach $\mathcal{B}$. En introduisant des conditions de concentration et d'entropie de Kolmogorov, on a le résultat suivant :

Théorème 5

$$\sup_{x\in S_{\mathcal{F}}}||\widehat{m}(x)-m(x)|| = O(h^b)+O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h)}}\right)+O\left(\frac{1}{n\phi(h)}\right)^{1-1/p},$$

où m est la fonction de régression définie par

$$m(x)=\mathbb{E}\left[Y\mid X=x\right],\quad \forall x\in\mathcal{F},$$

et $\widehat{m}(x)$ est l'estimateur de $m(x)$ donné par

$$\widehat{m}(x)=\frac{\sum_{i=1}^n K(h^{-1}d(x,X_i))Y_i}{\sum_{i=1}^n K(h^{-1}d(x,X_i))},\ \forall x\in\mathcal{F},$$

et où p désigne le type de l'espace $\mathcal{B}$.

La démonstration et le détail des conditions imposées font l'objet d'un article récemment soumis et qui est présenté dans le chapitre 5 de cette thèse.

Réponse fonctionnelle et cas β-mélangeant

Afin de généraliser des problèmes non standards de la statistique non paramétrique (par exemple, la prévision des séries temporelles), nous modélisons la notion de dépendance en considérant une suite d'observations β-mélangentes. Ainsi, en renforçant les hypothèses de concentration sur la loi conjointe (X_i, Y_i) ainsi que les hypothèses sur les coefficients de mélange, on montre le résultat suivant de convergence ponctuelle :

Théorème 6

$$||\widehat{m}(x) - m(x)|| = O(h^b) + O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\delta}\phi_x(h)}}\right) + O\left(\frac{1}{n^{1-\delta}\phi_x(h)}\right)^{1-1/p}.$$

Ce résultat sera établi dans le chapitre 6 de cette thèse.

Références

Antoniadis, I., Sapatinas, T. (2003). Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. J. Multivariate Anal. **87**, 133-158.

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

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

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

BENHENNI, K., FERRATY, F., RACHDI, M., VIEU, P. (2007). Locally smoothing regression with functional data. Computational. Statistics. **22**, 353-370.

BERLINET, A., GANNOUN, A., MATZNER-LOBER, E. (1998). Propriétés asymptotiques d'estimateurs convergents des quantiles conditionnels. C. R. Acad. Sci., Paris, Sér. I, Math. **326**, No.5, 611-614.

BESSE, P. C., CARDOT, H. (1996). Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. Canad. J. Statist. **24**, No.4, 467-487.

BIAU, G., CADRE, B. (2004). Nonparametric Spatial Prediction. J. Statistical Inference for Stochastic Processes. **7**, 327-349.

BOSQ, D. (1989). Propriétés des opérateurs de covariance empiriques d'un processus stationnaire hilbertien. C. R. Acad. Sci. Paris Sér. I Math. **309**, No.14, 873-875.

BOSQ, D. (1990). Modèle autorégressif hilbertien. Application à la prévision du comportement d'un processus à temps continu sur un intervalle de temps donné. C. R. Acad. Sci. Paris Sér. I Math. **310**, No.11, 787-790.

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. **335**, Kluwer Acad. Publ., Dordrecht.

BOSQ, D. (2000). Linear processes in function spaces. Theory and Applications, Lecture Notes in Statistics. **149**, Springer-Verlag.

BOSQ, D., LECOUTRE, J. P.(1987). Théorie de l'estimation fonctionnelle. ECONOMICA, Paris.

BURBA, F., FERRATY, F., VIEU, P. (2008). Convergence de l'estimateur à noyau des k plus proches voisns en régression fonctionnelle non-paramétrique. C. R. Acad. Sci. Paris. **346**, 339-342.

Carbon, M., Francq, C., Tran, L. T. (2007). Kernel regression estimation for random fields. J. Statist. Plann. Inference. **137**, 778-798.

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

Collomb, G. (1981). Estimation non paramétrique de la regression : Revue bibliographique. Inter. Statist. Revue. **49**, No.1, 75-93

Collomb, G. (1983). Méthodes non paramétriques en régression, analyse de séries temporelles, prédiction et discrimination. Doctorat d'état, Toulouse 3.

Collomb, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau. Z. W. Gebiete. **66**, 441-460.

Collomb, G. (1985). Nonparametric regression : an up to date bibliography Statistics. **16**, 309-324.

Collomb, G., Härdle, W., Hassani, S. (1987). A note on prediction via conditional mode estimation. J. Statist. Plann. and Inf. **15**, 227-236.

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

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

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

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

Delsol, L., Ferraty, F., Vieu, P. (2011). Structural test in regression on functional variables. J. Multivariate Anal. (to appear)

Devroye, L. (1978). The uniform convergence of the Nadaraya-Watson regression function estimate. Canad. J. Statist. **6**, No.2, 179-191.

Ezzahrioui, M. (2007). Prévision dans les modèles conditionnels en dimension infinie. PhD Thesis. Univ. du Littoral Côte d'Opale.

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

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

Ferraty, F. (2010). Special issue on statistical methods and problems in infinite dimensional spaces. J. Multivariate Analysis. **101**(2), 305-490.

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

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

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

Ferraty, F., Rabhi, A., Vieu, P. (2008). Estimation non paramétrique de la fonction de hasard avec variable explicative fonctionnelle. Rom. J. Pure & Applied Math. **52**, 1-18.

Ferraty, F., Romain, Y. (2011). The Oxford Handbook of Functional Data Analysis. Oxford University Press.

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

Ferraty, F., Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. Comput. Statist. and Data Anal. **17**, 545-564.

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

Ferraty, F., Vieu, P. (2006). Nonparametric functional data analysis. Theory and Practice. Springer-Verlag.

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

Galileo Galilei (1632). Dialogue concerning the two chief world systems. Modern library, 2001. Translated by Stillman Drake.

Gannoun, A., Saracco, J., Yu, K. (2003), Nonparametric prediction by conditional median and quantiles. J. Stat. Plann. Inference. **117**, No.2, 207-223.

Gauss, J. F. (1809). Theoria motus corporum coelestium. Hamburg : Perthes et Besser.

Györfi, L., Härdle, W., Sarda, P. et Vieu, P. (1989). Nonparametric curve estimation for time series, Lecture Notes in Statistics. **60**, Springer-Verlag.

Kolmogorov, A. N., Tikhomirov, V. M. (1959). ϵ-entropy and ϵ-capacity. Uspekhi Mat. Nauk. **14**, 3-86., **2**, 277-364 (1961).

Laksaci, A. (2007). Convergence en moyenne quadratique de l'estimateur à noyau de la densité conditionnelle avec variable explicative fonctionnelle. Ann. I.S.U.P. **51**, 69-80.

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

Laksaci, A., Madani, F., Rachdi, M. (2010). Kernel conditional density estimation when the regressor is valued in a semi metric space. International Statistical Review. (In press).

Laksaci, A., Mechab, B. (2010). Estimation non parametrique de la fonction de hasard avec variable explicative fonctionnelle cas des donnees spatiales. Rev : Roumaine, Math Pures Appl. **55**, 35-51.

Laukaitis, A. (2007). An Empirical Study for the Estimation of Autoregressive Hilbertian Processes by Wavelet Packet Method Nonlinear Analysis : Modelling and Control **12** (1).

Lecoutre, J. P., Ould-Saïd, E. (1993). Estimation de la fonction de hasard pour un processus fortement mélangeant avec censure. Publ. Inst. Statist. Univ. Paris. **37**, No.1-2, 59-69.

Legendre, A. M. (1805). Nouvelle méthode pour la détermination des orbites des comètes. Paris : Courcier.

Lemdani, M., Ould-Saïd, E., Poulin, N. (2009). Asymptotic properties of a conditional quantile estimator with randomly truncated. J. Multivariate Anal.**100**, 546-559.

Li, W. V., Lin, Z. (2007). Asymptotic normality for L^1-norm kernel estimator of conditional median under association dependence. J. Multivariate Anal. **98**, 1214-1230.

LI, W. V., TRAN, L. T. (2009). Nonparametric estimation of conditional expectation. J. Statist. Plann. Inference. **139**, 164-175.

LOANNIDES, D., MATZNER-LOBER, E. (2004). A note on asymptotic normality of convergent estimates of the conditional mode with errors-in-variables. J. Nonparametr. Stat. **16**, 515-524.

LOUANI, D., OULD-SAÏD, E. (1999). Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis. J. Nonparametric Statist. **11**, No.4, 413-442.

LU, Z., CHEN, X. (2004). Spatial kernel regression estimation : Weak concistency. Stat and Probab. Lett, **68**, 125-136.

MANTEIGA, W. G., VIEU, P. (2007). Statistics for Functional Data. Computational Statistics & Data Analysis. **51**, 4788-4792.

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

NADARAYA, E. (1964). On estimating regression. Theory Prob. Appl. **10**, 186-196.

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., CAI, Z. (2005). Strong uniform consistency of nonparametric estimation of the censored conditional mode function. J. Nonparametric Statist. **17**, 797-806.

QUINTELA-DEL-RIO, A. (2010). On non-parametric techniques for area-characteristic seismic hazard parameters. Geophys. J. Int. **180**, 339-346.

QUINTELA-DEL-RIO, A., VIEU, P. (1997). A nonparametric conditional mode estimate. J. Nonparametr. Statist. **8**, No.3, 253-266.

RACHDI, M., VIEU, P. (2007). Nonparametric regression functional data : Automatic smoothing parameter selection. J. Statist. Plan. Inf. **137**, 2784-2801.

RAMSAY, J. O., SILVERMAN, B. W. (1997). Functional data analysis. Springer, New York.

ROUSSAS, G. G. (1991). Kernel estimates under association : strong uniform consistency. Statist. Probab. Lett. **12**, No.5, 393-403.

SAMANTA, M. (1989). Non-parametric estimation of conditional quantiles. Stat. Probab . Lett. **7**, No.5, 407-412.

SAMANTA, M., THAVANESWARAN, A. (1990). Non-parametric estimation of conditional mode. Comm. Statist. Theory and Meth. **16**, 4515-4524.

SARDA, P., VIEU, P. (2000). Kernel regression. In : M. Schimek (ed.) Smoothing and regression ; Approaches, Computation, and Application. Wiley Series in Pprobability and Statistics, Wiley, New York.

SCHIMEK, M. (2000). Smoothing and Regression : Approaches, computation, and application, Ed. M.G. Schimek, Wiley Series in Probability and Statistics.

STONE, C. (1977). Consistent nonparametric regression. With discussion a nd a reply by the author. Ann. Statist. **5**, no. 4, 595-645.

STONE, C. (1980). Optimal rates of convergence for nonparametric estimators. Ann. Statist. **8**, No.6, 1348-1360.

STONE, C. (1982). Optimal global rates of convergence for nonparametric regression. Ann. Statist. **10**, 1040-1053.

TUKEY, J.W. (1961). Curve as parameters, and touch estimation. Proceedings of the 4th Symposium on Mathematics, Statistics and Probability, 681-694, Berkeley, CA, USA.

Nonparametric estimation. II. Statistically Equivalent Blocks and Tolerance Regions., Ann. Math. Statist. **18**, 529-539.

Vieu, P. (1991). Quadratic errors for nonparametric estimates under dependence. J. Multivariate Anal. **39** (2), 324-347.

Watson, G. S. (1964). Smooth regression analysis. Sankhya Ser. A. **26**, 359-372.

Zhou, Y., Liang, H. (2003). Asymptotic properties for L_1 norm kernel estimator of conditional median under dependance. J. Nonparametr. Stat. **15**, 205-219.

Première partie

Convergence uniforme : réponse réelle

Chapitre 2

Réponse réelle et cas i.i.d.

Ce chapitre fait l'objet d'une publication dans *Journal of Statistical Planning and Inference.*

Rate of uniform consistency for nonparametric estimates with functional variables.

Frédéric FERRATY[a], Ali LAKSACI[b], Amel TADJ[c,*] and Philippe VIEU[a]
[a]Université Paul Sabatier, Toulouse,
[b]Université Djillali Liabes, Sidi Bel Abbes
[c,*]Corresponding author : Université de Sidi Bel Abbès, BP 89 Sidi Bel Abbès 22000. Algérie . ameltdz@yahoo.fr

abstract In this paper we investigate nonparametric estimation of some functionals of the conditional distribution of a scalar response variable Y given a random variable X taking values in a semi-metric space. These functionals include the regression function, the conditional cumulative distribution, the conditional density and some other ones. The literature on nonparametric functional statistics is only concerning pointwise consistency results, and our main aim is to prove the uniform almost complete convergence (with rate) of the kernel estimators of these nonparametric models. Unlike in standard multivariate cases, the gap between pointwise and uniform results is not immediate. So, suitable topological considerations are needed, implying changes in the rates of convergence which are quantified by entropy considerations. These theoretical uniform consistency results are (or will be) key tools for many further developments in functional data analysis.
keyword : Uniform almost complete convergence, kernel estimators, functional data, entropy, semi-metric space.

2.1 Introduction

Studying the link between a scalar response variable Y given a new value for the explanatory variable X is an important subject in nonparametric statistics, and there are several ways to explain this link. For examples, the conditional expectation, the conditional distribution, the conditional density and the conditional hazard function. The purpose of this paper is to give some contribution to the nonparametric estimation of these various conditional quantities when the explanatory variable is functional. This investigation is motivated by the fact that there is an increasing number of examples coming from different fields of applied sciences for which the data are curves and there are many nonparametrical statistical problems which occur in the functional setting (see, Ferraty and Vieu, 2006 for an extensive discussion on nonparametric statistics for functional data). Note that the modelization of functional variable is becoming more and more popular since the publication of the monograph of Ramsay and Silverman (1997) on functional data analysis. However, the first results concerning the nonparametric models (mainly the regression function) were obtained by Ferraty and Vieu (2000). They established the almost complete pointwise consistency [1] of kernel estimators of the regression function when the explanatory variable is functional and the observations are i.i.d. Their study is extended to nonstandard regression problems such that time series prediction (see, Ferraty *et al.*, 2002). Niang and Rhomari (2003) stated the convergence in L^p norm of the kernel estimator of this model, and Delsol (2007) states the exact asymptotic expression for L^p errors. The asymptotic normality result for the same estimator in the strong mixing case has been obtained by Masry (2005) and extended by Delsol (2009). The kernel type estimation of some characteristics of the conditional cumulative distribution function and the successive derivatives of the conditional density have been introduced by Ferraty *et al.* (2006). They established the almost complete consistency for i.i.d observations. The strong mixing case has been studied by Ferraty *et al.* (2005). Recently, Ferraty *et al.* (2007) gave the asymptotic expansion of the mean squared error of the kernel estimator of the regression function. Pointwise asymptotic properties of a kernel estimate of the conditional hazard function have been investigated by

1. Let $(z_n)_{n\in\mathbb{N}^*}$ be a sequence of real variables; we say that z_n converges almost completly (a.co.) to zero if and only if, $\forall\epsilon > 0$, $\sum_{n=1}^{\infty} P(|z_n| > \epsilon) < \infty$. 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 $\exists\epsilon > 0$, $\sum_{n=1}^{\infty} P(|z_n| > \epsilon u_n) < \infty$. This kind of convergence implies both almost sure convergence and convergence in probability.

Ferraty *et al.* (2008). Among the lot of papers concerning the nonparametric models related with the conditional distribution of a real variable given a random variable taking values in infinite dimensional spaces, we only refer to papers by Niang and Laksaci (2007) and Ezzahrioui and Ould-Saïd (2008).

While this literature is only concerning pointwise asymptotic, our interest in this paper is to establish the uniform almost complete convergence of the nonparametric estimates of the various conditional quantities mentioned above. Uniform consistency results have been successfully used in standard nonparametric setting in order to derive asymptotic properties for data-driven bandwidth choice, additive modelling, multi-step estimation, ... So, it is natural in this setting of functional data analysis (FDA) to investigate in a systematic way uniform consistency properties. Indeed, because the FDA topic youth one can expect in a near future that all these results will be useful in numerous functional statistical methodologies like data-driven procedures, additive modelling, ... (see, Section 7 for more detailed motivations or related bibliography). Moreover, this work completes the results obtained in Ferraty and Vieu (2006) where the pointwise almost complete consistency with rate is given for such models. It is worth noting that the uniform convergence is not a direct extension of the previous pointwise results. Indeed, it requires additional topological conditions, expressed here in terms of Kolmogrov's entropy. We will see that, unlike the standard nonparametric statistics, these infinite dimensional topological considerations may lead in some cases (see for instance, Example 3 in Section 2.2) to rates which are slower for uniform results than for pointwise ones. At last, all these asymptotic results are established under conditions related to some concentration properties expressed in terms of small balls probabilities of the underlying explanatory variable. We note that our hypotheses and results unify both cases of finite and infinite dimension of the regressors, which permits to overcome the curse of dimensionality problem. Section 2 focuses on topological considerations via the Kolmogorov's entropy, whereas Section 3 deals with a general regression model. In Section 4, we study the conditional cumulative distribution while

the conditional density estimation is developed in Section 5. In Section 6, we emphasize the repercussion of the previous results on the estimation of the conditional mode and conditional hazard function. Finally, in Section 7, we comment the obtained results and their potential impact on the statistical literature as key tools for many further advances in FDA.
Throughout this paper, we consider a sample of independent pairs $(X_i, Y_i)_{1\leq i\leq n}$ identically distributed as (X, Y) which is a random vector valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a semi-metric space, d denoting the semi-metric and we will use the notation $B(x, h) = \{x' \in \mathcal{F}/d(x', x) \leq h\}$.

2.2 Topological considerations

2.2.1 Kolmogorov's entropy

The purpose of this section is to emphasize the topological components of our study. Indeed, as indicated in Ferraty and Vieu (2006), all the asymptotic results in nonparametric statistics for functional variables are closely related to the concentration properties of the probability measure of the functional variable X. Here, we have moreover to take into account the uniformity aspect. To this end, let $S_{\mathcal{F}}$ be a fixed subset of $\mathcal{F}$; we consider the following assumption :

(H1) $\forall x \in S_{\mathcal{F}},\ 0 < C\phi(h) \leq P\left(X \in B(x, h)\right) \leq C'\phi(h) < \infty.$

We can say that the first contribution of the topological structure of the functional space can be viewed through the function ϕ controling the concentration of the measure of probability of the functional variable on a small ball. Moreover, for the uniform consistency, where the main tool is to cover a subset $S_{\mathcal{F}}$ with a finite number of balls, one introduces an other topological concept defined as follows :

Definition 1
Let S be a subset of a semi-metric space $\mathcal{F}$, and let $\epsilon > 0$ be given. A finite

set of points $x_1, x_2, \ldots, x_N$ *in* $\mathcal{F}$ *is called an* ϵ*-net for* S *if* $S \subset \bigcup_{k=1}^{N} B(x_k, \epsilon)$*.* *The quantity* $\psi_S(\epsilon) = \log(N_\epsilon(S))$*, where* $N_\epsilon(S)$ *is the minimal number of open balls in* $\mathcal{F}$ *of radius* ϵ *which is necessary to cover* S*, is called the Kolmogorov's* ϵ*-entropy of the set* S*.*

This concept was introduced by Kolmogorov in the mid-1950's (see, Kolmogorov and Tikhomirov, 1959) and it represents a measure of the complexity of a set, in sense that, high entropy means that much information is needed to describe an element with an accuracy ϵ. Therefore, the choice of the topological structure (with other words, the choice of the semi-metric) will play a crucial role when one is looking at uniform (over S) asymptotic results. More precisely, we will see thereafter that a good semi-metric can increase the concentration of the probability measure of the functional variable X as well as minimize the ϵ-entropy of the subset $S_{\mathcal{F}}$. In an earlier contribution (see, Ferraty *et al.*, 2006) we highlighted the phenomenon of concentration of the probability measure of the functional variable by computing the small ball probabilities in various standard situations. We will devote Section 2.2 to discuss the behaviour of the Kolmogorov's ϵ-entropy in these standard situations. Finally, we invite the readers interested in these two concepts (entropy and small ball probabilities) or/and the use of the Kolmogorov's ϵ-entropy in dimensionality reduction problems to refer to respectively, Kuelbs and Li (1993) or/and Theodoros and Yannis (1997).

2.2.2 Same examples

We will start (Example 1) by recalling how this notion behaves in unfunctional case (that is when $\mathcal{F} = \mathbb{R}^{\mathcal{P}}$). Then, Examples 2 and 3 are covering special cases of functional process. More interestingly (from statistical point of view) is Example 4 since it allows to construct, in any case, a semi-metric with reasonably "small" entropy.
Example1 : Compact subset in finite dimensional space : A standard theorem

of topology guaranties that for each compact subset S of R^p and for each $\epsilon > 0$ there is a finite ϵ-net and we have for any $\epsilon > 0$,

$$\psi_S(\epsilon) \leq Cp \log\left(\frac{1}{\epsilon}\right).$$

More precisely, Chate and Courbage (1997) have shown that, for any $\epsilon > 0$ the regular polyhedron in $\mathbb{R}^p$ with length r can be covered by $\left(\left[\frac{2r\sqrt{p}}{\epsilon}\right]+1\right)^p$ balls, where $[m]$ is the largest integer which is less than or equal to m. Thus, the Kolmogorov's ϵ-entropy of a polyhedron P_r in $\mathbb{R}^p$ with length r is

$$\forall \epsilon > 0, \qquad \psi_{P_r}(\epsilon) \sim p \log\left(\left[\frac{2r\sqrt{p}}{\epsilon}\right]+1\right).$$

Example2 : Closed ball in a Sobolev space : Kolmogorov and Tikhomirov (1959) obtained many upper and lower bounds for the ϵ-entropy of several functional subsets. A typical result is given for the class of functions $f(t)$ on $T = [0, 2p)$ with periodic boundary conditions and

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

The ϵ-entropy of this class, denoted $W_2^m(r)$, is

$$\psi_{W_2^m(r)}(\epsilon) \leq C\left(\frac{r}{\epsilon}\right)^{1/m}.$$

Example3 : Unit ball of the Cameron-Martin space : Recently, Van der Vaart and Van Zanten (2007) characterized the Cameron-Martin space associated to a Gaussian process viewed as map in $\mathcal{C}[0,1]$ with the spectral measure μ satisfying

$$\int \exp(\delta|\lambda|)\mu(d\lambda) < \infty,$$

by

$$H = \left\{t :\mapsto Re\left(\int e^{-it\lambda}h(\lambda)d\mu(\lambda)\right) : \quad h \in L_2(\mu)\right\},$$

and they show that Kolmogorov's ϵ-entropy of the unit ball B^{CMW} of this space with respect to the supremum norm $\|.\|_\infty$ is

$$\psi_{B^{CMW}_{\|\cdot\|_\infty}}(\epsilon) \sim \left(\log\left(\frac{1}{\epsilon}\right)\right)^2, \text{ as } \epsilon \to 0.$$

Example 4 : Compact subset in a Hilbert space with a projection semi-metric : The projection-based semi-metrics are constructed in the following way. Assume that $\mathcal{H}$ is a separable Hilbert space, with inner product $< .,. >$ and with orthonormal basis $\{e_1, \ldots, e_j, \ldots\}$, and let k be a fixed integer, $k > 0$. As shown in Lemma 13.6 of Ferraty and Vieu (2006), a semi-metric d_k on $\mathcal{H}$ can be defined as follows :

$$d_k(x, x') = \sqrt{\sum_{j=1}^{k} < x - x', e_j >^2}. \tag{2.1}$$

Let χ be the operator defined from $\mathcal{H}$ into $\mathbb{R}^k$ by

$$\chi(x) = (< x, e_1 >, \ldots, < x, e_k >),$$

and let d_{eucl} be the euclidian distance on $\mathbb{R}^k$, and let us denote by $B_{eucl}(.,.)$ an open ball of $\mathbb{R}^k$ for the associated topology. Similarly, let us note by $B_k(.,.)$ an open ball of $\mathcal{H}$ for the semi-metric d_k. Because χ is a continuous map from $(\mathcal{H}, d_k)$ into $(\mathbb{R}^k, d_{eucl})$, we have that for any compact subset S of $(\mathcal{H}, d_k)$, $\chi(S)$ is a compact subset of $\mathbb{R}^k$. Therefore, for each $\epsilon > 0$ we can cover $\chi(S)$ with balls of centers $z_i \in \mathbb{R}^k$:

$$\chi(S) \subset \cup_{i=1}^{d} B_{eucl}(z_i, r), \text{ with } dr^k = C \text{ for some } C > 0. \tag{2.2}$$

For $i = 1, \ldots, d$, let x_i be an element of $\mathcal{H}$ such that $\chi(x_i) = z_i$. The solution of the equation $\chi(x) = z_i$ is not unique in general, but just take x_i to be one of these solutions. Because of (2.1), we have that :

$$\chi^{-1}(B_{eucl}(z_i, r)) = B_k(x_i, r). \tag{2.3}$$

Finally, (2.2) and (2.3) are enough to show that the Kolmogorov's ϵ-entropy of S is

$$\psi_S(\epsilon) \approx Ck \log\left(\frac{1}{\epsilon}\right).$$

2.3 Estimation of the regression function

In this Section, we consider the problem of the estimation of a generalized regression function defined as follows :

$$m_\varphi(x) = \mathbb{E}\left[\varphi(Y) \mid X = x\right], \quad \forall x \in \mathcal{F}, \tag{2.4}$$

where φ is a known real-valued Borel function. The model (2.4) has been widely studied, when the explicative variable X is real and $\varphi(Y) = Y$ while Deheuvels and Mason (2004) provide recent advances for general function φ. This model covers and includes many important nonparametric models such as the classical regression function, the conditional distribution,....
From the kernel estimator of the classical regression function (see, Ferraty and Vieu, 2006), we propose the estimate of $\widehat{m_\varphi}(x)$ of $m_\varphi(x)$ defined as

$$\widehat{m_\varphi}(x) = \frac{\sum_{i=1}^n K(h_K^{-1} d(x, X_i))\varphi(Y_i)}{\sum_{i=1}^n K(h_K^{-1} d(x, X_i))}, \ \forall x \in \mathcal{F},$$

where K is a kernel function and $h_K = h_{K,n}$ is a sequence of positive real numbers which goes to zero as n goes to infinity.
Our aim is to establish the uniform almost complete convergence of $\widehat{m_\varphi}$ on some subset $S_\mathcal{F}$ of $\mathcal{F}$. To do that we denote by C or/and C' some real generic constants supposed strictly positive and we assume that

(H2) There exists $b > 0$ such that $\forall x_1, x_2 \in S_\mathcal{F}, \ |m_\varphi(x_1) - m_\varphi(x_2)| \leq Cd^b(x_1, x_2)$,

(H3) $\forall m \geq 2, \ E(|\varphi(Y)|^m | X = x) < \delta_m(x) < C < \infty$ with $\delta_m(.)$ continuous on $S_\mathcal{F}$,

(H4) K is a bounded and Lipschitz kernel on its support $[0, 1)$, and if $K(1) = 0$, the kernel K has to fulfill the additional condition $-\infty < C < K'(t) < C' < 0$,

(H5) The functions ϕ and $\psi_{S_\mathcal{F}}$ are such that :

(H5a) $\exists C > 0, \exists \eta_0 > 0, \ \forall \eta < \eta_0, \ \phi'(\eta) < C$, and if $K(1) = 0$, the function $\phi(.)$ has to fulfill the additional condition :

$$\exists C > 0, \exists \eta_0 > 0, \ \forall 0 < \eta < \eta_0, \ \int_0^\eta \phi(u)\, du \ > \ C\, \eta\, \phi(\eta),$$

(H5b) For n large enough, $\dfrac{(\log n)^2}{n\,\phi(h_K)} < \psi_{S_{\mathcal{F}}}\left(\dfrac{\log n}{n}\right) < \dfrac{n\,\phi(h_K)}{\log n}$,

(H6) The Kolmogorov's ϵ-entropy of $S_{\mathcal{F}}$ satisfies

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

Conditions (H2)-(H4) are very standard in the nonparametric setting. Concerning (H5a), the boundness of the derivative of ϕ around zero allows to consider ϕ as a Lipschitzian function. In addition, from a theoretical point of view, one has to separate the case when $K(.)$ is a continuous kernel (i.e. $K(1) = 0$) and the case when $K(.)$ is not continuous (which contains, for instance, the uniform kernel). The case when $K(1) = 0$ is more delicate and one has to introduce an additional assumption acting on the behaviour of ϕ around zero (see the proof in appendix). Hypothesis (H5b) deals with topological considerations by controling the entropy of $S_{\mathcal{F}}$. For a radius not too large, one requires that $\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)$ is not too small and not too large. Moreover, (H5b) implies that $\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)/(n\,\phi(h_K))$ tends to 0 when n tends to $+\infty$. As remarked in Section 2, in some "usual" cases, one has $\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right) \sim C\log n$ and (H5b) is satisfied as soon as $(\log n)^2 = O(n\,\phi(h_K))$. In a different way, Assumption (H6) acts on the Kolmogorov's ϵ-entropy of $S_{\mathcal{F}}$. However, if one considers the same particular case as previously, it is easy to see that (H6) is verified as soon as $\beta > 2$.

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

$$\sup_{x\in S_{\mathcal{F}}} |\widehat{m_\varphi}(x) - m_\varphi(x)| = O(h_K^b) + O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h_K)}}\right). \tag{2.5}$$

2.4 Conditional cumulative distribution estimation

In this section, we assume that the regular version of the conditional probability of Y given X exists and we study the uniform almost complete convergence of a kernel estimator of the conditional cumulative distribution function, denoted by F^x. A straightforward way to estimate the function F^x is to treat this function as particular case of m_φ with $\varphi(t) = \mathbb{I}_{]-\infty,y]}(t)$ (for $y \in \mathbb{R}$). Thus, we estimate F^x by

$$\widehat{F}^x(y) = \sum_{i=1}^n W_{ni}(x)\mathbb{I}_{\{Y_i \leq y\}}, \qquad \forall y \in \mathbb{R},\ \forall x \in \mathcal{F},$$

where

$$W_{ni}(x) = \frac{K(h_K^{-1}d(x, X_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))}.$$

The estimate of the conditional cumulative distribution function has been investigated, in the real case, by several authors (see, Roussas, 1969, Samanta, 1989, among others). In the functional case, Ferraty *et al.* (2006) established the almost complete convergence of a double kernel estimator of the conditional cumulative distribution function.

Clearly, the previous result stated in Section 3 allows to conclude the almost complete convergence of $\widehat{F}^x$ uniformly in the functional argument x. Indeed, it suffices to apply Theorem 1 to get :

Corollary 1 *Under the hypotheses (H1),(H2) and (H4)-(H6), we have*

$$\sup_{x \in S_\mathcal{F}} |\widehat{F}^x(y) - F^x(y)| = O(h_K^b) + O_{a.co.}\left(\sqrt{\frac{\psi_{S_\mathcal{F}}\left(\frac{\log n}{n}\right)}{n\phi(h_K)}}\right).$$

But, in order to derive the uniform consistency on both (functional and real arguments), we fix a compact subset $S_\mathbb{R}$ of $\mathbb{R}$ and we consider the following additional assumptions :

(H2′) $\forall (y_1, y_2) \in S_\mathbb{R} \times S_\mathbb{R},\ \forall (x_1, x_2) \in S_\mathcal{F} \times S_\mathcal{F}$,
$|F^{x_1}(y_1) - F^{x_2}(y_2)| \leq C\left(d(x_1, x_2)^{b_1} + |y_1 - y_2|^{b_2}\right)$,

(H6′) The Kolmogorov's ϵ-entropy of $S_\mathcal{F}$ satisfies

$$\sum_{n=1}^{\infty} n^{1/2b_2} \exp\left\{(1-\beta)\psi_{S_\mathcal{F}}\left(\frac{\log n}{n}\right)\right\} < \infty, \text{ for some } \beta > 1.$$

Theorem 2 *Under the hypotheses (H1), (H2'),(H4), (H5) and (H6'), we have*

$$\sup_{x \in S_{\mathcal{F}}} \sup_{y \in S_{\mathbb{R}}} |\widehat{F}^x(y) - F^x(y)| = O(h_K^{b_1}) + O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h_K)}} \right). \tag{2.6}$$

2.5 Conditional density estimation

In this section, similar results will be derived for the kernel estimator of the conditional density of Y given X. We assume that the conditional probability of Y given X is absolutely continuous with respect to the Lebesgues measure on $\mathbb{R}$ and we will denote by f^x the conditional density of Y given $X = x$. We define the kernel estimator $\widehat{f^x}$ of f^x as follows :

$$\widehat{f^x}(y) = \frac{h_H^{-1} \sum_{i=1}^n K(h_K^{-1} d(x, X_i)) H(h_H^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1} d(x, X_i))}, \qquad \forall y \in \mathbb{R},\ \forall x \in \mathcal{F}, \tag{2.7}$$

where the functions K and H are kernels and $h_K = h_{K,n}$ (resp. $h_H = h_{H,n}$) is a sequence of positive real numbers. Note that a similar estimate was already introduced in the special case when X is a real random variable by Rosenblatt (1969) and by Youndjé (1996) among other authors.

In order to establish the uniform almost complete convergence of this estimator, we consider the following additional assumptions :

(H2″) $\forall (y_1, y_2) \in S_{\mathbb{R}} \times S_{\mathbb{R}},\ \forall (x_1, x_2) \in S_{\mathcal{F}} \times S_{\mathcal{F}},$
$|f^{x_1}(y_1) - f^{x_2}(y_2)| \leq C\left(d^{b_1}(x_1, x_2) + |y_1 - y_2|^{b_2}\right),$

(H5b″) For some $\gamma \in (0, 1)$, $\lim_{n \to +\infty} n^{\gamma} h_H = \infty$, and for n large enough :

$$\frac{(\log n)^2}{n^{1-\gamma}\, \phi(h_K)} < \psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right) < \frac{n^{1-\gamma}\, \phi(h_K)}{\log n},$$

(H6″) The Kolmogorov's ϵ-entropy of $S_{\mathcal{F}}$ satisfies

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

(H7) H is bounded Lipschitzian continuous function, such that $\int |t|^{b_2} H(t)dt < \infty$ and $\int H^2(t)dt < \infty$.

Theorem 3 *Under the hypotheses (H1), (H2′),(H4),(H5a),(H5b″), (H6″) and (H7), we have*

$$\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}}|\widehat{f^x}(y)-f^x(y)| = O(h_K^{b_1})+O(h_H^{b_2})+O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n^{1-\gamma}\phi(h_K)}}\right). \quad (2.8)$$

2.6 Some direct consequences

2.6.1 Estimation of the conditional hazard function

This section is devoted to the almost complete convergence of the kernel estimator of the conditional hazard function of Y given X uniformly on fixed subset $S_{\mathcal{F}}\times S_{\mathbb{R}}$ of $\mathcal{F}\times\mathbb{R}$. We return to Ferraty *et al.* (2008) for the pointwise almost complete convergence of this model, in the functional case. Recall that the conditional hazard function is defined by :

$$h^x(y) = \frac{f^x(y)}{1-F^x(y)},\ \forall y,\ F^x(y)<1,\ \forall x\in\mathcal{F}.$$

Naturally, the conditional hazard function estimator is closely linked to the conditional survival function estimate. Consider the kernel estimates of the functions F^x and f^x defined in the previous Section, and adopt the kernel estimator $\widehat{h^x}(y)$ of the conditional hazard function h^x defined by :

$$\widehat{h^x}(y) = \frac{\widehat{f^x}(y)}{1-\widehat{F^x}(y)}.$$

In addition to the previous assumptions, used to establish the convergence rates of (2.6) and (2.8), the following is needed :

(H8) $\exists\delta_1>0,\ \delta_2>0$ such that $\forall x\in S_{\mathcal{F}},\ \forall y\in S_{\mathbb{R}},\ F^x(y)<\delta_1<1$ and $f^x(y)\le\delta_2$.

Theorem 4 *Under the hypotheses of Theorem 3 and if (H6′) and (H8) hold, then*

$$\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}}|\widehat{h}^x(y)-h^x(y)| = O(h_K^{b_1})+O(h_H^{b_2})+O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n^{1-\gamma}\phi(h_K)}}\right). \quad (2.9)$$

2.6.2 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 fixed compact subset $S_{\mathcal{F}}$ of $\mathcal{F}$. For this, we assume that $\theta(x)$ satisfies on $S_{\mathcal{F}}$ the following uniform uniqueness property (see, Ould-said and Cai, 2005, for the multivariate case).

(H9) $\forall\epsilon_0>0\ \exists\eta>0,\ \forall r:S\to S_{\mathbb{R}}$,

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

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 with respect to y on $S_{\mathbb{R}}$ and

(H10) $f^{x(l)}(\theta(x))=0$ if $1\leq l<j$ and $f^{x(j)}(\cdot)$ is uniformly continuous on $S_{\mathbb{R}}$ such that $|f^{x(j)}(\theta(x))|>C>0$,

where $f^{x(j)}$ is the j^{th} order derivative of the conditional density f^x.
We estimate the conditional mode $\theta(x)$ with a random variable $\widehat{\theta(x)}$ such that

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

From Theorem 3 we derive the following Corollary.

Corollary 2 *Under the hypotheses of Theorem 3 and if the conditional density f^x satisfies (H9) and (H10), we have*

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

□

2.7 Comments

2.7.1 Impact of the results

This paper has stated uniform consistency results in functional setting. They are not only nice extensions of pointwise results but they have great impacts (both from theoretical and practical point of view). First of all, the natural practical interest of getting uniform consistency is for prediction. Look for instance at Theorem 2. The fact to be able to state results on the quantity

$$\sup_{x\in S_{\mathcal{F}}} |\widehat{m_\varphi}(x) - m_\varphi(x)|$$

allows directly to obtain result on quantity

$$|\widehat{m_\varphi}(X) - m_\varphi(X)|,$$

where X is a new random functional element valued in $S_{\mathcal{F}}$. The same kind of remark can be done for the other problems treated in Theorem 8, 11 and 14 (i.e. conditional cumulative distribution, conditional density and conditional hazard function). More generally, as in multivariate statistics, it can be useful for estimating the solution of general equations with applications for detecting peaks, valleys, change points (see, for instance, Boularan et *al.*, 1995). Secondly, and this is maybe the main point, uniform consistency results are indispensable tools for the study of more sophisticated models in which multi-stage procedures are involved. This occurs in a very wide scope of situation in standard multivariate nonparametric analysis. In functional setting, the needs of uniform consistency results have been recently pointed out in some multi-stage models such as additive modelling (Ferraty and Vieu, 2009), partial functional linear model (Aneiros and Vieu, 2008), and single functional models (Ait Saïdi *et al.*, 2008). Other functional data methodologies are also using such kind of uniform tools, such as for instance data-driven bandwidth choice (Benhenni *et al.*, 2007), or bootstrapping (Ferraty *et al.*, 2009). The scope of functional applications of our theoretical uniform results will increase in the near future following the progress of FDA.

2.7.2 General comments on the hypotheses

In addition to the comments given in Section 3, we go back to complete this discussion by comparing the structural assumptions of the uniform convergence to those of the pointwise ones studied by Ferraty and Vieu (2006).

On the functional variable. Unlike the pointwise case, the uniform consistency requires a concentration property of the probability measure uniformly over $S_{\mathcal{F}}$ (see, (H1)). So, it is important to give here general situations when such an assumption is fulfilled. From a probabilistic point of view, this can be done by introducing the Onsager-Machlup function (see, Onsager and Machlup, 1953) defined as :

$$\forall (x,z) \in S_{\mathcal{F}},\ \mathcal{F}_X(x,z) = \log\left(\lim_{h\to 0} \frac{P(B(x,h))}{P(B(z,h))}\right).$$

Then, (H1) is verified if the Onsager-Machlup function of the probability measure of the functional variable is such that

$$\forall x \in S_{\mathcal{F}}, \qquad |\mathcal{F}_X(x,0)| \leq C < \infty. \tag{2.10}$$

The Onsager-Machlup function has been intensively studied in the literature as well as the quantities $P(X \in B(0,h))$. Their respective explicit expression for several continuous time processes can be found in Bogachev (p.186, 1999), which produces thereby examples of subsets and functional variables that satisfy (H1). This pure probabilistic point of view focuses on small ball probabilities with standard topologies. But, from a statistical point of view, the practitioner can choose the semi-metric. In particular, Example 4 in Section 2 gives an interesting family of semi-metrics allowing to fulfill (H1) for a large set of functional variables (see, Lemma 13.6 in Ferraty and Vieu, 2006). In fact, as a statistician, an important task consists in building a semi-metric adapted to the functional variable. Here, the word “adapted” can mean that this semi-metric allows to satisfy (H1) but other properties for the semi-metric can be required and this issue will be certainly investigated in further works.

On the regularity constraints of the model. Regularity-type conditions on the functional objects to be estimated are given via assumptions (H2), (H2'), (H2"). In comparison with the pointwise case (see, Ferraty *et al.*, 2006), we do not assume that the constants depend on the conditioning point. Moreover, these assumptions are sufficient but not necessary. For instance, one can replace (H2) by a new one thanks to the function

$$L_x(z) = E\left[\varphi(Y) - m_\varphi(x) | X = z\right].$$

Consider $\mathcal{F}$ as a vector semi-normed space and let us assume that it exists a linear operator A_x such that

$$L_x(z) = A_x(z - x) + o(\|x - z\|),$$

where A_x is bounded uniformly over $S_\mathcal{F}$ (which amounts to assume that L_x is differentiable since $L_x(x) = 0$). From an asymptotic point of view, the only change is appearing in the bias. So, by using the first-order expansion of L_x, we can get also under this alternative assumption that

$$\sup_{x \in S_\mathcal{F}} |E\widehat{g_\varphi}(x) - m_\varphi(x)| = O(h).$$

As a conclusion, there are several ways to introduce regularity constraints in the functional nonparametric models. The alternative assumption used here preserves the same rate of convergence. However, one has to keep in mind that considering other smoothness conditions on the model can lead to different convergence rates for the bias.

2.7.3 Comments on convergence rates

It is well known that in finite dimensional case (that is $\mathcal{F} = \mathbb{R}^p$), the uniform rates of convergence (over compact sets) are the same as the pointwise

ones. The main point of this paper is to show how this is not so obvious in functional settings. To fix ideas, let us just look at the regression case (the other ones could be discussed similarly).
For fixed point $x \in \mathcal{F}$ and for $\varphi \equiv Identity$, Ferraty and Vieu (2006, Theorem 6.11) obtained the pointwise result :

$$\widehat{m_\varphi}(x) - m_\varphi(x) = O(h_K^b) + O\left(\sqrt{\frac{\log n}{n\phi(h_K)}}\right), \ a.co.,$$

while in Theorem 2 of this paper we stated the following result

$$\sup_{x \in S_\mathcal{F}} |\widehat{m_\varphi}(x) - m_\varphi(x)| = O(h_K^b) + O\left(\sqrt{\frac{\psi_{S_\mathcal{F}}\left(\frac{\log n}{n}\right)}{n\phi(h_K)}}\right), \ a.co.$$

To see how the uniform point of view leads to a deterioration of the rates of convergence, one may look at examples discussed in Section 2.2. For instance, if one considers the standard gaussian process with usual metric topology (Example 3), the loss is of $\log n$ since one gets

$$\psi_{S_\mathcal{F}}\left(\frac{\log n}{n}\right) = O\left((\log n)^2\right).$$

However, if ones focuses on Example 4, we can see the interest of using a semi-metric. Indeed, by building a suitable projection-based semi-metric, the entropy function becomes :

$$\psi_{S_\mathcal{F}}\left(\frac{\log n}{n}\right) = O(\log n).$$

So, such a new topological choice (as described in Example 4) allows to avoid the deterioration of the rates of convergence.

2.8 Appendix : Proofs

In the following, we will denote, for all $i = 1, \ldots, n$, by

$$K_i(x) = K(h_K^{-1} d(x, X_i)) \qquad \text{and} \qquad H_i(y) = H(h_H^{-1}(y - Y_i)).$$

First of all, according to (H1) and (H4), it is clear that if $K(1) > C > 0$,

$$\forall x \in S_{\mathcal{F}},\ \exists 0 < C < C' < \infty,\ C\phi(h_K) < E[K_1(x)] < C'\phi(h_K). \qquad (2.11)$$

In the situation when $K(1) = 0$, the combination of (H1) and (H5a) allows to get the same result (see, p44, Lemma 4.4. in Ferraty and Vieu, 2006). From now on, in order to simplify the notation, we set $\epsilon = \dfrac{\log n}{n}$.

Proof of Theorem 1 We consider the decomposition :

$$\widehat{m_\varphi}(x) - m_\varphi(x) = \frac{1}{\widehat{f}(x)}\Big[\widehat{g_\varphi}(x) - E\widehat{g_\varphi}(x)\Big]$$

$$+ \frac{1}{\widehat{f}(x)}\Big[E\widehat{g_\varphi}(x) - m_\varphi(x)\Big] + \Big[1 - \widehat{f}(x)\Big]\frac{m_\varphi(x)}{\widehat{f}(x)} \qquad (2.12)$$

where

$$\widehat{f}(x) = \frac{1}{nE[K(h_K^{-1}d(x, X_1))]}\sum_{i=1}^{n} K(h_K^{-1}d(x, X_i))$$

and

$$\widehat{g_\varphi}(x) = \frac{1}{nE[K(h_K^{-1}d(x, X_1))]}\sum_{i=1}^{n} K(h_K^{-1}d(x, X_i))\varphi(Y_i).$$

Therefore, Theorem 1 is a consequence of the following intermediate results.

Lemma 1 *Under the hypotheses (H1) and (H4)-(H6), we have*

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

Corollary 3 *Under the hypotheses of Lemma 1, we have*

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

Lemma 2 *Under the hypotheses (H1),(H2) and (H4)-(H6), we have*

$$\sup_{x\in S_{\mathcal{F}}} |E\widehat{g_\varphi}(x) - m_\varphi(x)| = O(h_K^b).$$

Lemma 3 *Under the assumptions of Theorem 1, we have*

$$\sup_{x\in S_{\mathcal{F}}} |\widehat{g_\varphi}(x) - E\widehat{g_\varphi}(x)| = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h_K)}}\right).$$

□

Proof of Lemma 1. Let $x_1, \ldots, x_{N_\epsilon(S_{\mathcal{F}})}$ be an ϵ-net for $S_{\mathcal{F}}$ (see, Definition 2) and for all $x \in S_{\mathcal{F}}$, one sets $k(x) = \arg\min_{k\in\{1,2,\ldots,N_\epsilon(S_{\mathcal{F}})\}} d(x, x_k)$. One considers the following decomposition :

$$\sup_{x\in S_{\mathcal{F}}} |\widehat{f}(x) - E\widehat{f}(x)| \leq \underbrace{\sup_{x\in S_{\mathcal{F}}} |\widehat{f}(x) - \widehat{f}(x_{k(x)})|}_{F_1} + \underbrace{\sup_{x\in S_{\mathcal{F}}} |\widehat{f}(x_{k(x)}) - E\widehat{f}(x_{k(x)})|}_{F_2}$$
$$+ \underbrace{\sup_{x\in S_{\mathcal{F}}} |E\widehat{f}(x_{k(x)}) - E\widehat{f}(x)|}_{F_3}.$$

• Let us study F_1. By using (2.11) and the boundness of K, one can write :

$$\begin{aligned} F_1 &\leq \sup_{x\in S_{\mathcal{F}}} \frac{1}{n}\sum_{i=1}^n \left| \frac{1}{E[K_1(x)]} K_i(x) - \frac{1}{E[K_1(x_{k(x)})]} K_i(x_{k(x)}) \right| \\ &\leq \frac{C}{\phi(h_K)} \sup_{x\in S_{\mathcal{F}}} \frac{1}{n}\sum_{i=1}^n \left| K_i(x) - K_i(x_{k(x)}) \right| \mathbb{1}_{B(x,h_K)\cup B(x_{k(x)},h_K)}(X_i). \end{aligned}$$

Let us first consider the case $K(1) = 0$. Because K is Lipschitz on $[0, 1]$ in this case, it comes

$$F_1 \leq \sup_{x\in S_{\mathcal{F}}} \frac{C}{n}\sum_{i=1}^n Z_i \text{ with } Z_i = \frac{\epsilon}{h_K\,\phi(h_K)} \mathbb{1}_{B(x,h_K)\cup B(x_{k(x)},h_K)}(X_i),$$

with, uniformly on x,
$Z_1 = O\left(\frac{\epsilon}{h_K\phi(h_K)}\right)$, $EZ_1 = O\left(\frac{\epsilon}{h_K}\right)$ and $Var(Z_1) = O\left(\frac{\epsilon^2}{h_k^2\phi(h_K)}\right)$.

A standard inequality for sums of bounded random variables (see, Corollary A.9 in Ferraty and Vieu, 2006) with (H5b) allows to get

$$F_1 \;=\; O\left(\frac{\epsilon}{h_K}\right) + O_{a.co.}\left(\frac{\epsilon}{h_K}\sqrt{\frac{\log n}{n\,\phi(h_K)}}\right),$$

and it suffices to combine (H5a) and (H5b) to get

$$F_1 \;=\; O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\,\phi(h_K)}}\right).$$

Now, let $K(1) > C > 0$. In this situation K is Lipschitz on $[0,1)$. One has to decompose F_1 into three terms as follows :

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

with

$$\begin{aligned}
F_{11} &= \frac{1}{n\phi(h_K)}\sum_{i=1}^{n}\left|K_i(x) - K_i(x_{k(x)})\right| \mathbb{1}_{B(x,h_K)\cap B(x_{k(x)},h_K)}(X_i),\\
F_{12} &= \frac{1}{n\phi(h_K)}\sum_{i=1}^{n} K_i(x)\mathbb{1}_{B(x,h_K)\cap\overline{B(x_{k(x)},h_K)}}(X_i),\\
F_{13} &= \frac{1}{n\phi(h_K)}\sum_{i=1}^{n} K_i(x_{k(x)})\mathbb{1}_{\overline{B(x,h_K)}\cap B(x_{k(x)},h_K)}(X_i).
\end{aligned}$$

One can follow the same steps (i.e. case $K(1) = 0$) for studying F_{11} and one gets the same result :

$$F_{11} \;=\; O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\,\phi(h_K)}}\right).$$

Following same ideas for studying F_{12}, one can write :

$$F_{12} \;\leq\; \frac{C}{n}\sum_{i=1}^{n} W_i \text{ with } W_i = \frac{1}{\phi(h_K)}\mathbb{1}_{B(x,h_K)\cap\overline{B(x_{k(x)},h_K)}}(X_i),$$

and by using again (H5a) and the same inequality for sums of bounded random variables, one has :

$$F_{12} \;=\; O\left(\frac{\epsilon}{\phi(h_K)}\right) + O_{a.co.}\left(\sqrt{\frac{\epsilon\,\log n}{n\,\phi(h_K)^2}}\right).$$

Similarly, one can state the same rate of convergence for F_{13}. To end the study of F_1, it suffices to put together all the intermediate results and to use again (H5b) for getting :

$$F_1 \;=\; O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h_K)}}\right).$$

• Now, concerning F_2, we have, for all $\eta > 0$,

$$\begin{aligned}
P&\left(F_2 > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h_K)}}\right) \\
&\qquad = P\left(\max_{k\in\{1,\cdots,N_\epsilon(S_{\mathcal{F}})\}} |\widehat{f}(x_{k(x)}) - E\widehat{f}(x_{k(x)})| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h_K)}}\right) \\
&\qquad \leq N_\epsilon(S_{\mathcal{F}}) \max_{k\in\{1,\cdots,N_\epsilon(S_{\mathcal{F}})\}} P\left(|\widehat{f}(x_k) - E\widehat{f}(x_k)| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h_K)}}\right).
\end{aligned}$$

Let $\Delta_{ki} = \dfrac{1}{E[K_1(x_k)]}\left(K_i(x_k) - E[K_i(x_k)]\right)$. We show, under (H1) and (H4), that $\forall k = 1, \ldots, N_\epsilon(S_{\mathcal{F}})$, $\forall i = 1, \ldots, n$,

$$\Delta_{ki} = O\left(\phi(h_K)^{-1}\right) \qquad \text{and also} \qquad Var(\Delta_{ki}) = O\left(\phi(h_K)^{-1}\right).$$

So, one can apply the Bernstein-type inequality (see, Corollary A.9 in Ferraty and Vieu, 2006) which gives directly :

$$\begin{aligned}
P\left(|\widehat{f}(x_k) - E\widehat{f}(x_k)| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h_K)}}\right) \;&=\; P\left(\frac{1}{n}\left|\sum_{i=1}^{n}\Delta_{ki}\right| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h_K)}}\right) \\
&\leq\; 2\exp\{-C\eta^2\psi_{S_{\mathcal{F}}}(\epsilon)\}.
\end{aligned}$$

Thus, by using the fact that $\psi_{S_{\mathcal{F}}}(\epsilon) = \log N_\epsilon(S_{\mathcal{F}})$ and by choosing η such that $C\eta^2 = \beta$, we have

$$N_\epsilon(S_{\mathcal{F}}) \max_{k\in\{1,\cdots,N_\epsilon(S_{\mathcal{F}})\}} P\left(|\widehat{f}(x_k) - E\widehat{f}(x_k)| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h_K)}}\right) \leq C' N_\epsilon(S_{\mathcal{F}})^{1-\beta}. \tag{2.13}$$

Because $\sum_{n=1}^{\infty} N_\epsilon(S_{\mathcal{F}})^{1-\beta} < \infty$, we obtain that

$$F_2 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h_K)}}\right).$$

• For F_3, it is clear that $F_3 \leq E(\sup_{x\in S_{\mathcal{F}}} |\widehat{f}(x) - \widehat{f}(x_{k(x)})|)$ and by following a similar proof to the one used for studying F_1, it comes

$$F_3 \;=\; O\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h_K)}}\right).$$

$\square$

Proof of Corollary 3. It is easy to see that,

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

We deduce from Lemma 1 that

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

Consequently ,

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

$\square$

Proof of Lemma 2. One has

$$\begin{aligned} |E\widehat{g}_\varphi(x) - m_\varphi(x)| &= \left|\frac{1}{nE[K_1(x)]} E\left[\sum_{i=1}^{n} K_i(x)\varphi(Y_i)\right] - m_\varphi(x)\right| \\ &\leq \left|\frac{1}{E[K_1(x)]} E[K_1(x)\varphi(Y_1)] - m_\varphi(x)\right| \\ &\leq \frac{1}{E[K_1(x)]}\Big[E[K_1(x)\,|m_\varphi(X_1) - m_\varphi(x)|]\Big]. \end{aligned}$$

Hence, we get

$$\forall x \in S_{\mathcal{F}}, \qquad |E\widehat{g_\varphi}(x) - m_\varphi(x)| \leq \frac{1}{E[K_1(x)]}[EK_1(x)\,|m_\varphi(X_1) - m_\varphi(x)|].$$

Thus, with hypotheses (H1), (H2) and (2.11) we have

$$\forall x \in S_{\mathcal{F}}, \quad |E\widehat{g_\varphi}(x) - m_\varphi(x)| \leq C\frac{1}{E[K_1(x)]}[EK_1(x)\mathbb{1}_{B(x,h_K)}(X_1)d^b(X_1,x)] \leq Ch_K^b,$$

this last inequality yields the proof, since C does not depend on x.

$\square$

Proof of Lemma 3. This proof follows the same steps as the proof of Lemma 1. For this, we keep these notations and we use the following decomposition

$$\sup_{x\in S_{\mathcal{F}}} |\widehat{g_\varphi}(x) - E\widehat{g_\varphi}(x)| \leq \underbrace{\sup_{x\in S_{\mathcal{F}}} |\widehat{g_\varphi}(x) - \widehat{g_\varphi}(x_{k(x)})|}_{G_1} + \underbrace{\sup_{x\in S_{\mathcal{F}}} |\widehat{g_\varphi}(x_{k(x)}) - E\widehat{g_\varphi}(x_{k(x)})|}_{G_2}$$
$$+ \underbrace{\sup_{x\in S_{\mathcal{F}}} |E\widehat{g_\varphi}(x_{k(x)}) - E\widehat{g_\varphi}(x)|}_{G_3}.$$

The condition (H1) and the result (2.11) allow to write directly, for G_1 and G_3

$$\begin{aligned} G_1 &= \sup_{x\in S_{\mathcal{F}}} \left| \sum_{i=1}^n \frac{1}{nE[K_1(x)]} K_i(x)\varphi(Y_i) - \frac{1}{nE[K_1(x_{k(x)})]} K_i(x_{k(x)})\varphi(Y_i) \right| \\ &\leq \sup_{x\in S_{\mathcal{F}}} \frac{1}{n\,\phi(h_K)} \sum_{i=1}^n |\varphi(Y_i)| \left| K_i(x) - K_i(x_{k(x)}) \right| \mathbb{1}_{B(x,h_K)\cup B(x_{k(x)},h_K)}. \end{aligned}$$

Now, as for F_1, one considers $K(1) = 0$ (i.e. K Lipschitz on $[0,1]$) and one gets :

$$G_1 \;\leq\; \frac{C}{n}\sum_{i=1}^n Z_i \text{ with } Z_i = \frac{\epsilon\,\varphi(Y_i)}{h_K\,\phi(h_K)} \sup_{x\in S_{\mathcal{F}}} \mathbb{1}_{B(x,h_K)\cup B(x_{k(x)},h_K)}.$$

The main difference with the study of F_1 is that one uses here an exponential inequality for unbounded variables. Note that one has

$$E[|\varphi^m(Y)|] = E\left[E[|\varphi(Y)|^m|X]\right] = \int \delta_m(x)dP_X < C < \infty,$$

which implies that

$$E(|Z_1|^m) \;\leq\; \frac{C\epsilon^m}{h_K^m\,\phi(h_K)^{m-1}}.$$

So, by applying Corollary A.8 in Ferraty and Vieu (2006) with $a^2 = \frac{\epsilon}{h_K\,\phi(h_K)}$, one gets $G_1 \;=\; O_{a.co.}\left(\sqrt{\frac{\epsilon\,\log n}{n\,h_K\,\phi(h_K)}}\right)$.

Now, (H5b) allows to get

$$G_1 \;=\; O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\,\phi(h_K)}}\right). \tag{2.14}$$

If one considers the case $K(1) > C > 0$ (i.e. K Lipschitz on $[0,1)$), one has to split G_1 into three terms as for F_1 and by using similar arguments, one can state the same rate of almost complete convergence. Similar steps allow to get

$$G_3 = O\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h_K)}}\right). \tag{2.15}$$

For G_2, similarly to the proof of Lemma 1, we have, $\forall \eta > 0$,

$$\begin{aligned}
&P\left(G_2 > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h_K)}}\right)\\
&\qquad = P\left(\max_{k\in\{1,\cdots,N_\epsilon(S_{\mathcal{F}})\}}|\widehat{g}_\varphi(x_k) - E\widehat{g}_\varphi(x_k)| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h_K)}}\right)\\
&\qquad \leq N_\epsilon(S_{\mathcal{F}})\max_{k\in\{1,\cdots,N_\epsilon(S_{\mathcal{F}})\}}P\left(|\widehat{g}_\varphi(x_k) - E\widehat{g}_\varphi(x_k)| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h_K)}}\right).
\end{aligned}$$

The rest of the proof is based on the exponential inequality given by Corollary A.8.ii in Ferraty and Vieu (2006).

Indeed, let

$$\Gamma_{ki} = \frac{1}{E[K_1(x_k))]}\left[K_i(x_k)\varphi(Y_i) - \mathbb{E}\left[K_i(x_k)\varphi(Y_i)\right]\right].$$

The same arguments as those invoked for proving Lemma 6.3 in Ferraty and Vieu (2006, p.65) can be used to show that $E\,|\Gamma_{ki}|^m = O(\phi(h_K)^{-m+1})$ which gives by applying the exponential inequality mentioned above, for all $\eta > 0$

$$P\left(|\widehat{g}_{\varphi}(x_k) - \mathbb{E}\widehat{g}_{\varphi}(x_k)| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\,\phi(h_K)}}\right) \leq 2N_{\epsilon}(S_{\mathcal{F}})^{-C\eta^2}.$$

Therefore, by suitable choice of $\eta > 0$, we have

$$N_{\epsilon}(S_{\mathcal{F}}) \max_{k\in\{1,\ldots,N_{\epsilon}(S_{\mathcal{F}})\}} P\left(|\widehat{g}_{\varphi}(x_k) - \mathbb{E}\widehat{g}_{\varphi}(x_k)| > \eta\sqrt{\frac{\log N_{\epsilon}(S_{\mathcal{F}})}{n\,\phi(h_K)}}\right) \leq C' N_{\epsilon}(S_{\mathcal{F}})^{1-\beta}.$$

As $\sum_{n=1}^{\infty} N_{\epsilon}(S_{\mathcal{F}})^{1-\beta} < \infty$, we obtain that

$$G_2 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\,\phi(h_K)}}\right). \tag{2.16}$$

Now, Lemma 3 can be easily deduced from (2.14), (2.15) and (2.16).

□

Proof of Theorem 2 Similarly to (2.12), we have

$$\begin{aligned}\widehat{F}^x(y) - F^x(y) = \frac{1}{\widehat{f}(x)}\Big[(\widehat{F}_N^x(y) - E\widehat{F}_N^x(y)) - (F^x(y) - E\widehat{F}_N^x(y))\Big]\\ + \frac{F^x(y)}{\widehat{f}(x)}\Big[E\widehat{f}(x) - \widehat{f}(x)\Big],\end{aligned} \tag{2.17}$$

where

$$\widehat{F}_N^x(y) = \frac{1}{nE[K(h_K^{-1}d(x,X_1))]}\sum_{i=1}^{n} K\left(h_K^{-1}d(x,X_i)\right)\mathbb{1}_{\{Y_i\leq y\}}.$$

Then, Theorem 2 can be deduced from the following intermediate results, together with Lemma 1 and Corollary 3.

Lemma 4 *Under the hypotheses (H2′) and (H4), one has*

$$\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}} |F^x(y) - E\widehat{F}_N^x(y)| = O(h_K^{b_1}).$$

Lemma 5 *Under the assumptions of Theorem 2, we have*

$$\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}}|\widehat{F}_N^x(y)-E\widehat{F}_N^x(y)| = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h_K)}}\right).$$

$\square$

Proof of Lemma 4. It is clear that (H4) implies $\forall(x,y)\in S_{\mathcal{F}}\times S_{\mathbb{R}}$,

$$E\left[\widehat{F}_N^x(y)\right]-F^x(y)=\frac{1}{E\left[K_1(x)\right]}E\left[\left(K_1(x)1\!\!1_{B(x,h_K)}(X_1)\right)\left(F^{X_1}(y)-F^x(y)\right)\right]. \tag{2.18}$$

The Lipschitz condition (H2′) allows us to write that

$$\forall(x,y)\in S_{\mathcal{F}}\times S_{\mathbb{R}},\quad 1\!\!1_{B(x,h_K)}(X_1)|F^{X_1}(y)-F^x(y)|\le Ch_K^{b_1},$$

then,

$$\forall(x,y)\in S_{\mathcal{F}}\times S_{\mathbb{R}},\quad \left|E\left[\widehat{F}_N^x(y)\right]-F^x(y)\right|\le Ch_K^{b_1}.$$

$\square$

Proof of Lemma 5. We keep the notation of the Lemma 1 and we use the compactness of $S_{\mathbb{R}}$, we can write that, for some, $t_1,t_2,\ldots,t_{z_n}\in S_{\mathbb{R}}$,

$$S_{\mathbb{R}}\subset\bigcup_{j=1}^{z_n}(t_j-l_n,t_j+l_n)$$

with $l_n=n^{-\frac{1}{2b_2}}$ and $z_n\le n^{\frac{1}{2b_2}}$. Taking

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

Thus, we have the following decomposition :

$$\begin{aligned}\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}}\left|\widehat{F}_N^x(y)-E\widehat{F}_N^x(y)\right| \le\; & \underbrace{\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}}\left|\widehat{F}_N^x(y)-\widehat{F}_N^{x_{k(x)}}(y)\right|}_{F_1'}\\ & +\underbrace{\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}}\left|\widehat{F}_N^{x_{k(x)}}(y)-E\widehat{F}_N^{x_{k(x)}}(y)\right|}_{F_2'}\\ & +\underbrace{\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}}\left|E\widehat{F}_N^{x_{k(x)}}(y)-E\widehat{F}_N^x(y)\right|}_{F_3'}.\end{aligned}$$

Concerning (F_1') and (F_3'), by following the same lines as for studying the terms F_1 and F_3, it comes

$$F_1' = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\ \phi(h_K)}}\right) \text{ and } F_3' = O\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\,\phi(h_K)}}\right). \tag{2.19}$$

Concerning (F_2'), the monotony of the functions $E[\widehat{F}_N^x(\cdot)]$ and $\widehat{F}_N^x(\cdot)$ permits to write, for all $j \le z_n$ and for all $x \in S_{\mathcal{F}}$,

$$\begin{aligned} E\widehat{F}_N^{x_{k(x)}}(t_j - l_n) &\le \sup_{y\in(t_j-l_n,t_j+l_n)} E\widehat{F}_N^{x_{k(x)}}(y) \le E\widehat{F}_N^{x_{k(x)}}(t_j + l_n), \\ \widehat{F}_N^{x_{k(x)}}(t_j - l_n) &\le \sup_{y\in(t_j-l_n,t_j+l_n)} \widehat{F}_N^{x_{k(x)}}(y) \le \widehat{F}_N^{x_{k(x)}}(t_j + l_n). \end{aligned} \tag{2.20}$$

Next, we use the Hölder's condition on F^x and we show that, for any y_1, y_2 $\in S_{\mathbb{R}}$ and for all $x \in S_{\mathcal{F}}$,

$$\begin{aligned} \left|E\widehat{F}_N^x(y_1) - E\widehat{F}^x(y_2)\right| &= \left|\frac{1}{E[K_1(x)]}E[K_1(x)F^{X_1}(y_1)] - \frac{1}{E[K_1(x)]}E[K_1(x)F^{X_1}(y_2)]\right| \\ &\le C|y_1 - y_2|^{b_2}. \end{aligned} \tag{2.21}$$

Then we get by (2.20) and (2.21) and because $l_n = n^{-\frac{1}{2b_2}}$:

$$F_2' \le F_4' + O\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\,\phi(h_K)}}\right),$$

where

$$F_4' = \max_{k\in\{1,\dots,N_\epsilon(S_{\mathcal{F}})\}} \max_{1\le j\le z_n} \max_{s_j=t_j-l_n,t_j+l_n} \left|\widehat{F}_N^{x_{k(x)}}(s_j) - E\widehat{F}_N^{x_{k(x)}}(s_j)\right|.$$

Thus, it remains to study F_4'. By using similar arguments as those invoked for studying F_2 and combined with (H6′), one has $F_4' = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\,\phi(h_K)}}\right)$, which implies that

$$\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}}\left|\widehat{F}_N^{x_{k(x)}}(s_{j(y)}) - E\widehat{F}_N^{x_{k(x)}}(s_{j(y)})\right| = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\,\phi(h_K)}}\right). \tag{2.22}$$

So, Lemma 5 can be easily deduced from (2.19) and (2.22).

□

Proof of Theorem 3 The proof is based on the following decomposition

$$\widehat{f}^x(y) - f^x(y) = \frac{1}{\widehat{f}(x)}\Big[(\widehat{f}_N^x(y) - E\widehat{f}_N^x(y)) - (f^x(y) - E\widehat{f}_N^x(y))\Big] + \frac{f^x(y)}{\widehat{f}(x)}\Big[E\widehat{f}(x) - \widehat{f}(x)\Big] \tag{2.23}$$

where

$$\widehat{f}_N^x(y) = \frac{1}{nh_H E[K(h_K^{-1}d(x,X_1))]}\sum_{i=1}^n K\left(h_K^{-1}d(x,X_i)\right)H\left(h_H^{-1}(y-Y_i)\right).$$

Theorem 3 can be deduced from the following intermediate results, together with Lemma 1 and Corollary 3.

Lemma 6 *Under the hypotheses (H2′),(H4) and (H7), we have*

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

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

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

□

Proof of Lemma 6. One has

$$\begin{aligned} E\widehat{f}_N^x(y) - f^x(y) &= \frac{1}{nh_H EK_1(x)}E\left[\sum_{i=1}^n K_i(x)H_i(y)\right] - f^x(y) \\ &= \frac{1}{EK_1(x)}E\left(K_1(x)\left[h_H^{-1}E(H_1(y)|X_1) - f^x(y)\right]\right). \end{aligned}$$

Moreover, by change of variable

$$h_H^{-1}E\left(H_1(y)|X_1\right) = \frac{1}{h_H}\int_{\mathbb{R}} H(\frac{y-z}{h_H})f^{X_1}(z)dz = \int_{\mathbb{R}} H(t)f^{X_1}(y-h_Ht)dt,$$

we arrive at

$$|h_H^{-1}E(H_1(y)|X_1) - f^x(y)| \leq \int_{\mathbb{R}} H(t)|f^{X_1}(y - h_H t) - f^x(y)|dt.$$

Finally, the use of (H2″) implies that

$$|h_H^{-1}E(H_1(y)|X_1) - f^x(y)| \leq C \int_{\mathbb{R}} H(t)(h_K^{b_1} + |t|^{b_2} h_H^{b_2})dt.$$

This inequality is uniform on (x, y) in $S_{\mathcal{F}} \times S_{\mathbb{R}}$ and the use of (H7) states Lemma 6.

□

Proof of Lemma 7. Let us keep the definition of $k(x)$ (resp. $j(y)$) as in Lemma 1 (resp. in Lemma 5). The compactness of $S_{\mathbb{R}}$ permits to write that

$$S_{\mathbb{R}} \subset \bigcup_{j=1}^{z_n} (t_j - l_n, t_j + l_n)$$

with $l_n = n^{-\frac{3}{2}\gamma - \frac{1}{2}}$ and $z_n \leq C\, n^{\frac{3}{2}\gamma + \frac{1}{2}}$. We have the following decomposition :

$$\begin{aligned}
\left|\widehat{f}_N^x(y) - E\widehat{f}_N^x(y)\right| \quad \leq & \underbrace{\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}}\left|\widehat{f}_N^x(y) - \widehat{f}_N^{x_{k(x)}}(y)\right|}_{T_1} + \underbrace{\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}}\left|\widehat{f}_N^{x_{k(x)}}(y) - \widehat{f}_N^{x_{k(x)}}(t_{j(y)})\right|}_{T_2} \\
& + \underbrace{\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}}\left|\widehat{f}_N^{x_{k(x)}}(t_{j(y)}) - E\widehat{f}_N^{x_{k(x)}}(t_{j(y)})\right|}_{T_3} \\
& + \underbrace{\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}}\left|E\widehat{f}_N^{x_{k(x)}}(t_{j(y)}) - E\widehat{f}_N^{x_{k(x)}}(y)\right|}_{T_4} \\
& + \underbrace{\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}}\left|E\widehat{f}_N^{x_{k(x)}}(y) - E\widehat{f}_N^{x}(y)\right|}_{T_5}.
\end{aligned}$$

Similarly to the study of the term F_1 and by replacing (H5b) with (H5b″), it comes

$$T_1 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n^{1-\gamma}\phi(h_K)}}\right) \text{ and } T_5 = O\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n^{1-\gamma}\,\phi(h_K)}}\right). \tag{2.24}$$

Concerning the term T_2, by using the Lipschitz's condition on the kernel H, one can write

$$\begin{aligned}\left|\widehat{f}_N^{x_{k(x)}}(y) - \widehat{f}_N^{x_{k(x)}}(t_{j(y)})\right| &\leq C\frac{1}{n\, h_H\, \phi(h_K)}\sum_{i=1}^{n} K_i(x_{k(x)})\left|H_i(y) - H_i(t_{j(y)})\right| \\ &\leq \frac{C}{n}\sum_{i=1}^{n} Z_i,\end{aligned}$$

where $Z_i = \dfrac{l_n\, K_i(x_{k(x)})}{h_H^2\, \phi(h_K)}$. Once again, a standard exponential inequality for a sum of bounded variables allow us to write

$$\widehat{f}_N^{x_{k(x)}}(y) - \widehat{f}_N^{x_{k(x)}}(t_{j(y)}) = O\left(\frac{l_n}{h_H^2}\right) + O_{a.co.}\left(\frac{l_n}{h_H^2}\sqrt{\frac{\log n}{n\, \phi(h_K)}}\right).$$

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

$$T_2 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_\mathcal{F}}(\epsilon)}{n^{1-\gamma}\phi(h_K)}}\right) \text{ and } T_4 = O\left(\sqrt{\frac{\psi_{S_\mathcal{F}}(\epsilon)}{n^{1-\gamma}\phi(h_K)}}\right). \tag{2.25}$$

By using analogous arguments as for Lemma 1, we can show for all $\eta > 0$,

$$\begin{aligned}&P\left(T_3 > \eta\sqrt{\frac{\psi_{S_\mathcal{F}}(\epsilon)}{n\, h_H\, \phi(h_K)}}\right) \\ &\quad = P\left(\max_{j\in\{1,2,\ldots,z_n\}}\max_{k\in\{1,\ldots,N_\epsilon(S_\mathcal{F})\}}\left|\widehat{f}_N^{x_k}(t_j) - E\widehat{f}_N^{x_k}(t_j)\right| > \eta\sqrt{\frac{\psi_{S_\mathcal{F}}(\epsilon)}{n\, h_H\, \phi(h_K)}}\right) \\ &\quad \leq z_n N_\epsilon(S_\mathcal{F})\max_{j\in\{1,2,\ldots,z_n\}}\max_{k\in\{1,\ldots,N_\epsilon(S_\mathcal{F})\}} P\left(\left|\widehat{f}_N^{x_k}(t_j) - E\widehat{f}_N^{x_k}(t_j)\right| > \eta\sqrt{\frac{\psi_{S_\mathcal{F}}(\epsilon)}{n\, h_H\, \phi(h_K)}}\right).\end{aligned}$$

Let

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

and apply Bernstein exponential inequality (see, Corollary A.9 in Ferraty and Vieu, 2006). For that, we must calculate the asymptotic behavior of $E|\Delta_i|$ and $E\Delta_i^2$. Firstly, it follows from the fact that the kernel K and H

are bounded that $E|\Delta_i| \leq C\,(h_H\,\phi(h_K))^{-1}$. Secondly, the use of the same analytic arguments as of Lemma 6, allows us to get :

$$\lim_{n\to\infty} \frac{1}{h_H} E[H_1^2(y)|X_1] = f^{X_1} \int_{\mathbb{R}} H^2(t)dt,$$

which implies that

$$E|\Delta_i|^2 \leq \frac{C}{h_H\,\phi(h_K)}.$$

Thus, we are now in position to apply the Bernstein exponential inequality and we get :

$$\forall j \leq z_n,\ P\left(\left|\widehat{f}_N^{x_k}(t_j) - E\widehat{f}_N^{x_k}(t_j)\right| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\,h_H\,\phi(h_K)}}\right)$$
$$\leq 2\exp\{-C\eta^2\psi_{S_{\mathcal{F}}}(\epsilon)\}.$$

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

$$z_n N_\epsilon(S_{\mathcal{F}}) \max_{j\in\{1,2,\ldots,z_n\}} \max_{k\in\{1,\ldots,N_\epsilon(S_{\mathcal{F}})\}} P\left(\left|\widehat{f}_N^{x_k}(t_j) - E\widehat{f}_N^{x_k}(t_j)\right| > \eta\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\,h_H\,\phi(h_K)}}\right)$$
$$\leq C' z_n N_\epsilon(S_{\mathcal{F}})^{1-C\eta^2}.$$

By using the fact that $\lim_{n\to+\infty} n^\gamma\,h_H = \infty$ and (H6″), one obtains

$$T_3 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n^{1-\gamma}\,\phi(h_K)}}\right). \tag{2.26}$$

So, Lemma 7 can be easily deduced from (2.24)-(2.26).

□

Proof of Theorem 4 The proof is based on the same kind of decomposition as (2.23)

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

Consequently, Theorem 4 is deduced from Theorems 2 and 3, and from the next result which is a consequence of Theorem 2.

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

$$\exists \delta > 0 \ \ \textit{such that} \ \sum_{n=1}^{\infty} P\left\{\inf_{x\in S_{\mathcal{F}}}\inf_{y\in S_{\mathbb{R}}} |1-\widehat{F}^x(y)| < \delta\right\} < \infty.$$

□

Proof of Corollary 4. It is clear that

$$\inf_{x\in S_{\mathcal{F}}}\inf_{y\in S_{\mathbb{R}}} |1-\widehat{F}^x(y)| \leq (1-\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}} F^x(y))/2$$

$$\Rightarrow \sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}} |\widehat{F}^x(y)-F^x(y)| \geq (1-\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}} F^x(y))/2,$$

which implies that

$$\sum_{n=1} P\left(\inf_{x\in S_{\mathcal{F}}}\inf_{y\in S_{\mathbb{R}}} |1-\widehat{F}^x(y)| \leq (1-\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}} F^x(y))/2\right)$$

$$\leq \sum_{n=1} P\left(\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}} |\widehat{F}^x(y)-F^x(y)| \geq (1-\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}} F^x(y))/2\right) < \infty.$$

We deduce from Theorem 2 that

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

This proof is achieved by taking $\delta = (1-\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}} F^x(y))/2$ which is strictly positive.

□

Proof of Corollary 2. By a simple manipulation, we show that

$$|f^x(\widehat{\theta(x)}) - f^x(\theta(x))| \leq 2\sup_{y\in S}|\widehat{f}^x(y)-f^x(y)|. \tag{2.27}$$

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 $\widehat{\theta(x)}$. Clearly, it follows by (H9), (2.27) and Theorem 3 that

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

Moreover, by means of (H10), we obtain that

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

Hence, as for Corollary 3, we can get $\tau > 0$ such that

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

and we have

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

By combining this result with Theorem 3, we obtain the claimed result.

□

Acknowledgements. The authors would like to thank both referees whose comments and suggestions have improved significantly the presentation of this work. All the participants of the working group STAPH on Functional and Operatorial Statistics in Toulouse are also gratified for their continuous supports and comments.

References

Ait Saïdi, A., Ferraty, F., Kassa, R., Vieu, P. (2008). Cross-validated estimations in the single-functional index model. *Statistics*, **42**, 475-494.

Aneiros Perez, G., Vieu, P. (2006). Semi-functional partial linear regression. *Statist. & Proba. Let.*, **76**, 1102-1110.

Benhenni, K., Ferraty, F., Rachdi, M., Vieu, P. (2007). Local smoothing regression with functional data. *Comput. Statistics,* **22**, 353-370.

Bogachev, V.I. (1999). *Gaussian measures.* Math surveys and monographs, **62**, Amer. Math. Soc.

BOULARAN, J., FERRÉ, L., VIEU, P. (1995). Location of particular points in nonparametric regression analysis. *Austral. J. Statist.* **37**, 161-168.

CHATE, H., COURBAGE, M. (1997). Lattice systems. *Physica D*, **103** 1 -612.

DABO-NIANG, S., RHOMARI, N. (2003). Estimation non paramétrique de la régression avec variable explicative dans un espace métrique. *C. R., Math., Acad. Sci. Paris.* **336**, 75-80.

DABO-NIANG, S., LAKSACI, A. (2007). Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle. *C. R., Math., Acad. Sci. Paris.* **344**, 49-52.

DEHEUVELS, P., MASON, D. (2004). General asymptotic confidence bands based on kernel type function estimators. *Statistical Inference for Stochastic Processes*, **7**, 225-277.

DELSOL, L. (2007). Régression nonparamétrique fonctionnelle : expression asymptotique des moments. *Annales de L'ISUP*, **LI**(3), 43-67.

DELSOL, L. (2009). Advances on asymptotic normality in nonparametric functional time series analysis. *Statistics*, **43**, 13-33.

EZZAHRIOUI, M., OULD-SAÏD (2008). Asymptotic normality of nonparametric estimator of the conditional mode for functional data. *Journal of Nonparametric Statistics*, **20**(1), 3-18.

FERRATY, F., VIEU, P. (2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. *C. R., Math., Acad. Sci. Paris,* **330**, 139-142.

FERRATY, F., GOIA, A., VIEU, P. (2002). Functional Nonparametric Model for Time Series : A Fractal Approach for Dimension Reduction. TEST, **11**, 317-344.

FERRATY, F., RABHI, A., VIEU, P. (2005b). Conditional quantiles for functionally dependent data with application to the climatic El Phenomenon. *Sankhya,* **67**, 378-399.

FERRATY, F., LAKSACI, A., VIEU, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statistics Inference for Stochastics Processes,* **9**, 47-76.

FERRATY, F., VIEU, P. (2006). *Nonparametric functional data analysis. Theory and Practice.* Springer-Verlag.

FERRATY, F., MAS, A., VIEU, P. (2007). Nonparametric regression on functional data : inference and practical aspects. *Aust. N. Z. J. Stat.*, **49**, 267-286.

FERRATY, F., RABHI, A., 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., VIEU, P. (2009). On the validity of the bootstrap in nonparametric functional regression. *Scand. J. Stat. (In press).*

FERRATY, F., VIEU, P. (2009). Additive prediction and boosting for functionnal data. *Computational Statistics & Data Analysis*, **53** 1400-1413.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, **58**, 13-30.

KOLMOGOROV, A. N., TIKHOMIROV, V .M. (1959). ϵ-entropy and ϵ-capacity. *Uspekhi Mat. Nauk* **14** 3-86. (*Engl Transl. Amer. Math. Soc. Transl. Ser*), **2** 277-364 (1961).

KUELBS, J., LI, W. (1993). Metric entropy and the small ball problem for Gaussian measures. *J. Funct. Anal.*, **116**, 133-157.

MASRY, E. (2005). Nonparametric regression estimation for dependent functional data : asymptotic normality. *Stochastic Processes and their Applications,* **115**,155-177.

ONSAGER, L., MACHLUP, S. (1953). Fluctuations and irreversible proceses, I-II. *Phys. Rev.*, **91**, 1505-1512 ; 1512-1515.

OULD-SAÏD, E., CAI Z. (2005). Strong uniform consistency of nonparametric estimation of the censored conditional mode function. *Nonparametric Statistics,* **17**, 797-806.

QUINTELA-DEL-RIO, A. (2006). Nonparametric estimation of the maximum hazard under dependence conditions. *Statist. Probab. Lett.* , **76**, 1117-1124.

RAMSAY, J. O., SILVERMAN, B. W. (1997). *Functional data analysis.* Springer, New York.

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

ROUSSAS, G. G. (1968). On some properties of nonparametric estimates of probability density functions. *Bull. Soc. Math. Greece (N.S.)* **9**, 29-43.

SAMANTA, M. (1989). Nonparametric estimation of conditional quantiles, *Stat. Probab . Lett.*, **7**, 407-412.

THEODOROS, N., YANNIS G. Y. (1997) Rates of convergence of estimate, Kolmogorov entropy and the dimensioanlity reduction principle in regression. *The Annals of Statistics*, **25**, No. 6, 2493–2511.

VAN DER VAART, A. W., VAN ZANTEN, J. H. (2007). Bayesian inference with rescaled Gaussian process priors. *Electronic Journal of Statistics.*, **1**, 433-448.

WATSON, G. S., LEADBETTER, M. R. (1964). Hazard analysis. I. *Biometrika*, **51**, 175-184.

YOUNDJÉ, E., (1996). Propriétés de convergence de l'estimateur à noyau de la densité conditionnelle. *Rev. Roumaine Math. Pures Appl.*, **41**, 535-566.

Chapitre 3

Applications

L'objectif de ce chapitre est de montrer, à l'aide de données simulées, comment on peut implémenter facilement et rapidement la fonction de régression par la méthode du noyau et que le choix du paramètre de lissage joue un rôle crucial dans le comportement asymptotique de l'estimateur en insistant sur une méthode de sélection du paramètre de lissage.

3.1 Régression fonctionnelle : un exemple

Les exemples d'applications sur l'estimation de la fonction de régression sont très nombreux comme en témoigne la littérature abondante accordée à ce type de modèle et ses implications dans différents domaines, tels la classification des courbes (Ferraty et Vieu 2004), la prévision non paramétrique (Ferraty et Vieu 2006) et la prévision de séries temporelles fonctionnelles (Ferraty *et al.* (2002), Delsol (2007)).

Dans cette partie, on revient à l'utilité de la fonction de régression dans ce contexte de prévision. Pour cela, on considère un jeu de données assez compliqué pour lequel les observations (Y_i, X_i) sont liées par le modèle de régression suivant :

$$Y_i = r(X_i) + \epsilon_i, \qquad i = 1, \dots n,$$

où $r(x) = 4\log\left\{1/\left(\int(x'(t))^2dt + [\int(x'(t))^2dt]^2\right)\right\} + 3$ et $\epsilon_1, \ldots, \epsilon_n$ n v.a.r. i.i.d. de loi normale $N(0,1)$. Pour les variables explicatives X_i, on suppose que ces observations sont issues d'un processus stochastique $X(t) = a\cos(b + \pi Wt) + c\sin(d + 2\pi Wt)$ où W, b et d sont indépendantes de lois normales respectivement $N(0,1)$, $N(0,0.3)$ et $N(0,0.5)$. Les variables a et c sont de lois de Bernoulli $\mathcal{B}(0.5)$. Les courbes X_i sont régulièrement observées sur une grille de discrétisation de 100 points dans l'intervalle $[0,1]$ et sont représentées sur le graphe suivant :

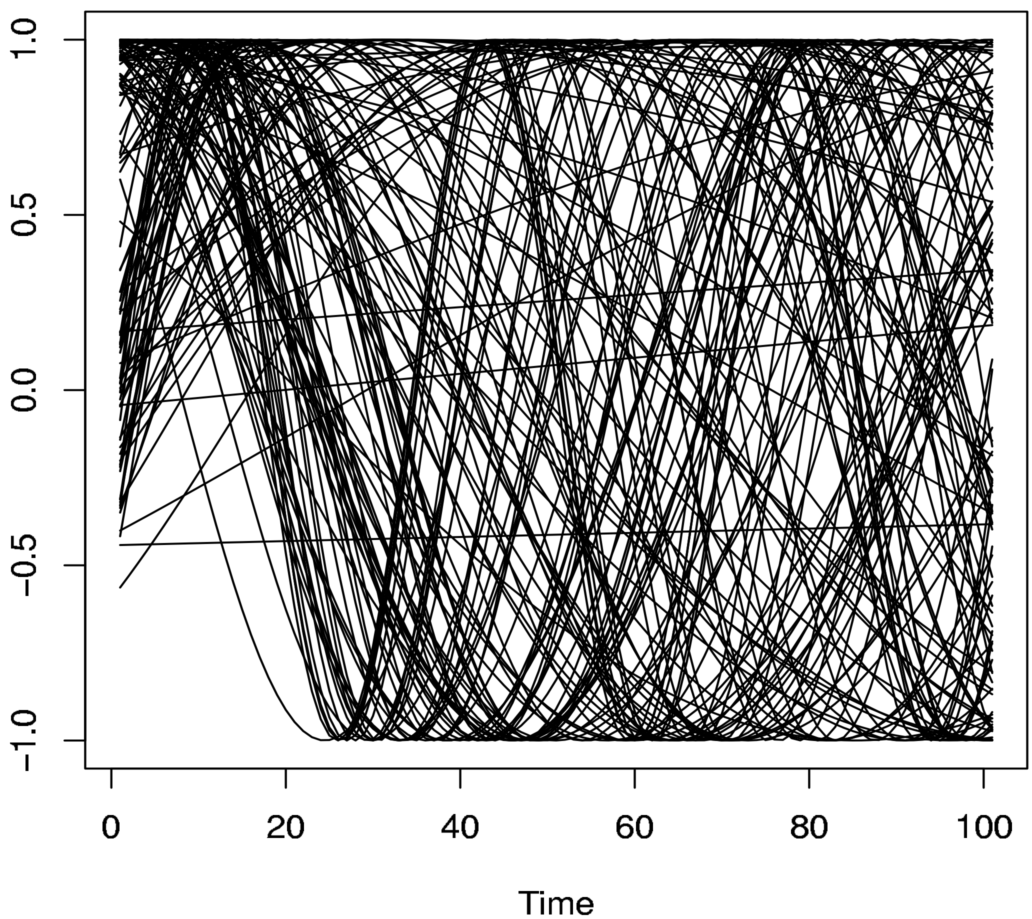

Fig. 1. Les courbes X_i

Notre objectif principal est de montrer l'efficacité de notre méthode d'estimation dans un contexte de prévision. Pour cela, nous considérons 110 observations réparties sur deux paquets : échantillon d'apprentissage $(X_i, Y_i)_{i\in A}$, $card(A) = 90$, échantillon de test $(X_i, Y_i)_{i\in T}$, $card(T) = 20$ et nous utilisons un noyau quadratique défini par

$$K(x) = \begin{cases} \frac{3}{2}(1 - x^2) & \text{si } x \in [0, 1], \\ 0 & \text{ailleurs,} \end{cases} \tag{3.1}$$

ainsi qu'une semi-métrique définie par

$$\forall (x_1, x_2) \in \mathcal{F} \times \mathcal{F};\ d_q^{\text{deriv}}(x_1, x_2) = \sqrt{\int (x_1^{(q)}(t) - x_2^{(q)}(t))^2 dt}, \quad q \in \mathbb{N},$$

où $x^{(q)}$ est la dérivée d'ordre q. Le choix de cette famille de semi-métriques est guidé par la forme du modèle de régression choisi et est adapté à la régularité des courbes $(X_i)_{i=1,\ldots n}$ Dans notre procédure d'estimation, nous avons utilisé plusieurs valeurs de q ($q \in \{1, 2, 3, 5, 7\}$) et nous avons constaté que lorsque $q = 1$ les résultats sont légèrement meilleurs. De même, le choix du paramètre de lissage h_K joue un rôle très important dans cette estimation par la méthode du noyau. Dans ce contexte fonctionnel, Rachdi et Vieu (2007) et Benhenni *et al.* (2007) ont élaboré des méthodes pratiques pour le choix automatique (local et/ou global) de ce paramètre. Typiquement, ces méthodes sont basées sur la minimisation des quantités suivantes :

$$LCV_x(h_K) = n^{-1} \sum_{i=1}^{n} (Y_i - \hat{r}^{-i}(X_i))^2 W_{n,x}(X_i) \text{ pour le choix local,}$$

et

$$GCV(h_K) = n^{-1} \sum_{i=1}^{n} (Y_i - \hat{r}^{-i}(X_i))^2 W(X_i) \text{ pour le choix global,}$$

où W et $W_{n,x}$ sont des fonctions convenablement choisies par l'utilisateur et où $\hat{r}^{-i}$ est l'estimateur basé sur l'échantillon privé de la $i^{\text{ème}}$ observation. Le paramètre de lissage obtenu par cette procédure est optimal selon les critères suivants :

$$ASE(h_K) = n^{-1}\sum_{i=1}^{n}(\hat{r}(X_i) - r(X_i))^2 W(X_i),$$

$$ISE(h_K) = \int (\hat{r}(x) - r(x))^2 W(x) dP_X(x)$$

et

$$MISE(h_K) = \int E(\hat{r}(x) - r(x))^2 W(x) dP_X(x),$$

où P_X est la mesure de probabilité de la variable fonctionnelle X. D'un point de vue pratique, l'estimation de la fonction de régression r avec ces deux types de choix a été déjà programmé par Ferraty et Vieu (2006) (voir les différents programmes sur le site STAPH http ://www.math.univ-toulouse.fr/staph/ npfda). Dans cet exemple, nous comparons le choix local au choix global en considérant les deux fonctions $funopare.knn.lcv$ et $funopare.knn.gcv$ pour prévoir les valeurs réponses $(Y_i)_{i\in T}$ dans l'échantillon. Notons que dans la fonction $funopare.knn.lcv$ (resp. la fonction $funopare.knn.gcv$) on prend $W_{n,x} = 1\!\!1_{B(x,h)}$ (resp. $W = 1$) et que la fenêtre optimale est sélectionnée localement (resp. globalement) en terme de nombre de voisins les plus proches (resp. sur l'ensemble des quantiles de la matrice $a_{ij} = d(X_i, X_j)$). Nous testons l'efficacité de ces méthodes d'estimation en calculant l'erreur suivante :

$$RMSE = \frac{\sum_{i\in T}(Y_i - \hat{r}(X_i))^2}{Var(Y_i)},$$

où $\hat{r}$ est l'estimateur de r. Les résultats de prévision pour les deux méthodes sont donnés sur le graphe suivant :

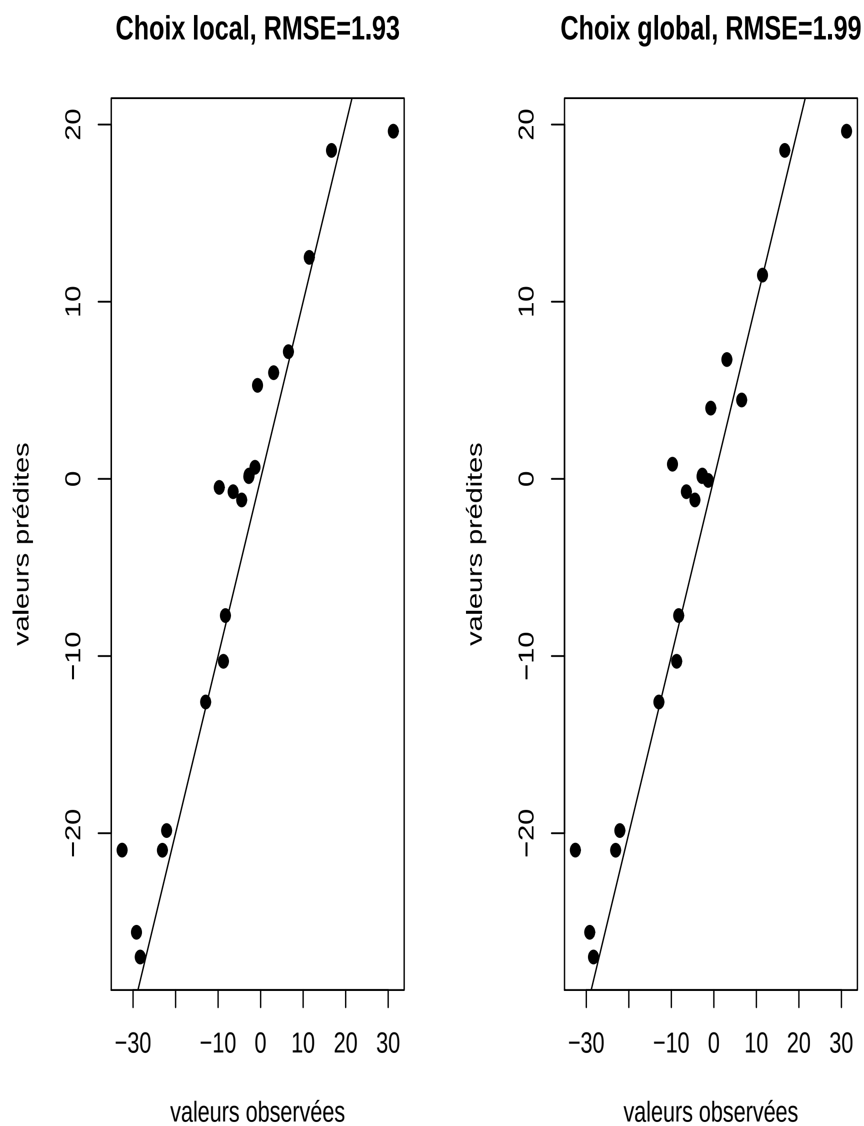

Fig. 2. Qualité de prédiction

Ces résultats numériques montrent que notre procédure d'estimation est performante en terme de prévision et que les critères de choix du paramètre de lissage ont un bon comportement en pratique. Notons aussi que dans cet exemple, l'estimation utilisant un choix local du paramètre de lissage est légèrement meilleure ($RMSE = 1.93$) par rapport au cas du choix global où l'erreur de prévision est $RMSE = 1.99$.

3.2 Sur la mise en oeuvre de l'estimateur de la densité conditionnelle

La densité conditionnelle est une alternative à l'espérance conditionnelle pour décrire la relation entre deux variables aléatoires. Du point de vue pratique, la plupart des travaux en statistique non paramétrique fonctionnelle traite la densité conditionnelle comme une étape préliminaire à l'estimation du mode conditionnel (outil de prévision alternatif à la régression ; voir, par exemple la monographie de Ferraty et Vieu (2006)). Dans cette perspective prédictive, les paramètres h_K et h_H sont obtenus en minimisant le critère suivant :

$$\sum_{i=1}^{n}(Y_i - \hat{\theta}^{-i}(X_i))^2, \tag{3.2}$$

où $\hat{\theta}^{-i}(X_i) = \arg\sup_y \widehat{f^{X_i}}^{-i}(y)$ représente l'estimateur du mode conditionnel basé sur l'ensemble des données à l'exception de (X_i, Y_i). Les programmes sont accessibles sur le site du groupe STAPH (http ://www.math.univ-toulouse.fr/staph/npfda). On peut aussi estimer la densité conditionnelle pour diverses finalités autres que le mode conditionnel (estimation de la fonction de hasard conditionnelle, test de multi-modalités,...). Pour cela, on pourrait être amené à considérer le critère de validation croisée suivant :

$$CV(h_K, h_H) = \frac{1}{n}\sum_{i=1}^{n} W_1(X_i)\int(\widehat{f^{X_i}}^{-i}(z))^2 W_2(z)dz - \frac{2}{n}\sum_{i=1}^{n}\widehat{f^{X_i}}^{-i}(Y_i)W_1(X_i)W_2(Y_i),$$

où

$$\widehat{f^{x}}^{-i}(y) = \frac{h_H^{-1}\sum_{j\neq i}^{n} K(h_K^{-1}d(x, X_j))H(h_H^{-1}(y - Y_j))}{\sum_{j\neq i}^{n} K(h_K^{-1}d(x, X_j))},$$

W_1 et W_2 étant des fonctions convenablement choisies par l'utilisateur. Ce critère est une extension au cadre fonctionnel du critère introduit par Youndjé

(1993), mais son étude théorique dans le cadre fonctionnel reste un problème ouvert.

3.3 A propos de la mise en oeuvre de l'estimateur de la fonction de répartition conditionnelle

De même que le cas de la densité conditionnelle, la fonction de répartition conditionnelle est souvent traitée en statistique non paramétrique fonctionnelle comme une étape préliminaire à l'estimation de la médiane conditionnelle ou bien celle des quantiles conditionnels (voir Ferraty et Vieu (2006)). Le critère du choix des paramètres de lissage est donné par

$$(h_{K_{opt}}, h_{H_{opt}}) = \arg \min_{h_K, h_H} \sum_{i=1}^{n} (Y_i - \inf_u \{ \widehat{F^{X_i}}^{-i}(u) \geq .5\})^2,$$

où $\widehat{F^{X_i}}^{-i}(u)$ est l'estimateur de la fonction de répartition conditionnelle défini par

$$\widehat{F^{X_i}}^{-i}(u) = \frac{\displaystyle\sum_{j \neq i}^{n} K(h_K^{-1} d(X_i, X_j)) H^*(h_H^{-1}(u - Y_j))}{\displaystyle\sum_{j \neq i}^{n} K(h_K^{-1} d(X_i, X_j))},$$

avec $H^*(t) = \int_0^t K(t)dt$. Ce critère permet d'implémenter facilement la fonction de répartition conditionnelle pour s'en servir dans un but prévisionnel. Les programmes sont accessibles sur le site du groupe STAPH (http ://www.math.univ-toulouse.fr/staph/npfda). Les questions liées au choix du paramètre de lissage restent en particulier ouvertes dans le cadre fonctionnel.

3.4 Sur la mise en oeuvre de l'estimateur de la fonction de hasard conditionnelle

Très peu de travaux se sont intéressés à la mise en oeuvre de l'estimation de la fonction de hasard conditionnellement à une variable fonctionnelle, à l'exception de l'article de Quintela-del-Rio (2010). Cependant, les aspects théoriques ont été largement étudiés (voir, Ferraty *et al.* (2008), Ezzahroui et Ould-Saïd (2010), Laksaci et Mechab (2010)). Les questions liées au choix du paramètre de lissage restent en particulier ouvertes dans le cadre fonctionnel.

3.5 Conclusion

Notre objectif dans ce chapitre a été de montrer l'applicabilité des modèles conditionnels étudiés dans cette première partie de thèse. En conclusion, on peut dire que l'estimateur utilisé pour la régression est très facile à mettre en oeuvre en pratique et que l'aspect fonctionnel des données n'a pas trop d'influence sur la rapidité du programme utilisé tout en exploitant la nature fonctionnelle des données à travers les différentes métriques qu'on peut utiliser. Nous avons constaté que le bon comportement de cet estimateur est étroitement liée au choix de la métrique et du paramètre de lissage. En ce qui concerne le choix optimal du paramètre associé à une famille de semi-métriques, il est encore une question ouverte.

Références

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

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

Ezzahrioui, M., Ould-Saïd (2010). Some asymptotic results of a nonparametric conditional mode estimator for functional time-series data. Stat. Neerl. **64**, 171-201.

Ferraty, F., Goia, A., Vieu, P. (2002). Functional Nonparametric Model for Time Series : A Fractal Approach for Dimension Reduction. TEST, **11**, 317-344.

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

Ferraty, F., Vieu, P. (2006). Nonparametric functional data analysis. Theory and Practice. Springer-Verlag.

Laksaci, A., Mechab, M. (2010). Estimation non paramétrique de la fonction de hasard avec variable explicative fonctionnelle cas des donnees spatiales. Rev : Roumaine, Math Pures Appl. **55**, 35-51.

Quintela-del-Rio, A. (2010). On non-parametric techniques for area-characteristic seismic hazard parameters. Geophys. J. Int. **180**, 339-346.

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

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

Deuxième partie

Convergence uniforme : réponse fonctionnelle

Chapitre 4

Les variables aléatoires Banachiques

Dans ce chapitre, nous ouvrons une courte parenthèse pour donner quelques définitions et quelques outils probabilistes pour variables aléatoires à valeurs dans un espace de Banach. Dans ce contexte, de nombreux travaux sont disponibles dans la littérature (à titre d'exemple, on peut citer Bru et Heinich (1980) et Li et Queffélec (2004)). Nous renvoyons à Beck (1958) qui fait déjà état de développements nombreux et variés sur ce thème. En ce qui concerne les rappels qui suivent, nous nous référons essentiellement à l'ouvrage de Bosq (2000) qui dresse le bilan actuel des diverses connaissances acquises en la matière.

4.1 Espérance de variable Banachique

Afin d'établir la définition de l'espérance d'une variable aléatoire prenant leurs valeurs dans un espace de Banach, nous commençons par définir les variables aléatoires dans un espace de Banach.

Définition 1 *(Beck, 1958)*
Soit (Γ, Σ, μ) *un espace mesuré et* $\mathcal{B}$ *un espace de Banach. Si une application*

Y de Γ dans $\mathcal{B}$ est telle que :

i) il existe un ensemble $\Gamma_0 \in \Sigma$ tel que $\mu(\Gamma - \Gamma_0) = 0$ et que $Y(\Gamma_0) = \{Y(t),\ t \in \Gamma_0\}$ est séparable,

ii) Pour tout ensemble de Borel $A \subset \mathcal{B}$, $Y^{-1}(A) = \{t/Y(t) \in A\} \in \Sigma$,

nous disons que Y est une fonction fortement mesurable de (Γ, Σ, μ) dans $(\mathcal{B}, \mathrm{B}_{\mathcal{B}})$.

Définition 2 *(Beck, 1958)*
Soit Y une fonction fortement mesurable de (Γ, Σ, μ) dans $\mathcal{B}$. S'il existe un élément $y \in \mathcal{B}$ tel que pour toute fonctionnelle linéaire bornée y^* définie sur $\mathcal{B}$, on ait

$$\int_t y^*(Y(t))\mu(dt) = y^*(y),$$

nous disons que Y est fortement intégrable et y est l'intégrale de Y. Nous appellerons y l'espérance mathématique de Y et nous la noterons EY.

Définition 3 *(Beck, 1958)*
Si $\mathcal{B}$ est un espace de Banach séparable, alors $Y :\ (\Omega, \mathcal{A}) \to (\mathcal{B}, \mathrm{B}_{\mathcal{B}})$ est une $\mathcal{B}$-variable aléatoire si et seulement si $y^*(Y)$ est une variable aléatoire réelle pour tout $y^* \in \mathcal{B}^*$, où $\mathrm{B}_{\mathcal{B}}$ est la tribu de Borel et $\mathcal{B}^*$ est l'espace dual de $\mathcal{B}$.

Remarque 1 *(Bru et Heinich, 1980)*
Toute fonction fortement mesurable de Ω dans $\mathcal{B}$ est une $\mathcal{B}$-variable aléatoire.

Exemple 1 *(Bosq, 2000)*
Soit $\mathcal{B} = L^p([0,1], \mathrm{B}_{[0,1]}, \lambda)$, où $1 < p < \infty$, muni de la norme

$$||y|| = \left(\int_0^1 |y(t)|^p\right)^{1/p},\ y \in \mathcal{B},$$

λ est la mesure de Lebesgue sur $[0,1]$, alors $\mathcal{B}$ est un espace de Banach, son espace dual est

$$\mathcal{B}^* = L^q([0,1], \mathrm{B}_{[0,1]}, \lambda) \text{ avec } \frac{1}{p} + \frac{1}{q} = 1.$$

Si $\xi = \{\xi_t,\ 0 \leq t \leq 1\}$ *est un processus mesurable dans* $L^p([0,1], \mathrm{B}_{[0,1]}, \lambda)$, *alors* $\int_0^1 \xi_t z(t)dt$ *est une variable aléatoire pour tout* $z \in \mathcal{B}^*$. *D'où,* ξ *est une* $\mathcal{B}$*-variable aléatoire.*

Définition 4 *(Beck, 1958)*
Soit $(Y_i)_{i \in A}$ *une suite de* $\mathcal{B}$*-variables aléatoires, si pour tout entier positif* m, *pour tout* $(Y_i,\, Y_{i_1}, \cdots, Y_{i_m})$ *et pour tout* $(K_1, \ldots, K_m) \subset \mathcal{B}^m$, *on a :*

$$P(Y_{i_1} \in K_1, \cdots, Y_{i_m} \in K_m) = \prod_{j=1}^{m} P(Y_{i_j} \in K_j),$$

alors les Y_i *sont dites indépendantes.*

Définition 5 *(Bosq, 2000)*
Soit Y *une variable aléatoire Banachique ;* Y *est dite intégrable si*
$E(||Y||) < +\infty$.

Remarque 2 *(Bosq, 2000)*

i) $L^1_{\mathcal{B}}(P)$ *l'espace des classes d'équivalence* ($X \Leftrightarrow Y$ *lorsque* $X = Y$ *p.s) des* $\mathcal{B}$*-variables aléatoires intégrables muni de la norme* $||Y||_{1,\mathcal{B}} = E||Y||$ *est un espace de Banach.*

ii) *L'espérance mathématique* E *est un opérateur linéaire continu défini de* $L^1_{\mathcal{B}}(P)$ *dans* $\mathcal{B}$ *tel que :*

$$||EY|| \leq E||Y||.$$

Exemple 2 *(Bosq, 2000)*
Soit $C[0,1]$ *l'espace des fonctions réelles continues dans* $[0,1]$, *alors* $C[0,1]$ *muni de la norme sup :*

$$||y|| = \sup_{t \in [0,1]} |y(t)|,\ y \in C[0,1]$$

est un espace de Banach séparable. Soit $\xi = \{\xi_t,\ 0 \leq t \leq 1\}$ *un processus aléatoire réel à trajectoire continue ;* ξ *est donc une* $C[0,1]$*-variable aléatoire.*

Si ξ est intégrable (i.e. $E||\xi|| < +\infty$), alors

$$E\left(\int_0^1 \xi_t d\mu(t)\right) = \int_0^1 (E\xi)(t)d\mu(t),\ \mu \in \mathcal{M}([0,1]),$$

où $\mathcal{M}([0,1])$ est l'espace dual de $C[0,1]$ qui est l'espace des mesures signées sur $([0,1], \mathrm{B}_{[0,1]})$. Ceci implique :

$$(E\xi)(t) = E(\xi(t)),\ t \in [0,1].$$

4.2 Type et cotype d'un espace de Banach

La notion de type (cotype) d'un espace de Banch a été introduite par Hoffman-Jorgensen (1974). On trouvera dans la littérature plusieurs versions de cette notion. On pourra voir à ce sujet les travaux de Hamedani et Mandrekar (1978), Bastero et Uriz (1986) ainsi que Bosq (2000). On adopte la définition suivante :

Définition 6 *Soit $p \in [1,2]$ et ϵ_i une suite de variables aléatoires de Rademacher indépendantes (c'est-à-dire ϵ_i est une variable aléatoire qui prend les valeurs 1 ou -1 avec probabilité $1/2$). On dit que $\mathcal{B}$ est de type p si il existe une constante C telle que pour toute suite finie (x_i) d'éléments de $\mathcal{B}$, on a :*

$$E\left\|\sum_i \epsilon_i x_i\right\|^p \leq \sum_i \|x_i\|^p.$$

Cette approche nous permet d'étudier plusieurs propriétés fondamentales liées à la structure et à la géométrie de l'espace de Banach (voir Johnson and Lindenstrause (2003)). Dans la théorie des éléments aléatoires dans un espace de Banach, ce concept a été employé pour caractériser plusieurs propriétés fondamentales tel le théorème de limite centrale, la loi du logarithme itérée, la loi des grands nombres,..... A ce sujet De Acosta (1985) a montré que les espaces de Banach de type p sont caractérisés par la loi forte des grands nombres de Marcinkiewicz et Zygmund. Dans notre travail, on a utilisé cette notion pour préciser la vitesse de convergence de nos estimateurs. Notre démarche repose sur les propriétés suivantes :

Proposition 1 *(Bosq, 2000)*
Soit $p \in [1,2]$. Si $\mathcal{B}$ est de type p, il existe une constante C telle que, pour toute suite finie (X_i) de variables aléatoires centrées appartenant à L_p :

$$E\left\|\sum_i X_i\right\|^p \leq \sum_i E\|X_i\|^p.$$

Par ailleurs, nous donnons quelques exemples d'espaces Banachiques pour lesquels ce concept est établi.

1. **Espace de dimension finie :** On considère sur $\mathbb{R}^n$ la norme euclidienne définie par $\|x\|_2 = \sqrt{\sum_i x_i^2}$; l'espace $(\mathbb{R}^n, \|.\|_2)$ est de type 2.
2. **Espace de Hilbert :** Soit H un espace de Hilbert muni du produit scalaire $< \cdot, \cdot >$. L'espace H muni de la norme induite par le produit scalaire est de type 2. De plus si un espace $\mathcal{B}$ est un espace de Banach de type 2, alors il existe un espace de Hilbert isométrique à l'espace $\mathcal{B}$ (Mikhail et Vladimir, 1997, Théorème 5.3.4, p.69).
3. **Les espaces L_p :** L'espace des fonctions p-intégrables est de type p (Bosq, 2000).

4.3 Quelques inégalités pour variables Banachiques

D'un point de vue théorique, nos résultats asymptotiques s'appuient sur des inégalités de type exponentiel prenant en compte deux situations : le cas d'observations indépendantes et le cas d'observations dépendantes. Un autre type d'inégalité, appelée inégalité C_r, s'avérera aussi très utile.

Proposition 2 *(Bosq, 2000)*
Soit $Z_1, \ldots, Z_n$ une suite de $\mathcal{B}$-variables aléatoires indépendantes et identiquement distribuées. Si il existe des constantes positives $b = b(n)$, $l = l(n)$,

telles que :

$$\sum_{i=1}^{n} E||Z_i||^k \leq \frac{k!}{2} l^2 b^{k-2}, \forall k \geq 2,$$

alors pour tout $t > 0$, on a

$$P(|||S_n|| - E||S_n||| \geq t) \leq 2\exp(-\frac{t^2}{2l^2 + 2bt}),$$

où $S_n = \sum_{i=1}^{n} Z_i$.

Tous d'abord, nous avons besoin de donner la définition d'une suite β-mélangeante.

Définition 7 *(Rhomari, 2002)*
Soit $\{Z_i, i = 1, 2, ...\}$ une suite de variables aléatoires. Pour tout entier n, on pose $\beta(n)$ le coefficient de mélange défini par

$$\beta(n) = \sup_k \left\{ \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} |P(A_i \cap B_j) - P(A_i)P(B_j)| : \right.$$

$$\left. A_i \in \mathcal{F}_1^k(Z) \text{ et } B_j \in \mathcal{F}_{k+n}^{\infty}(Z),\ I,\ J \subset \mathbb{N}^* \text{ et } k \in \mathbb{N}^* \right\},$$

où $\mathcal{F}_i^k(Z)$ est la tribu engendrée par $\{Z_j, i \leq j \leq k\}$.
Si $\beta(n) \to 0$ lorsque $n \to \infty$, nous disons que la suite Z_i est β-mélangeante.

Proposition 3 *(Rhomari, 2002)*
Soit $Z_1, \ldots, Z_n$ une suite β-mélangeante où les $(Z_i)_{i=1,\cdots,n}$ sont centrées. Si pour $m \geq 2$, $\max_j \sum_{i=1}^{q} E||Z_{j+(i-1)p}||^m \leq \frac{1}{2} q m! \sigma^2 M^{m-2}$, alors pour tout $1 \leq j \leq p \leq n$ et tout $\epsilon > pe^*(q)$, on a

$$P(||\sum_{l=1}^{n} Z_l|| \geq \epsilon) \leq p\exp\left(\frac{(\epsilon - pe^*(q))^2}{2p[(n+p)C + (\epsilon - pe^*(q))M]}\right) + n\beta(p),$$

où C est une constante positive, M est un nombre positif, $e^*(q) = \max_j e_j^*(q)$ et $e_j^*(q) = E(\sum_{i=1}^q E||Z_{j,i}^*||)$, et où $(Z_{j,i}^*)_{i=1,\cdots,q}$ sont des v.a.i.i.d. ayant même lois que $(Z_{j+(i-1)p})_{i=1,\cdots,q}$.

Inégalité C_r

Proposition 4 *(Loève, 1963)*
Soit Z une $\mathcal{B}$-variable aléatoire, où $\mathcal{B}$ est un espace de Banach de type p $(p \geq 1)$; alors on a :

$$E\|Z - EZ\|^p \leq 2^{p-1}\left[E\|Z\|^p + \|EZ\|^p\right].$$

Références

De Acosta, A. (1985). Inequalities for B-valued random variables with applications to the strong law of large numbers, Ann. Probab. **9**, 157-161.

Bastero, J., Uriz, Z. (1986). Cotype for p-Banach spaces. Compositio Mathematica. **57**, 73-80.

Beck, A. (1958). Une loi forte des grands nombres dans des espaces de Banach uniformément convexes. Annales de l'institut Henri Poincaré. **16**, 35-45.

Bosq, D. (2000). Linear processs in function spaces. Theory and Application. Lectures Notes in Statistics. Vol 149, Springer Verlag.

Bru, B., Heinich, H. (1980). Sur l'espérance des variables aléatoires vectorielles. Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques. **16**, 177-196.

Hamedani, G. G., Mandrekar, V. (1978). Lévy-khinchine representation and Banach spaces of type and cotype. Studia Math. **Vol. 63**.

Hoffmann-Jørgensen, J. (1974). Sums of independent Banach space valued random variables. Studia Math. **52**, 159-186.

LI, D., QUEFFÉLEC, H. (2004). Introduction à l'étude des espaces de Banach. Analyse et probabilités. Vol 12, Eyrolles.

LOÈVE, M. (1963). Probability Theory. Third Edition. Van Nostranr Princeton.

MIKHAIL I. KADETS, VLADIMIR M. KADETS. (1997). Series in Banach Spaces : Conditional and Unconditional Convergence (Operator Theory : Advances and Applications). **94**, Birkhäuser.

RHOMARI, N. (2002). Approximation et inégalités exponentielles pour les sommes de vecteurs aléatoires dépendants. C. R. Acad. Sci. Paris. **334**, 149-154.

Chapitre 5

Réponse fonctionnelle et cas i.i.d.

Ce chapitre fait Ce chapitre fait l'objet d'une publication dans Electronic Journal of Statistics.

Kernel regression with functional response

Frédéric FERRATY[a], Ali LAKSACI[b], Amel TADJ[c,*] and Philippe VIEU[a]
[a]Université Paul Sabatier, Toulouse,
[b]Université Djillali Liabes, Sidi Bel Abbes
[c,*]Corresponding author : Université de Sidi Bel Abbès, BP 89 Sidi Bel Abbès 22000. Algérie . ameltdz@yahoo.fr

abstract We consider kernel regression estimate when both the response variable and the explanatory one are functional. The rates of uniform almost complete convergence are stated as function of the small ball probability of the predictor and as function of the entropy of the set on which uniformity is obtained.

keyword : Uniform almost complete convergence, kernel estimators, functional data, entropy, semi-metric space.

5.1 Introduction

Regression model is a main tool to examine the relationship between a response variable and an explanatory one. We are interested in estimating the nonparametric regression when both variables (response and explanatory) are functional. The study of statistical models adapted to such kind of data received a lot of attention in recent literature (see, Ramsay and Silverman (2005), Bosq (2000) and Ferraty and Vieu (2006) for recent monographies and Ferraty and Romain (2011) for an handbook on statistics in infinite dimensional spaces). Since the previous paper by Ferraty and Vieu (2002) the literature on nonparametric regression (see the survey by Ferraty and Vieu (2011)) starts to be rather important when the response variable is scalar, but there are very few advances in this direction when this response is functional (see however Bosq and Delecroix (1985) and Lecoutre (1990) for earlier references). Various advances on this field have been already provided but under some linear assumption on the regression operator (see for instance Müller *et al* (2008), He *et al* (2009), in standard i.i.d. case ; see also Bosq (2000) in the specific time series context).

This work presents some asymptotic property for a kernel-type regression estimator when both response and explanatory variables are functional. Precisely we state the uniform almost complete convergence rate of this doubly functional kernel estimate. As far as we know, our result is the first one stating uniform asymptotic results in nonparametric doubly functional regression problems. As usually in functional statistics, the topological structures on the infinite-dimensional spaces play a prominent role, and we present the rates of convergence in such a way to highlight these topological effects. Firstly, the problems linked with the high (high because infinite) dimensionality of the explanatory variable are dealt with by means of small ball probability considerations (and this is directly linked with the topological structure). Secondly, the uniformity of the convergence is obtained by means of entropy notions (which are, once again, direct topological considerations). Finally, the type of the Banach space on which the response variable takes its values acts also directly on the rates of convergence.

Section 2 is dedicated to some probability tools for functional variable valued in a Banach space. The doubly functional model and its estimate are presented in Section 3 and the uniform rates of convergence are stated therein. Some technical proofs are defered to the Appendix.

Before closing this introduction it is worth being stressed that, even if we have deliberately chosen to present a short theoretical paper, there exists a wide scope of applied scientific fields for which our approach could be of interest. In the next future, our works will be concentrated on the implementation of this doubly functional nonparametric method. To fix the ideas, one can for instance find examples in Biometrics, Genetics or Environmetrics in He *et al* (2009), Müller *et al* (2008) and Hlubinka and Prchal (2007) respectively.

5.2 Some probability tools for functional variables

We present two general tools for random variables valued in Banach spaces. The topological complexity of the Banach space $\mathcal{B}$ will appear through the following following notion.

Definition 2 *A Banach space $\mathcal{B}$ is of type $p \in]1,2]$ if there exists a strictly positive constant c such that for any finite sequence $Z_1, \dots, Z_n$ of independent $\mathcal{B}$-random variables such that $E||Z_i||^p < \infty$ and $EZ_i = 0$, we have*

$$E||\sum_{i=1}^{n} Z_i||^p \leq c \sum_{i=1}^{n} E||Z_i||^p.$$

Remark 1 *Clearly, $\mathbb{R}^k$ $(k \geq 1)$ or more generally any Hilbert space is a Banach space of type 2.*

In addition, and as it is usual in nonparametric statistics, one needs some kind of exponential inequality for getting rates of convergence. We will make use later in this paper of the following Bernstein's type inequality for sums of Banach-valued random elements :

Lemma 8 *Let $Z_1, \dots, Z_n$ be independent $\mathcal{B}$-random variables. If we have*

$$\sum_{i=1}^{n} E||Z_i||^k \leq \frac{k!}{2} l^2 b^{k-2}, \forall k \geq 2,$$

for some positive constants $b = b(n)$ and $l = l(n)$, then :

$$\forall t > 0,\ P(|||S_n|| - E||S_n||| \geq t) \leq 2 \exp(-\frac{t^2}{2l^2 + 2bt}),$$

where $S_n = \sum_{i=1}^{n} Z_i$.

This inequality can be found at page 49 in Bosq (2000), and a deeper discussion on $\mathcal{B}$-random variables can also be found in this book.

5.3 The functional regression with functional response

5.3.1 The doubly functional nonparametric setting

Let us consider a sample of independent pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ identically distributed as (X, Y) which is a random pair valued in $\mathcal{F} \times \mathcal{B}$, where $(\mathcal{F}, d)$ is a semi-metric space and $(\mathcal{B}, ||.||)$ is a Banach space of type $p \in]1, 2]$. Recall that a semi-metric (sometimes called pseudo-metric) is just a metric violating the property : $[d(x, y) = 0] \Rightarrow [x = y]$. The functional regression operator is defined by

$$m(x) = E\left[Y \mid X = x\right], \quad \forall x \in \mathcal{F}. \tag{5.1}$$

Nonparametric estimates of the operator m are constructed by local weighting ideas, as for instance the following doubly functional kernel estimate :

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} K(h^{-1} d(x, X_i)) Y_i}{\sum_{i=1}^{n} K(h^{-1} d(x, X_i))}, \ \forall x \in \mathcal{F},$$

where K is a kernel function and $h = h_n$ is a sequence of positive real numbers which goes to zero as n goes to infinity. We stay here with this simple nonparametric smoother, but alternative ones could be introduced such as the local functional one previously studied by Barrientos-Marin *et al.* (2010) for scalar Y.

5.3.2 The general hypotheses

Now $S_{\mathcal{F}}$ is a fixed subset of $\mathcal{F}$, and for $\eta > 0$ we consider the following η-neighborhood of $S_{\mathcal{F}}$:

$$S_{\mathcal{F}}^{\eta} = \{x \in \mathcal{F}, \exists x' \in S_{\mathcal{F}}, d(x, x') \leq \eta\}.$$

We will use the notation $B(x, h) = \{x' \in \mathcal{F}, \, d(x', x) \leq h\}$ for the closed ball in the space $\mathcal{F}$. The model consists in assuming that the probability distribution of X is such that there exists a non-decreasing function ϕ such that :

(H1) $\exists (C_1, C_2),\ \forall x \in S_{\mathcal{F}},\ \forall \epsilon > 0,\ 0 < C_1\phi(\epsilon) \leq P\left(X \in B(x,\epsilon)\right) \leq C_2\phi(\epsilon) < \infty$,

while the joint distribution of (X, Y) has to satisfy :

(H2) $\exists C_3 < \infty,\ \exists b > 0, \exists \eta > 0, \forall x, x' \in S_{\mathcal{F}}^{\eta},\ ||m(x) - m(x')|| \leq C_3 d^b(x, x')$,

(H3) $\exists C_4, \forall r \geq 1,\ E(||Y||^r|X) < C_4 r! < \infty$, a.s.,

where $r! = r(r-1)\ldots(r-[r]+1)$, $[r]$ being the largest integer smaller than r. We also need the following technical conditions on the kernel K :

(H4) The kernel function has to be such that :

(i) K is a bounded and Lipschitz continuous function with support $[0, 1)$,

and if $K(1) = 0$ it has also to fulfill, jointly with $\phi(.)$, the conditions :

(ii) $\exists (C_5, C_6),\ -\infty < C_5 < K'(t) < C_6 < 0$;

(iii) $\exists C_7 > 0, \exists \eta_0 > 0,\ \forall \eta < \eta_0,\ \int_0^{\eta} \phi(u)\, du > C_7\, \eta\, \phi(\eta)$.

While (H1)-(H4) are standard conditions to get pointwise rates of convergence, the next assumptions are directly linked with our wishes to have uniform rates over the set $S_{\mathcal{F}}$. These conditions will make appear the topological complexity of the set $S_{\mathcal{F}}$ which will act through the Kolmogorov's entropy of $S_{\mathcal{F}}$ defined for any $\epsilon > 0$ by :

$$\psi_{S_{\mathcal{F}}}(\epsilon) = \log(N_\epsilon(S_{\mathcal{F}}))$$

where $N_\epsilon(S_{\mathcal{F}})$ is the minimal number of open balls in $\mathcal{F}$ of radius ϵ which are necessary to cover $S_{\mathcal{F}}$. Of course these conditions will also cross restrictions on the small ball probability function ϕ introduced in (H1). Assume that :

(H5) $\exists C_8 > 0, \exists \eta_0 > 0,\ \forall \eta < \eta_0,\ 0 \leq \phi'(\eta) < C_8$,

(H6) $\exists n_0, \forall n > n_0,\ \dfrac{(\log n)^2}{n\phi(h)} < \psi_{S_{\mathcal{F}}}\left(\dfrac{\log n}{n\phi(h)}\right) < \dfrac{n\,\phi(h)}{\log n}$,

(H7) $\exists \beta > 1,\ \sum_{n=1}^{\infty} \exp\left\{(1-\beta)\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)\right\} < \infty$.

It is important to stress that, despite of its rather intricate form, this set of assumptions is not too much restrictive. Excepted (H3), all other conditions are related with the explanatory variable X and they have been discussed in various previous papers. The reader can look for instance at Chapter 13 in Ferraty and Vieu (2006) to see how all these conditions can be shown to be true pending to suitable topological structure on the space $\mathcal{F}$ (that is, pending to suitable choice of the semi-metric d). It is out of purpose to provide detailled discussion in this paper, because this is definitively not linked with the functional nature of the response (which is the main point that we wish to address) and also because such a discussion appears already in various other papers. Finally, the only condition which is specific to the functional response Y is the rather unrestrictive conditional moment existency assumed in (H3).

5.3.3 Uniform rates of convergence

The following theorem states the rate of convergence of $\widehat{m}$, uniformly over the set $S_{\mathcal{F}}$. The asymptotics are stated in terms of almost complete convergence (denoted by *a.co.*) which is known to imply both weak and strong convergences (see, among other, Section A-1 in Ferraty and Vieu, 2006). The topological structure on the space $\mathcal{F}$ acts directly on these rates through the functions ϕ and $\psi_{S_{\mathcal{F}}}$, while the topological complexity of the space $\mathcal{B}$ will act through its type p.. As discussed before, this is the first result of this kind in regression setting when both X and Y are functional. In the special simpler case when Y is real this result was given in Ferraty *et al.* (2010). To fix the ideas and to highlight the wide generality of the apparently highly technical assumptions (H1)-(H7) a few special cases will be considered later in Remark 2, while Remark 3 will present an interesting direct consequence of this general result.

Theorem 5 *Under the hypotheses (H1)-(H7), we have*

$$\begin{aligned}\sup_{x\in S_{\mathcal{F}}} \|\widehat{m}(x)-m(x)\| &= O(h^b)+O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h)}}\right)\\ &+ O_{a.co.}\left(\frac{1}{n\phi(h)}\right)^{1-\frac{1}{p}}. \end{aligned} \tag{5.2}$$

Proof. We consider the decomposition

$$\widehat{m}(x)-m(x)=\frac{(\widehat{g}(x)-E\widehat{g}(x))}{\widehat{f}(x)}+\frac{(E\widehat{g}(x)-m(x))}{\widehat{f}(x)}+\frac{(1-\widehat{f}(x))m(x)}{\widehat{f}(x)} \tag{5.3}$$

where

$$\widehat{f}(x)=\frac{\sum_{i=1}^n K(h^{-1}d(x,X_i))}{nE[K(h^{-1}d(x,X_1))]} \quad\text{and}\quad \widehat{g}(x)=\frac{\sum_{i=1}^n K(h^{-1}d(x,X_i))Y_i}{nE[K(h^{-1}d(x,X_1))]}. \tag{5.4}$$

The denominator $\widehat{f}(x)$ does not involve the functional response, and therefore the following results stated in Ferraty *et al.* (2010) remain true :

$$\sup_{x\in S_{\mathcal{F}}}|\widehat{f}(x)-1|=O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h)}}\right)\quad\text{and}\quad\sum_{n=1}^{\infty}P\left(\inf_{x\in S_{\mathcal{F}}}\widehat{f}(x)<\frac{1}{2}\right)<\infty.$$

So finally, Theorem 5 will be true as long as both following lemmas will be proved. The proofs of these lemmas are reported to the Appendix.

Lemma 9 *Under the hypotheses (H1),(H2) and (H4)-(H7), we have*

$$\sup_{x\in S_{\mathcal{F}}}\|E\widehat{g}(x)-m(x)\|=O(h^b).$$

Lemma 10 *Under the assumptions of Theorem 5, we have*

$$\sup_{x\in S_{\mathcal{F}}}\|\widehat{g}(x)-E\widehat{g}(x)\|=O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h)}}\right)+O_{a.co.}\left(\frac{1}{n\phi(h)}\right)^{1-1/p}.$$

□

Remark 2 *For any Hilbert space, it is clear that $p = 2$ (see, Remark 1) so the rate (5.2) becomes :*

$$\sup_{x \in S_{\mathcal{F}}} ||\widehat{m}(x) - m(x)|| = O(h^b) + O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h)}}\right).$$

In the special case when $\mathcal{B}$ is the euclidian space $\mathbb{R}^k$, when $S_{\mathcal{F}}$ is compact and when X has a density with respect to the Lebesgue measure, then (5.2) becomes the usual multivariate nonparametric rate :

$$\sup_{x \in S_{\mathcal{F}}} ||\widehat{m}(x) - m(x)|| = O(h^b) + O_{a.co.}\left(\sqrt{\frac{\log n}{nh^k}}\right).$$

Remark 3 *Uniform consistency allows to replace a fixed x with a random element X. Indeed, as soon as $P(X \in S_{\mathcal{F}}) = 1$, one gets*

$$||\widehat{m}(X) - m(X)|| \leq \sup_{x \in S_{\mathcal{F}}} ||\widehat{m}(x) - m(x)||.$$

So, results on $\sup_{x \in S_{\mathcal{F}}} ||\widehat{m}(x) - m(x)||$ remain valid for $||\widehat{m}(X) - m(X)||$. This can be of great practical interest for other problems (automatic bandwidth choice, semi-parametric modelling, ...) which are out of the scope here.

5.4 Proof of technical lemmas

In the following, we will denote $K_i(x) = K(h^{-1}d(x, X_i))$. First of all, one has

$$\exists 0 < C_9 \leq C_{10} < \infty, \forall x \in S_{\mathcal{F}},\ C_9\phi(h) < E[K_1(x)] < C_{10}\phi(h). \tag{5.5}$$

The result (5.5) is obvious when $K(1) > 0$ and can be extended to continuous kernel K satisfying (H4) as shown in Lemma 4.4, page 44, in Ferraty and Vieu (2006). From now on, we will denote by C is a generic nonnegative real constant, and we will take

$$\epsilon = \frac{\log n}{n}.$$

Note that condition (H5) implies that for n large enough :

$$0 \leq \phi(h) \leq Ch,$$

in such a way that (H6) implies both that

$$\lim_{n\to\infty} \frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h)} = 0 \text{ and } \lim_{n\to\infty} \frac{\log n}{nh} = 0. \tag{5.6}$$

Proof of Lemma 9. One has

$$\begin{aligned} ||E\widehat{g}(x) - m(x)|| &= \frac{1}{E[K_1(x)]} \left|\left| E\frac{1}{n}\left[\sum_{i=1}^{n} K_i(x)Y_i\right] - m(x)E[K_1(x)]\right|\right| \\ &= \frac{1}{E[K_1(x)]} \left|\left| E[K_1(x)Y_1] - m(x)E[K_1(x)]\right|\right|. \end{aligned}$$

Hence, we get

$$\forall x \in S_{\mathcal{F}}, \qquad ||E\widehat{g}(x) - m(x)|| = \frac{1}{E[K_1(x)]}[EK_1(x)(||m(X_1) - m(x)||)].$$

Thus, with hypotheses (H1), (H2) and (5.5) we have, for n large enough :

$$\forall x \in S_{\mathcal{F}}, \quad ||E\widehat{g}(x) - m(x)|| \leq C\frac{1}{E[K_1(x)]}[EK_1(x)\mathbb{1}_{B(x,h)}(X_1)d^b(X_1, x)] \leq Ch^b.$$

This last inequality yields the proof, since C does not depend on x.

$\square$

Proof of Lemma 10. Let $x_1, \ldots x_{N_\epsilon(S_{\mathcal{F}})}$ be such that the following covering of $S_{\mathcal{F}}$ holds :

$$S_{\mathcal{F}} \subset \cup_{k=1}^{N_\epsilon(S_{\mathcal{F}})} B(x_k, \epsilon),$$

and define

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

One considers now the following decomposition

$$\sup_{x\in S_{\mathcal{F}}} ||\widehat{g}(x) - E\widehat{g}(x)|| \leq \underbrace{\sup_{x\in S_{\mathcal{F}}} ||\widehat{g}(x) - \widehat{g}(x_{k(x)})||}_{G_1} + \underbrace{\sup_{x\in S_{\mathcal{F}}} ||\widehat{g}(x_{k(x)}) - E\widehat{g}(x_{k(x)})||}_{G_2}$$

$$+ \underbrace{\sup_{x\in S_{\mathcal{F}}} ||E\widehat{g}(x_{k(x)}) - E\widehat{g}(x)||}_{G_3}.$$

We will first deal with the terms G_1 and G_3 which are the simplest ones in the sense that they are not linked with the functional nature of Y and so one can make use of previous literature for scalar response Y to treat them. The term G_2 will need more specific attention.

i) *Study of the term* G_1. We get directly from (H1) and (5.5) :

$$\begin{aligned} G_1 &= \sup_{x\in S_{\mathcal{F}}} \left|\left|\sum_{i=1}^{n}\left(\frac{1}{nE[K_1(x)]}K_i(x)Y_i - \frac{1}{nE[K_1(x_{k(x)})]}K_i(x_{k(x)})Y_i\right)\right|\right| \\ &\leq \sup_{x\in S_{\mathcal{F}}} \frac{1}{n\,\phi(h)}\sum_{i=1}^{n} ||Y_i||\,\left|\left|K_i(x) - K_i(x_{k(x)})\right|\right|\,\mathbb{1}_{B(x,h)\cup B(x_{k(x)},h)}(X_i). \end{aligned}$$

In a first attempt, assume that $K(1) = 0$ (i.e. K is Lipschitz on $[0, 1]$) in order to get :

$$G_1 \;\leq\; \frac{C}{n}\sum_{i=1}^{n} Z_i \text{ with } Z_i = \frac{\epsilon\,\|Y_i\|}{h\,\phi(h)}\mathbb{1}_{B(x,h)\cup B(x_{k(x)},h)}(X_i).$$

Clearly, we get from (H3) :

$$E[||Y||^m] = E\left[E[||Y||^m|X]\right] < Cm! < \infty,$$

which implies that

$$E(|Z_1|^m) \;\leq\; \frac{Cm!\epsilon^m}{h^m\,\phi(h)^{m-1}}. \tag{5.7}$$

Moreover, by using the second result in (5.6) together with the definition of ϵ we have for n large enough :

$$\frac{\epsilon}{h} \leq C.$$

Both last results yeld directly to

$$E(|Z_1|^m) \;\leq\; \frac{Cm!\epsilon^{m-1}}{h^{m-1}\,\phi(h)^{m-1}}.$$

So, by applying Corollary A.8 in Ferraty and Vieu (2006) with $a^2 = \frac{\epsilon}{h\,\phi(h)}$, one gets

$$\frac{1}{n}\sum_{i=1}^{n} Z_i = EZ_1 + \; O_{a.co.}\left(\sqrt{\frac{\epsilon\,\log n}{n\,h\,\phi(h)}}\right).$$

Finally, applying again (5.7) for $m = 1$ one gets

$$G_1 = O_{a.co.}\left(\sqrt{\frac{\epsilon\,\log n}{n\,h\,\phi(h)}}\right) + O(\frac{\epsilon}{h}).$$

Now, using (H6) together with the second part of (5.6) and with the definition of ϵ, we get :

$$G_1 \;=\; O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\,\phi(h)}}\right). \tag{5.8}$$

The proof of (5.8) for the case $K(1) > C > 0$ (i.e. K Lipschitz on $[0, 1)$) is not presented here since one can proceed exactly as in Lemma 6 in Ferraty *et al.* (2010) by splitting again G_1 into three terms.

ii) *Study of the term G_3.* By definition of G_3 we have :

$$\begin{aligned} G_3 \;&\leq\; \sup_{x\in S_{\mathcal{F}}} E\left[||\widehat{g}(x_{k(x)}) - \widehat{g}(x)||\right] \\ &\leq\; E\left[\sup_{x\in S_{\mathcal{F}}} ||\widehat{g}(x_{k(x)}) - \widehat{g}(x)||\right], \end{aligned}$$

the first inequality coming from the contractive property of the expectation operator (see Bosq, 2000, page 29). So we have finally $G_3 \leq EG_1$ which, combined with (5.8), leads directly to

$$G_3 = O\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h)}}\right). \tag{5.9}$$

iii) *Study of the term G_2.* This part is the most technical because it involves directly the functional response Y. This is the main specificity of our work and so G_2 cannot be treated by the same techniques as if Y was real (as for G_1 and G_3). The proof will use the exponential inequality for Banach space valued random variables (see, Lemma 8). Let :

$$\Gamma_{ki} = \frac{1}{nE[K_1(x_{k(x)})]}\left[K_i(x_{k(x)})Y_i - E\left[K_i(x_{k(x)})Y_i\right]\right],$$

and

$$W_n = \|\sum_i^n \Gamma_{ki}\| - E\left(\|\sum_i^n \Gamma_{ki}\|\right).$$

It is clear that, $\forall \eta > 0$,

$$\begin{aligned}
P\left(G_2 > \eta\right) = P&\left(\max_{k\in\{1,\cdots,N_\epsilon(S_{\mathcal{F}})\}}\|\widehat{g}(x_k) - E\widehat{g}(x_k)\| > \eta\right)\\
&\leq N_\epsilon(S_{\mathcal{F}})\max_{k\in\{1,\cdots,N_\epsilon(S_{\mathcal{F}})\}} P\left(\|\widehat{g}(x_k) - E\widehat{g}(x_k)\| > \eta\right)\\
\leq N_\epsilon(S_{\mathcal{F}})&\max_{k\in\{1,\cdots,N_\epsilon(S_{\mathcal{F}})\}} P\left(\|\sum_i^n \Gamma_{ki}\| > \eta\right)\\
\leq N_\epsilon(S_{\mathcal{F}})&\max_{k\in\{1,\cdots,N_\epsilon(S_{\mathcal{F}})\}} P\left(|W_n| > \eta - E\left(\|\sum_i^n \Gamma_{ki}\|\right)\right).
\end{aligned}$$

Choosing now

$$\eta = \eta_0\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h)}} + E\left(\|\sum_i^n \Gamma_{ki}\|\right), \tag{5.10}$$

we have

$$P\left(G_2 > \eta\right) \leq N_\epsilon(S_{\mathcal{F}})\max_{k\in\{1,\cdots,N_\epsilon(S_{\mathcal{F}})\}} P\left(|W_n| > \eta_0\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h)}}\right). \tag{5.11}$$

To apply the inequality of Lemma 8, one must evaluate the quantities

$$\sum_{i=1}^n E\|\Gamma_{ki}\|^m \quad\text{and}\quad E(\|\sum_{i=1}^n \Gamma_{ki}\|). \tag{5.12}$$

Using the condition (H3) we have for all $j \leq m$:

$$\begin{aligned}
E\left\|Y_1K_1(x)\right\|^j &= E\left[K_1^j(x)E\left[\|Y_1\|^j|X_1\right]\right]\\
&\leq Cj!E[K_1^j(x)],
\end{aligned}$$

leading finally, by using the result (5.5) and the boundedness of K (see (H4)), to

$$E\|Y_1K_1(x)\|^j \leq Cj!\phi(h). \tag{5.13}$$

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

$$\begin{aligned}\|\Gamma_{ki}\|^m &\leq \frac{1}{n^m(E[K_1(x)])^m}\left(\|K_i(x)Y_i\| + \|E[K_i(x)Y_i]\|\right)^m \\ &= \frac{1}{n^m(E[K_1(x)])^m}\sum_{k=0}^{m} C_{k,m}\|Y_1K_1(x)\|^k \|E[Y_1K_1(x)]\|^{m-k},\end{aligned}$$

where $C_{k,m} = \frac{m!}{k!(m-k)!}$. We get by (5.13) :

$$\begin{aligned}E\|\Gamma_{ki}\|^m &\leq \tfrac{1}{n^m(E[K_1(x)])^m}\textstyle\sum_{k=0}^{m} C_{k,m}E\|Y_1K_1(x)\|^k \|E[Y_1K_1(x)]\|^{m-k} \\ &\leq Cn^{-m}m!\phi(h)^{-m+1}.\end{aligned}$$

Note that we have used the fact that

$$\forall j \leq m, \qquad C\phi(h) \leq E[K_1^j(x)].$$

which is again a consequence of the boundednes of K (see (H4)) and of the result (5.5). It follows that for all $m \geq 1$,

$$\sum_{i=1}^{n} E\|\Gamma_{ki}\|^m \leq Cm!(n\phi(h))^{-m+1}. \tag{5.14}$$

On the other hand, we use the C_r inequality (see, Loève, 1963, p.155) for the second quantity of (5.12) and we have

$$E\|Y_1K_1(x) - E[Y_1K_1(x)]\|^p \leq 2^{p-1}\left[E\|Y_1K_1(x)\|^p + [\|E[Y_1K_1(x)]\|]^p\right].$$

Noting that (5.13) can be obtained by the same route for non necessary integer power (that is changing j into p) one gets

$$n^p\phi^{p-1}(h)E\|\Gamma_{k1}\|^p \leq C.$$

Now, we use that $\mathcal{B}$ is a space of type $p \in]1,2]$ and we have

$$\left(E\left(\|\sum_{i}^{n}\Gamma_{ki}\|\right)\right)^p \leq E\left(\|\sum_{i}^{n}\Gamma_{ki}\|\right)^p \leq C\sum_{i=1}^{n}E\|\Gamma_{ki}\|^p,$$

and hence

$$E\left(\|\sum_{i}^{n}\Gamma_{ki}\|\right)=O((n\phi(h))^{-1+1/p}). \tag{5.15}$$

Because of (5.14), we are now in position for applying Lemma 8, by taking

$$l^2=b=\frac{C}{n\phi(h)} \text{ and } t=\eta_0\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h)}},$$

and we arrive at

$$\begin{aligned} P\left(\left|\|\sum_{i}^{n}\Gamma_{ki}\|-E\left(\|\sum_{i}^{n}\Gamma_{ki}\|\right)\right|>\eta_0\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h)}}\right) &\leq 2\exp\left(-\frac{t^2}{2l^2+2bt}\right)\\ &\leq 2\exp\left(-C\frac{t^2}{2l^2}\right)\\ &\leq 2N_\epsilon(S_{\mathcal{F}})^{-C\eta_0^2}, \end{aligned}$$

the second inequality coming from the first result in (5.6). Therefore, by using (5.15), (5.11) and (5.10), we have :

$$P\left(G_2>\eta_0\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\,\phi(h)}}+\left(\frac{1}{n\phi(h)}\right)^{1-1/p}\right)\leq CN_\epsilon(S_{\mathcal{F}})^{1-\beta}.$$

Because of (H7) there exists some $\beta>1$ such that

$$\sum_{n=1}^{\infty}N_\epsilon(S_{\mathcal{F}})^{1-\beta}<\infty,$$

and we obtain that

$$G_2=O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\,\phi(h)}}\right)+O_{a.co.}\left(\frac{1}{n\phi(h)}\right)^{1-1/p}. \tag{5.16}$$

Now, Lemma 10 can be easily deduced from (5.8), (5.9) and (5.16).

□

Acknowledgements

The authors thank the participants in the working group STAPH in Toulouse (*http ://www.math.univ-toulouse.fr/staph*) for continuous support. Special thanks go also to the Editor, an Associate Editor and two referees for the speed and the quality of the reviewing of a previous draft of our work. The strong relevance of their comments had a real impact on the final quality of this paper.

Bibliographie

Barrientos-Marin, J., Ferraty, F. and Vieu, P. (2010). Locally modelled regression and functional data. *J. Nonparametr. Stat.*, **22** 617-632.

Bosq, D. (2000). *Linear processs in function spaces. Theory and Application. Lectures Notes in Statistics*, Vol 149. Springer-Verlag.

Bosq, D. and Delecroix, M. (1985). Nonparametric prediction of a Hilbert-space valued random variable. *Stochastic Process. Appl.* **19** 271-280.

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

Ferraty, F. and Romain, Y. (2011). *The Oxford Handbook of Functional Data Analysis.* Oxford University Press.

Ferraty, F. and Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Comput. Statist.*, **17** 545-564.

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis.*

Theory and Practice. Springer-Verlag.

FERRATY, F. and VIEU, P. (2011). Kernel regression estimation for functional data. In *The Oxford Handbook of Functional Data Analysis (Ed. F. Ferraty and Y. Romain)*. Oxford University Press.

HE, G., MÜLLER, H.G., WANG, J.L., YANG, W. (2009). Functional Linear Regression Via Canonical Analysis. *Bernoulli*. **16**, 705-729.

HLUBINKA, D., PRCHAL, L. (2007). Changes in atmospheric radiation from the statistical point of view. *Comput. Statist. Data Anal.*. **51**,(10), 4926-4941.

LECOUTRE, J. P. (1990). Uniform consistency of a class of regression function estimators for Banach-space valued random variable. *Statist. Probab. Lett.* **10** 145-149.

LOÈVE, M. (1963). *Probability Theory*, 3rd ed. Van Nostranr Princeton.

MÜLLER, H.G., CHIOU, J.M., LENG, X. (2008). Infering gene expression dynamics via functional regression analysis. *BMC Bioinformatics*. **9**, 60.

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New-York.

Troisième partie

Cas dépendant

Chapitre 6

Convergence ponctuelle : réponse fonctionnelle

Ce chapitre fait l'objet d'un projet de la note CRAS.

Estimation de la fonction de régression pour variables explicative et réponse fonctionnelles dépendantes

Frédéric Ferraty[a], Ali Laksaci[b], Amel Tadj[c,*] and Philippe Vieu[a]
[a]Université Paul Sabatier, Toulouse,
[b]Université Djillali Liabes, Sidi Bel Abbes
[c,*]Corresponding author : Université de Sidi Bel Abbès, BP 89 Sidi Bel Abbès 22000. Algérie . ameltdz@yahoo.fr

Résumé Nous considérons l'estimateur à noyau de la fonction de régression lorsque la variable réponse est dans un espace de Banach et la variable explicative est à valeurs dans un espace semi-métrique. En considérant des observations β-mélangeantes, on établit la vitesse de convergence presque complète de l'estimateur construit.

Abstract We consider a kernel estimate of the regression when the response variable is in a Banach space and the explanatory variable takes its values in a semi-metric space. Our main result states the almost complete convergence (with rate) of the constructed estimate when the sample considered is a β-mixing sequence.

6.1 Introduction

Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathcal{F} \times \mathcal{B}$, où $\mathcal{B}$ est un espace de Banach de type $\tau \in]1, 2]$ et $\mathcal{F}$ est un espace semi-métrique. On note $\|.\|$ la norme sur $\mathcal{B}$ et on note $d(., .)$ la semi métrique sur $\mathcal{F}$. On considère une suite de n observations de même loi que le couple (X, Y). Pour $x \in \mathcal{F}$, on suppose que l'espérance conditionnelle de Y sachant $X = x$, notée $m(x) = E[Y \mid X = x]$, existe. L'estimateur naturel de $m(x)$ par la méthode du noyau est

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} K(h^{-1}d(x, X_i))Y_i}{\sum_{i=1}^{n} K(h^{-1}d(x, X_i))}, \ \forall x \in \mathcal{F}, \tag{6.1}$$

où $K(.)$ est un noyau et $h := h_n$ est une suite de nombres réels positifs. Le but de cette Note est d'étudier l'estimateur non-paramétrique $\widehat{m}(x)$ de $m(x)$ lorsque les deux variables X et Y sont à valeurs dans un espace de dimension infinie et les observations $(Y_i, X_i)_{i=1,\cdots,n}$ sont dépendantes. La modélisation statistique des données fonctionnelles dépendantes est motivée par un grand nombre de problèmes concrets liés à l'étude de phénomènes temporels (voir, pour des exemples, Ramsay et Silverman (2002), Bosq (2000) pour les modèles paramétriques et Ferraty et Vieu (2006) dans le cas non-paramétrique). Typiquement, les séries chronologiques fonctionnelles sont issues d'un découpage d'un processus à temps continu sur des intervalles. Ceci nous permet d'étudier le processus en tenant compte des morceaux de trajectoires continus. D'une manière générale, les premiers résultats sur l'estimation non-paramétrique de la fonction de régression à variable réponse fonctionnelle ont été obtenus par Bosq et Delecroix (1985). En 1990, Lecoutre a étudié la convergence uniforme presque complète d'un estimateur de type régressogramme de la fonction de régression lorsque la variable réponse est dans un espace de Banach et la variable explicative est réelle. Récemment, Ferraty *et al.* (2011) ont obtenu la convergence uniforme presque complète d'un estimateur à noyau d'une variable réponse Banachique conditionnée par une variable explicative dans un espace semi-métrique lorsque les observations sont indépendantes et identiquement distribuées.

L'apport principal de cette Note est de généraliser, au cas dépendant, les résultats de Ferraty *et al.* (2011). Plus précisément, on établit la vitesse de convergence presque complète de $\hat{m}$ lorsque les observations sont β-mélangeantes. L'expression de la vitesse de convergence exploite la structure topologique de la variable explicative et de la variable réponse, ainsi que la structure de dépendance des observations.

Nos hypothèses sont présentées au paragraphe 6.2 alors que le résultat principal est donné au paragraphe 6.3.

6.2 Notations et Hypothèses

Nous commençons par rappeler la notion de β-mélange pour une suite de variables aléatoires.

Définition 8 *Soit $\{Z_i,\ i = 1, 2, ...\}$ une suite stationnaire de variables aléatoires. Pour tout entier n, on note $\beta(n)$ le coefficient de mélange défini par*

$$\beta(n) = \sup_k \Bigg\{ \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} |P(A_i \cap B_j) - P(A_i)P(B_j)| :$$

$$A_i \in \mathcal{F}_1^k(Z) \ \text{et} \ B_j \in \mathcal{F}_{k+n}^{\infty}(Z),\ I,\ J \subset \mathbb{N}^* \ \ \text{et} \ k \in \mathbb{N}^* \Bigg\},$$

où $\mathcal{F}_i^k(Z)$ est la tribu engendrée par $\{Z_j, i \leq j \leq k\}$. Nous disons que la suite (Z_i) est β-mélangeante si le coefficient de mélange $\beta(n) \to 0$ lorsque $n \to \infty$.

On trouvera dans la littérature plusieurs exemples de processus stochastiques vérifiant cette condition de mélange. Par exemple, dans notre contexte fonctionnel, Allam et Mourid (2002) ont donné les conditions suffisantes pour qu'un processus autorégressif Banachique soit β-mélangeant. Par la suite, on note par $B(x, h) = \{x' \in \mathcal{F}/d(x', x) < h\}$ la boule de centre x de rayon h et on introduit les hypothèses suivantes :

(H1) $P(X \in B(x, r)) = \phi_x(r) > 0$ pour tout $r > 0$,

(H2) $\exists b > 0, \exists C > 0\, \exists \mu > 0, \forall x_1, x_2 \in B(x, \mu),\ \|m(x_1) - m(x_2)\| \leq C d^b(x_1, x_2)$,

(H3) $(X_i, Y_i)_{i\geq 1}$ est une suite β-mélangante dont le coefficient de mélange vérifie

$$\exists a > 2,\ \beta(n) = O\left(n^{-a}\right),$$

(H4) $\exists C > 0, \forall r > 1,\ E(\|Y\|^r|X) \leq C\, r! < \infty$ p.s., avec $r! = r(r-1)\cdots(r-[r]+1)$ et où $[.]$ désigne la partie entière,

(H5) K est une fonction de support $[0,1[$ vérifiant $0 < C\mathbf{1}_{[0,1[} < K(t) < C'\mathbf{1}_{[0,1[} < \infty$ et si $K(1)=0$, la fonction $\phi_x(.)$ vérifie la condition supplémentaire

$$\exists C > 0,\ \exists \eta_0 > 0,\ \forall 0 < \eta < \eta_0,\ \int_0^\eta \phi_x(u)\, du \ >\ C\,\eta\,\phi_x(\eta),$$

(H6) $\exists \delta \in (2/a, 1),\ \lim_{n\to\infty}(n^{1-\delta}\,\phi_x(h)/\log n) = \infty$.

6.3 La convergence presque complète

On établit le résultat suivant

Théorème 7 *Sous les conditions (H1)-(H6), on a*

$$\|\widehat{m}(x) - m(x)\| = O(h^b) + O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\delta}\phi_x(h)}}\right) + O\left(\frac{1}{n^{1-\delta}\phi_x(h)}\right)^{1-1/\tau}.$$

Schéma de la preuve. La démonstration de ce Théorème est basée sur la décomposition suivante :

$$\widehat{m}(x)-m(x) = \frac{1}{\widehat{f}(x)}\Big[\widehat{g}(x)-E\widehat{g}(x)\Big]+\frac{1}{\widehat{f}(x)}\Big[E\widehat{g}(x)-m(x)\Big]+\Big[1-\widehat{f}(x)\Big]\,\frac{m(x)}{\widehat{f}(x)},$$

où

$$\widehat{f}(x) = \frac{1}{nE[K(h^{-1}d(x,X_1))]}\sum_{i=1}^n K(h^{-1}d(x,X_i))$$

et

$$\widehat{g}(x) = \frac{1}{nE[K(h^{-1}d(x,X_1))]}\sum_{i=1}^n K(h^{-1}d(x,X_i))Y_i.$$

Le Théorème (3.1) est une conséquence des résultats préliminaires suivants :

Lemme 1 *Sous les hypothèses (H1),(H2), (H5) et (H6), on a*

$$\|E\widehat{g}(x) - m(x)\| = O(h^b).$$

Lemme 2 *Sous les hypothèses (H1) et (H3)-(H6), on a*

$$\|\widehat{g}(x) - E\widehat{g}(x)\| = O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\delta}\phi_x(h)}}\right) + O\left(\frac{1}{n^{1-\delta}\phi_x(h)}\right)^{1-1/\tau}. \tag{6.2}$$

Lemme 3 *Sous les hypothèses (H1), (H3), (H5) et (H6), on a*

$$\widehat{f}(x) - 1 = O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\delta}\phi_x(h)}}\right). \tag{6.3}$$

Remarque 3 *Une des idées essentielles de la preuve consiste à utiliser une inégalité exponentielle pour variables aléatoires Banachiques mélangeantes due à Rhomari (2002). Cette inégalité est la suivante :*

Soit $Z_1, \ldots, Z_n$ une suite β-mélangeante où les $(Z_i)_{i=1,\cdots,n}$ sont centrées. Soit j et p des entiers tels que $1 \leq j \leq p \leq n$. Si pour $m \geq 2$,
$\max_j \sum_{i=1}^q E||Z_{j+(i-1)p}||^m \leq qm!\sigma^2 M^{m-2/2}$, alors pour tout $\epsilon > pe^*(q)$, on a

$$P(||\sum_{l=1}^n Z_l|| \geq \epsilon) \leq p \exp\left(\frac{(\epsilon - pe^*(q))^2}{2p[(n+p)C + (\epsilon - pe^*(q))M]}\right) + n\beta(p),$$

où C est une constante positive, M est un nombre positif, $e^*(q) = \max_j e_j^*(q)$ et $e_j^*(q) = E(\sum_{i=1}^q E||Z_{j,i}^*||)$, où $(Z_{j,i}^*)_{i=1,\cdots,q}$ sont des v.a.i.i.d. ayant même lois que $(Z_{j+(i-1)p})_{i=1,\cdots,q}$.

Références

Allam, A., Mourid, T. (2002). Geometric absolute regularity of Banach space-valued autoregressive processes. Statist. Probab. Lett. **60**, 241-252.

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

BOSQ, D., DELECROIX, M. (1985). Nonparametric prediction of a Hilbert-space valued random variable. Stochastic Process. Appl. **19** , 271-280.

FERRATY, F., LAKSACI, A., TADJ, A. VIEU, P. (2011). Kernel regression whith functional response. *Electronic Journal of Statistics.* **5**, 159-171.

FERRATY, F., VIEU, P. (2006). Nonparametric functional data analysis. Theory and Practice. Springer-Verlag.

LECOUTRE, J. (1990). Uniform consistency of a class of regression function estimators for Banach-space valued random variable. Statist. Probab. Lett. **10**, 145-149

RAMSAY, J.O., SILVERMAN, B.W. (2002). Applied functional data analysis, Methods and case studies. Springer-Verlag, New York.

RHOMARI, N. (2002). Approximation et inégalités exponentielles pour les sommes de vecteurs aléatoires dépendants. C. R. Acad. Sci. Paris. **334**, 149-154.

Chapitre 7

Preuves détaillées du résultat du chapitre précédent

Pour tout $i = 1, \ldots, n$, on pose : $K_i(x) = K(h^{-1}d(x, X_i))$. Dans un premier temps, on remarque que les conditions (H1) et (H5) permettent d'écrire pour tout réel $j > 0$

$$\exists 0 < C < C' < \infty,\ C\phi_x(h) < E[K_1^j(x)] < C'\phi_x(h). \tag{7.1}$$

Voir Ferraty et Vieu, 2006, p44, Lemme 4.4, pour la preuve de ce résultat.
Preuve du Lemme 1.

$$\begin{aligned}
\|E\widehat{g}(x) - m(x)\| &= \frac{1}{E[K_1(x)]}\left\|E\frac{1}{n}\left[\sum_{i=1}^{n} K_i(x)Y_i\right] - m(x)E[K_1(x)]\right\| \\
&= \frac{1}{E[K_1(x)]}\left\|E[K_1(x)Y_1] - m(x)E[K_1(x)]\right\|
\end{aligned}$$

Ce qui implique

$$\|E\widehat{g}(x) - m(x)\| \leq \frac{1}{E[K_1(x)]}[EK_1(x)\,\|m(X_1) - m(x)\|].$$

Sous les hypothèses (H1), (H2) et (H5), on obtient :

$$\|E\widehat{g}(x) - m(x)\| \leq C\frac{1}{E[K_1(x)]}[EK_1(x)\mathbf{1}_{B(x,h)}(X_1)d^b(X_1, x)] \leq Ch^b.$$

$\square$

Dans toute la suite, $(p_n)_n$ et $(q_n)_n$ désignent deux suites d'entiers positifs.

Preuve du Lemme 2. On pose, pour $i = 1, \ldots, n$:

$$\Gamma_i = \frac{1}{E[K_1(x))]} \left[K_i(x)Y_i - E\left[K_i(x)Y_i\right]\right].$$

Alors, on a

$$\widehat{g}(x) - E[\widehat{g}(x)] = \frac{1}{n}\sum_{i=1}^{n}\Gamma_i.$$

Pour appliquer l'inégalité de type exponentiel (voir Rhomari, 2002) on doit calculer asymptotiquement $\sum_{i=1}^{q_n} E\|\Gamma_i\|^m$ pour tout entier $m \geq 2$ ainsi que $e^*(q_n) = \max_i E\,\|\sum_{i=1}^{q_n}\Gamma_i^*\|$, où $(\Gamma_i^*)_i$ est une suite de variables indépendantes de même loi que Γ_1 et q_n est une suite de nombres positifs, définie par

$$q_n = \left[\frac{n}{p_n}\right] + 1,$$

[.] étant la partie entière.

i) Il est clair que $E\left(\|Y_1K_1(x)\|^j\right) = E\left[K_1^j(x)E\left[\|Y_1\|^j\,|X_1\right]\right]$. En utilisant (H4) et (7.1), on a pour tout réel $j > 1$:

$$E\left(\|Y_1K_1(x)\|^j\right) \leq C\,j!\,\phi_x(h), \tag{7.2}$$

De plus, pour tout entier $m > 0$, on a

$$\begin{aligned} \|\Gamma_i\|^m &\leq \frac{1}{\left(E[K_1(x)]\right)^m}\left(\|K_i(x)Y_i\| + \|E[K_i(x)Y_i]\|\right)^m \\ &\leq \frac{1}{\left(E[K_1(x)]\right)^m}\sum_{k=0}^{m} C_{k,m}\,\|Y_1K_1(x)\|^k\,\|E\left[Y_1K_1(x)\right]\|^{m-k}, \end{aligned} \tag{7.3}$$

où $C_{k,m} = m!/(k!(m-k)!)$. Par ailleurs, il est facile de voir que

$$\begin{aligned} \left(\|E(Y_1\,K_1(x))\|\right)^{m-k} &\leq \left(E\,K_1(x)E\left[\|Y_1\|\,|X_1\right]\right)^{m-k} \\ &\leq C\,\left(\phi_x(h)\right)^{m-k}. \end{aligned} \tag{7.4}$$

La combinaison de (7.2)-(7.4) permet d'obtenir pour tout entier $m > 0$:

$$E\|\Gamma_i\|^m \ \leq \ C\, m!\, (\phi_x(h))^{-m+1}$$

Il vient finalement pour tout entier $m > 0$:

$$\sum_{i=1}^{q_n} E\,\|\Gamma_i\|^m \leq C q_n (\phi_x(h))^{-m+1}. \tag{7.5}$$

ii) Evaluons maintenant la quantité $e^*(q_n)$. On applique le fait que l'espace de Banach $\mathcal{B}$ est de type $\tau \in]1,2]$ pour calculer asymptotiquement cette quantité. En effet :

$$\left(E\left(\|\sum_{i=1}^{q_n}\Gamma_i^*\|\right)\right)^\tau \leq E\left(\|\sum_{i=1}^{q_n}\Gamma_i^*\|\right)^\tau \leq C\sum_{i=1}^{q_n} E\|\Gamma_i^*\|^\tau = C\sum_{i=1}^{q_n} E\|\Gamma_i\|^\tau. \tag{7.6}$$

Tout d'abord, on a besoin d'évaluer la quantité $\sum_{i=1}^{q_n} E\|\Gamma_i\|^\tau$ pour tout $\tau \in]1, 2]$. Pour cela, on applique l'inégalité C_r (voir Loève, 1963, P.155) pour écrire :

$$E\|Y_1 K_1(x) - E[Y_1 K_1(x)]\|^\tau \leq 2^{\tau-1}\left[E\|Y_1 K_1(x)\|^\tau + [\|E[Y_1 K_1(x)]\|]^\tau\right].$$

En combinant l'équation (7.1) et (7.2) avec la dernière inégalité, on arrive à $E\|\Gamma_i\|^\tau \leq C\,\phi_x^{-\tau+1}(h)$. Il suffit de combiner ce résultat avec (7.6), pour obtenir :

$$e^*(q_n) = \max_i E\left(\|\sum_{i=1}^{q_n}\Gamma_i^*\|\right) = O(q_n^{1/\tau}(\phi_x(h))^{-1+1/\tau}).$$

Comme $q_n = O(n/p_n)$, on a

$$e^*(q_n) = O((n/p_n)^{1/\tau}(\phi_x(h))^{-1+1/\tau}). \tag{7.7}$$

Maintenant, nous sommes en mesure d'appliquer l'inégalité exponentielle de Rhomari (2002). Donc, pour tout $\eta > 0$

$$P\left(\left\|\sum_{i=1}^{n}\Gamma_i\right\| > n\eta\right) \leq p_n \exp\left(-\frac{(n\eta - p_n e^*(q_n))^2}{2p_n\left[(n+p_n)C + (n\eta - p_n e^*(q_n))M\right]}\right) + n\beta(p_n) \tag{7.8}$$

où p_n et q_n sont telles que $q_n = \left[\frac{n}{p_n}\right] + 1$. En posant $M = \phi_x(h)^{-1}$, et en choisissant

$$\eta = \eta_0 \left(\sqrt{\frac{p_n \log n}{n\phi_x(h)}} + n^{-1} p_n e^*(q_n) \right) > 0,$$

on obtient :

$$P\left(\frac{1}{n} \left\| \sum_{i=1}^{n} \Gamma_i \right\| > \eta_0 \left(\sqrt{\frac{p_n \log n}{n\phi_x(h)}} + n^{-1} p_n e^*(q_n) \right) \right)$$

$$\leq p_n \exp\left(-\frac{n\eta_0^2 \log n}{2\left[C(n+p_n)\phi_x(h) + \eta_0\sqrt{np_n \log n}\right]} \right) + n\beta(p_n).$$

Maintenant, en choisissant $p_n = O(n^\delta)$, on a

$$P\left(\frac{1}{n} \left\| \sum_{i=1}^{n} \Gamma_i \right\| > \eta_0 \left(\sqrt{\frac{p_n \log n}{n\phi_x(h)}} + n^{-1} p_n e^*(q_n) \right) \right) \leq n^{\delta - C\eta_0} + n\beta(n^\delta). \tag{7.9}$$

En utilisant (H3) et le fait que $\delta > 2/a$ (voir condition (H6)), il découle que pour η_0 choisi suffisamment grand, le terme de gauche de (7.9) est celui d'une série convergente.

□

Preuve du Lemme 3. Il suffit d'appliquer le résultat précédent avec :

$$\begin{cases} Y_i = 1 \quad \forall\, i, \\ \mathcal{B} = \mathbb{R}. \end{cases}$$

Dans ce cas là, le type de l'espace est $p = 2$ et par conséquent le deuxième terme de la vitesse de convergence de lemme 2 devient négligeable.

□

Quatrième partie

Commentaires et perspectives

Chapitre 8

Discussions et Perspectives

8.1 Sur la convergence uniforme en statistique fonctionnelle

En statistique non paramétrique, la convergence uniforme est considérée comme étape préalable pour l'obtention de résultats plus pointus. Par exemple, ce mode de convergence s'avère particulièrement utile dès qu'on souhaite valider théoriquement des procédures de choix automatique de paramètre. C'est le cas pour la sélection du paramètre de lissage (voir Rachdi et Vieu, 2007 et Benhenni *et al.* 2007) dont l'étude théorique utilise de manière déterminante les résultats de convergence uniforme. Il en va de même de l'étude de modèles fonctionnels plus spécifiques tel le modèle additif (Ferraty et Vieu, 2009), le modèle partiellement linéaire (Aneiros-Perez et Vieu, 2007) ou le modèle à indice simple (Aït saïdi *et al.*, 2008). Le choix de la semi-métrique, qui est lui aussi un point crucial, reste un problème ouvert pour lequel les résultats uniformes obtenus dans cette thèse devraient s'avérer utiles.

8.2 Sur la régression non paramétrique fonctionnelle avec réponse fonctionnelle

Les premiers résultats donnés dans cette thèse lorsque la variable réponse est aussi fonctionnelle traitent de la régression classique. Cela peut ouvrir plusieurs perspectives. Par exemple, on peut également envisager d'étendre au cas d'une variable explicative fonctionnelle les autres modèles non paramétriques conditionnels étudiés dans la première partie de cette thèse. Les choix du paramètre de lissage et de la semi-métrique sont encore des problèmes ouverts lorsque la réponse est fonctionnelle. Là aussi, les résultats uniformes de cette thèse pourront s'avérer utiles. Les test de structures dans les modèles de régression non paramétriques fonctionnels présentent un grand intérêt en statistique fonctionnelle, et le cas de variable réponse fonctionnelle reste encore non traité à ce jour.

8.3 Cas d'une réponse fonctionnelle en situation de dépendance

A notre connaissance, c'est la situation la moins développée dans la littérature et les résultats de la dernière partie de cette thèse ouvrent la porte à de nombreuses extensions, d'ordre théorique (uniformité des résultats, allégement de l'hypothèse de dépendance au cadre α-mélangeant,...) ou bien d'ordre appliqué (choix du paramètre de lissage en particulier). Concernant le cadre du α-mélange, les résultats obtenus précédemment devraient se généraliser à condition toutefois de bénéficier d'inégalités exponentielles adéquates ce qui, dans l'état actuel de nos connaissances, amènerait à considérer des variables Hilbertiennes plutôt que Banachiques.

Références

AIT SAÏDI, A., FERRATY, F., KASSA, R., VIEU, P. (2008). Cross-validated estimations in the single-functional index model. Statistics, **42**, 475-494.

ANEIROS-PEREZ, G., VIEU, P. (2007). Semi-functional partial linear regression. *Statist. & Proba. Let.*, **76**, 1102-1110.

BENHENNI, K., FERRATY, F., RACHDI, M., VIEU, P. (2007). Locally smoothing regression with functional data. Computational. Statistics. **22**, 353-370.

FERRATY, F., VIEU, P. (2009). Additive prediction and boosting for functionnal data. *Computational Statistics & Data Analysis*, **53** 1400-1413.

RACHDI, M., VIEU, P. (2007). Nonparametric regression functional data : Automatic smoothing parameter selection. J. Statist. Plan. Inf. **137**, 2784-2801.

Chapitre 9

Bibliographie générale

De Acosta, A. (1985). Inequalities for B-valued random variables with applications to the strong law of large numbers, Ann. Probab. **9**, 157-161.

Ait Saïdi, A., Ferraty, F., Kassa, R., Vieu, P. (2008). Cross-validated estimations in the single-functional index model. Statistics, **42**, 475-494.

Allam, A., Mourid, T. (2002) . Geometric absolute regularity of Banach space-valued autoregressive processes. Statist. Probab. Lett., **60**, 241-252.

Aneiros-Perez, G., Vieu, P. (2007). Semi-functional partial linear regression. *Statist. & Proba. Let.*, **76**, 1102-1110.

Antoniadis, I., Sapatinas, T. (2003). Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. J. Multivariate Anal. **87**, 133-158.

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

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

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

Bastero, J., Uriz, Z. (1986). Cotype for p-Banach spaces. Compositio Mathematica. **57**, 73-80.

Beck, A. (1958). Une loi forte des grands nombres dans des espaces de Banach uniformément convexes. Annales de l'institut Henri Poincaré. **16**, 35-45.

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

Berlinet, A., Gannoun, A., Matzner-Lober, E. (1998). Propriétés asymptotiques d'estimateurs convergents des quantiles conditionnels. C. R. Acad. Sci., Paris, Sér. I, Math. **326**, No.5, 611-614.

Besse, P. C., Cardot, H. (1996). Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. Canad. J. Statist. **24**, No.4, 467-487.

Biau, G., Cadre, B. (2004). Nonparametric Spatial Prediction. J. Statistical Inference for Stochastic Processes. **7**, 327-349.

Bogachev, V.I. (1999). *Gaussian measures.* Math surveys and monographs, **62**, Amer. Math. Soc.

Bosq, D. (1989). Propriétés des opérateurs de covariance empiriques d'un processus stationnaire hilbertien. C. R. Acad. Sci. Paris Sér. I Math. **309**, No.14, 873-875.

Bosq, D. (1990). Modèle autorégressif hilbertien. Application à la prévision du comportement d'un processus à temps continu sur un intervalle de temps donné. C. R. Acad. Sci. Paris Sér. I Math. **310**, No.11, 787-790.

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. **335**, Kluwer Acad. Publ., Dordrecht.

Bosq, D. (2000). Linear processs in function spaces. Theory and Application. Lectures Notes in Statistics. Vol 149, Springer Verlag.

Bosq, D., Delecroix, M. (1985). Nonparametric prediction of a Hilbert-space valued random variable. *Stochastic Process. Appl.* 19, 271-280.

Bosq, D., Lecoutre, J. P.(1987). Théorie de l'estimation fonctionnelle. ECONOMICA, Paris.

Boularan, J., Ferré, L., Vieu, P. (1995). Location of particular points in nonparametric regression analysis. *Austral. J. Statist.* **37**, 161-168.

Bru, B., Heinich, H. (1980). Sur l'espérance des variables aléatoires vectorielles. Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques. **16**, 177-196.

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

Carbon, M., Francq, C., Tran, L. T. (2007). Kernel regression estimation for random fields. J. Statist. Plann. Inference. **137**, 778-798.

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

Chate, H., Courbage, M. (1997). Lattice systems. *Physica D*, **103** 1 -612.

Collomb, G. (1981). Estimation non paramétrique de la regression : Revue bibliographique. Inter. Statist. Revue. **49**, No.1, 75-93

Collomb, G. (1983). Méthodes non paramétriques en régression, analyse de séries temporelles, prédiction et discrimination. Doctorat d'état, Toulouse 3.

Collomb, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau. Z. W. Gebiete. **66**, 441-460.

Collomb, G. (1985). Nonparametric regression : an up to date bibliography Statistics. **16**, 309-324.

Collomb, G., Härdle, W., Hassani, S. (1987). A note on prediction via conditional mode estimation. J. Statist. Plann. and Inf. **15**, 227-236.

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

Dabo-Niang, S., Rhomari, N. (2003). Estimation non paramétrique de la régression avec variable explicative dans un espace métrique. *C. R., Math., Acad. Sci. Paris.* **336**, 75-80.

Dabo-Niang, S., Laksaci, A. (2007). Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle. *C. R., Math., Acad. Sci. Paris.* **344**, 49-52.

Deheuvels, P., Mason, D. (2004). General asymptotic confidence bands based on kernel type function estimators. *Statistical Inference for Stochastic Processes*, **7**, 225-277.

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

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

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

Delsol, L., Ferraty, F., Vieu, P. (2011). Structural test in regression on functional variables. J. Multivariate Anal. (to appear)

Devroye, L. (1978). The uniform convergence of the Nadaraya-Watson regression function estimate. Canad. J. Statist. **6**, No.2, 179-191.

Ezzahrioui, M. (2007). Prévision dans les modèles conditionnels en dimension infinie. PhD Thesis. Univ. du Littoral Côte d'Opale.

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

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

Ezzahrioui, M., Ould-Saïd (2008). Asymptotic normality of nonparametric estimator of the conditional mode for functional data. *Journal of Nonparametric Statistics*, **20**(1), 3-18.

Ezzahrioui, M., Ould-Saïd (2010). Some asymptotic results of a nonparametric conditional mode estimator for functional time-series data. Stat. Neerl. **64**, 171-201.

Ferraty, F. (2010). Special issue on statistical methods and problems in infinite dimensional spaces. J. Multivariate Analysis. **101**(2), 305-490.

Ferraty, F., Goia, A., Vieu, P. (2002). Functional Nonparametric Model for Time Series : A Fractal Approach for Dimension Reduction. TEST, **11**, 317-344.

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

Ferraty, F., Laksaci, A., Tadj, A. Vieu, P. (2011). Kernel regression whith functional response. *Electronic Journal of Statistics.* **5**, 159-171.

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

FERRATY, F. LAKSACI, A., VIEU, P. (2006). Estimation of some characteristics of the conditional distribution in nonparametric functional models. Statistical Inference for Stochastic Processes. **9**, 47-76.

FERRATY, F., MAS, A., VIEU, P. (2007). Nonparametric regression on functional data : inference and practical aspects. *Aust. N. Z. J. Stat.*, **49**, 267-286.

FERRATY, F., RABHI, A., VIEU, P. (2005b). Conditional quantiles for functionally dependent data with application to the climatic El Phenomenon. *Sankhya,* **67**, 378-399.

FERRATY, F., RABHI, A., 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., ROMAIN, Y. (2011). The Oxford Handbook of Functional Data Analysis. Oxford University Press.

FERRATY, F., VAN KEILEGOM, I., VIEU, P. (2009). On the validity of the bootstrap in nonparametric functional regression. *Scand. J. Stat. (In press).*

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

FERRATY, F., VIEU, P. (2002). The functional nonparametric model and application to spectrometric data. Comput. Statist. and Data Anal. **17**, 545-564.

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

FERRATY, F., VIEU, P. (2006). Nonparametric functional data analysis.

Theory and Practice. Springer-Verlag.

FERRATY, F., VIEU, P. (2009). Additive prediction and boosting for functionnal data. *Computational Statistics & Data Analysis*, **53** 1400-1413.

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

GALILEO GALILEI (1632). Dialogue concerning the two chief world systems. Modern library, 2001. Translated by Stillman Drake.

GANNOUN, A., SARACCO, J., YU, K. (2003), Nonparametric prediction by conditional median and quantiles. J. Stat. Plann. Inference. **117**, No.2, 207-223.

GAUSS, J. F. (1809). Theoria motus corporum coelestium. Hamburg : Perthes et Besser.

GYÖRFI, L., HÄRDLE, W., SARDA, P. ET VIEU, P. (1989). Nonparametric curve estimation for time series, Lecture Notes in Statistics. **60**, Springer-Verlag.

HAMEDANI, G. G., MANDREKAR, V. (1978). Lévy-khinchine representation and Banach spaces of type and cotype. Studia Math. **Vol. 63**.

HE, G., MÜLLER, H.G., WANG, J.L., YANG, W. (2009). Functional Linear Regression Via Canonical Analysis. *Bernoulli*. **16**, 705-729.

HLUBINKA, D., PRCHAL, L. (2007). Changes in atmospheric radiation from the statistical point of view. *Comput. Statist. Data Anal.* **51**(10), 4926-4941.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13-30.

HOFFMANN-JØRGENSEN, J. (1974). Sums of independent Banach space valued random variables. Studia Math. **52**, 159-186.

KOLMOGOROV, A. N., TIKHOMIROV, V .M. (1959). ϵ-entropy and ϵ-capacity. *Uspekhi Mat. Nauk* **14** 3-86. (*Engl Transl. Amer. Math. Soc. Transl. Ser*), **2** 277-364 (1961).

KUELBS, J., LI, W. (1993). Metric entropy and the small ball problem for Gaussian measures. *J. Funct. Anal.*, **116**, 133-157.

LAKSACI, A. (2007). Convergence en moyenne quadratique de l'estimateur à noyau de la densité conditionnelle avec variable explicative fonctionnelle. Ann. I.S.U.P. **51**, 69-80.

LAKSACI, A., LEMDANI, M., OULD-SAÏD, E. (2009). A generalized L^1-approach for a kernel estimator of conditional quantile with functional regressors : consistency and asymptotic normality. Statist. Probab. Lett. **79**, 1065-1073.

LAKSACI, A., MADANI, F., RACHDI, M. (2010). Kernel conditional density estimation when the regressor is valued in a semi metric space. International Statistical Review. (In press).

LAKSACI, A., MECHAB, M. (2010). Estimation non parametrique de la fonction de hasard avec variable explicative fonctionnelle cas des donnees spatiales. Rev : Roumaine, Math Pures Appl. **55**, 35-51.

LAUKAITIS, A. (2007). An Empirical Study for the Estimation of Autoregressive Hilbertian Processes by Wavelet Packet Method Nonlinear Analysis : Modelling and Control **12** (1).

LECOUTRE, J. P. (1990). Uniform consistency of a class of regression function estimators for Banach-space valued random variable. *Statist. Probab. Lett.* 10, 145-149.

LECOUTRE, J. P., OULD-SAÏD, E. (1993). Estimation de la fonction de hasard pour un processus fortement mélangeant avec censure. Publ. Inst. Statist. Univ. Paris. **37**, No.1-2, 59-69.

LEGENDRE, A. M. (1805). Nouvelle méthode pour la détermination des orbites des comètes. Paris : Courcier.

Lemdani, M., Ould-Saïd, E., Poulin, N. (2009). Asymptotic properties of a conditional quantile estimator with randomly truncated. J. Multivariate Anal.**100**, 546-559.

Li, W. V., Lin, Z. (2007). Asymptotic normality for L^1-norm kernel estimator of conditional median under association dependence. J. Multivariate Anal. **98**, 1214-1230.

Li, D., Queffélec, H. (2004). Introduction à l'étude des espaces de Banach. Analyse et probabilités. Vol 12, Eyrolles.

Li, W. V., Tran, L. T. (2009). Nonparametric estimation of conditional expectation. J. Statist. Plann. Inference. **139**, 164-175.

Loannides, D., Matzner-Lober, E. (2004). A note on asymptotic normality of convergent estimates of the conditional mode with errors-in-variables. J. Nonparametr. Stat. **16**, 515-524.

Loève, M. (1963). Probability Theory. Third Edition. Van Nostranr Princeton.

Louani, D., Ould-Saïd, E. (1999). Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis. J. Nonparametric Statist. **11**, No.4, 413-442.

Lu, Z., Chen, X. (2004). Spatial kernel regression estimation : Weak concistency. Stat and Probab. Lett, **68**, 125-136.

Manteiga, W. G., Vieu, P. (2007). Statistics for Functional Data. Computational Statistics & Data Analysis. **51**, 4788-4792.

Masry, E. (2005). Nonparametric regression estimation for dependent functional data : asymptotic normality. *Stochastic Processes and their Applications,* **115**, 155-177.

Mikhail I. Kadets, Vladimir M. Kadets.(1997). Series in Banach Spaces :

Conditional and Unconditional Convergence (Operator Theory : Advances and Applications). **94**, Birkhäuser.

Müller, H.G., Chiou, J.M., Leng, X. (2008). Infering gene expression dynamics via functional regression analysis. *BMC Bioinformatics*. **9**, 60.

Nadaraya, E. (1964). On estimating regression. Theory Prob. Appl. **10**, 186-196.

Onsager, L., Machlup, S. (1953). Fluctuations and irreversible proceses, I-II. *Phys. Rev.*, **91**, 1505-1512 ; 1512-1515.

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., Cai Z. (2005). Strong uniform consistency of nonparametric estimation of the censored conditional mode function. *Nonparametric Statistics,* **17**, 797-806.

Quintela-del-Rio, A. (2006). Nonparametric estimation of the maximum hazard under dependence conditions. *Statist. Probab. Lett.* , **76**, 1117-1124.

Quintela-del-Rio, A. (2010). On non-parametric techniques for area-characteristic seismic hazard parameters. Geophys. J. Int. **180**, 339-346.

Quintela-del-Rio, A., Vieu, P. (1997). A nonparametric conditional mode estimate. J. Nonparametr. Statist. **8**, No.3, 253-266.

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

Ramsay, J. O., Silverman, B. W. (1997). Functional data analysis. Springer, New York.

Ramsay, J. O., Silverman, B. W. (2002). Applied functional data analysis : Methods and case studies Spinger-Verlag, New York.

Ramsay, J. O., Silverman, B. W. (2005). Functional Data Analysis, Springer, New-York, 2*nd* Edition.

Rhomari, N. (2002). Approximation et inégalités exponentielles pour les sommes de vecteurs aléatoires dépendants. C. R. Acad. Sci. Paris. **334**, 149-154.

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

Roussas, G. G. (1968). On some properties of nonparametric estimates of probability density functions. *Bull. Soc. Math. Greece (N.S.)* **9** 29-43.

Roussas, G. G. (1991). Kernel estimates under association : strong uniform consistency. Statist. Probab. Lett. **12**, No.5, 393-403.

Samanta, M. (1989). Non-parametric estimation of conditional quantiles. Stat. Probab . Lett. **7**, No.5, 407-412.

Samanta, M., Thavaneswaran, A. (1990). Non-parametric estimation of conditional mode. Comm. Statist. Theory and Meth. **16**, 4515-4524.

Sarda, P., Vieu, P. (2000). Kernel regression. In : M. Schimek (ed.) Smoothing and regression ; Approaches, Computation, and Application. Wiley Series in Pprobability and Statistics, Wiley, New York.

Schimek, M. (2000). Smoothing and Regression : Approaches, computation, and application, Ed. M.G. Schimek, Wiley Series in Probability and Statistics.

Stone, C. (1977). Consistent nonparametric regression. With discussion a nd a reply by the author. Ann. Statist. **5**, no. 4, 595-645.

Stone, C. (1980). Optimal rates of convergence for nonparametric estimators. Ann. Statist. **8**, No.6, 1348-1360.

Stone, C. (1982). Optimal global rates of convergence for nonparametric regression. Ann. Statist. **10**, 1040-1053.

THEODOROS, N., YANNIS G. Y. (1997) Rates of convergence of estimate, Kolmogorov entropy and the dimensioanlity reduction principle in regression. *The Annals of Statistics*, **25**, No.6, 2493–2511.

TUKEY, J.W. (1961). Curve as parameters, and touch estimation. Proceedings of the 4th Symposium on Mathematics, Statistics and Probability, 681-694, Berkeley, CA, USA.

VAN DER VAART, A. W., VAN ZANTEN, J. H. (2007). Bayesian inference with rescaled Gaussian process priors. *Electronic Journal of Statistics.*, **1**, 433-448.

VIEU, P. (1991). Quadratic errors for nonparametric estimates under dependence. J. Multivariate Anal. **39** (2), 324-347.

WATSON, G. S. (1964). Smooth regression analysis. Sankhya Ser. A. **26**, 359-372.

WATSON, G. S., LEADBETTER, M. R. (1964). Hazard analysis. I. *Biometrika*, **51**, 175-184.

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

YOUNDJÉ, E. (1996). Propriétés de convergence de l'estimateur à noyau de la densité conditionnelle. *Rev. Roumaine Math. Pures Appl.*, **41**, 535-566.

ZHOU, Y., LIANG, H. (2003). Asymptotic properties for L_1 norm kernel estimator of conditional median under dependance. J. Nonparametr. Stat. **15**, 205-219.

Résumé

La problématique abordée dans cette thèse est l'estimation non paramétrique des modèles conditionnels à variable explicative fonctionnelle en traitant deux cas : le cas où la variable réponse est réelle et le cas d'une variable réponse fonctionnelle. On établit la convergence uniforme presque complète d'estimateurs non paramétriques pour certains modèles conditionnels.

Dans un premier temps, nous considérons une suite d'observations i.i.d. et nous construisons des estimateurs par la méthode du noyau pour la fonction de régression généralisée, la fonction de répartition conditionnelle, la densité conditionnelle, la fonction de hasard conditionnelle et le mode conditionnel. Nous étudions la convergence uniforme presque complète de ces estimateurs en précisant leurs vitesses. A titre illustratif, nous donnons des exemples d'applications sur des données simulées.

Dans un second temps, on généralise nos résultats au cas d'une variable réponse fonctionnelle (appartenant à un espace de Banach) et on estime la régression classique. Cette généralisation a été étudiée dans les deux cas : les observations i.i.d. ainsi que le cas dépendant. Dans ce dernier, nous avons fixé comme objectif la convergence presque complète ponctuelle lorsque les observations sont β-mélangeantes.

Nos résultats asymptotiques exploitent bien la structure topologique de l'espace fonctionnel de nos observations et le caractère fonctionnel de nos modèles. En effet, toutes nos vitesses de convergence sont quantifiées en fonction de la concentration de la mesure de probabilité de la variable fonctionnelle, de l'entropie de Kolmogorov et du degré de régularité des modèles. Notons également que dans le cas où la variable réponse est aussi fonctionnelle, nos vitesses de convergence contiennent un terme additionnel qui dépend du type de l'espace de Banach de la variable réponse.