

UNIVERSITE PAUL SABATIER TOULOUSE III

U.F.R. MATHEMATIQUE INFORMATIQUE GESTION

**Thèse**

POUR OBTENIR LE GRADE DE

**Docteur de l'université Toulouse III**

Discipline : Mathématiques Appliquées

Option : Statistique

Présentée et soutenue par

**Ali Laksaci**

le 28 Novembre 2005

**Contribution aux modèles non paramétriques conditionnels  
pour variables explicatives fonctionnelles**

Composition du Jury

A. Cuevas	Univ. Autónoma, Madrid	Rapporteur
F. Ferraty	Univ. Toulouse III	Directeur de thèse
A. Mas	Univ. Montpellier II	Examinateur
E. Ould-Saïd	Univ. Littoral	Examinateur
J. Saracco	Univ. Bourgogne	Rapporteur
Ph. Vieu	Univ. Toulouse III	Directeur de thèse

Dédicace  
*À tous ceux qui me sont proches.*

## Remerciements

En premier lieu, je tiens à exprimer mon profond sentiment de respect et de reconnaissance à mes directeurs de thèse les Professeurs **Frédéric Ferraty** et **Philippe Vieu** pour leur encadrement et leurs encouragements durant toute la période de la réalisation de ce travail. Par la même occasion, qu'ils trouvent en ces mots l'expression de mon admiration pour leurs grandes qualités scientifiques et humaines.

Je tiens à exprimer aussi ma reconnaissance à Monsieur le Professeur **Antonio Cuevas** pour avoir accepté d'être rapporteur de cette thèse et ainsi me faire bénéficier de son immense expérience.

Je remercie également Monsieur le Professeur **Jérôme Saracco** pour avoir accepté d'être rapporteur de cette thèse.

Pour la pertinence et la justesse de leur commentaires, remarques et suggestions, tant sur le fond que sur la forme, encore merci aux rapporteurs.

Pour l'intérêt qu'il a bien voulu accorder à mes recherches en acceptant de participer au jury, je tiens à exprimer mes vifs remerciements à Monsieur le Professeur **André Mas**.

Je voudrais aussi remercier vivement Monsieur le Professeur **Elias Ould-Saïd** pour sa participation au jury, pour l'intérêt qu'il porte à mes travaux et pour les nombreuses discussions fructueuses que nous avons eues.

Je tiens à exprimer particulièrement ma vive reconnaissance à Monsieur le professeur **Abderrahmane Yousfate** en tant qu'enseignant pour m'avoir initié au domaine de la statistique tant sur le plan académique que sur le plan de la recherche.

Un grand merci à ma famille pour son soutien constant et chaleureux pendant toute la réalisation de ce travail.

Enfin, je remercie les membres du groupe **STAPH**, ainsi que l'ensemble des mes collègues qui m'ont apportés aide, soutien, sympathie et amitié ainsi que tous ceux pour qui le mot science rime encore avec conscience.

J'adresse mes remerciements à tous ceux qui ont participé de près ou de loin à la bonne réalisation de la soutenance de cette thèse.



# Table des matières

<b>1</b>	<b>Présentation</b>	<b>7</b>
1.1	Résumé . . . . .	7
1.2	Summary . . . . .	8
1.3	Liste des travaux . . . . .	9
1.4	Contexte bibliographique . . . . .	10
1.4.1	Données et variables fonctionnelles . . . . .	10
1.4.2	Prévision non paramétrique via certaines caractéristiques conditionnelles . . . . .	12
1.5	Description de la thèse . . . . .	13
1.6	Brève présentation des résultats . . . . .	14
1.6.1	Notations . . . . .	14
1.6.2	Les Modèles non paramétriques . . . . .	15
1.6.3	Résultats : Cas i.i.d. . . . .	16
1.6.4	Résultats : Cas $\alpha$ mélangeant . . . . .	17
1.6.5	Applications . . . . .	17
<b>2</b>	<b>Cas i.i.d.</b>	<b>19</b>
2.1	Introduction . . . . .	21
2.2	General notations and conditions . . . . .	23
2.3	Some conditional nonparametric functional problems . . . . .	25
2.3.1	Conditional cumulative distribution estimation . . . . .	25
2.3.2	Estimating derivatives of the conditional density . . . . .	27
2.4	Estimation of some real conditional parameters . . . . .	28
2.4.1	Conditional mode estimation . . . . .	28
2.4.2	Conditional quantiles estimation . . . . .	30
2.5	The continuous time processes framework . . . . .	33
2.5.1	The small ball probability effects . . . . .	33
2.5.2	Application to diffusion processes . . . . .	34

2.5.3	Application to Gaussian paths . . . . .	37
2.5.4	A statistical solution to the curse of infinite dimension . . . . .	39
2.6	Application to chemometrical functional data . . . . .	41
2.7	Appendix . . . . .	43
2.7.1	Proofs of lemmas of Section 2.3.1 . . . . .	43
2.7.2	Proofs of lemmas of Section 2.3.2 . . . . .	47
2.7.3	Proofs of lemmas of Section 2.4 . . . . .	50
<b>3</b>	<b>Cas dépendant</b>	<b>55</b>
3.1	Résultats principaux . . . . .	56
3.1.1	Introduction . . . . .	58
3.1.2	Model . . . . .	58
3.1.3	Main result . . . . .	59
3.2	Preuves . . . . .	64
3.2.1	Proof of Theorem 3.1.1 . . . . .	64
3.2.2	Proof of Lemma 3.1.1 . . . . .	65
3.2.3	Proof of Corollary 3.1.1 . . . . .	68
3.2.4	Proof of Lemma 3.1.2 . . . . .	68
3.2.5	Proof of Lemma 3.1.3 . . . . .	69
3.2.6	Proof of Corollary 3.1.2 . . . . .	72
3.2.7	Proof of Lemma 3.1.4 . . . . .	72
3.3	Application . . . . .	72
3.3.1	La prévision en série temporelle . . . . .	72
3.3.2	Mise en oeuvre sur des données réelles . . . . .	73
<b>4</b>	<b>Commentaires et perspectives</b>	<b>83</b>
4.1	Quelques situations particulières . . . . .	83
4.1.1	Processus à temps continu . . . . .	83
4.1.2	Cas fractal . . . . .	85
4.1.3	Cas de dimension finie . . . . .	85
4.2	Quelques questions pratiques . . . . .	86
4.2.1	Choix de la semi-métrie . . . . .	86
4.2.2	Choix du paramètre de lissage . . . . .	87
4.2.3	A propos de l'existence de la version régulière de la probabilité conditionnelle . . . . .	87
4.3	Perspectives . . . . .	88

# Chapitre 1

## Présentation

### 1.1 Résumé

Dans cette thèse, nous nous proposons d'étudier le problème de la modélisation non paramétrique lorsque les données statistiques sont des courbes. Plus précisément, nous nous intéressons à des problèmes de prévision à partir d'une variable explicative à valeurs dans un espace de dimension infinie, et nous cherchons à développer des alternatives à la méthode de régression. En effet, on suppose qu'on dispose d'une variable aléatoire réelle (réponse), notée  $Y$ , et d'une variable fonctionnelle (explicative), notée  $X$ . Le modèle non paramétrique utilisé pour étudier le lien entre  $X$  et  $Y$  concerne la distribution conditionnelle dont la fonction de répartition (resp. densité), notée  $F$  (resp.  $f$ ), est supposée appartenir à un espace fonctionnel approprié.

Dans un premier temps, on considère une suite d'observations i.i.d. Dans ce contexte, nous construisons des estimateurs par la méthode du noyau pour la fonction de répartition conditionnelle, la densité conditionnelle et ses dérivées. On établit la vitesse de convergence presque complète de ces estimateurs. On déduit des estimateurs précédents ceux pour estimer le mode conditionnel et les quantiles conditionnels, pour lesquels, on donne la vitesse de convergence presque complète.

Dans un second temps, nous supposons que les observations sont fortement mélangeantes et nous nous fixons comme objectif l'estimation du mode conditionnel. Nous étudions les propriétés asymptotiques de cet estimateur, en donnant l'expression de sa vitesse de convergence. Ce résultat peut être utilisé pour le problème de la prévision en série chronologique.

Notre étude met en évidence le phénomène de concentration de la mesure de probabilité de la variable fonctionnelle sur des petites boules. Plus précisément, des hypothèses portant sur les probabilités de petites boules nous permettent de proposer une solution originale au problème du fléau de la dimension et ainsi de généraliser à la dimension infinie de nombreux résultats asymptotiques existant dans le cas multivarié. De plus, en utilisant les nombreux résultats récents en théorie des probabilités sur les petites boules, on précise nos résultats pour de nombreux processus à temps continu.

L'originalité de cette thèse est qu'elle aborde et développe aussi bien des aspects pratiques que théoriques. Nos méthodes sont appliquées à des données réelles de type spectrométrique ou de pollution.

## 1.2 Summary

In this thesis, we study the problem of a nonparametric modelization when the data are curves. Indeed, we consider real random variable (named response variable) noted  $Y$ , and a functional variable (explanatory variable) noted  $X$ . The nonparametric model used to study the relation between  $X$  and  $Y$  is the conditional distribution function noted  $F$  which has a density  $f$ . Both  $F$  and  $f$  are supposed to belong to some suitable functional spaces.

Firstly, we consider a sequence of i.i.d observations. In this context, we build kernel estimators of the conditional distribution function, the conditional density and its successive derivatives. We establish the almost complete convergence rate of these estimators. We use these results in order to study the conditional mode and the conditional quantiles and we give also the almost complete convergence rate of their estimators.

Secondly, we suppose that the observations are  $\alpha$ -mixing and we focus on the estimate of the conditional mode. We quantify the asymptotic properties of this estimator, by giving the convergence rate. This result can be used to the prediction problem in functional time series.

Our study highlights the phenomenon of concentration properties on small balls of the probability measure of the functional variable. More precisely, these ideas are used to give a statistical solution to curse of dimension and to generalize to infinite dimension many asymptotic results existing in the multivariate case. Moreover, by using recent results in the probability



theory of small balls we can see that our results include many time continuous processes.

The originality of this thesis is in the fact, that it considers both practical and theoretical aspects. We apply our prediction methods to the spectrometric data and the pollution data. We give a comparison between the prediction via the regression model and the prevision via the conditional mode in both cases : depending and not depending.

## 1.3 Liste des travaux

### Publications

1. Ferraty ; F., Laksaci ; A. and Vieu ; Ph. (2006) Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat. Inference Stoch. Process.* Vol 9 (1) (En cours d'impression),
2. Ferraty, F., Laksaci, A. and Vieu, Ph., (2005) Functional times series prediction via conditional mode. *C. R., Math., Acad. Sci. Paris* **340**, No.5, 389-392,
3. Dabo-Niang, S. Laksaci, A. Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle (Soumis pour publication).

### Communications

1. Estimation de certains paramètres conditionnels pour des variables explicatives fonctionnelles. *MSS'04 16-17 Avril 2004 U.S.T.H.B. Alger (ALGERIE)*.
2. Statistique non paramétrique fonctionnelle : Eude asymptotique et application d'estimateurs des modes et quantiles conditionnels fonctionnels. *XXXVIèmes journées de statistique 24-28 mai 2004 MONTPELLIER, (FRANCE)*
3. Estimating the conditional mode and application in functional time series. *MOAD'04 28-29 Novembre 2004, Saida (ALGERIE)*.
4. La distribution conditionnelle pour variable explicative fonctionnelle : Théorie et Application, *Journées de statistique Appliquées Biskra, 26-27- février 2005 (ALGERIE)*

5. Analyse non paramétrique de données fonctionnelles. *3èmes Journées de Statistique Fonctionnelle et Opératoirelle, 13-14 juin 2005, Université Paul Sabatier, Toulouse, (FRANCE)*

## 1.4 Contexte bibliographique

Le problème de la prévision est une question très fréquente en statistique. En statistique non paramétrique, l'outil principal pour répondre à cette question est le modèle de régression. Cet outil a pris un essor considérable de par le nombre de publications qui lui sont consacrées, que les variables explicatives soient uni, multi ou infiniment dimensionnées. Cependant, cet outil de prévision n'est pas très adapté pour certaines situations. À titre d'exemple, citons le cas de densité conditionnelle dissymétrique ou encore le cas où elle comporte plusieurs pics avec l'un des pics strictement plus important que les autres. Dans ces différents cas, on peut espérer que le mode, la médiane ou les quantiles conditionnels prévoient mieux que la régression. L'objectif de la thèse est d'étudier cette alternative de prévision non paramétrique dans le cas de données fonctionnelles. Ce paragraphe est divisé en deux sous paragraphes : le premier sera consacré aux données et variables fonctionnelles alors que le deuxième sera axé sur la prévision non paramétrique via les caractéristiques conditionnelles dans le cas de variables explicatives multivariées (dimension finie).

### 1.4.1 Données et variables fonctionnelles

Les problèmes statistiques liés à la modélisation et à l'étude de variables aléatoires fonctionnelles connaissent depuis longtemps un grand intérêt en statistique. Les premiers travaux sont basés sur la discrétisation de ces observations fonctionnelles afin de pouvoir adapter des techniques statistiques multivariées classiques. Mais, grâce aux progrès de l'outil informatique permettant la récupération de données de plus en plus volumineuses, une alternative a été récemment élaborée consistant à traiter ce type de données dans sa propre dimension, c'est-à-dire en conservant le caractère fonctionnel. En effet, depuis les années 60, la manipulation des observations sous forme de trajectoires a été l'objet de plusieurs études dans différentes disciplines scientifiques tels Obhukov (1960), Holmstrom (1961) en climatologie, Deville (1974) en économétrie, Molenaar et Boomsma (1987) puis Kirkpatrick et

Heckman (1989) en génétique.

Les modèles fonctionnels de régression (paramétrique ou non paramétrique) sont des thèmes qui ont été privilégiés ces dernières années. Dans le cadre linéaire, la contribution de Ramsay et Silverman, (1997,2002) présente une collection importante de méthodes statistiques pour les variables fonctionnelles. De même, notons que Bosq (2000) a significativement contribué au développement de méthodes statistiques dans le cadre de processus d'autorégression fonctionnel linéaire. En utilisant l'analyse en composantes principales fonctionnelles, Cardot et *al.* (1999) ont construit un estimateur pour le modèle de la régression linéaire Hilbertien similaire à l'estimateur de Bosq (1991) pour la cas de processus autorégressifs Hilbertiens. Cet estimateur est défini à l'aide des propriétés spectrales de la version empirique de l'opérateur de variance-covariance de la variable explicative fonctionnelle. Ils ont obtenu la convergence en probabilité et la convergence presque complète de l'estimateur construit. La convergence en norme  $L^2$  pour une version régularisée (spline) de l'estimateur précédent a été établie par les mêmes auteurs en 2000. Récemment, Cardot et *al.* (2004) ont introduit, toujours par une méthode de régularisation, un estimateur pour les quantiles conditionnels, vu comme forme linéaire continue définie sur un espace de Hilbert. Sous des conditions sur les valeurs propres de l'opérateur de covariance de la variable explicative et sur la densité de la loi conditionnelle, ils ont donné la vitesse de convergence en norme  $L^2$  de l'estimateur ainsi construit. Nous renvoyons à Cardot et *al.* (2003) et à Cuevas et *al.* (2004) pour le problème du test dans le modèle linéaire fonctionnel. Plusieurs auteurs se sont intéressés au cas où la variable réponse est qualitative. Nous en citons, par exemple, Hastie et *al.* (1995), Hall et *al.* (2002), ...

L'étude des modèles non paramétriques de régression est beaucoup plus récente que celle du cas linéaire. Les premiers résultats ont été fournis par Ferraty et Vieu (2000). Ces résultats ont été prolongés par Ferraty et *al.* (2002), aux problèmes non-standards de la régression telle la prévision dans le contexte des séries chronologiques. En reprenant l'estimateur de Ferraty et Vieu (2004) et en utilisant la propriété de concentration de la mesure de probabilité de la variable explicative fonctionnelle, Niang et Rhomari (2004) ont étudié la convergence en norme  $L^p$  de l'estimateur de la régression. Ils ont appliqué leur résultat à la discrimination et la classification des courbes. D'autres auteurs se sont intéressés au cas où la variable réponse est fonctionnelle (Bosq et Delecroix (1985), Besse et *al.* (2000)). Enfin, plus récemment,

des premiers travaux portant sur des modèles présentant à la fois des aspects linéaires et non paramétriques ont été réalisés par Ferraty et *al.* (2003), Ait-Saidi et *al.* (2005), Ferré et Villa (2005).

Les premiers travaux sur l'estimation de distribution des variables fonctionnelles ont été donnés par Geffroy (1974). Plus récemment, Gasser et *al.* (1998) puis Hall et Heckman (2002) se sont intéressés à l'estimation non paramétrique du mode de la distribution d'une variable fonctionnelle. L'estimation de la médiane de la distribution d'une variable aléatoire qui prend ses valeurs dans un espace de Banach a été étudiée par Cadre (2001). Niang (2002) donne un estimateur de la densité dans un espace de dimension infinie et a établi des résultats asymptotiques de cet estimateur, tels la convergence en moyenne quadratique, la convergence presque sûre et la normalité asymptotique d'un estimateur de type histogramme. On trouvera aussi dans cet article une application donnant l'expression de la vitesse de convergence dans le cas de l'estimation de la densité d'un processus de diffusion par rapport à la mesure de Wiener. Ferraty et *al.* (2004) ont étudié l'estimateur non paramétrique du mode de la densité d'une variable aléatoire à valeurs dans un espace vectoriel semi-normé de dimension infinie. Ils établissent sa convergence presque sûre et ils appliquent également ce résultat au cas où la mesure de probabilité de la variable vérifie une condition de concentration. Plusieurs auteurs se sont intéressés à l'application de la modélisation statistique par des variables fonctionnelles sur des données réelles. A titre d'exemple, Ferraty et Vieu (2002, 2003) se sont intéressés à des données spectrométriques et à des enregistrements vocaux, Besse et *al.* (2002) à des données météorologiques, Gasser et *al.* (1998) ont considéré des données médicales, ...

#### 1.4.2 Prévision non paramétrique via certaines caractéristiques conditionnelles

Concernant l'étude de la fonction de répartition conditionnelle, la densité conditionnelle, le mode conditionnel et les quantiles conditionnels, de nombreux travaux sont disponibles dans le cadre multivarié (i.e. dimension finie). Roussas (1968) semble être le premier à aborder le problème de l'estimation de la fonction de répartition conditionnelle par la méthode à noyau. Il a établi la convergence en probabilité de cet estimateur en considérant des données à caractère Markovien. La densité conditionnelle a été largement étudiée par Youndjé (1993) dans le cas de données mélangeants ou indépendantes.

Récemment, Laksaci et Yousfate (2002) ont établi la convergence en norme  $L^p$  de l'estimateur à noyau de la densité conditionnelle d'un processus markovien stationnaire. La littérature sur les autres caractéristiques conditionnelles est aussi très importante. Citons, par exemple, Collomb et *al.* (1987) ont montré la convergence uniforme sur tout compact de  $\mathbb{R}^p$  pour un processus  $\phi$ -mélangeant. Quintela et Vieu (1997) ont estimé le mode conditionnel comme étant le point annulant la dérivée d'ordre un de l'estimateur de la densité conditionnelle et ont établi la convergence presque complète de cette estimateur pour le processus  $\alpha$ -mélangeant. Ould-said (1997) a montré la validité du résultat de Collomb et *al.* (1987) pour des processus ergodiques. Nous renvoyons à Berlinet et *al.* (1998), Louani et Ould-Said (1999) pour le cas de la normalité asymptotique du mode conditionnel dans le cas fortement mélangeant. D'autres auteurs ont étudié l'estimation de la médiane conditionnelle et/ou les quantiles conditionnels ainsi que des applications à la prévision (voir Stone (1977), Samanta (1989) Gannoun (1990), Boente et Fraiman (1995), Berlinet et *al.* (1998), Gannoun et *al.* (2003), Saracco (2003)).

## 1.5 Description de la thèse

Le travail développé dans cette thèse se situe à l'intersection entre les deux thématiques importantes : variables fonctionnelles et prévision non paramétrique via la distribution conditionnelle. On obtient ainsi deux alternatives à la régression non paramétrique fonctionnelle, l'une basée sur l'estimation du mode conditionnel, l'autre axée autour de la médiane conditionnelle. La thèse se décompose en quatre chapitres.

Après avoir donner une brève présentation de la thèse, on s'intéresse dans le deuxième chapitre au cas où on dispose de réalisations de variables aléatoires i.i.d. Dans ce but, on construit des estimateurs à noyau pour la fonction de répartition conditionnelle, la densité conditionnelle et ses dérivées. Nous établissons la convergence presque complète de ces estimateurs et nous appliquons ces résultats pour estimer le mode conditionnel et les quantiles conditionnels. On insistera sur le fait que les propriétés mathématiques des estimateurs sont étroitement liées à la question de la concentration de la variable explicative dans les petites boules. Ainsi, on peut préciser ces résultats dans plusieurs situations pour lesquelles la variable fonctionnelle est un processus à temps continu tels les processus de

diffusion ou les processus gaussien en général. A travers ces phénomènes de concentration, le lien sera fait avec le problème du fléau de la dimension. Cette étude sera ponctuée par une application sur des données spectrométriques issues du secteur agro-alimentaire. Dans le troisième chapitre on étend les résultats précédents au cas de variables aléatoires dépendantes. Ce chapitre est découpé en trois sections. La première section présente une généralisation des résultats obtenus pour le mode conditionnel. De nouveaux résultats de convergence seront également explicités dans ce chapitre, lesquels sont encore liés aux probabilités de petites boules de la loi de la variable explicative fonctionnelle et aussi à la régularité de la densité conditionnelle. La deuxième section est consacrée aux démonstrations des résultats élaborés dans la section précédente. Un exemple d'application sur la prévision de la concentration de l'ozone sera étudié dans la troisième section, ainsi qu'une comparaison avec l'approche standard de régression. Pour conclure, dans le chapitre 4 nous donnerons des commentaires sur l'ensemble des résultats asymptotiques obtenus. On insistera sur le fait que nos résultats se comportent bien dans plusieurs situations. On achèvera ce chapitre par quelques perspectives de recherche.

## 1.6 Brève présentation des résultats

Nous donnons ci-après une brève présentation des résultats obtenus dans la thèse.

### 1.6.1 Notations

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 sur  $\mathcal{F}$ . Pour tout  $x \in \mathcal{F}$  on définit la fonction de répartition conditionnelle de  $Y$  sachant  $X = x$ , notée  $F^x$ , 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 et, comme  $f^x = f^{x(0)}$ , tous nos résultats seront énoncés en fonction de  $j$ .

Soit  $(X_i, Y_i)_{i=1, \dots, n}$  des couples ayant la même loi que  $(X, Y)$ . On construit un estimateur pour la fonction de répartition, noté  $\widehat{F}_n^x$ , défini par

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

où  $K$  est un noyau,  $H$  est une fonction de répartition et  $h_K = h_{K,n}$  (resp.  $h_H = h_{H,n}$ ) est une suite de réels positifs. De cet estimateur on déduit un estimateur pour la densité conditionnelle et ses dérivées, noté  $\widehat{f}^{x(j)}$ , défini par

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

Pour l'estimation du mode conditionnel on suppose qu'il existe un compact  $S$  où le mode est unique. On note par  $\theta$  ce mode. On estime le mode conditionnel  $\theta$  par la variable aléatoire  $\widehat{\theta}$  solution du problème d'optimisation suivant :

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

Enfin, étant donné un  $\alpha \in ]0, 1[$ , le quantile conditionnel d'ordre  $\alpha$ , noté  $t_\alpha$ , est tel que

$$F^x(t_\alpha) = \alpha.$$

Pour assurer l'existence et l'unicité de  $t_\alpha$  on suppose que la fonction de répartition conditionnelle  $F^x$  est strictement croissante. On estime le quantile conditionnel  $t_\alpha$  par  $\widehat{t}_\alpha$  tel que

$$\widehat{F}^x(\widehat{t}_\alpha) = \alpha.$$

## 1.6.2 Les Modèles non paramétriques

Pour tout  $x$  fixé dans  $\mathcal{F}$ , on note par  $N_x$  un voisinage de ce point. Soit  $S$  un compact de  $\mathbb{R}$ . Nos modèles non paramétriques sont définis à partir de contraintes de régularité agissant sur les objets (fonctions) que l'on souhaite estimer :

- Pour la fonction de répartition conditionnelle elle sera supposée vérifier la condition suivante :

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

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

$$\forall (y_1, y_2) \in S \times S \quad \forall (x_1, x_2) \in N_x \times N_x, |f^{x_1}(y_1) - f^{x_2}(y_2)| \leq C_x (d(x_1, x_2)^{b_1} + |y_1 - y_2|^{b_2}).$$

### 1.6.3 Résultats : Cas i.i.d.

Dans cette partie, on suppose que les observations sont indépendantes. Si la fonction de répartition conditionnelle et la densité conditionnelle vérifient respectivement les modèles non paramétriques ci-dessus, alors, sous une condition de concentration de la mesure de probabilité de la variable fonctionnelle  $X$  et des conditions techniques standards portant sur l'estimateur on a les résultats suivants.

**Theorem 1.6.1** *Pour tout compact  $S$ , on a*

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

où  $\phi_x(h_K)$  est la concentration la mesure de probabilité de la variable fonctionnelle  $X$  dans la boule de centre  $x$  et de rayon  $h_K$ .

**Theorem 1.6.2** *Pour tout compact  $S$ , on a*

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

**Theorem 1.6.3**

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

**Theorem 1.6.4**

$$\widehat{t}_\alpha - t_\alpha = O(h_K^{\frac{b_1}{j}}) + O(h_H^{\frac{b_2}{j}}) + O\left(\left(\sqrt{\frac{\log n}{nh_H\phi_x(h_K)}}\right)^{\frac{1}{j}}\right), \quad p.co.$$

La démonstration de ces résultats et le détail des conditions imposées seront donnés au chapitre 2.



### 1.6.4 Résultats : Cas $\alpha$ mélangeant

Pour l'extension de nos résultats à des problèmes non standard de la statistique non paramétrique, tel la prévision en série temporelle, nous modélisons la notion de dépendance en considérant une suite d'observations  $\alpha$ -mélangeante (voir Doukhan (1994)). Si on renforce les hypothèses faites précédemment pour des variables indépendantes, par des conditions sur la concentration de la loi conjointe des couples  $(X_i, X_j)$  et par quelques hypothèses sur les coefficients de mélange on montre le résultat suivant.

#### Theorem 1.6.5

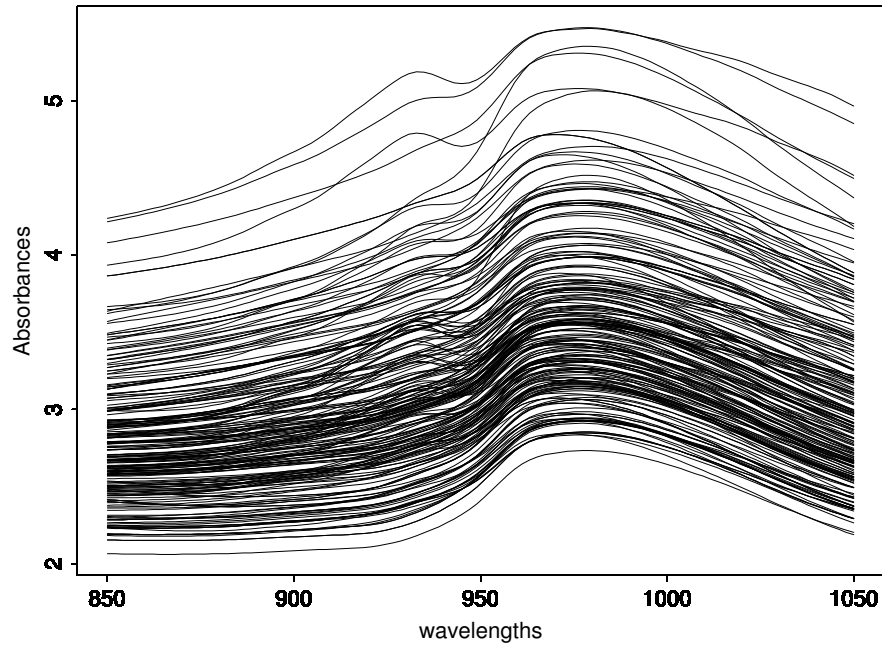
$$\hat{\theta} - \theta = O(h_K^{\frac{b_1}{j}}) + O(h_H^{\frac{b_2}{j}}) + O\left(\left(\sqrt{\frac{\log n}{nh_H \mathcal{X}_x(h_K)}}\right)^{\frac{1}{j}}\right), \quad p.co.$$

où  $\mathcal{X}_x(h_K)$  est le maximum de concentration entre la loi marginale et les lois conjointes de chaque couple d'observations fonctionnelles dans la boule de centre  $x$  et de rayon  $h_K$ .

Les hypothèses imposées et la preuve du résultat ci-dessus seront données au chapitre 3

### 1.6.5 Applications

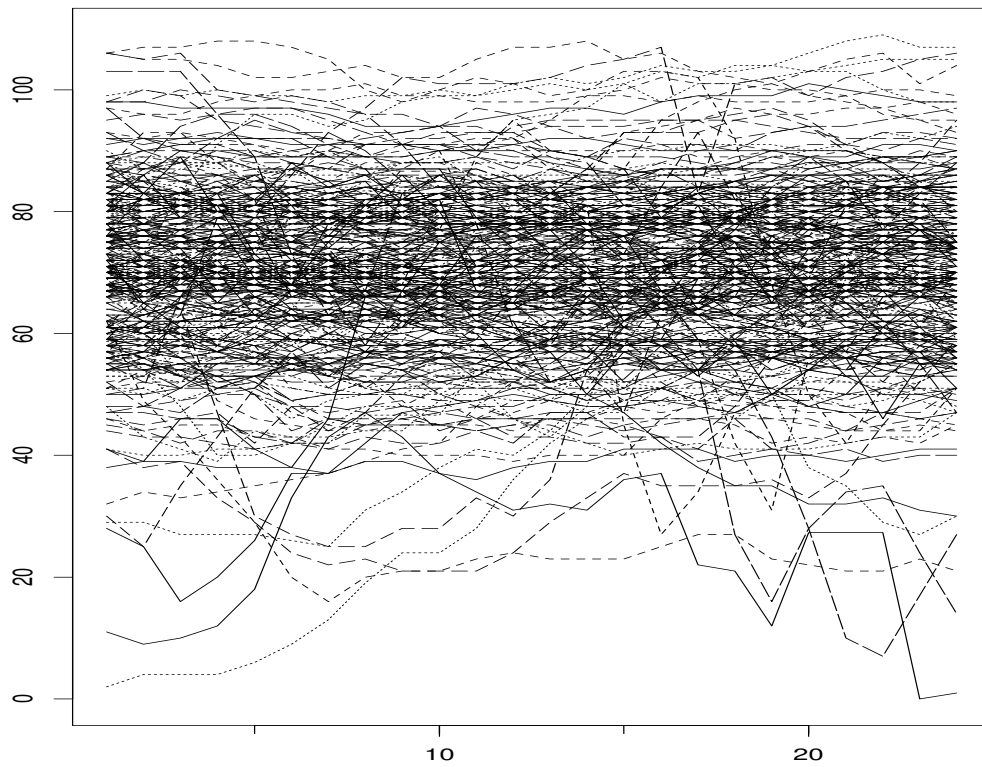
Deux exemples d'applications sont étudiés. Le premier correspond au cas i.i.d.. Il concerne l'industrie agro- alimentaire. Les courbes spectrométriques  $X_i$  correspondant à 215 morceaux de viande sont présentées ci-dessous.



Données spectrométrique

L'objectif de cette étude est de prévoir le taux de matière grasse ( $Y$ ) dans un morceau de viande sachant sa courbe spectrométrique ( $X$ ). On a mis en oeuvre deux méthodes de prédiction : le mode conditionnel et la régression, l'estimateur du mode donnant de meilleurs résultats sur cet exemple. Cet exemple est traité en détail dans le chapitre 2.

L'autre exemple considéré correspond au cas dépendant, il concerne un problème de pollution. Il s'agit plus particulièrement de l'étude de la courbe de la concentration de l'ozone sur le pôle nord (les données sont récupérés sur le site : <http://www.nilu.no/projects/cc/sitedescriptions/>). L'objectif est de prévoir le maximum de la concentration de l'ozone dans une journée sachant la courbe de concentration de l'ozone de la journée précédente. En faisant un découpage journalier de la courbe de l'année on obtient les courbes suivantes :



Données de pollution

Une comparaison est faite entre les deux approches suivantes : la prévision par le régression (Ferraty et Vieu (2002)) et la prévision via le mode conditionnel. Nous avons constaté que la prévision par le mode conditionnel est meilleure dans ce cas. Cet exemple est traité en détail dans le chapitre 3.



# Chapitre 2

## Cas i.i.d.

Ce chapitre fait l'objet d'une publication dans *Stat. Inference Stoch. Process.*

# Estimating some characteristics of the conditional distribution in nonparametric functional models

Frédéric FERRATY <sup>\*</sup>, Ali LAKSACI <sup>†</sup> and Philippe VIEU <sup>‡</sup>

20 avril 2005

## Abstract

This paper deals with a scalar response conditioned by a functional random variable. The main goal is to estimate nonparametrically some characteristics of this conditional distribution. Kernel type estimators for the conditional cumulative distribution function and the successive derivatives of the conditional density are introduced. Asymptotic properties are stated for each of these estimates, and they are applied to the estimations of the conditional mode and conditional quantiles.

Our asymptotic results highlightes the importance of the concentration properties on small balls of the probability measure of the underlying functional variable. So, a special section is devoted to show how our results behave in several situations when the functional variable is a continuous time process, with special attention to diffusion processes and Gaussian processes. Even if the main purpose of our paper is theoretical, an application to some chemometrical data set coming from food industry is presented in a short final section. This example illustrates the easy implementation of our method as well as its good behaviour for finite sample sizes.

**Key Words :** Conditional mode, conditional quantile, Conditional cumulative distribution, Derivatives of conditional density, Nonparametric estimation, Kernel estimator, Functional random variable, Small ball probability, Gaussian processes, Diffusion processes, Chemometrics.

---

<sup>\*</sup>Corresponding author : Labo. Statist. Prob., Université P. Sabatier, 31062 Toulouse, France. ferraty@cict.fr

<sup>†</sup>Université Djillali Liabes, Sidi Bel Abbès

<sup>‡</sup>Université Paul Sabatier, Toulouse

## 2.1 Introduction

The modelization of variables taking values in infinite dimensional spaces had received a lot of attention in the last few years, as well from mathematical, probabilistic, statistical or physicist point of views. The infatuation for this topic is certainly linked with many fields of applications in which the data to be treated are curves. In this general context, the study of statistical models adapted to such type of infinite dimensional data has been the subject of several works in the recent statistical literature. Previous works were concerning parametric (mainly linear) models, and good overviews about this literature can be found in Ramsay and Silverman (1997 and 2002) or Bosq (2000). The study of nonparametric models is much more recent, and only a few theoretical results have been obtained until now. For instance Ferraty and Vieu (2004) are studying regression estimation when the explanatory variable is functional and the response is scalar, and their study is extended to nonstandard regression problems such as time series prediction (see Ferraty *et al.*, 2002) or curves discrimination (see Ferraty and Vieu, 2002b). In an other hand, Niang (2002) is interested with the topic of functional density estimation (including various other related problems such as application to diffusion processes for instance). Additional bibliography about recent advances in functional statistics can be found through Staph (2002).

Even if we will mainly deal in this paper with theoretical features of nonparametric functional problems, it is worth to note that there are an increasing sources of potential applications of these functional models, as well in chemometrics (see Ferraty and Vieu, 2002), environmetrics (see Damon and Guillas, 2002, Fernández *et al.*, 2003), meteorological sciences (see Besse *et al.*, 2000, Hall and Heckman, 2002), speech recognition data (see Ferraty and Vieu, 2002b), radar range profile studies (see Hall *et al.*, 2001), medical data (see Gasser *et al.*, 1998), . . . One real data example will be quickly treated at the end of this paper.

This paper is concerned with several estimation problems, all related with the conditional distribution of a real random variable given some random variable which is supposed to be of functional feature in the sense that it is assumed to take its values in some abstract semi-metric space. More precisely, we will start in Section 2.3.1 by introducing some nonparametric conditional distribution estimator, and we will give asymptotic results (with rates). In

Section 2.3.2 we will look at the problem of estimating the conditional density and more generally its successive derivatives by mean of some nonparametric kernel type estimate for which we will also give asymptotic results with rates. Then we will discuss how we can apply these previous results to the estimation of some real characteristics of the conditional distribution that play an important role in prediction problems, and so Sections 2.4.1 and 2.4.2 will investigate respectively conditional mode and conditional quantile estimation problems. The main difficulty, when dealing with functional variables, relies on the difficulty for choosing some appropriate measure of reference in infinite dimensional spaces (see the first chapter in Niang, 2002b, for a deep discussion on this point). The main feature of our approach is to build estimates and to derive their asymptotic properties without any notion of density for the functional variable  $X$ . This approach allows us to avoid the use of a reference measure in such functional spaces. In each of the above described sections, we will give general asymptotic results without assuming existence of such a density, and each of these results will be discussed in relation with earlier literature existing in the usual finite dimensional case.

Indeed, the effects of the infinite dimension of the space of the functional variables will be seen through the concentration properties on small balls of the probability measure of this functional variable. Both to highlight this important feature and also to show the large scope of possible applications of our nonparametric methodology, we have devoted the special Section 2.5 to the study of several well-known continuous time processes. More precisely, we will pay special attention to diffusion processes and Gaussian processes, and we will show how our general methodology can be directly applied to these kind of processes.

Even if the main purpose of our paper is to develop theoretical asymptotic support for functional nonparametric estimates, the short final Section 2.6 will be devoted to the treatment of some chemometrical real data set. This functional data set comes from food industry problem. This Section 2.6 will give evidence on how our methods, that look rather complicated, are indeed easy to implement. Moreover, it will be seen on this kind of spectrometric data, how conditional functional mode is an efficient prediction tool.



## 2.2 General notations and conditions

All along the paper, when no confusion will be possible, we will denote by  $C$ ,  $C'$  or/and  $C_x$  some generic constant in  $\mathbb{R}^{*+}$ , and in the following, any real function with an integer in brackets as exponent denotes its derivative with the corresponding order.

We consider a random pair  $(X, Y)$  where  $Y$  is valued in  $\mathbb{R}$  and  $X$  is valued in some semi-metric space  $(\mathcal{F}, d(\cdot, \cdot))$  which can be of infinite dimension. We will say that  $X$  is a functional random variable and we will use the abbreviation *frv*. From a sample of independant pairs  $(X_i, Y_i)$ , each having the same distribution as  $(X, Y)$ , our aim is to build nonparametric estimates of several functions related with the conditional probability distribution (*ch2-cond-cdf*) of  $Y$  given  $X$ . For  $x \in \mathcal{F}$ , we will denote the *ch2-cond-cdf* of  $Y$  given  $X = x$  by

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

If this distribution is absolutely continuous with respect to the Lebesgues measure on  $\mathbb{R}$ , then we will denote by  $f^x$  (*resp.*  $f^{x(j)}$ ) the conditional density (*resp.* its  $j^{\text{th}}$  order derivative) of  $Y$  given  $X = x$ . In Sections 2.3 and 2.4, we will give almost complete convergence<sup>1</sup> results (with rates) for nonparametric estimates of both functions  $F^x$  and  $f^{x(j)}$ . Since  $f^x = f^{x(0)}$ , we will deduce immediately the convergence of the conditional density estimate from the general results concerning  $f^{x(j)}$ .

In the following  $x$  will be a fixed point in  $\mathcal{F}$ ,  $N_x$  will denote a fixed neighborhood of  $x$ ,  $S$  will be a fixed compact subset of  $\mathbb{R}$ , and we will use the notation  $B(x, h) = \{x' \in \mathcal{F} / d(x', x) < h\}$ . Our nonparametric models will be quite general in the sense that we will just need the following simple assumption for the marginal distribution of  $X$  :

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

---

<sup>1</sup>Let  $(z_n)_{n \in \mathbb{N}^*}$  be a sequence of real variables; we say that  $z_n$  converges almost completely (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.

together with some usual smoothness condition on the function to be estimated. According to the type of estimation problem to be considered, we will assume either

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

or, for some  $j \geq 0$ ,

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

There will be the only conditions related with the probabilistic distribution of  $(X, Y)$  that we will need. They are sufficiently weak to make unnecessary the introduction of the notion of density for the *frv*  $X$ , and so, the concentration condition (H1) will play an important role. This kind of condition is closely linked with the semi-metric  $d(., .)$  and is common in any nonparametric model for *frv*. Note that a right choice for  $d(., .)$  can provide (through the corresponding function  $\phi_x(., .)$ ) a solution to the curse of dimensionality. Note that (H1) is less restrictive than what it is usually assumed in most of previous works in nonparametric statistics for functional data (see *e.g.* Gasser *et al.*, 1998, Ferraty and Vieu, 2000, ...). This condition will be deeply studied in Section 2.5 through several different examples of *frv* coming from continuous time stochastic processes, and it will be linked with recent developments in small ball probability theory for infinite dimensional variables.

At this stage let us note that, as far as we know, it does not exist in the literature any work about such functional nonparametric models, as well for conditional density as for conditional cumulative distribution function, while in counterpart the literature in the special case when  $X$  is real is rather important (key references will be discussed for that in each of the above sections). Indeed, we will attack this problem by mean of kernel type estimators that will be constructed by combining ideas related with the existing literature in the real case together with the recent techniques developed in functional nonparametric regression (see Section 2.1).

## 2.3 Some conditional nonparametric functional problems

### 2.3.1 Conditional cumulative distribution estimation

Our first purpose in this section is to estimate the *cond-cdf*  $F^x$ . We introduce a kernel type estimator  $\widehat{F}^x$  of  $F^x$  as follows :

$$\widehat{F}^x(y) = \frac{\sum_{i=1}^n K(h_K^{-1}d(x, X_i)) H(h_H^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))}, \quad (2.1)$$

where  $K$  is a kernel,  $H$  is a *cdf* and  $h_K = h_{K,n}$  (*resp.*  $h_H = h_{H,n}$ ) is a sequence of positive real numbers. Note that using similar ideas, Roussas (1969) introduced some related estimate but in the special case when  $X$  is real, while Samanta (1989) produces wide asymptotic study for it. In our functional context, we need the following conditions for our estimate :

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

$$(H5) \quad K \text{ is a function with support } (0, 1) \text{ such that } 0 < C_1 < K(t) < C_2 < \infty,$$

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

$$(H7) \quad \lim_{n \rightarrow \infty} h_H = 0 \text{ with } \lim_{n \rightarrow \infty} n^\alpha h_H = \infty \text{ for some } \alpha > 0.$$

**Theorem 2.3.1** *Under the hypotheses (H1), (H.2) and (H4)-(H.7), we have*

$$\sup_{y \in S} |\widehat{F}^x(y) - F^x(y)| = O(h_K^{b_1}) + O(h_H^{b_2}) + O\left(\sqrt{\frac{\log n}{n\phi_x(h_K)}}\right), \quad a.co. \quad (2.2)$$

**Proof of Theorem 2.3.1.** For  $i = 1, \dots, n$ , we consider the quantities  $K_i = K(h_K^{-1}d(x, X_i))$ ,  $H_i(y) = H(h_H^{-1}(y - Y_i))$  and let  $\widehat{F}_N^x(y)$  (*resp.*  $\widehat{F}_D^x$ ) be

defined as

$$\widehat{F}_N^x(y) = \frac{1}{n EK_1} \sum_{i=1}^n K_i H_i(y) \quad (\text{resp. } \widehat{F}_D^x = \frac{1}{n EK_1} \sum_{i=1}^n K_i).$$

This proof is based on the decomposition

$$\begin{aligned} \widehat{F}^x(y) - F^x(y) &= \frac{1}{\widehat{F}_D^x} \left\{ \left( \widehat{F}_N^x(y) - E\widehat{F}_N^x(y) \right) - \left( F^x(y) - E\widehat{F}_N^x(y) \right) \right\} + \\ &\quad \frac{F^x(y)}{\widehat{F}_D^x} \left\{ E\widehat{F}_D^x - \widehat{F}_D^x \right\} \end{aligned} \quad (2.3)$$

and on the following intermediate results.

**Lemma 2.3.1** *Under the hypotheses (H1) and (H5)-(H6), we have*

$$\widehat{F}_D^x - E\widehat{F}_D^x = O \left( \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right), \quad a.co. \quad (2.4)$$

**Corollary 2.3.1** *Under the hypotheses of Lemma 2.3.1, we have*

$$\sum_{n=1}^{\infty} P \left( \widehat{F}_D^x < 1/2 \right) < \infty. \quad (2.5)$$

**Lemma 2.3.2** *Under the hypotheses (H1), (H2) and (H4)-(H6), we have*

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |F^x(y) - E\widehat{F}_N^x(y)| = O(h_K^{b_1}) + O(h_H^{b_2}), \quad a.co. \quad (2.6)$$

**Lemma 2.3.3** *Under the hypotheses (H1), (H2) and (H4)-(H7), we have*

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{F}_N^x(y) - E\widehat{F}_N^x(y)| = O \left( \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right), \quad a.co., \quad (2.7)$$

■

### 2.3.2 Estimating derivatives of the conditional density

Now we focus on the estimation of the  $j^{\text{th}}$  order derivative of the conditional density  $f^{x(j)}$  of  $Y$  given  $X = x$ . We propose to define the estimator  $\widehat{f}^{x(j)}$  of  $f^{x(j)}$  as follows :

$$\widehat{f}^{x(j)}(y) = \frac{h_H^{-j-1} \sum_{i=1}^n K(h_K^{-1}d(x, X_i)) H^{(j+1)}(h_H^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))}. \quad (2.8)$$

Similar estimate was already introduced in the special case when  $X$  is a real random variable by Rosenblatt (1969), and widely studied since this time (see for instance Youndjé, 1996, for several asymptotic results and references). In addition to the conditions introduced along the previous section, we need the following ones :

$$(H8) \quad \begin{cases} \forall (y_1, y_2) \in \mathbb{R}^2, |H^{(j+1)}(y_1) - H^{(j+1)}(y_2)| \leq C|y_1 - y_2|, \\ \exists \nu > 0, \forall j' \leq j, \lim_{y \rightarrow \infty} |y|^{1+\nu} |H^{(j'+1)}(y)| = 0, \end{cases}$$

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

The next result concerns the asymptotic behaviour of the kernel functional estimator  $\widehat{f}^{x(j)}$ .

**Theorem 2.3.2** *Under the hypotheses (H1), (H3)-(H5) and (H7)-(H9) we have*

$$\sup_{y \in S} |\widehat{f}^{x(j)}(y) - f^{x(j)}(y)| = O(h_K^{b_1}) + O(h_H^{b_2}) + O\left(\sqrt{\frac{\log n}{n h_H^{2j+1} \phi_x(h_K)}}\right), \quad a.co. \quad (2.9)$$

**Proof of Theorem 2.3.2.** This result is based on the same kind of decomposition as (2.3). Indeed, we can write :

$$\begin{aligned} \widehat{f}^{x(j)}(y) - f^{x(j)}(y) &= \frac{1}{\widehat{F}_D^x} \left\{ \left( \widehat{f}_N^{x(j)}(y) - E \widehat{f}_N^{x(j)}(y) \right) - \left( f^{x(j)}(y) - E \widehat{f}_N^{x(j)}(y) \right) \right\} + \\ &\quad \frac{f^{x(j)}(y)}{\widehat{F}_D^x} \left\{ E \widehat{F}_D^x - \widehat{F}_D^x \right\} \end{aligned} \quad (2.10)$$

where

$$\widehat{f}_N^{x(j)}(y) = \frac{1}{n h_H^{(j+1)} E K_1} \sum_{i=1}^n K(h_K^{-1} d(x, X_i)) H^{j+1}(h_H^{-1}(y - Y_i)).$$

Then, Theorem 2.3.2 can be deduced from both following lemmas, together with Lemma 2.3.1 and Corollary 2.3.1.

**Lemma 2.3.4** *Under the hypotheses (H1), (H3)-(H5), (H8) and (H9) we have*

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |f^{x(j)}(y) - E \widehat{f}_N^{x(j)}(y)| = O(h_K^{b_1}) + O(h_H^{b_2}), \quad a.co. \quad (2.11)$$

**Lemma 2.3.5** *Under the hypotheses (H1), (H3)-(H5) and (H7)-(H9), we have*

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{f}_N^{x(j)}(y) - E \widehat{f}_N^{x(j)}(y)| = O\left(\sqrt{\frac{\log n}{n h_H^{2j+1} \phi_x(h_K)}}\right), \quad a.co. \quad (2.12)$$

Now, let  $\widehat{f}^x = \widehat{f}^{x(0)}$ ; it is obvious that the previous theorem allows us to get the following corollary. ■

**Corollary 2.3.2** *Suppose that hypotheses (H3) and (H8)-(H9) are verified for  $j = 0$ . Then, if (H1), (H4)-(H5) and (H7) are satisfied, we have*

$$\sup_{y \in S} |\widehat{f}^x(y) - f^x(y)| = O(h_K^{b_1}) + O(h_H^{b_2}) + O\left(\sqrt{\frac{\log n}{n h_H \phi_x(h_K)}}\right), \quad a.co. \quad (2.13)$$

## 2.4 Estimation of some real conditional parameters

### 2.4.1 Conditional mode estimation

In this section we will consider the problem of the estimation of the conditional mode denoted by  $\theta$ . Saying that, it is implicitly assumed that the compact set  $S$  is chosen such that the mode  $\theta$  is uniquely defined. More precisely, we will assume that

$$(H10) \quad \exists \xi > 0, f^x \nearrow \text{ on } (\theta - \xi, \theta) \text{ and } f^x \searrow \text{ on } (\theta, \theta + \xi).$$

In the following we will use as compact set  $S = (\theta - \xi, \theta + \xi)$ . The literature on conditional mode estimation is quite important when the explanatory variable  $X$  is real (see for instance Collomb *et al.*, 1987, for previous works and Quintela and Vieu, 1997, Berlinet *et al.*, 1998, or Louani and Ould-Saïd, 1999, for recent advances and references). Concerning the functional case, Gasser *et al.* (1998) proposed an estimate of the mode of a distribution of random curves (i.e. *frv*), but nothing about conditional distribution. Our estimate is based on the previous functional conditional density estimate. We propose the estimate  $\hat{\theta}$  of the conditional mode  $\theta$  defined as

$$\hat{f}^x(\hat{\theta}) = \sup_{y \in S} \hat{f}^x(y). \quad (2.14)$$

Note that the estimate  $\hat{\theta}$  is not necessarily unique, and if this is the case all the remaining of our paper will concern any value  $\hat{\theta}$  satisfying (2.14). The difficulty of the problem is naturally linked with the flatness of the function  $f^x$  around the mode  $\theta$ . This flatness can be controled by the number of vanishing derivatives at point  $\theta$ , and this parameter will also have a great influence on the asymptotic rates of our estimate. More precisely, we introduce the following additional smoothness condition. We suppose that there exists some integer  $j > 0$  such that :

$$(H11) \quad f^x \text{ is } j\text{-times continuously differentiable with respect to } y \text{ on } (\theta - \xi, \theta + \xi),$$

and

$$(H12) \quad \begin{cases} f^{x^{(l)}}(\theta) = 0, \text{ if } 1 \leq l < j, \\ \text{and } 0 < |f^{x^{(j)}}(\theta)|. \end{cases}$$

While these additional conditions will have great influence on the rates of convergence of our estimate (see Theorem 2.4.1 below), let us note that the convergence of the estimator can be obtained under the minimal assumption (H10) (see Lemma 2.4.1 below).

**Theorem 2.4.1** *If the conditions of Theorem 2.3.2 hold together with (H10)-*

(H12), we have

$$\hat{\theta} - \theta = O\left(h_K^{\frac{b_1}{j}} + h_H^{\frac{b_2}{j}}\right) + O\left(\left(\frac{\log n}{n h_H \phi_x(h_K)}\right)^{\frac{1}{2j}}\right), \quad a.co. \quad (2.15)$$

**Proof of Theorem 2.4.1.** Note first that by simple analytic arguments, we can show that

$$|f^x(\hat{\theta}) - f^x(\theta)| \leq 2 \sup_{y \in S} |\hat{f}^x(y) - f^x(y)|. \quad (2.16)$$

Let us now write the following Taylor expansion of the function  $f^x$  :

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

for some  $\theta^*$  between  $\theta$  and  $\hat{\theta}$ . Because of (2.16), as long as we could be able to check that

$$\exists \tau > 0, \quad \sum_{n=1}^{n=\infty} P\left(f^{x(j)}(\theta^*) < \tau\right) < \infty, \quad (2.17)$$

we would have

$$(\hat{\theta} - \theta)^j = O\left(\sup_{y \in S} |\hat{f}^x(y) - f^x(y)|\right), \quad a.co., \quad (2.18)$$

and by combining (2.18) together with Theorem 2.3.2 we would get the claimed result. So it suffices to check (2.17), and this is done directly by using the second part of (H12) together with the following lemma.

**Lemma 2.4.1** *If the conditions of Theorem 2.3.2 hold together with (H10), then we have :*

$$\hat{\theta} - \theta \rightarrow 0, \quad a.co. \quad (2.19)$$

■

## 2.4.2 Conditional quantiles estimation

Now we will consider the problem of the estimation of the conditional quantiles. Let  $\alpha$  be fixed,  $0 < \alpha < 1$ , and consider the problem of estimating the parameter  $t_\alpha$  defined by

$$F^x(t_\alpha) = \alpha.$$

To insure existence and unicity of this quantile, we will assume that



(H13)  $F^x(\cdot)$  is strictly increasing.

Naturally, we will estimate this quantile by mean of the conditional distribution estimator studied in previous sections. Here also, as far as we know, the literature on (conditional and/or unconditional) quantile estimation is quite important when the explanatory variable  $X$  is real (see for instance Samanta, 1989, for previous results and Berline *et al.*, 2001, for recent advances and references). In the functional case, as far as we know, the only work existing concerns (unconditional) median estimation and was provided by Cadre (2001). We propose to estimate  $t_\alpha$  by the estimate  $\hat{t}_\alpha$  defined as

$$\widehat{F}^x(\hat{t}_\alpha) = \alpha. \quad (2.20)$$

In order to insure unicity of  $\hat{t}_\alpha$  we will make the following, quite unrestrictive, assumption :

(H14)  $H$  is strictly increasing.

As for the mode estimation problem discussed before, the difficulty occurring in estimating the conditional quantile  $t_\alpha$  is linked with the flatness of the curve of the conditional distribution  $F^x$  around  $t_\alpha$ . More precisely, we will suppose that there exists some integer  $j > 0$  such that :

(H15)  $F^x(\cdot)$  is  $j$ -times continuously differentiable in some neighborhood of  $t_\alpha$ ,

and

$$(H16) \begin{cases} F^{x(l)}(t_\alpha) = 0, \text{ if } 1 \leq l < j, \\ 0 < |F^{x(j)}(t_\alpha)|. \end{cases}$$

**Theorem 2.4.2** *If the conditions of Theorem 2.3.2 hold together with (H13)-(H16), we have*

$$\hat{t}_\alpha - t_\alpha = O\left(h_K^{\frac{b_1}{j}} + h_H^{\frac{b_2}{j}}\right) + O\left(\left(\frac{\log n}{n \phi_x(h_K)}\right)^{\frac{1}{2j}}\right), \quad a.co. \quad (2.21)$$

**Proof of Theorem 2.4.2.** Let us write the following Taylor expansion of the function  $\widehat{F}^x$  :

$$\begin{aligned}\widehat{F}^x(t_\alpha) - \widehat{F}^x(\hat{t}_\alpha) &= \sum_{l=1}^{j-1} \frac{(t_\alpha - \hat{t}_\alpha)^l}{l!} \widehat{F}^{x(l)}(t_\alpha) \\ &+ \frac{(t_\alpha - \hat{t}_\alpha)^j}{j!} \widehat{F}^{x(j)}(t^*),\end{aligned}$$

where  $t^*$  is some point between  $t_\alpha$  and  $\hat{t}_\alpha$ . It suffices now to use the first part of condition (H16) to be able to rewrite this expression as :

$$\begin{aligned}\widehat{F}^x(t_\alpha) - \widehat{F}^x(\hat{t}_\alpha) &= \sum_{l=1}^{j-1} \frac{(t_\alpha - \hat{t}_\alpha)^l}{l!} \left( \widehat{f}^{x(l-1)}(t_\alpha) - f^{x(l-1)}(t_\alpha) \right) \\ &+ \frac{(t_\alpha - \hat{t}_\alpha)^j}{j!} \widehat{f}^{x(j-1)}(t^*).\end{aligned}$$

As long as we could be able to check that

$$\exists \tau > 0, \sum_{n=1}^{n=\infty} P \left( \widehat{f}^{x(j-1)}(t^*) < \tau \right) < \infty, \quad (2.22)$$

we would have

$$\begin{aligned}(t_\alpha - \hat{t}_\alpha)^j &= O \left( (\widehat{F}^x(t_\alpha) - F^x(t_\alpha)) \right) \\ &+ O \left( \sum_{l=1}^{j-1} (t_\alpha - \hat{t}_\alpha)^l (\widehat{f}^{x(l-1)}(t_\alpha) - f^{x(l-1)}(t_\alpha)) \right), \text{ a.co.}\end{aligned} \quad (2.23)$$

By comparing the rates of convergence given in Theorems 2.3.1 and 2.3.2, we see that the leading term in right hand side of equation (2.23) is the first one. So we have

$$(t_\alpha - \hat{t}_\alpha)^j = O \left( \widehat{F}^x(t_\alpha) - F^x(t_\alpha) \right), \text{ a.co.} \quad (2.24)$$

Because of Theorem 2.3.2, this is enough to get the claimed result, and so (2.22) is the only result that remains to check. This will be done directly by using the continuity of the function  $f^{x(j-1)}$  given by (H15) together with the second part of (H12) and with the following lemma.

**Lemma 2.4.2** *If the conditions of Theorem 2.3.1 hold together with (H13) and (H14), then we have :*

$$\hat{t}_\alpha - t_\alpha \rightarrow 0, \text{ a.co.} \quad (2.25)$$

■

## 2.5 The continuous time processes framework

### 2.5.1 The small ball probability effects

Our aim in this section, is to discuss the hypothesis that we had introduced before for our general methodology. All the results before have the same structure, since our rates of convergence are splitted into two parts. Because of this common structure, we will only discuss here the Theorem 2.3.1. The same kind of discussion can trivially be made also for which concerns Theorems 2.3.2, 2.4.1 and 2.4.2, as well as for other results existing in regression and density estimation (see Ferraty and Vieu, 2004 or Niang, 2002b). So, let us look carefully at the results of Theorem 2.3.1.

First of all, note that the statistical model studied in Section 2.3.1 is a doubly infinite structure, since both the target (*i.e.* the function  $F^x(y)$  to be estimated) and the observed data (*i.e.* the variable  $X$ ) are supposed to belong to some infinitely dimensioned space. As usually in nonparametric problems, the infinite dimension of the target function is controled by mean of a smoothness condition (namely here, the condition (H2)), and this has only an influence on the bias component of the rates of convergence (namely, the terms  $O(h_K^{b_1})$  and  $O(h_H^{b_2})$  in Theorem 2.3.1). The other term appearing in the rates in Theorem 2.3.1 comes directly from dispersion effects (see Lemmas 2.3.1 and 2.3.3), and it is therefore naturally linked with the concentration properties of the probability measure of the variable  $X$ . More precisely, these terms are in

$$O(\sqrt{(\log n)/(n\phi_x(h_K))}) \text{ a.co.},$$

where the small ball probabilities are quantified and controled by mean of the function  $\phi_x(\cdot)$  defined in condition (H1). Of course, the less concentrated is the measure of the process  $X$  then the slower will be the rate of convergence. In the special multi- (but finite) -dimensional situation where  $X \in \mathbb{R}^d$ , the

condition (H1) is satisfied under standard assumptions with  $\phi_x(h) \sim C_x h^d$  and this is commonly known as the curse of dimensionality (see Ferraty and Vieu, 2003, for details). Here we would rather speak about the *curse of infinite dimension*, or to be more precise about the *effects of small ball probabilities*.

This small ball probability effects are really inherent to our infinite dimensional context, and we will devote the following of this section to discuss how our methodology can be applied to several different continuous time processes for which the small ball probabilities are known. We will concentrate our discussion on two kinds of situations, namely on diffusion processes and on Gaussian processes. Our discussion will be only based on (H1) since it is the only condition really linked with the infinite dimension of  $X$  (as we told before, (H2) is only linked with the nonparametric structure of the model while (H3)-(H7) do not concern the statistical model but only the parameters of our estimate). Our nonparametric free-distribution approach allows to apply our results to many parametric processes, but to highlight our point of view we will start each family of examples (diffusion processes and Gaussian processes) with a standard example (Ornstein-Uhlenbeck Process and Fractional Brownian Motion).

### 2.5.2 Application to diffusion processes

→ Example 1 : The standard Ornstein-Uhlenbeck process. Let us consider the space  $\mathcal{C}([0, 1], \mathbb{R})$  of all the real valued continuous functions defined on  $[0, 1]$ , and consider the metric  $d(., .)$  associated with the supremum norm

$$\forall x \in \mathcal{C}([0, 1], \mathbb{R}), \quad \|x\|_{sup} = \sup_{t \in [0, 1]} |x(t)|.$$

Denote by  $P^W$  the Wiener measure on  $\mathcal{C}([0, 1], \mathbb{R})$ , and let us take as functional space the Cameron-Martin space<sup>2</sup> of  $\mathcal{C}([0, 1], \mathbb{R})$ , that we will denote by :

---

<sup>2</sup>Let us note at this stage that there are many ways for defining the Cameron Martin subspace  $\mathcal{X}^{CM}$  (also called the reproducing kernel Hilbert space) of a locally convex space  $\mathcal{X}$  with some gaussian measure  $\gamma$  (see for instance Bogachev, 1999, Chapter 2). For our purpose, we will use here the following characterization, proved in Theorem 2.4.5 of Bogachev (1999, p.61) :  $\mathcal{X}^{CM} = \{x \in \mathcal{X}, \gamma_x \sim \gamma\}$ , where  $\gamma_x$  denotes the translated measure  $\gamma_x(\cdot) = \gamma(\cdot - x)$ .

$$\mathcal{F} = \mathcal{C}([0, 1], \mathbb{R})^{CM}.$$

We denote by  $w$  the standard Wiener process, and we consider the Ornstein-Uhlenbeck process  $\zeta^{OU}$  defined by  $\zeta_0^{OU} = 0$  and by

$$d\zeta_t^{OU} = dw_t - \frac{1}{2}\zeta_t^{OU}, \forall t, 0 < t \leq 1.$$

Note that the small centered ball Wiener measures are known to be of the form (Bogachev, 1999, p.187) :

$$P^W(\|x\|_{sup} \leq h) \sim \frac{4}{\pi} e^{-\frac{\pi^2}{8h^2}}.$$

In addition, by construction and characterization of the Cameron-Martin space (see discussion before) we can extend this result to any small ball probability measure, and we have :

$$\forall x_0 \in \mathcal{F}, P^W(\|x - x_0\|_{sup} \leq h) \sim C_{x_0} e^{-\frac{\pi^2}{8h^2}}.$$

Because the process  $\zeta^{OU}$  has a probability measure which is absolutely continuous with respect to  $P^W$ , we can directly write that

$$\forall x_0 \in \mathcal{F}, P(\zeta^{OU} \in \mathcal{B}(x_0, h)) \sim C_{x_0} e^{-\frac{\pi^2}{8h^2}}.$$

Finally, it is easy to see that our crucial condition (H1) is trivially satisfied for this Ornstein-Uhlenbeck process by choosing our function  $\phi_x$  of the form

$$\phi_x^{OU}(h) \sim C_x e^{-\frac{\pi^2}{8h^2}}. \quad (2.26)$$

Now, it suffices to choose as parameters for our estimate :

$$h_K \sim \eta_H (\log n)^{-\frac{1}{2}}, \text{ for some } \eta_H \text{ large enough,}$$

and

$$h_H \sim n^{-\epsilon}, \text{ for some } \epsilon > 0,$$

to see that the conditions (H6) and (H7) are fulfilled. Finally, we arrive by this way to a rate of convergence in our Theorem 2.3.1 in

$$O(\log n)^{-\frac{b_1}{2}}, \text{ a.co.} \quad (2.27)$$

→ Example 2 : General diffusion processes. The same ideas as those developed in the previous example can be used in much more general diffusion settings. Because of the similarity with the steps described in Example 1 before, and also because the reader can find much more details in Niang (2002b) who investigated these difusions processes in a different functional estimation context, we will be very short in the description of this example. Let us consider the space  $\mathcal{C}([0, 1], \mathbb{R}^p)$  and its Cameron-Martin associated space :

$$\mathcal{F} = \mathcal{C}([0, 1], \mathbb{R}^p)^{CM},$$

where the metric  $d(., .)$  is still the one associated with the supremum norm. Let us consider some diffusion process  $\zeta^{Diff}$  that can be written on the following usual form :

$$\zeta_t^{Diff} = \int_0^t \beta(s, \zeta^{Diff}) ds + w_t,$$

where  $w$  is the standard Wiener process and where  $\beta$  is such that the solution of the above equation has an unique solution  $\zeta^{Diff}$  (examples of such functions  $\beta$  can be found for instance in Niang, 2002b, p. 47 or in Banon, 1978). In Lipster and Shirayev (1977), it is shown that if the condition

$$\int_0^1 \beta^2(s, \zeta^{Diff}) dt < \infty, w.p.1,$$

holds, then the process  $\zeta^{Diff}$  is absolutely continuous with respect to the Wiener measure  $P^W$ . So finally, we can follow exactly the same steps as in Example 1 before, to show that our condition (H1) is always satisfied with a function  $\phi_x$  such that

$$\phi_x^{Diff}(h) \sim C_x e^{-Ch^{-2}}, \quad (2.28)$$

and by suitable bandwidth choices we arrive to a rate of convergence in our Theorem 2.3.1 of the form

$$O(\log n)^{-\frac{b_1}{2}}, a.co. \quad (2.29)$$

→ Complementary discussion. For seak of shortness, we have only discussed the case of the supremum norm. Indeed, similar applications could be carried out with metrics  $d(., .)$  associated with other kind of norms, such as for instance the  $L_q$  norm,  $0 < q < \infty$  :

$$\|x\|_{L_q} = \left( \int_0^1 |x(t)|^q \right)^{\frac{1}{q}},$$

or the Hölder norm of order  $\alpha$

$$\|x\|_\alpha = \sup_{t \neq s} \frac{|x(t) - x(s)|}{|s - t|^\alpha},$$

or the more general  $f$ -norm associated with a real valued function  $f$  defined by

$$\|x\|_f = \sup_{t \neq s} \frac{|x(t) - x(s)|}{f(|s - t|)}.$$

Indeed, by following the same steps as before and by using the small ball probability expressions given in Bogachev (1999, p. 202), we can apply our general methodology with these new distances leading to rates of convergence of order

$$O(\log n)^{-\frac{b_1}{2}}, \text{ a.co., for any } \|\cdot\|_{L_q} \text{ metric with } q \geq 1, \quad (2.30)$$

or of order

$$O(\log n)^{-\frac{(1-2\alpha)b_1}{2}}, \text{ a.co., for any } \|\cdot\|_\alpha \text{ metric with } 0 < \alpha < \frac{1}{2}. \quad (2.31)$$

Note also finally that similar application can be possible for some weighted version of the supremum norm and by using small ball probabilities as given by Shi (1996).

### 2.5.3 Application to Gaussian paths

→ Example 3 : Fractional Brownian Motion. Let consider the space  $\mathcal{C}([0, 1], \mathbb{R})$  with the supremum norm, and consider its Cameron-Martin associated space

$$\mathcal{F} = \mathcal{C}([0, 1], \mathbb{R})^{CM}.$$

We will look at the usual Fractional Brownian Motion  $\zeta^{FBM}$  of parameter  $\delta$ ,  $0 < \delta < 2$ . The small ball probabilities in this setting have been widely studied, and for instance according to Li and Shao (2001, Theorems 3.1 and 4.6), we have that

$$\forall x_0 \in \mathcal{F}, C'_{x_0} e^{h^{-\frac{2}{\delta}}} \leq P(\|\zeta^{FBM} - x_0\|_{sup} \leq h) \leq C_{x_0} e^{h^{-\frac{2}{\delta}}}.$$

Finally, it is easy to see that our crucial condition (H1) is trivially satisfied for the Fractional Brownian Motion process by choosing our function  $\phi_x$  of the form

$$\phi_x^{FBM}(h) \sim C_x e^{h^{-\frac{2}{\delta}}}. \quad (2.32)$$

Now, it suffices to choose as parameters for our estimate :

$$h_K \sim \eta_H (\log n)^{-\frac{\delta}{2}}, \text{ for } \eta_H \text{ large enough,}$$

and

$$h_H \sim n^{-\epsilon}, \text{ for some } \epsilon > 0,$$

to see that the conditions (H6) and (H7). Finally, we arrive by this way to a rate of convergence in our Theorem 2.3.1 which is of order

$$O(\log n)^{-\frac{\delta b_1}{2}}, \text{ a.co.} \quad (2.33)$$

→ Example 4 : General Gaussian Process. The same ideas as those developed in the previous example can be used for general Gaussian processes. Let us consider the space  $\mathcal{C}([0, 1]^p, \mathbb{R})$  and its Cameron-Martin associated space :

$$\mathcal{F} = \mathcal{C}([0, 1]^p, \mathbb{R})^{CM},$$

where the metric  $d(., .)$  is still the one associated with the supremum norm. Let us consider some Gaussian process  $\zeta^{GP}$  such that

$$\forall t \neq s \in [0, 1]^p, E|\zeta_t^{GP} - \zeta_s^{GP}| \leq C \|t - s\|_{\mathbb{R}^p}^{\delta}.$$

The small centered ball probabilities are given for instance in Corollary 4.10.7 of Bogachev (1999) and can be easily extended to uncentered balls by using the general translation result given in Theorem 4.6 of Li and Shao (2001), in such a way that we have :

$$\forall x_0 \in \mathcal{F}, P(\|\zeta^{GP} - x_0\|_{sup} \leq h) \geq C_{x_0} e^{h^{-\frac{2d}{\delta}}}.$$

So finally, we can follow exactly the same steps as in Example 3 before, to show that our condition (H1) is always satisfied with a function  $\phi_x$  such that

$$\phi_x^{GP}(h) \sim C_x e^{-Ch^{-\frac{2p}{\delta}}}, \quad (2.34)$$

and by suitable bandwidth choices we arrive to a rate of convergence in our Theorem 2.3.1 of the form

$$O(\log n)^{-\frac{\delta b_1}{2p}}, \text{ a.co.} \quad (2.35)$$

Obviously, this result allows to find again our two simple introductory examples (see Examples 1 and 3), but it also applies to many other known processes



such for instance the Lévy Fractional Brownian Motion of order  $\delta$ ,  $0 < \delta < 2$  (see *eg* Li and Shao, 2001), the fractional Ornstein-Uhlenbeck differenced process of parameter  $\delta$  (see *eg* Bogachev, 1999, Example 4.10.9), . . .

→ Complementary discussion. As before for diffusion processes, we could directly apply our methodology with other metrics. For instance, by using the small ball probability expression given in Theorem 4.10.6 of Bogachev (1999) we would arrive to Holder-norms rates of convergence of order

$$O(\log n)^{-b_1 \frac{(\delta-2\alpha)}{2p}}, \text{ a.co.} \quad (2.36)$$

for any  $\|\cdot\|_\alpha$  metric such that  $0 < \alpha < \frac{\delta}{2}$ , and more generally we would arrive to f-norm rates of order

$$O(\log n)^{-b_1 \frac{(\delta-\delta')}{2p}}, \text{ a.co.}, \quad (2.37)$$

for any  $\|\cdot\|_f$  metric associated with a function  $f$  satisfying

$$\exists u_0 > 0, \exists C_f > 0, \forall u \in (0, u_0), f(u) \geq C|u|^{2\delta'}, \text{ for some } 0 < \delta' < \delta.$$

#### 2.5.4 A statistical solution to the curse of infinite dimension

After the previous probabilistic point of views, let us go back to the statistical aspects of our methodology. In all the examples before (see the results (2.27), (2.29), (2.30), (2.31), (2.33), (2.35), (2.36) or (2.37)), we find rates of convergence as power of  $\log n$ . This is not surprising in this doubly infinitely dimensional context, since all the previous small ball probabilities have exponential forms. In fact, for most of known processes the expressions of the small ball probabilities would lead to such a slow rate of convergence (see Li and Shao, 2001, for a survey on existing small ball probability results). This is even less surprising if we keep in mind the fact that, in the finite dimensional case, the rates of convergence are dramatically deteriorating with the dimension (see discussion and references in Section 2.5.1). Nevertheless, the small ball probability  $P(X \in B(x, h)) = \phi_x(h)$  depends crucially on the semi-metric  $d(\cdot, \cdot)$ . Hence, from a statistical point of view, he could be interesting to look for some semi-metric that increases the concentration of  $\phi_x$

around 0. We will provide below some examples of semi-metrics for which this *curse of infinite dimension* can receive some satisfactory solution. Indeed, the idea below would not be usable by staying in the usual Hilbertian or Banach context, and this point will highlight the interest of the semi-metric modelization that we have introduced.

→ Example 5 : Gaussian processes and Karhunen-Loeve projection semi-metric.

Let us consider some centered Gaussian process  $\zeta^{GP} = \{\zeta_t^{GP}, 0 \leq t \leq 1\}$ . This process can be expanded in terms of the following Karhunen-Loeve decomposition :

$$\zeta_t^{GP} = \sum_{i=1}^{\infty} \sqrt{\lambda_i} W_i f_i(t),$$

where the  $\lambda_i$ 's are the eigenvalues of the covariance operator of  $\zeta^{GP}$ , where the  $f_i$ 's are the associated orthonormal eigenfunctions and where the  $W_i$ 's are independent real random variables with  $\mathcal{N}(0, 1)$  distribution. For any fixed  $k \in \mathbb{N}^*$ , let  $\Pi_k$  the orthogonal projection onto the subspace spanned by the eigenfunctions  $\{f_1, \dots, f_k\}$ . We define a semi-metric by :

$$d^2(x, y) = \int_0^1 [\Pi_k(x - y)(t)]^2 dt.$$

Using the Karhunen-Loeve expansion, we can write

$$d^2(\zeta^{GP}, x) = \sum_{i=1}^k (\sqrt{\lambda_i} W_i - x_i)^2 = \sum_{i=1}^k Z_i^2,$$

where  $x_i = \int_0^1 x(t) f_i(t) dt$ . Thus, the quantity  $d^2(\zeta^{GP}, x)$  can be expressed in terms of the usual euclidian norm on  $\mathbb{R}^k$  of a vector  $Z = (Z_1, \dots, Z_k)$ . Because the  $Z_i$ 's are independent real random variables, each having a density with respect to Lebesgues measure, it is easy to see that

$$P(d^2(\zeta^{GP}, x) < h) \sim C_x h^k.$$

Considering now this small ball probability expression and with suitable choices for  $h_H$  and  $h_K$ , the rate of convergence in Theorem 2.3.1 becomes :

$$O\left(\left(\frac{n}{\log n}\right)^{-2b_1/(2b_1+k)}\right), \text{ a.co.}$$

→ Example 6 : General family of projection semi-metrics. Let us now consider  $X$  as a *frv* valued in a separable Hilbert space  $\mathcal{H}$  with inner product  $\langle \cdot, \cdot \rangle$  and let  $\{f_i\}_{i \in \mathbb{N}^*}$  be an orthonormal basis. We can construct a family of semi-metrics  $d_k(\cdot, \cdot)$  based on the projections onto the subspaces spanned by  $\{f_i, \dots, f_k\}$  for any  $k \in \mathbb{N}^*$  and defined by :

$$d_k^2(x, y) = \sum_{i=1}^k \langle x - y, f_i \rangle^2 .$$

By similar argument as for Example 5, we can consider as semi-metric space

$$(\mathcal{F}, d) = (\mathcal{H}, d_k),$$

and we can derive the following rate of convergence in Theorem 2.3.1 :

$$O\left(\left(\frac{n}{\log n}\right)^{-2b_1/(2b_1+k)}\right), \text{ a.co.}$$

## 2.6 Application to chemometrical functional data

The aim of this short section is to show through some chemometrical real data set that, in addition to the nice asymptotic behaviour described before, our methods are easy to implement and may provide interesting result on finite sample. Because practical aspects are not the main purpose of our paper, this section will be quite short and we will just discuss the interest of the conditional functional mode methodology.

Spectrometric is an usual tool to analyse chemical composition of some element. The result of some spectrometric analysis consists in a (continuous) spectrum of light absorbance for different wavelenghts. It is well-known that, both because of the smooth shape of the spectra and because of the high number of discretized wavelenghts measurements, such type of data has to be treated as curve data. This is quite clear in the data set to be considered here, and which is given by a set of continuous spectrum observed for 215 different pieces of meat (see Figure 2.1).

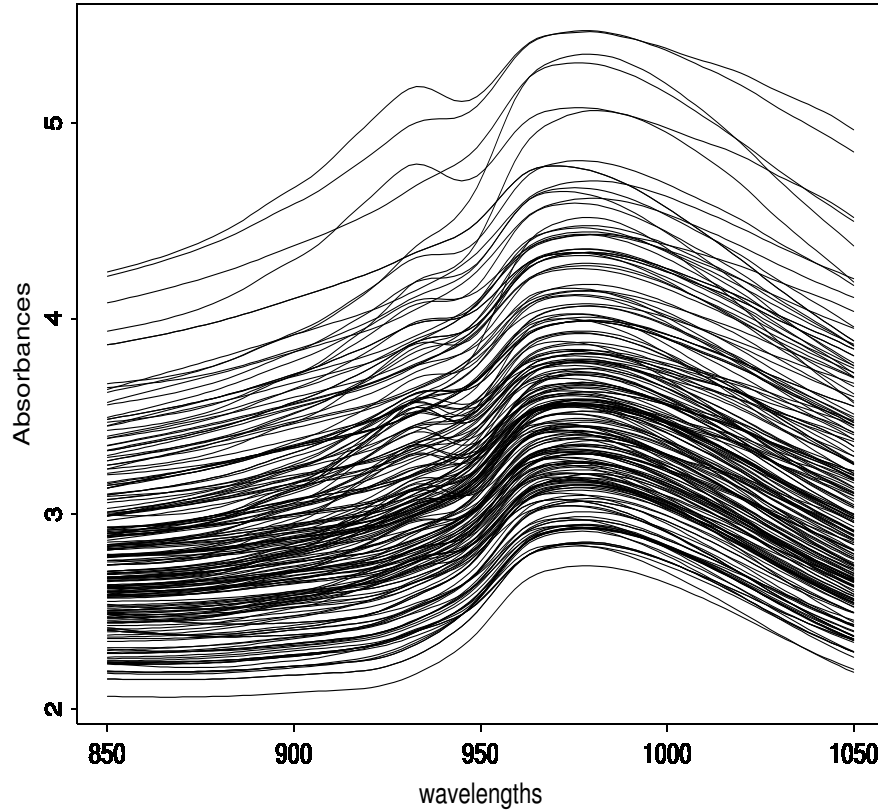


FIG. 2.1 – The 215 Spectrometric curves

Let us denote by  $X_i = \{X_i(\lambda), \lambda \in (850, 1050)\}$  each among these functional data. Together with each spectrum  $X_i$ , we have also at hand the value of some real response variable  $Y_i$  which is the percentage of fat in the corresponding piece of meat. The question which is posed in force in practice is the following one : *given some new spectra  $x_{new}$ , can we predict the corresponding percentage of fat  $y_{new}$  ?* This is a typical prediction problem from infinite dimensional explanatory variable.

This problem is usually attacked by mean of regression method, and Ferraty and Vieu (2002) have shown how a nonparametric estimate (denoted

from now on by  $\widehat{R}$ ) of the functional operator

$$R(x) = E(Y|X = x),$$

may improve substantially upon classical finite dimensional approaches. With this notation, the prediction of  $y_{new}$  obtained by this functional regression approach will be denoted by

$$\widehat{y}_{new} = \widehat{R}(x_{new}).$$

However, an other way to solve this prediction problem is to use the conditional mode approach described before in our paper. More precisely, from the functional conditional density estimate defined in (2.8), we can predict  $y_{new}$  by computing the quantity

$$\widetilde{y}_{new} = \arg \max_y \widehat{f}^{x_{new}}(y).$$

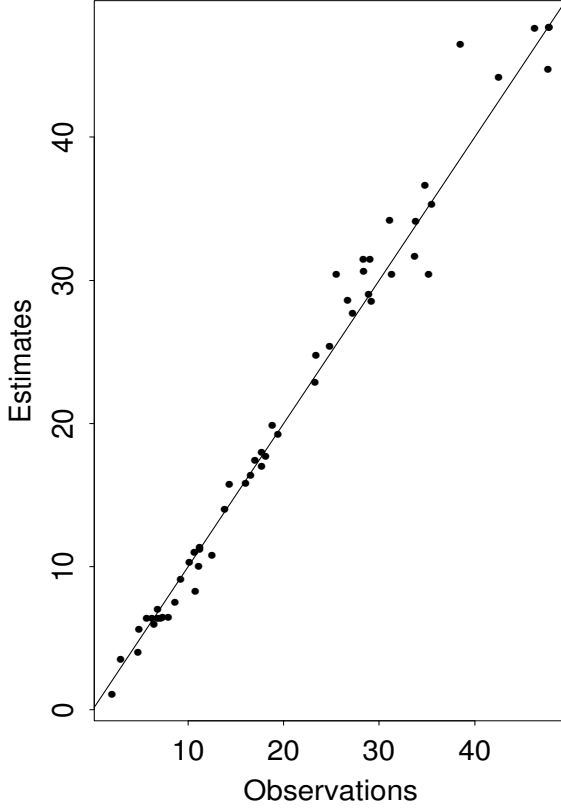
For the same reasons as in finite dimensional data, the conditional approach can be expected to be more robust than the regression approach.

In order to compare these two predictors, we have randomly splitted our data into two subsets. The first sample, of size  $n = 165$ , will be our statistical sample from which the estimates are calculated<sup>3</sup>. From this subsample, we compute the predicted values  $\widehat{y}_{new}$  and  $\widetilde{y}_{new}$ , corresponding to the 50 remaining curves (i.e. the 50  $x_{new}$ 's). The results are presented in Figure 2.2 below. The left part of Figure 2.2 is plotting the predicted values  $\widetilde{y}_{new}$  versus the true values  $y_{new}$ , while the right part of Figure 2.2 is plotting  $\widehat{y}_{new}$  versus  $y_{new}$ . The continuous line represents what would be the perfect prediction. Clearly, the comparison of both scatterplots is indicating that the conditional mode approach gives slightly better results than the regression approach. This is confirmed by the sum of squared residuals (RSS) whose gain is around 20 per 100.

---

<sup>3</sup>For both methods, the optimal bandwidths were chosen by cross-validation procedure and the kernel  $K$  is the quadratic and  $H' = K$ . In addition, the semimetric  $d$  is the  $L_2$  distance between the second derivatives of the curves (see Ferraty and Vieu, 2002, for motivations of this choice).

Cond. Expect.: RSS=192



Cond. Mode: RSS=145

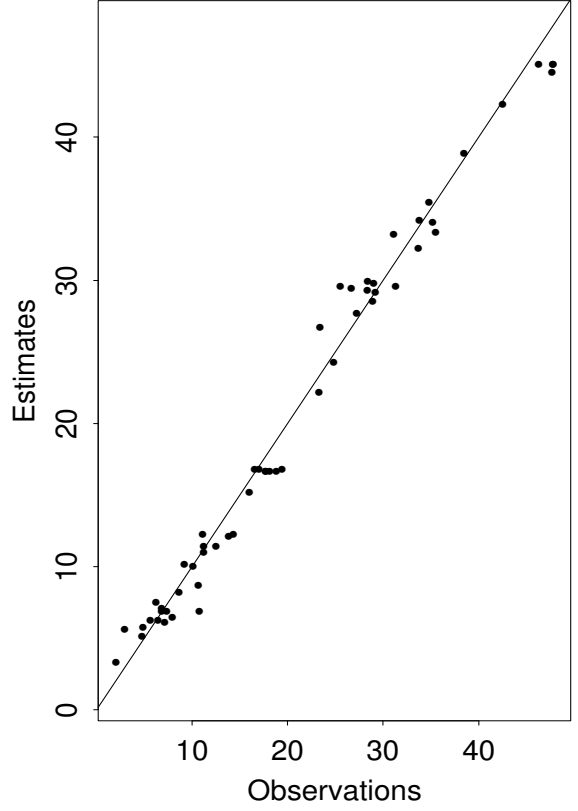


FIG. 2.2 – Comparative prediction between regression and conditional mode

## 2.7 Appendix

### 2.7.1 Proofs of lemmas of Section 2.3.1

**Proof of Lemma 2.3.1.** We have  $\hat{F}_D^x - E\hat{F}_D^x = \frac{1}{n} \sum_{i=1}^n \left( \frac{K_i}{EK_1} - 1 \right) = \frac{1}{n} \sum_{i=1}^n (\tilde{\Delta}_i - 1)$ . Because of (H.1) and (H.5) we can write

$$C\phi_x(h_K) < EK_1 < C'\phi_x(h_K).$$

So, we can get directly that :

$$\left| \tilde{\Delta}_i \right| < C/\phi_x(h_K) = \delta_1 \text{ and } E \left| \tilde{\Delta}_i \right|^2 < C'/\phi_x(h_K) = \delta_2.$$

Thus, the use of the classical Bernstein's inequality (Uspensky, 1937, p205) allows us to write for all  $\eta \in (0, \delta_2/\delta_1)$  :

$$P \left| \hat{F}_D^x - E\hat{F}_D^x \right| > \eta \sqrt{\frac{\log n}{n \phi_x(h_K)}} \leq C'n^{-C\eta^2}.$$

■

**Proof of Corollary 2.3.1.** We have

$$P \left( \left| \hat{F}_D^x \right| \leq 1/2 \right) \leq P \left( \left| \hat{F}_D^x - 1 \right| > 1/2 \right).$$

Noting that  $\hat{F}_D^x - 1 = \hat{F}_D^x - E\hat{F}_D^x$  and by applying the previous Lemma 2.3.1, we can write  $\sum_{n=1}^{\infty} P \left( \hat{F}_D^x < 1/2 \right) < \infty$ . ■

**Proof of Lemma 2.3.2.**

$$E\hat{F}_N^x(y) - F^x(y) = \frac{1}{EK_1} E (K_1 [E (H_1(y)|X) - F^x(y)]). \quad (2.38)$$

Moreover, we have

$$E (H_1(y)|X) = \int_{\mathbb{R}} H (h_H^{-1}(y - z)) f^X(z) dz,$$

and integrating by parts and using the fact that  $H$  is a *cdf*, we obtain that

$$E (H_1(y)|X) = \int_{\mathbb{R}} H^{(1)}(t) F^X(y - h_H t) dt.$$

Thus, we have

$$\left| E (H_1(y)|X) - F^x(y) \right| \leq \int_{\mathbb{R}} H^{(1)}(t) \left| F^X(y - h_H t) - F^x(y) \right| dt.$$

Now, (H2) allows to write

$$\left| E (H_1(y)|X) - F^x(y) \right| \leq C_x \int_{\mathbb{R}} H^{(1)}(t) (h_K^{b_1} + |t|^{b_2} h_H^{b_2}) dt. \quad (2.39)$$

Because this inequality is uniform on  $y$  and because of (H.4), (2.6) is a direct consequence of (2.38), (2.39) and of Corollary 2.3.1. ■

**Proof of Lemma 2.3.3.** Using the compactness of  $S$ , we can write that  $S \subset \bigcup_{k=1}^{s_n} S_k$  where  $S_k = (t_k - l_n, t_k + l_n)$ . Taking  $t_y = \arg \min_{t \in \{t_1, \dots, t_{s_n}\}} |y - t|$ , we have

$$\begin{aligned} \frac{1}{\widehat{F}_D^x} \sup_{y \in S} \left| \widehat{F}_N^x(y) - E\widehat{F}_N^x(y) \right| &\leq \underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in S} \left| \widehat{F}_N^x(y) - \widehat{F}_N^x(t_y) \right|}_{T_1} + \underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in S} \left| \widehat{F}_N^x(t_y) - E\widehat{F}_N^x(t_y) \right|}_{T_2} + \\ &\quad \underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in S} \left| E\widehat{F}_N^x(t_y) - E\widehat{F}_N^x(y) \right|}_{T_3}. \end{aligned} \quad (2.40)$$

- Concerning  $(T_1)$  :

$$\begin{aligned} \frac{1}{\widehat{F}_D^x} \sup_{y \in S} \left| \widehat{F}_N^x(y) - \widehat{F}_N^x(t_y) \right| &\leq \frac{1}{\widehat{F}_D^x} \sup_{y \in S} \frac{1}{nEK_1} \sum_{i=1}^n |H_i(y) - H_i(t_y)| K_i, \\ &\leq \frac{1}{\widehat{F}_D^x} \sup_{y \in S} \frac{C|y - t_y|}{h_H} \left( \frac{1}{nEK_1} \sum_{i=1}^n K_i \right), \\ &\leq C \frac{l_n}{h_H}. \end{aligned} \quad (2.41)$$

The second inequality is obtained by considering a Lipschitz argument whereas the last one comes from the definition of  $\widehat{F}_D^x$ . Take now

$$l_n = n^{-\alpha-1/2},$$

and note that, because of (H7), we have

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

Thus, for  $n$  large enough, we can write

$$P\left(\frac{1}{\widehat{F}_D^x} \sup_{y \in S} \left| \widehat{F}_N^x(y) - \widehat{F}_N^x(t_y) \right| > \frac{\eta}{3} \sqrt{\frac{\log n}{n\phi_x(h_K)}}\right) = 0. \quad (2.42)$$



- Concerning  $(T_2)$  :

$$\begin{aligned}
& P \left( \sup_{y \in S} \left| \widehat{F}_N^x(t_y) - E\widehat{F}_N^x(t_y) \right| > \frac{\eta}{3} \sqrt{\frac{\log n}{n \phi_x(h_K)}} \right) \\
&= P \left( \max_{t_y \in \{t_1, \dots, t_{s_n}\}} \left| \widehat{F}_N^x(t_y) - E\widehat{F}_N^x(t_y) \right| > \frac{\eta}{3} \sqrt{\frac{\log n}{n \phi_x(h_K)}} \right) \\
&\leq s_n \max_{t_y \in \{t_1, \dots, t_{s_n}\}} P \left( \left| \widehat{F}_N^x(t_y) - E\widehat{F}_N^x(t_y) \right| > \frac{\eta}{3} \sqrt{\frac{\log n}{n \phi_x(h_K)}} \right).
\end{aligned}$$

Let  $\Lambda_i = \{K_i H_i(t_y) - E(K_i H_i(t_y))\} / EK_1$ . By using similar arguments as for the proof of Lemma 2.3.1 and because  $H \leq 1$ , we deduce that  $E|\Lambda_i| \leq C/\phi_x(h_K)$  and  $E\Lambda_i^2 \leq C'/\phi_x(h_K)$ . We apply now again the Bernstein's exponential inequality to get

$$P \left( \left| \widehat{F}_N^x(y) - E\widehat{F}_N^x(y) \right| > \frac{\eta}{3} \sqrt{\frac{\log n}{n \phi_x(h_K)}} \right) \leq 2 \exp\{-C\eta^2 \log n\}.$$

Thus, by choosing  $\eta$  such that  $C\eta^2 = 3/2 + 2\alpha$ , we obtain

$$\begin{aligned}
s_n \max_{t_y \in \{t_1, \dots, t_{s_n}\}} P \left( \left| \widehat{F}_N^x(t_y) - E\widehat{F}_N^x(t_y) \right| > \frac{\eta}{3} \sqrt{\frac{\log n}{n \phi_x(h_K)}} \right) &\leq C s_n n^{-3/2-2\alpha}, \\
&\leq \frac{C}{l_n} n^{-3/2-2\alpha}.
\end{aligned}$$

Because  $l_n = n^{-\alpha-1/2}$ , we deduce from the previous inequality that

$$P \left( \sup_{y \in S} \left| \widehat{F}_N^x(t_y) - E\widehat{F}_N^x(t_y) \right| > \frac{\eta}{3} \sqrt{\frac{\log n}{n \phi_x(h_K)}} \right) \leq C n^{-1-\alpha}.$$

Finally, using Corollary 2.3.1, we obtain

$$P \left( \frac{1}{\widehat{F}_D^x} \sup_{y \in S} \left| \widehat{F}_N^x(t_y) - E\widehat{F}_N^x(t_y) \right| > \frac{\eta}{3} \sqrt{\frac{\log n}{n \phi_x(h_K)}} \right) \leq C n^{-1-\alpha}. \quad (2.43)$$

- Concerning  $(T_3)$  : because of (2.41) we have :

$$\sup_{y \in S} \left| E\widehat{F}_N^x(y) - E\widehat{F}_N^x(t_y) \right| \leq C \frac{l_n}{h_H}.$$

Using analogous arguments as for  $T_1$ , we can show for  $n$  large enough :

$$P \left( \frac{1}{\widehat{F}_D^x} \sup_{y \in S} \left| E \widehat{F}_N^x(y) - E \widehat{F}_N^x(t_y) \right| > \frac{\eta}{3} \sqrt{\frac{\log n}{n \phi_x(h_K)}} \right) = 0. \quad (2.44)$$

Now, Lemma 2.3.3 can be easily deduced from (2.40), (2.42), (2.43) and (2.44). ■

### 2.7.2 Proofs of lemmas of Section 2.3.2

**Proof of Lemma 2.3.4.** Let  $H_i^{(j+1)}(y) = H^{(j+1)}(h_H^{-1}(y - Y_i))$ , and note that we have

$$E \widehat{f}_N^{x(j)}(y) - f^{x(j)}(y) = \frac{1}{h_H^{j+1} E K_1} E \left( K_1 \left[ E \left( H_1^{(j+1)}(y) | X \right) - h_H^{j+1} f^{x(j)}(y) \right] \right). \quad (2.45)$$

Moreover,

$$\begin{aligned} E \left( H_1^{(j+1)}(y) | X \right) &= \int_{\mathbb{R}} H^{(j+1)}(h_H^{-1}(y - z)) f^X(z) dz, \\ &= - \sum_{l=1}^j h_H^l \left[ H^{(j-l+1)}(h_H^{-1}(y - z)) f^{X^{(l-1)}}(z) \right]_{-\infty}^{+\infty} \\ &\quad + h_H^j \int_{\mathbb{R}} H^{(1)}(h_H^{-1}(y - z)) f^{X^{(j)}}(z) dz. \end{aligned} \quad (2.46)$$

Condition (H8) allows us to cancel the first term in the right side of (2.46) and we can write :

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

Finally, (H3) allows to write

$$\left| E \left( H_1^{(j+1)}(y) | X \right) - h_H^{j+1} f^{x(j)}(y) \right| \leq C_x h_H^{j+1} \int_{\mathbb{R}} H^{(1)}(t) (h_K^{b_1} + |t|^{b_2} h_H^{b_2}) dt. \quad (2.47)$$

Because this last inequality is uniform on  $y$ , the use of (H4), (2.45) and Corollary 2.3.1 achieve the proof of Lemma 2.3.4. ■

**Proof of Lemma 2.3.5.** The proof of this lemma is very similar to the one used for Lemma 2.3.3. To do this, replace in (2.40)  $\widehat{F}_N^x$  with  $\widehat{f}_N^{x(j)}$  and use the same notations.

- Concerning  $(T_1)$  and  $(T_3)$  : The same arguments as those invoked in (2.41) are used;  $H$  is replaced with  $H^{(j+1)}$  and we apply the Lipschitz condition (H8). This allows us to get :

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in S} \left| \widehat{f}_N^{x(j)}(y) - \widehat{f}_N^{x(j)}(t_y) \right| \leq C \frac{l_n}{h_H^{j+2}} \text{ and } \sup_{y \in S} \left| E \widehat{f}_N^{x(j)}(y) - E \widehat{f}_N^{x(j)}(t_y) \right| \leq C \frac{l_n}{h_H^{j+2}}. \quad (2.48)$$

Take now  $l_n = n^{-\frac{3\alpha}{2} - \frac{1}{2}}$  and note that (H7) implies

$$l_n/h_H^{j+2} = o \left( \sqrt{\frac{\log n}{n h_H^{2j+1} \phi_x(h_K)}} \right). \quad (2.49)$$

- Concerning  $(T_2)$  :

$$\widehat{f}_N^{x(j)}(t_y) - E \widehat{f}_N^{x(j)}(t_y) = \frac{1}{n} \sum_{i=1}^n \underbrace{\left\{ \frac{H_i^{(j+1)}(t_y) K_i}{h_H^{j+1} E K_1} - \frac{E(H_1^{(j+1)}(t_y) K_1)}{h_H^{j+1} E K_1} \right\}}_{\Lambda_i^*}. \quad (2.50)$$

We have clearly  $|\Lambda_i^*| \leq C h_H^{-j-1} \phi_x(h_K)^{-1}$ . Now, we show that

$$E \Lambda_i^{*2} = O \left( h_H^{-2j-1} \phi_x(h_K)^{-1} \right). \quad (2.51)$$

Firstly, we can write

$$E \Lambda_i^{*2} \leq \frac{E \left( H_1^{(j+1)}(t_y)^2 K_1^2 \right)}{\left( h_H^{j+1} E K_1 \right)^2} \text{ and } E \left( H_1^{(j+1)}(t_y)^2 K_1^2 \right) = E \left( K_1^2 E \left( H_1^{(j+1)}(t_y)^2 | X \right) \right).$$

The condition (H8) implies that  $\int_{\mathbb{R}} H^{(j+1)2}(y) dy < +\infty$ . So, we have :

$$\begin{aligned} & \left| \frac{1}{h_H} E \left( H_1^{(j+1)}(t_y)^2 | X \right) - f^X(t) \int_{\mathbb{R}} H^{(j+1)2}(y) dy \right| \\ &= \left| \int_{\mathbb{R}} \frac{1}{h_H} H^{(j+1)2} \left( \frac{u}{h_H} \right) \left( f^X(t_y - u) - f^X(t_y) \right) du \right|, \end{aligned}$$

$$\begin{aligned}
&\leq \left| \int_{|u| \leq A} \frac{1}{h_H} H^{(j+1)^2} \left( \frac{u}{h_H} \right) (f^X(t_y - u) - f^X(t_y)) du \right| + \\
&\quad \left| \int_{|u| > A} \frac{1}{h_H} H^{(j+1)^2} \left( \frac{u}{h_H} \right) (f^X(t_y - u) - f^X(t_y)) du \right|, \\
&\leq \underbrace{C \sup_{|u| \leq A} |f^X(t_y - u) - f^X(t_y)|}_{B_1} + \underbrace{\sup_{|y| > A/h_H} H^{(j+1)^2}(y)}_{B_2} + \\
&\quad \underbrace{f^X(t_y) \int_{|y| > A/h_H} H^{(j+1)^2}(y) dy}_{B_3}.
\end{aligned}$$

Now, we deduce from (H8) that

$$\forall \epsilon > 0, \forall A > 0, \exists n_{A,\epsilon}, \forall n \geq n_{A,\epsilon}, B_2 + B_3 < \epsilon.$$

Moreover, because the continuity of  $f^X$ , we have

$$\forall \epsilon > 0, \exists A_\epsilon, \forall A \leq A_\epsilon, B_1 < \epsilon.$$

Thus, we obtain that

$$\lim_{n \rightarrow +\infty} \frac{1}{h_H} E \left( H_1^{(j+1)}(t_y)^2 | X \right) = f^X(t) \int_{\mathbb{R}} H^{(j+1)^2}(y) dy.$$

Finally, because  $0 < C\phi_x(h_K) < EK_1$  and  $EK_1^2 < C'\phi_x(h_K)$ , we have (2.51). Now, the Bernstein's inequality leads us to :

$$\begin{aligned}
P \left( \left| \widehat{f}^{x(j)}(t_y) - E\widehat{f}^{x(j)}(t_y) \right| > \eta \sqrt{\frac{\log n}{n h_H^{2j+1} \phi_x(h_K)}} \right) \\
\leq 2 \exp \left\{ -nC \frac{\eta^2 \log n}{n h_H^{2j+1} \phi_x(h_K)} h_H^{2j+1} \phi_x(h_K) \right\}, \\
\leq C' n^{-C\eta^2}.
\end{aligned}$$

Once again, similar arguments as those invoked for proving Lemma 2.3.3 can be used, which allow us to obtain that :

$$P \left( \sup_{y \in S} \left| \widehat{f}_N^{x(j)}(t_y) - E\widehat{f}_N^{x(j)}(t_y) \right| > \eta \sqrt{\frac{\log n}{n h_H^{2j+1} \phi_x(h_K)}} \right) \leq \frac{C'}{l_n} n^{-C\eta^2}.$$

Choosing now  $\eta$  such that  $C\eta^2 = \frac{5\alpha}{3} + \frac{3}{2}$ , we get

$$\sup_{y \in S} \left| \widehat{f}^{x(j)}(t_y) - E \widehat{f}^{x(j)}(t_y) \right| = O \left( \sqrt{\frac{\log n}{n h_H^{2j+1} \phi_x(h_K)}} \right). \quad (2.52)$$

Finally, Lemma 2.3.5 is a consequence of (2.48), (2.49), (2.52) and Corollary 2.3.1. ■

### 2.7.3 Proofs of lemmas of Section 2.4

**Proof of Lemma 2.4.1.** Because the function  $f^x$  is continuous (see (H3)) and (H.10), we have :

$$\forall \epsilon > 0, \exists \delta(\epsilon) > 0, \forall y \in (\theta - \xi, \theta + \xi), |f^x(y) - f^x(\theta)| \leq \delta(\epsilon) \Rightarrow |y - \theta| \leq \epsilon.$$

This leads directly to

$$\forall \epsilon > 0, \exists \delta(\epsilon) > 0, P \left( |\hat{\theta} - \theta| > \epsilon \right) \leq P \left( |f^x(\hat{\theta}) - f^x(\theta)| \geq \delta(\epsilon) \right). \quad (2.53)$$

It suffices now to combine (2.53) together with (2.16), to get the claimed result. ■

**Proof of Lemma 2.4.2.** Because of (H4) and (H14) the function  $\widehat{F}^x$  is continuous and strictly increasing. So, we have :

$$\forall \epsilon > 0, \exists \delta(\epsilon) > 0, \forall y, |\widehat{F}^x(y) - \widehat{F}^x(t_\alpha)| \leq \delta(\epsilon) \Rightarrow |y - t_\alpha| \leq \epsilon.$$

This leads directly to

$$\begin{aligned} \forall \epsilon > 0, \exists \delta(\epsilon) > 0, P \left( |\hat{t}_\alpha - t_\alpha| > \epsilon \right) &\leq P \left( |\widehat{F}^x(\hat{t}_\alpha) - \widehat{F}^x(t_\alpha)| \geq \delta(\epsilon) \right) \\ &= P \left( |F^x(t_\alpha) - \widehat{F}^x(t_\alpha)| \geq \delta(\epsilon) \right). \end{aligned}$$

It suffices now to use the result of Theorem 2.3.1 to get the claimed result. ■

**Acknowledgments :** All the participants of the group STAPH on Functional Statistics in the Department of Statistics and Probability of Toulouse are greatly acknowledged for their numerous comments and suggestions (activities of this group are available on line at the adress *www.lsp.univ-tlse.fr/Fp/Ferraty/staph.html*). The Chief Editor is also greatly acknowledged for his encouragements in producing the examples given Section 2.5, as well as two referees for their pertinent comments and their encouragements for providing the short real curves data application presented in Section 2.6.

### References

- Banon, G. (1978). Nonparametric identification for diffusion processes. *Siam. Journ. Control and Optim.*, **16**, 380-395.
- Berlinet, A., Gannoun, A. and Matzner-Lober, E. (1998). Normalité asymptotique d'estimateurs convergents du mode conditionnel (french). *La Rev. Canad. de Statist.*, **26**, 365-380.
- Berlinet, A., Gannoun, A. and Matzner-Lober, E. (2001). Asymptotic normality of convergent estimates of conditional quantiles. *Statistics*, **35**, 139-169.
- Besse, P., Cardot, H. and Stephenson, D. (2000) Autoregressive forecasting of some functional climatic variations. *Scand. J. of Statist.*, **27**, 673-687.
- Bogachev, V.I. (1999). *Gaussian measures*. Math surveys and monographs, **62**, Amer. Math. Soc.
- Bosq, D. (2000). *Linear processes in function spaces*, Lecture Notes in Statistics, **149**, Springer-Verlag.
- Cadre, B. (2001). Convergent estimators for the  $L_1$ -median of a Banach valued random variable. *Statistics*, , 509-521.
- Collomb, G., Härdle, W. and Hassani, S. (1987). A note on prediction via conditional mode estimation. *J. Statist. Plann. and Inf.*, **15**, 227-236.
- Fernández de Castro, B. Guillas, S. and Gonzalez Manteiga, W. (2003). Functional samples and bootstrap for the prediction of  $SO_2$  levels, Preprint.
- Ferraty, F., Goia, A. and Vieu, P. (2002). Functional nonparametric model for time series : a fractal approach to dimension reduction. *TEST*, **11**, 317-344.
- Ferraty, F. and Vieu, P. (2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. *Compte Rendus Acad. Sci. Paris*, **330**, 403-406.
- Ferraty, F. and Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Computational Statistics*, **4**, 545-564.

- Ferraty, F. and Vieu, P. (2002b). Curves discrimination : a nonparametric functional approach. *Computational Statistics and Data Analysis*, In print.
- Ferraty, F. and Vieu, P. (2003). Functional nonparametric statistics : a double infinite dimensional framework. *Current advances and trends in nonparametric statistics (Editors : M. Akritas and D. Politis)*, Elsevier, 61-79.
- Ferraty, F. and Vieu, P. (2004). Nonparametric models for functional data, with applications in regression, time series prediction and curves discrimination. *J. of Nonparametric Statistics*, **16**, 111-127.
- Gasser, T., Hall, P. and Presnell, B. (1998). Nonparametric estimation of the mode of a distribution of random curves. *J. R. Statist. Soc. B*, **60**, 681-691.
- Hall, P. and Heckman, N. (2002). Estimating and depicting the structure of the distribution of random functions, *Biometrika*, **89**, 145-158.
- Hall, P., Poskitt, P. and Presnell, D. (2001). A functional data-analytic approach to signal discrimination. *Technometrics*, **35**, 140-143.
- Li, W.V. and Shao, Q.M. (2001). Gaussian processes : inequalities, small ball probabilities and applications. In *Stochastic processes : Theory and Methods*, Ed. C.R. Rao and D. Shanbhag. Handbook of Statistics, **19**, North-Holland, Amsterdam.
- Louani, D. and Ould-Saïd, E. (1999). Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis. *J. of Nonparametric Statistics*, **11** , 413-442.
- Niang, S. (2002). Estimation de la densité dans un espace de dimension infinie : Application aux diffusions (french). *Comptes Rendus Acad. Sci. Paris, Ser. I*, **334**, 213-216.
- Niang, S. (2002b). *Sur l'estimation fonctionnelle en dimension infinie : Application aux diffusions (french)*, PhD thesis, Paris 6, France.
- Quintela del Rio, A. and Vieu, P. (1997). A nonparametric conditionnal mode estimate. *Nonparametric Statistics*, **8**, 253-266.
- Ramsay, J. and Silverman, B. (1997). *Functional Data Analysis*. Springer-Verlag.
- Ramsay, J. and Silverman, B. (2002). *Appied functional data analysis*. Springer-Verlag.
- Rosenblatt, M. (1969). Conditional probability density and regression estimators. In *Multivariate Analysis II*, Ed. P.R. Krishnaiah. Acdaemic Press, New York and London.
- Roussas, G. (1969). Nonparametric estimation of the transition distribution function of a Markov process. *Ann. Math. Statist.*, **40**, 1386-1400.

- Samanta, M. (1989). Non-parametric estimation of conditional quantiles. *Statist. Proba. Letters*, **7**, 407-412.
- Shi, Z. (1996). Small ball probabilities for a Wiener process under weighted sup-norms, with an application to the supremum of Bessel local times. *J. Theoret. Probab.*, **9**, 915-929.
- Staph, (2002). Proceedings of the working group on Functional Statistics : 2001-2002. *Techn. report, Labo. Statist. Proba. Toulouse, France*, **2002-12**. (Available on line at [www.lsp.ups-tlse.fr/Fp/Ferraty/staph.html](http://www.lsp.ups-tlse.fr/Fp/Ferraty/staph.html)).
- Uspensky, J. V. (1937). *Introduction to Mathematical Probability*. McGraw-Hill Book Company, Inc., New York and London.
- Youndjé, E. (1996). Propriétés de convergence de l'estimateur à noyau de la densité conditionnelle (french). *Rev. Roumaine Math. Pures Appl.*, **41**, 535-566.



# Chapitre 3

## Cas dépendant

### 3.1 Résultats principaux

Cette section fait l'objet d'une note *C.R.A.S.*

# Functional Time Series Prediction via Conditional Mode Estimation.

Frédéric FERRATY <sup>1</sup>, Ali LAKSACI <sup>2</sup> and Philippe VIEU <sup>3</sup>

<sup>1</sup>Corresponding author: Labo. Statist. Prob., Université P. Sabatier, 31062  
Toulouse, France. ferraty@cict.fr

<sup>2</sup>Université Djillali Liabes, Sidi Bel Abbès

<sup>3</sup>Université Paul Sabatier, Toulouse

## Abstract

This Note focuses on an estimator of the conditional mode of a scalar response  $Y$  given a functional random variable  $X$ . We start by building a kernel estimator of the conditional density of  $Y$  given  $X$ ; the conditional mode is defined as the value which maximizes this conditional density. We establish the almost complete convergence for this estimate under  $\alpha$ -mixing assumption. *To cite this article: F. Ferraty, A. Laksaci, P. Vieu, C. R. Acad. Sci. Paris, Ser. I 336 (2004).*

## Résumé

On établit la convergence presque-complète de l'estimateur du mode de la distribution d'une variable réelle  $Y$  conditionnée par une variable fonctionnelle  $X$ . Le mode conditionnel est estimé par la valeur qui maximise l'estimateur à noyau de la densité conditionnelle de  $Y$  sachant  $X$ . Des résultats asymptotiques concernant cet estimateur sont établis sous l'hypothèse  $\alpha$ -mélangeante, rendant nos résultats opérationnels en prédiction de séries chronologiques. *Pour citer cet article : F. Ferraty, A. Laksaci, P. Vieu, C. R. Acad. Sci. Paris, Ser. I 336 (2004).*

### 3.1.1 Introduction

Often, prediction of a scalar response  $Y$  knowing an explained multivariate variable  $X$  is obtained by estimating the conditional expectation of  $Y$  given  $X$  (which is the standard regression function). However, this method is not efficient in some pathological situations. For instance, this is the case when the conditional density of  $Y$  given  $X$  is either unsymmetric or has several modes. In these last cases, a pertinent predictor of  $Y$  is obtained by estimating the conditional mode (see Collomb, 1987, pp 228-229). Recent advances concerning the conditional mode can be found for instance in Berlinet *et al.* (1998) and Louani and Ould-Saïd (1999).

On the other hand, more and more works deal with nonparametric approaches for variables valued in an infinite dimensional space. Gasser *et al.* (1998) proposed an estimate of the mode of a distribution of random curves whereas Ferraty *et al.* (2002) studied a kernel estimator in the functional regression setting. Note also that a kernel estimator for functional conditional distribution has been introduced by Ferraty *et al.* (2004). An overview on nonparametric methods for functional data can be found in Ferraty and Vieu (2003). The aim of this paper is to predict a scalar response  $Y$  given an explanatory functional variable  $X$  via the conditional mode estimate under  $\alpha$ -mixing assumption.

### 3.1.2 Model

Let  $(X_i, Y_i)_{i=1, \dots, n}$  be a stationary  $\alpha$ -mixing process valued in  $\mathfrak{S} \times \mathbb{R}$  where  $\mathfrak{S}$  is a semi-metric vector space,  $d(\cdot, \cdot)$  denoting the semi-metric. Assume there exists a regular version of the conditional probability of  $Y$  given  $X$ . Assume that for a given  $x$  there is some compact subset  $S = (\theta - \xi, \theta + \xi)$ ,  $\xi > 0$ , such that the conditional density of  $Y$  given  $X = x$  has an unique mode  $\theta$  on  $S$ . In the remaining of the paper  $x$  is fixed in  $\mathfrak{S}$  and  $N_x$  denotes a neighborhood of  $x$ . Let  $f^x$  (resp.  $f^{x(j)}$ ) be the conditional density (resp. the  $j^{\text{th}}$  order derivative of the conditional density) of the variable  $Y$  given  $X = x$ . We define the kernel estimator  $\hat{f}^x$  of  $f^x$  as follows :

$$\hat{f}^x(y) = \frac{h_H^{-1} \sum_{i=1}^n K(h_K^{-1}d(x, X_i)) H(h_H^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))}, \quad (1)$$

where  $K$  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 widely studied until now (see for instance Youndjé, (1996), for several asymptotic results and references). A natural extension of the kernel estimator  $\hat{\theta}$  of the conditional mode  $\theta$  to the functional framework is given by :

$$\hat{\theta} = \arg \sup_{y \in S} \hat{f}^x(y). \quad (2)$$

The estimate  $\hat{\theta}$  is not necessarily unique, and if this is the case, all the remaining of our paper concerns any value  $\hat{\theta}$  satisfying (2). This work establishes the almost complete convergence of the kernel estimate  $\hat{\theta}$  of  $\theta$  under  $\alpha$ -mixing hypothesis. As we will see next Section, the flatness of the function  $f^x$  around the mode  $\theta$  plays a major role. One way to control this flatness is to consider the number of vanishing derivatives at  $\theta$ .

### 3.1.3 Main result

We introduce now some assumptions that are needed to state our results :

- (H1)  $P(X \in B(x, r)) = \phi_x(r) > 0$ ,
- (H2)  $\sup_{i \neq j} P((X_i, X_j) \in B(x, r) \times B(x, r)) = \phi_x(r)\psi_x(r) > 0$ ,
- (H3) The coefficients of the  $\alpha$ -mixing sequence  $(X_i, Y_i)$  satisfy the following arithmetic condition :

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

- (H4)  $f^x$  is  $j$ -times continuously differentiable with respect to  $y$  on  $(\theta - \xi, \theta + \xi)$ ,
- (H5)  $\forall (y_1, y_2) \in S \times S, \forall (x_1, x_2) \in N_x \times N_x, |f^{x_1}(y_1) - f^{x_2}(y_2)| \leq C_x (d(x_1, x_2)^{b_1} + |y_1 - y_2|^{b_2})$ ,
- (H6)  $f^x$  is strictly increasing on  $(\theta - \xi, \theta)$  and strictly decreasing on  $(\theta, \theta + \xi)$ ,
- (H7)  $f^{x^{(l)}}(\theta) = 0$ , if  $1 \leq l < j$ , and  $0 < |f^{x^{(j)}}(\theta)| < \infty$ ,
- (H8)  $K$  is a function with support  $(0, 1)$  such that  $0 < C_1 < K(t) < C_2 < \infty$ ,
- (H9)  $\left\{ \forall (y_1, y_2) \in \mathbb{R}^2, |H(y_1) - H(y_2)| \leq C|y_1 - y_2|, \int |t|^{b_2} H(t) dt < \infty. \right.$

$$(H10) \lim_{n \rightarrow \infty} h_H = 0 \text{ and } \exists \beta_1 \leq \frac{4}{(a+1)(a-2)} \text{ such that } \lim_{n \rightarrow \infty} n^{\beta_1} h_H = \infty,$$

$$(H11) \begin{cases} \lim_{n \rightarrow \infty} h_K = 0, \lim_{n \rightarrow \infty} \frac{\log n}{n h_H \chi_x(h_K)} = 0, \\ \text{and } \exists \beta_2 > 0, c_1 > 0, c_2 > 0, c_2 n^{\left(\frac{3-a}{a+1}\right) + \beta_2} \leq \chi_x(h_K) \leq c_1 n^{\frac{1}{1-a}}. \end{cases}$$

In this last hypothesis, the function  $\chi_x$  was defined by  $\chi_x(h) = \max\{\phi_x(h), \psi_x(h)\}$ .

**Theorem 3.1.1** *Under the hypotheses (H1)-(H11) we have*

$$\hat{\theta} - \theta = O\left(h_K^{\frac{b_1}{j}} + h_H^{\frac{b_2}{j}}\right) + O\left(\left(\frac{\log n}{n h_H \chi_x(h_K)}\right)^{\frac{1}{2j}}\right), \quad a.co. \quad (3)$$

**Sketch of the proof :** Write a Taylor expansion of  $f^x$  in a neighbourhood of  $\theta$ . Because of (H4) and (H7), and because  $|f^x(\hat{\theta}) - f^x(\theta)| \leq 2 \sup_{y \in S} |\hat{f}^x(y) - f^x(y)|$ , we have for some  $\theta^*$  between  $\theta$  and  $\hat{\theta}$  :

$$|(\theta - \hat{\theta})|^j \leq \frac{j!}{f^{x(j)}(\theta^*)} \sup_{y \in S} |\hat{f}^x(y) - f^x(y)|.$$

Theorem 3.1.1 is deduced from the lemmas and corollaries below, and from the decomposition :

$$\forall y \in S, \quad \hat{f}^x(y) - f^x(y) = \frac{1}{\hat{f}_D^x} \left\{ \left( \hat{f}_N^x(y) - E \hat{f}_N^x(y) \right) - \left( f^x(y) - E \hat{f}_N^x(y) \right) \right\} + \frac{f^x(y)}{\hat{f}_D^x} \left\{ E \hat{f}_D^x - \hat{f}_D^x \right\},$$

$$\text{where } \hat{f}_N^x(y) = \frac{1}{n h_H E K (h_K^{-1} d(x, X_1))} \sum_{i=1}^n K(h_K^{-1} d(x, X_i)) H(h_H^{-1}(y - Y_i))$$

$$\text{and } \hat{f}_D^x = \frac{1}{n E K (h_K^{-1} d(x, X_1))} \sum_{i=1}^n K(h_K^{-1} d(x, X_i)).$$

**Lemma 3.1.1**

*Under the hypotheses (H1)-(H3), (H8) and (H11) we have*

$$\hat{f}_D^x - E \hat{f}_D^x = O\left(\sqrt{(\log n)/(n \chi_x(h_K))}\right), \quad a.co.$$

**Corollary 3.1.1**

*Under the hypotheses of Lemma 3.1.1, we have  $\sum_{n=1}^{\infty} P\left(\hat{f}_D^x < 1/2\right) < \infty$ .*

**Lemma 3.1.2**

Under the hypotheses (H1), (H5), (H8), (H9) and (H11) we have

$$\frac{1}{\widehat{f}_D^x} \sup_{y \in S} |f^x(y) - E \widehat{f}_N^x(y)| = O(h_K^{b_1}) + O(h_H^{b_2}), \quad a.co.$$

**Lemma 3.1.3**

Under the hypotheses (H1)-(H3), (H5), and (H8)-(H11) we have

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

**Corollary 3.1.2**

Under the hypotheses of Lemma 3.1.3, we have  $\sup_{y \in S} |\widehat{f}^x(y) - f^x(y)| \rightarrow 0$ , a.c.o.

**Lemma 3.1.4**

Under the hypotheses of Lemma 3.1.3 and (H6)-(H7), we have

$$\exists c > 0, \sum_{n=1}^{\infty} P\left(f^{x(j)}(\theta^*) < c\right) < \infty.$$

Remarks on the proof. The complete proofs of these lemmas are available on request. Let us just note that the results linked with the dependence (Lemmas (3.1.1) and (3.1.3)) need a special attention and use exponential inequalities for dependent variables (see Rio, 1999). For peoples wishing to apply these lemmas in other settings, note that the last part of condition (H11) is not necessary to get Lemmas 3.1.2 and 3.1.4.

Remarks on key hypothesis. The main novelty in our functional approach can be seen through expressions (H1) and (H2), that are not really restrictive. Indeed, as pointed out in [3] the expression of the function  $\phi_x$  appearing in (H1) can be specified for many usual continuous time processes, and is linked with small ball probability theory. To see the role of condition (H2) it suffices to think on the special case when  $X$  is real and has a density with respect to Lebesgues measure on  $\mathbb{R}$ . In this case, (H2) is true with  $\phi = \psi$  as long as the pairs  $(X_i, X_j)$  have a density with respect to the Lebesgues measure on  $\mathbb{R}^2$  (which is a quite often used assumption in the classical finite dimensional literature).

Perspective for applications. The result of this Note can be applied to the prediction of time series, just by cutting the past of some time series in

continuous paths. Details can be found in [3] in which this is done but with other functional prediction technique (based on conditional expectation estimation).

*Acknowledgements.* The authors would like to thank both referees whose comments and suggestions have improved 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.



# Bibliographie

- [1] Berline, A., Gannoun, A. and Matzner-Lober, E. (1998). Normalité asymptotique d'estimateurs convergents du mode conditionnel (french). *La Rev. Canad. de Statist.*, **26**, 365-380.
- [2] Collomb, G., Härdle, W. and Hassani, S. (1987). A note on prediction via conditional mode estimation. *J. Statist. Plann. and Inf.*, **15**, 227-236.
- [3] Ferraty, F., Goia, A. and Vieu, P. (2002). Functional nonparametric model for time series : a fractal approach to dimension reduction. *TEST*, **11** (2), 317-344.
- [4] Ferraty, F. and Vieu, P. (2003). Functional nonparametric statistics : a double infinite dimensional framework. In *Recent Advances and Trends in Nonparametric Statistics*, Ed. M. Akritas and D. Politis, **Elsevier**.
- [5] Ferraty, F., Laksaci, A. and Vieu, P. (2004). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statist. Inf. for Stoch. Proc.*. In print.
- [6] Gasser, T., Hall, P. and Presnell, B. (1998). Nonparametric estimation of the mode of a distribution of random curves. *J. R. Statist. Soc. B*, **60**, 681-691.
- [7] Louani, D. and Ould-Saïd, E. (1999). Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis. *J. Nonparametric Stat.*, **11** (4), 413-442.
- [8] Rosenblatt, M. (1969). Conditional probability density and regression estimators. In *Multivariate Analysis II*, Ed. P.R. Krishnaiah. Academic Press, New York and London.
- [9] Rio, E. (1999). Théorie asymptotique des processus faiblement dépendants. *Mathématiques & Applications*, **31**, Springer-SMAI.

## 3.2 Preuves

The proof is decomposed in several steps. In next Section 3.2.1 we show how the proof of the theorem can be decomposed in 6 intermediary results. Then, each of these 6 results will be treated in a special section. All along our proofs, when no confusion will be possible, we will denote by  $C$  or/and  $C'$  some generic constant in  $\mathbb{R}^{*+}$  and we will use the following notations :  $K_i = K(h_K^{-1}d(x, X_i))$ ,  $H_i(y) = H(h_H^{-1}(y - Y_i))$ .

### 3.2.1 Proof of Theorem 3.1.1

Note first that by simple analytic arguments, we can show that

$$|f^x(\hat{\theta}) - f^x(\theta)| \leq 2 \sup_{y \in S} |\hat{f}^x(y) - f^x(y)|. \quad (4)$$

Let us now write a Taylor expansion of the function  $f^x$  for obtaining because of (H4), (H7) and (4) :

$$|(\theta - \hat{\theta})|^j \leq \frac{j!}{f^{x(j)}(\theta^*)} \sup_{y \in S} |\hat{f}^x(y) - f^x(y)|,$$

for some  $\theta^*$  between  $\theta$  and  $\hat{\theta}$ . Consider now the following decomposition

$$\forall y \in S, \quad \hat{f}^x(y) - f^x(y) = \frac{1}{\hat{f}_D^x} \left\{ \left( \hat{f}_N^x(y) - E\hat{f}_N^x(y) \right) - \left( f^x(y) - E\hat{f}_N^x(y) \right) \right\} + \frac{f^x(y)}{\hat{f}_D^x} \left\{ E\hat{f}_D^x - \hat{f}_D^x \right\}$$

$$\text{where } \hat{f}_N^x(y) = \frac{1}{n h_H E K_1} \sum_{i=1}^n K(h_K^{-1}d(x, X_i)) H(h_H^{-1}(y - Y_i))$$

and

$$\hat{f}_D^x = \frac{1}{n E K_1} \sum_{i=1}^n K(h_K^{-1}d(x, X_i)).$$

Finally, Theorem 3.1.1 can be deduced from the lemmas ((3.1.1), (3.1.2), (3.1.3), (3.1.4)) and corollaries ((3.1.1), (3.1.2)) :

### 3.2.2 Proof of Lemma 3.1.1

We have

$$\widehat{f}_D^x - E\widehat{f}_D^x = \frac{1}{n} \sum_{i=1}^n \left( \frac{K_i}{EK_1} - 1 \right) := \frac{1}{nE(K_1)} \sum_{i=1}^n \widetilde{\Delta}_i,$$

where

$$\widetilde{\Delta}_i = K_i - E(K_i).$$

The main tool of this proof is the application of Fuk-Nagaev's inequality (Rio, 1999). For thus, we must calculate the asymptotic behavior of the  $s_n^2$  quantity defined as follows :

$$s_n^2 = \sum_{i=1}^n \sum_{j=1}^n |Cov(\widetilde{\Delta}_i, \widetilde{\Delta}_j)| := s_n^{2*} + \sum_{i=1}^n Var(\widetilde{\Delta}_i), \quad (5)$$

where

$$s_n^{2*} = \sum_{i \neq j}^n |Cov(\widetilde{\Delta}_i, \widetilde{\Delta}_j)|.$$

Firstly, we define the sets  $S_1$  and  $S_2$  as follows

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

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

where the sequence  $m_n$  is chosen such that

$$m_n \rightarrow \infty,$$

and we let  $J_{1,n}$  and  $J_{2,n}$  be the sum of covariances over  $S_1$  and  $S_2$ , respectively. Then

$$J_{1,n} = \sum_{S_1} |Cov(\widetilde{\Delta}_i, \widetilde{\Delta}_j)| = \sum_{S_1} |E(K_i K_j) - E(K_i)E(K_j)|.$$

Because of (H1), (H2) and (H8) we can write

$$J_{1,n} \leq Cnm_n(\phi_x(h_K)\psi_x(h_K) + (\phi_x(h_K))^2). \quad (6)$$

On the other hand, to study the sum over  $S_2$ , we use the Davydov-Rio's inequality (see Rio, 1999) in the  $L^\infty$  cases ((which is possible since  $\tilde{\Delta}_i$ ) <sub>$i$</sub>  is bounded). This leads, for all  $i \neq j$ , to

$$|Cov(\tilde{\Delta}_i, \tilde{\Delta}_j)| \leq C\alpha(|i - j|)$$

and therefore we get

$$J_{2,n} \leq n^2(\alpha(m_n)) \leq Cn^2m_n^{-a}. \quad (7)$$

By using both results (6)-(7), then by using the definition of  $\chi$  and finally by choosing

$$m_n = \frac{1}{\chi_x(h_K)},$$

we may write, for  $n$  large enough, the following calculus :

$$\begin{aligned} s_n^{2*} &= J_{1,n} + J_{2,n} \\ &\leq C [nm_n\chi_x^2(h_K) + n^2m_n^{-a}] \\ &\leq O(n\chi_x(h_K)). \end{aligned} \quad (8)$$

Now, for all  $i = 1, \dots, n$  we can write

$$Var(\tilde{\Delta}_i) = E(K_i)^2 - (E(K_i))^2,$$

and by using (H1) we get

$$Var(\tilde{\Delta}_i) \leq C(\phi_x(h_K) + (\phi_x(h_K))^2).$$

Finally, this last result combined with (5) and (8) leads directly to :

$$s_n^2 = o(n\phi_x(h_K)) + O(n\phi_x(h_K)).$$

Thus, we are now in position to apply the classical Fuk-Nagaev's inequality (Rio 1999) to get :  $\forall \lambda > 0, \forall r > 0$ ,

$$P \left\{ \left| \hat{f}_D^x - E\hat{f}_D^x \right| > 4 \frac{\lambda}{nEK_1} \right\} \leq P \left\{ \left| \hat{f}_D^x - E\hat{f}_D^x \right| > 4 \frac{\lambda}{n\chi_x(h_K)} \right\} \leq A_1 + A_2,$$

where

$$A_1 = 4 \left( 1 + \frac{\lambda^2}{rs_n^2} \right)^{-(r/2)}$$

$$A_2 = 4 \frac{cn}{r} \left( \frac{r}{\lambda} \right)^{(a+1)}.$$

Set  $\lambda = (\eta/4) \sqrt{n \chi_x(h_K) \log n}$  then  $A_2$

$$A_2 = cnr^a \lambda^{-(a+1)} = cnr^a (n \phi_x(h_K) \log n)^{-(a+1)/2}.$$

Taking  $r = C(\log n)^2$  it follows that

$$A_2 = C(\log n)^{(2a-(a+1)/2)} n^{(1-(a+1)/2)} \chi_x(h_K)^{-(a+1)/2}.$$

Next, using the left side of (H11) we obtain

$$\begin{aligned} A_2 &\leq C(\log n)^{(2a-(a+1)/2)} n^{(1-(a+1)/2)} n^{-(a+1)/2} \left( \frac{3-a}{a+1} + \beta_2 \right) \\ &\leq C(\log n)^{(2a-(a+1)/2)} n^{-1-\beta_2 \left( \frac{a+1}{2} \right)}. \end{aligned}$$

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

$$A_2 \leq Cn^{-1-\nu}. \quad (9)$$

Let us now look at  $A_1$  expressed by :

$$A_1 = 4 \left( 1 + \frac{\eta^2 n \chi_x(h_K) \log n}{16r s_n^2} \right)^{-r/2}.$$

Because  $r = C(\log n)^2$ , we can write

$$A_1 \leq C \exp\left(-\eta^2 \frac{\log n}{32}\right) = Cn^{-\eta^2/32}. \quad (10)$$

Finally, by combining the results (9) and (10), we get that for  $\eta$  large enough, there exists some  $\nu > 0$  so that :

$$P \left( \left| \widehat{f}_D^x - E \widehat{f}_D^x \right| > \eta \sqrt{\frac{\log n}{n \phi_x(h_K)}} \right) \leq C' n^{-1-\nu}.$$

The proof of this lemma is finished.  $\square$

### 3.2.3 Proof of Corollary 3.1.1

We have

$$P\left(\left|\widehat{f}_D^x\right| \leq 1/2\right) \leq P\left(\left|\widehat{f}_D^x - 1\right| > 1/2\right).$$

Noting that

$$\widehat{f}_D^x - 1 = \widehat{f}_D^x - E\widehat{f}_D^x$$

and by applying the previous Lemma 3.1.1, we can write

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

and the proof of this corollary is finished.  $\square$

### 3.2.4 Proof of Lemma 3.1.2

We have

$$E\widehat{f}_N^x(y) - f^x(y) = \frac{1}{h_H EK_1} E\left(K_1 \left[E\left(\frac{H_1(y)}{h_H} \middle| X\right) - f^x(y)\right]\right). \quad (11)$$

Moreover, we have

$$\frac{1}{h_H} E(H(y)|X) = \frac{1}{h_H} \int_{\mathbb{R}} H(h_H^{-1}(y-z)) f^X(z) dz,$$

and by usual integration by substitution, we obtain

$$\left|E\left(\frac{H_1(y)}{h_H} \middle| X\right) - f^x(y)\right| \leq \int_{\mathbb{R}} H(t) |f^X(y - h_H t) - f^x(y)| dt.$$

Now, (H5) allows to write

$$\left|E\left(H_1^{(1)}(y)|X\right) - f^x(y)\right| \leq C_x \int_{\mathbb{R}} H(t) (h_K^{b_1} + |t|^{b_2} h_H^{b_2}) dt. \quad (12)$$

Because this inequality is uniform on  $y$  and because of (H9), the claimed result follows directly from (11), (12) and Corollary 3.1.1.  $\square$

### 3.2.5 Proof of Lemma 3.1.3

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

$$\begin{aligned} \frac{1}{\widehat{f}_D^x} \sup_{y \in S} \left| \widehat{f}_N^x(y) - E \widehat{f}_N^x(y) \right| &\leq \underbrace{\frac{1}{\widehat{f}_D^x} \sup_{y \in S} \left| \widehat{f}_N^x(y) - \widehat{f}_N^x(t_y) \right|}_{T_1} + \underbrace{\frac{1}{\widehat{f}_D^x} \sup_{y \in S} \left| \widehat{f}_N^x(t_y) - E \widehat{f}_N^x(t_y) \right|}_{T_2} + \\ &\quad \underbrace{\frac{1}{\widehat{f}_D^x} \sup_{y \in S} \left| E \widehat{f}_N^x(t_y) - E \widehat{f}_N^x(y) \right|}_{T_3}. \end{aligned} \quad (13)$$

- Concerning  $(T_1)$  :

$$\begin{aligned} \frac{1}{\widehat{f}_D^x} \sup_{y \in S} \left| \widehat{f}_N^x(y) - \widehat{f}_N^x(t_y) \right| &\leq \frac{1}{\widehat{f}_D^x} \sup_{y \in S} \frac{1}{n h_H E K_1} \sum_{i=1}^n |H_i(y) - H_i(t_y)| K_i, \\ &\leq \frac{1}{\widehat{f}_D^x} \sup_{y \in S} \frac{C|y - t_y|}{h_H^2} \left( \frac{1}{n E K_1} \sum_{i=1}^n K_i \right), \\ &\leq C \frac{l_n}{h_H^2}. \end{aligned} \quad (14)$$

In the above calculus, the second inequality is obtained by considering a Lipschitz argument whereas the last one comes from the definition of  $\widehat{f}_D^x$ . Take now

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

and note that, because of (H10), we have

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

Thus, for  $n$  large enough, we can write

$$P\left(\frac{1}{\widehat{f}_D^x} \sup_{y \in S} \left| \widehat{f}_N^x(y) - \widehat{f}_N^x(t_y) \right| > \frac{\eta}{3} \sqrt{\frac{\log n}{n \chi_x(h_K) h_H}}\right) = 0. \quad (16)$$

- Concerning  $(T_2)$  :

$$\widehat{f}_N^x(t_y) - E \widehat{f}_N^x(t_y) = \frac{1}{n h_H E K_1} \sum_{i=1}^n \underbrace{\{H_i(t_y) K_i - E(H_i(t_y) K_i)\}}_{\Lambda_i^*}. \quad (17)$$

By following the same route as along the proof of the Lemma 3.1.1 we note now

$$s_n'^2 = \sum_{i,j=1}^n |Cov(\Lambda_i^*, \Lambda_j^*)|.$$

Firstly, we look for the asymptotic behavior of the quantity  $s_n'^{2*}$  defined as follows :

$$s_n'^{2*} = \sum_{i \neq j=1}^n |Cov(\Lambda_i^*, \Lambda_j^*)|.$$

We split the sum into two sets defined by

$$S_1' = \{(i, j) \text{ such that } 0 \leq |i - j| \leq m_n'\}$$

and

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

where  $m_n' \rightarrow \infty$ . Then

$$s_n'^{2*} \leq \sum_{S_1'} Cov(\Lambda_i^*, \Lambda_j^*) + \sum_{S_2'} Cov(\Lambda_i^*, \Lambda_j^*)$$

$K$ , and  $H$  are bounded, so

$$Cov(\Lambda_i^*, \Lambda_j^*) \leq C (P(X_i, X_j) \in B(x, h_{Kn}) \times B(x, h_{Kn}))$$

$$+ CP(X_i \in B(x, h_{Kn})P(X_j \in B(x, h_{Kn})) \leq C[\psi_x(h_K)\phi_x(h_K) + \phi_x^2(h_K)]$$

and

$$\forall i \neq j \quad Cov(\Lambda_i^*, \Lambda_j^*) \leq C\alpha(|i - j|).$$

Then

$$s_n'^{2*} \leq C[\sum_{S_1'} \psi_x(h_K)\phi_x(h_K) + \phi_x^2(h_K)] + \sum_{S_2'} \alpha(|i - j|).$$

We can take

$$m_n' = \frac{1}{h_H \chi_x(h_K)},$$

and we conclude as for the proof of Lemma 3.1.1 that :

$$s_n'^2 = O(nh_H \chi_x(h_K)). \quad (18)$$



Once again, similar arguments as those invoked for proving Lemma 3.1.1 can be used, and we obtain successively, for all  $\epsilon$  and  $r > 1$  :

$$\begin{aligned}
& P \left( \sup_{y \in S} \left| \widehat{f}_N^x(t_y) - E \widehat{f}_N^x(t_y) \right| > \frac{\epsilon}{3} \frac{1}{C_1 n h_H \chi_x(h_K)} \right) = \\
& P \left( \max_{t_y \in \{t_1, \dots, t_{\tau_n}\}} \left| \widehat{f}_N^x(t_y) - E \widehat{f}_N^x(t_y) \right| > \frac{\epsilon}{3} \frac{1}{C_1 n h_H \chi_x(h_K)} \right) \leq \\
& \tau_n \max_{t_y \in \{t_1, \dots, t_{\tau_n}\}} P \left( \left| \widehat{f}_N^x(t_y) - E \widehat{f}_N^x(t_y) \right| > \frac{\epsilon}{3} \frac{1}{C_1 n h_H \chi_x(h_K)} \right) \leq \\
& \tau_n \max_{t_y \in \{t_1, \dots, t_{\tau_n}\}} P \left( \left| \widehat{f}_N^x(t_y) - E \widehat{f}_N^x(t_y) \right| > \frac{\epsilon}{3} \frac{1}{C_1 n h_H E K_1} \right) \\
& \leq C l_n^{-1} (A'_1 + A'_2),
\end{aligned}$$

where

$$A'_1 = 4 \left( 1 + \frac{\epsilon^2}{r s_n'^2} \right)^{-r/2} \quad \text{and} \quad A'_2 = 4 c n r^{-1} \left( \frac{r}{\epsilon} \right)^{a+1}.$$

To state the desired result we take  $\epsilon = C_1 \eta \sqrt{\frac{\log n}{n h_H \chi_x(h_K)}}$  and  $r = c \log n^2$  and we use that  $s_n'^2 = O(n h_H \chi_x(h_K))$ . We note that  $A'_1$  and  $A'_2$  are exactly the same as the terms  $A_1$  and  $A_2$  appearing in the proof of Lemma 3.1.1. We also note that, the choice of  $l_n$  made in (15) and our conditions on  $h_K$  (resp. on  $\chi_x$ ) are immediately insuring the existence, for  $\eta$  large enough of some  $\nu > 0$  such that :

$$l_n^{-1} (A'_1 + A'_2) \leq C n^{-1-\nu}.$$

Finally, we arrive at

$$\sup_{y \in S} \left| \widehat{f}_N^x(t_y) - E \widehat{f}_N^x(t_y) \right| = O \left( \sqrt{\frac{\log n}{n h_H \chi_x(h_K)}} \right). \quad (19)$$

- Concerning  $(T_3)$  : because of (H8) and (H9) we have :

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

Using analogous arguments as for  $T_1$ , we can show for  $n$  large enough that

$$P \left( \frac{1}{\widehat{f}_D^x} \sup_{y \in S} \left| E \widehat{f}_N^x(y) - E \widehat{f}_N^x(t_y) \right| > \frac{\eta}{3} \sqrt{\frac{\log n}{n h_H \chi_x(h_K)}} \right) = 0. \quad (20)$$

Now, our lemma can be easily deduced from (13), (16), (19) and (20).  $\square$

### 3.2.6 Proof of Corollary 3.1.2

Corollary is a consequence of Lemmas 3.1.2 and 3.1.3.  $\square$

### 3.2.7 Proof of Lemma 3.1.4

In order to establish the lemma, it suffices to show that

$$\hat{\theta} - \theta \rightarrow 0, \text{ a.co.} \quad (21)$$

Indeed, because the function  $f^x$  is continuous (see (H4)) and (H6), we have :

$$\forall \epsilon > 0, \exists \delta(\epsilon) > 0, \forall y \in (\theta - \xi, \theta + \xi), |f^x(y) - f^x(\theta)| \leq \delta(\epsilon) \Rightarrow |y - \theta| \leq \epsilon.$$

This leads directly to

$$\forall \epsilon > 0, \exists \delta(\epsilon) > 0, P \left( |\hat{\theta} - \theta| > \epsilon \right) \leq P \left( |f^x(\hat{\theta}) - f^x(\theta)| \geq \delta(\epsilon) \right). \quad (22)$$

Then,  $\hat{\theta} - \theta$  tend to 0 by means of (13) and Corollary(0.2)

Finally, Lemma 3.1.4 follows directly by using the continuity of  $f^{(j)}$  together with (H7) and (21).  $\square$

## 3.3 Application

### 3.3.1 La prévision en série temporelle

On peut obtenir aisément un échantillon de variables fonctionnelles à partir d'un processus à temps continu. En effet, soit  $(Z_t)_{t \in [0, b]}$  un processus réel stationnaire à temps continu. A partir de  $Z_t$  on construit  $N$  variables fonctionnelles  $(X_i)_{i=1, \dots, N}$  définies de la façon suivante :

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

Si on souhaite prédire une caractéristique réelle (notée  $Y$ ) de  $X_N$  sachant la courbe  $X_{N-1}$ , on considère les observations  $(X_i, Y_i)$  où  $Y_i$  est la caractéristique qu'on veut prévoir à l'instant  $i$ . Par exemple :

- Si on veut prévoir la valeur du processus à l'instant  $t_0$  dans l'intervalle  $[(N-1)b, Nb]$ , sachant la courbe  $X_{N-1}$ , on pose  $Y_i = X_{i+1}(t_0)$ .
- Pour le sup, on pose  $Y_i = \sup_{t \in [0, b[} X_{i+1}(t)$ .
- Si on cherche le moment où le processus atteindra le maximum, on pose  $Y_i = \operatorname{argmax}_{t \in [0, b[} X_{i+1}(t)$ .
- Pour le taux d'accroissement, on prend  $Y_i = (X_{i+1}(b) - X_{i+1}(0))/X_{i+1}(0) \dots$

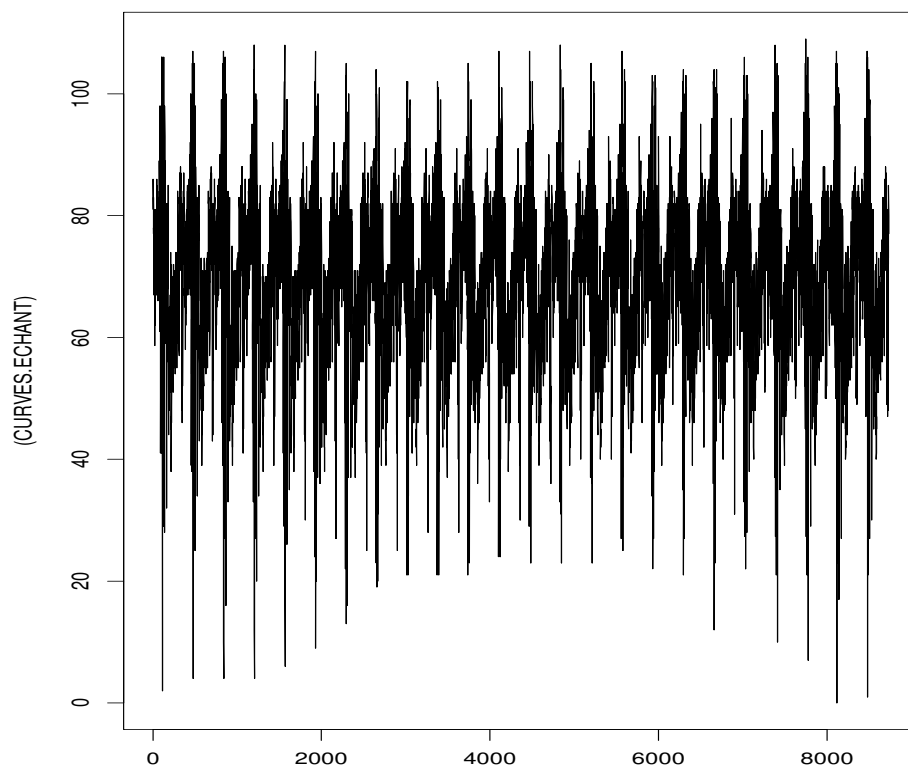
En utilisant le mode conditionnel comme outil de prévision, on peut prévoir  $Y$  par  $\hat{\theta}(X_{N-1})$ .

### 3.3.2 Mise en oeuvre sur des données réelles

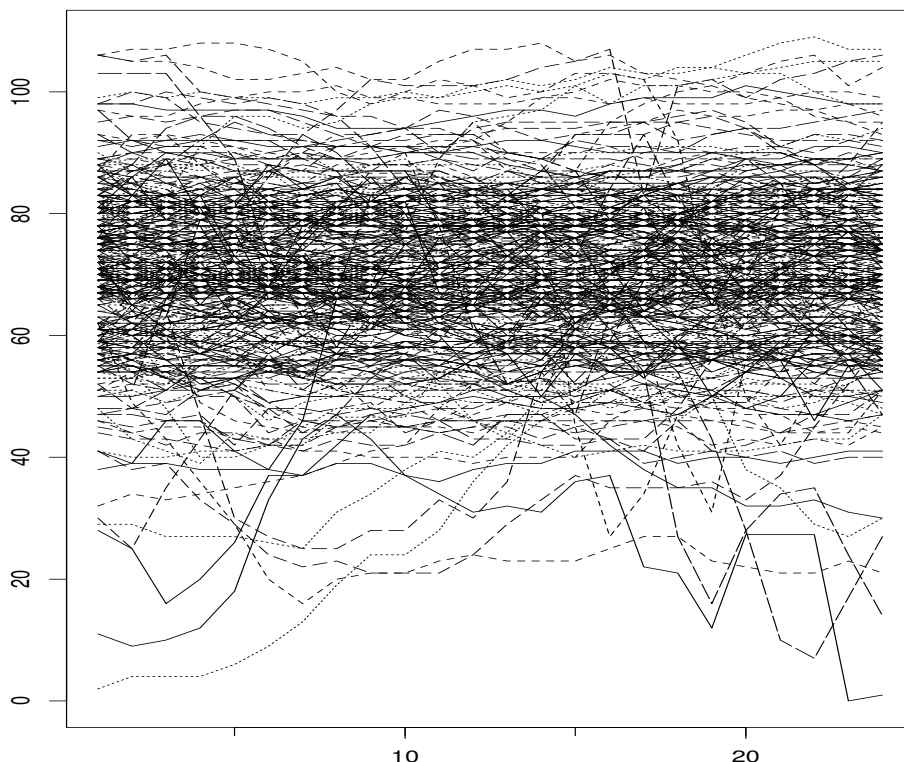
Dans ce paragraphe on examine un exemple pour lequel la prévision via le mode conditionnel est meilleure que celle obtenu par régression. Cet exemple concerne la prévision de la concentration maximale de l'ozone d'une journée au pôle nord sachant la courbe de cette concentration pendant la journée précédente. Plus précisément, Les observations ont été faites sur le site :

Country	Norway
Code	NO42
Database code	NO0042R
Geographical coordinates	78° 54'N, 11° 53'E
EMEP coordinates (50 km)	69.02, 104.35
Altitude above sea level	474
In operation since	September 1989
Closest climatological station	99910 Ny-Ålesund
Main wind direction	E-S-E

Nous avons récupéré ces données de la concentration de l'ozone de l'année 2003 (livrée par heure) sur le site <http://www.nilu.no/projects/cc/sitedescriptions/>. Elles sont représentées par le graphe suivant :



En reprenant les notations du paragraphe précédent, on peut représenter la concentration de l'ozone par un processus stochastique  $(Z_t)$ . En faisant un découpage par morceau (1 morceau=1 jour), on construit 365 variables aléatoires fonctionnelles notées  $X_{i=1,\dots,365}$ . Elle sont représentées ci-après :



Notre objectif est de prévoir la valeur maximale de la concentration de l'ozone sachant la courbe de la concentration d'ozone de la journée précédente. Ainsi la variable réponse, notée  $Y$ , est  $Y_i = \sup_{t \in [1, 24]} X_{i+1}(t)$ .

### Algorithme

Les estimations sont obtenues en choisissant le paramètre de lissage localement par la méthode de validation croisée sur le nombre des voisins les plus proches et on considère un noyau quadratique défini par :

$$K(x) = \begin{cases} \frac{3}{4}(1 - x^2) & \text{si } x \in [-1, 1] \\ 0 & \text{ailleurs.} \end{cases}$$

Le problème du choix de la métrique est aussi très important dans ce jeu des données. Dans notre cas nous avons choisi la norme  $L^2_{[1;24]}$  calculée en utilisant l'analyse en composantes principales fonctionnelles (Besse et al. (1997)). On procède par l'algorithme suivant

- *Etape 1.* On divise nos observations en deux paquets :

- $(X_j, Y_j)_{j=1, \dots, n}$  échantillon d'apprentissage,
- $(X_i, Y_i)_{i=1, \dots, I}$  échantillon de test.
- *Etape 2.* Pour chaque  $X_{new}$  dans l'échantillon de test, on prévoit sa valeur réponse  $Y_{new}$  par :
  - Méthode 1 : La régression

$$\widehat{Y}_{new}^{reg} = r_h(X_{new}),$$

avec

$$r_h(x) = \frac{\sum_{j=1}^n K(h_K^{-1}d(x, X_j))Y_j}{\sum_{j=1}^n K(h_K^{-1}d(x, X_j))}$$

- Méthode 2 : Le mode conditionnelle

$$\widehat{Y}_{new}^{mode} = \widehat{\theta}(X_{new}),$$

avec

$$\widehat{\theta}(x) = \text{Argmax}_y \widehat{f}_y^x.$$

- *Etape 3.* Pour chaque courbe  $X_i$  de l'échantillon de test, on pose :

$$i_{reg} = \text{Argmin}_{j=1, \dots, n} d(X_i, X_j)$$

et

$$i_{mode} = \text{Argmin}_{X_j \in \mathcal{J}} d(X_i, X_j),$$

où  $\mathcal{J}$  est l'ensemble des courbes de l'échantillon d'apprentissage appartenant à la même saison que celle de  $X_i$ .

- *Etape 4.* On prend

$$\widehat{Y}_{new}^{reg} = r_{h_{i_{reg}}^{opt}}(X_{new})$$

et

$$\widehat{Y}_{new}^{mode} = \widehat{\theta}(X_{i_{mode}}).$$

- *Etape 5.* L'erreur utilisée pour comparer les deux méthodes est

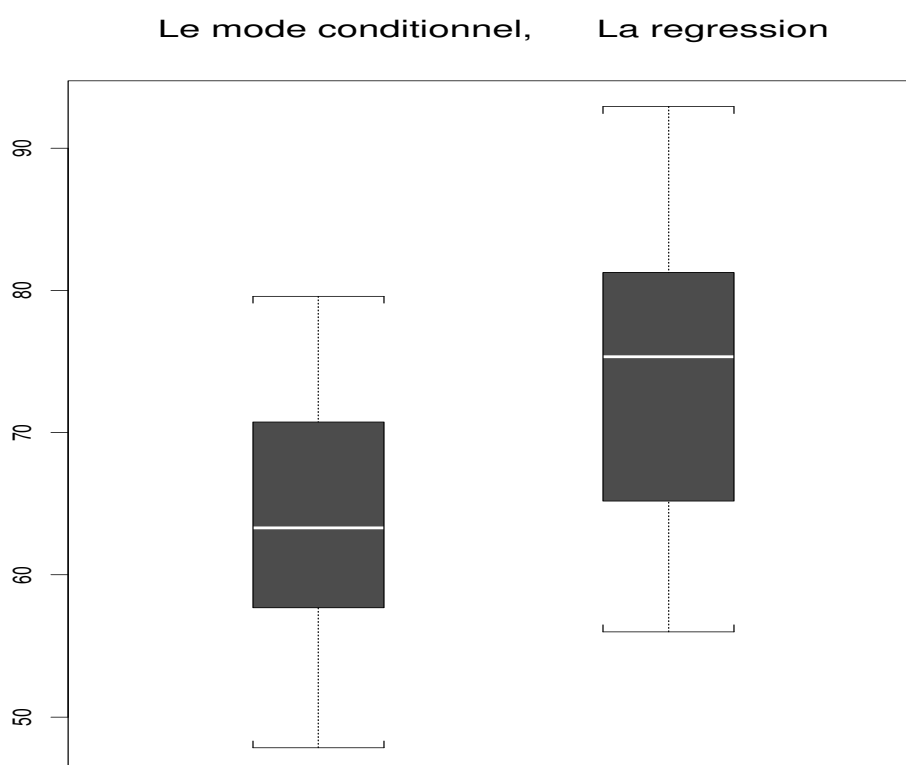
$$\frac{1}{n} \sum_{i=1}^n (Y_i - \widehat{T}(X_i))^2, \quad (23)$$

où  $\widehat{T}$  désigne l'estimateur utilisé : la régression ou le mode conditionnel.

- *Etape 6.* On repartage nos observations en deux paquets plusieurs fois et on répète les mêmes étapes 1-6
- *Etape 7.* On compare les deux méthodes en présentant notre résultat par un graphe de type boîte à moustaches.

## Résultats

On a répété 36 fois cette procédure et à chaque fois l'erreur de prédiction obtenue par les deux méthodes est comparable à celle obtenue récemment par Aneiros et *al.*(2004) dans un contexte différent (autres zones géographiques, variables explicatives additionnelles, ...). Nous avons constaté que les erreurs obtenues en faisant la prévision via le mode conditionnel sont dans l'ensemble plus faibles que celles de la régression (voir le graphe ci-dessous).



## Programmes

Cette section contient les programmes dont l'algorithme a été présenté précédemment. Ces programmes ont été réalisés par S-PLUS et ils sont présentés en deux sections. Dans la première section, on donne un programme permettant la prévision par le mode conditionnel en considérant une métrique quelconque. Dans la deuxième section, on considère la famille des semi-métriques de l'analyse en composantes principales et on donne le programme de la prévision via le mode conditionnel, en choisissant le nombre

optimale de composantes principales par la méthode de la validation croisée. De même, nous choisissons, aussi, le paramètre de lissage par la méthode de la validation croisée sur le nombre de voisins les plus proches. Nous renvoyons à Ferraty et Vieu (2005) pour le programme de la régression.

### Programme du mode pour un métrique quelconque

Ce programme donne la prévision via le mode conditionnel, pour n'importe quelle métrique.

```
funopare.mode.lcv <- function(A,B,Response,
                             CURVES,
                             ...,
                             Knearest=NULL,
                             PRED=CURVES,
                             kind.of.kernel = "quadratic",
                             semimetric = "deriv")
{
  sm <- get(paste("semimetric.", semimetric, sep=""))
  kernel <- get(kind.of.kernel)
  n <- nrow(CURVES)
  SMLEARN1 <- sm(CURVES,...)
  SML12.SOR <- apply(SMLEARN1, 2, sort)
  Resp.range <- range(Response)
  Response.grid <- seq(from=Resp.range[1]*0.9,to=Resp.range[2]*1.1,
                      length=100)
  # RESPMETRIC[i,j]=yi-yj with i in Response.grid and j in CURVES
  RESPMETRIC <- outer(Response.grid, Response, "-")
  RESPMET.SOR <- t(apply(abs(RESPMETRIC),1,sort))
  lgrid <- length(Response.grid)
  if(is.null(Knearest)){
    Knearest.min <- floor(n * .05)
    Knearest.max <- floor(n * .25)
    step <- ceiling((Knearest.max - Knearest.min) / 10)
    Knearest <- seq(Knearest.min, Knearest.max, by = step)
  }
  lknearest <- length(Knearest)
  BANDL12.CUR <- 0.5 * (SML12.SOR[Knearest,]
```



```

+ SML12.SOR[Knearest+1,])
BAND.RESP <- 0.5 * (RESPMET.SOR[,Knearest]
+ RESPMET.SOR[,Knearest+1])
CV <- matrix(0, nrow = lknearest^2, ncol = n)
MODE <- matrix(0, nrow = lknearest^2, ncol = n)
DENSITY.ESTIMATE <- array(0, dim=c(lknearest^2,n,lgrid))

count1 <- 0
count2 <- 0
for(kk in Knearest){
  count2 <- count2 + 1
  ARG <- t(t(SMLEARN1)/BANDL12.CUR[count2,])
  KERNEL.CURVES <- kernel(ARG)
  KERNEL.CURVES[KERNEL.CURVES<0] <- 0
  KERNEL.CURVES[KERNEL.CURVES>=1] <- 0
  Denom <- apply(KERNEL.CURVES, 2, sum)
  count3 <- 0
  for(hh in Knearest){
    count1 <- count1 + 1
    count3 <- count3 + 1
    KERNEL.RESP <- apply(abs(RESPMETRIC)/
                        BAND.RESP[,count3], 1, kernel)
    KERNEL.RESP[KERNEL.RESP<0] <- 0
    KERNEL.RESP[KERNEL.RESP>=1] <- 0
  }
  DENSITY.ESTIMATE[count1, ,] <- (t(KERNEL.CURVES)/Denom)
  %*% (KERNEL.RESP/BAND.RESP[,count3])
}
}

Ind.modes<-0
ind.knearest.opt<-0
ECrossvalidation<-0
Response.estimated<-0
for( j in 1:n){
for (i in 1:(lknearest^2)){
Modes <-peaks(DENSITY.ESTIMATE[i, j,])
Ind.modes <- order(abs(Response[j]-Response.grid[Modes]))[1]
MODE[i, j] <- Response.grid[Modes] [Ind.modes]
}
}

```

```

      CV[i,j] <- (Response[j]-MODE[i,j])^2
    }
    ind.knearest.opt[j] <- order(CV[,j])[1]
    ECrossvalidation[j] <- CV[ind.knearest.opt[j],j]
    Response.estimated[j] <- MODE[ind.knearest.opt[j],j]
  }

  Mse.estimated <- sum(ECrossvalidation)/n
  if(!missing(PRED)){
    SSMLEARN2NEW <- sm(CURVES,...,PRED)

    SMLEARN2NEW<-0
    Order.new<-0
    for (j in 1:nrow(PRED)){
      for ( i in 1:nrow(CURVES))
      { if (A[i]!=B[j]) SMLEARN2NEW[i]<- max (SSMLEARN2NEW)
        else SMLEARN2NEW[i]<-SSMLEARN2NEW[i,j]
      }
    }
    Order.new[j] <- order(SMLEARN2NEW)[1]
  }

  Response.predicted <- Response.estimated[Order.new]
  return(list(Estimated.values=Response.estimated,
             Predicted.values=Response.predicted,
             Response.values=Response,
             Mse=Mse.estimated))
}
else{
  return(list(Estimated.values=Response.estimated,
             Response.values=Response,
             Mse=Mse.estimated))
}
}

```

### Programme du mode avec un choix optimal de la métrique

Ce programme permettant la prévision par le mode conditionnel, en choisissant une métrique optimale parmi la famille des semi-métriques de l'ana-

lyse en composantes principales

```

funopare.metr-mode.lcv <- function(A,B,Response,
                                CURVES,
                                Knearest=NULL,
                                PRED=CURVES,
                                kind.of.kernel = "quadratic")
{
  kernel <- get(kind.of.kernel)
  Resp.range <- range(Response)
  n <- nrow(CURVES)
  Response.grid <- seq(from=Resp.range[1]*0.9,to=Resp.range[2]*1.1,
                      length=100)
  # RESPMETRIC[i,j]=yi-yj with i in Response.grid and j in CURVES
  RESPMETRIC <- outer(Response.grid, Response, "-")
  RESPMET.SOR <- t(apply(abs(RESPMETRIC),1,sort))
  lgrid <- length(Response.grid)

  if(is.null(Knearest)){
    Knearest.min <- floor(n * .05)
    Knearest.max <- floor(n * .25)
    step <- ceiling((Knearest.max - Knearest.min) / 10)
    Knearest <- seq(Knearest.min, Knearest.max, by = step)
  }
  lknearest <- length(Knearest)

  CV <- matrix(0, nrow = (24*(lknearest^2)), ncol = n)
  MODE <- matrix(0, nrow = (24*(lknearest^2)), ncol =n )
  DENSITY.ESTIMATE <- array(0, dim=c((24*(lknearest^2)),n,lgrid))
  count <- 0
  count1 <- 0
  for ( q in 1:24)
  {
    count <- count+1

    SMLEARN1 <- semimetric.pca(CURVES,q,CURVES)
    SML12.SOR <- apply(SMLEARN1, 2, sort)
    BANDL12.CUR <- 0.5 * (SML12.SOR[Knearest,]

```

```

+ SML12.SOR[Knearest+1,])
BAND.RESP <- 0.5 * (RESPMET.SOR[,Knearest]
+ RESPMET.SOR[,Knearest+1])

count2 <- 0
for(kk in Knearest){
count1 <-count1 +1
count2 <- count2 + 1
ARG <- t(t(SMLEARN1)/BANDL12.CUR[count2,])
KERNEL.CURVES <- kernel(ARG)
KERNEL.CURVES [KERNEL.CURVES<0] <- 0
KERNEL.CURVES [KERNEL.CURVES>=1] <- 0
Denom <- apply(KERNEL.CURVES, 2, sum)
count3 <- 0
for(hh in Knearest){
count1 <- count1 + 1
count3 <- count3 + 1
KERNEL.RESP <- apply(abs(RESPMETRIC)/
BAND.RESP[,count3], 1, kernel)
KERNEL.RESP [KERNEL.RESP<0] <- 0
KERNEL.RESP [KERNEL.RESP>=1] <- 0
DENSITY.ESTIMATE[count1,,] <- (t(KERNEL.CURVES)/Denom)%*%
(KERNEL.RESP/BAND.RESP[,count3])
}
}
}

Ind.modes<-0
ind.knearest.opt<-0
ECrossvalidation<-0
q.opt<-0
Response.estimated<-0
for( j in 1:n){
for (i in 1:(24*(lknearest^2))){
Modes <-peaks(DENSITY.ESTIMATE[i,j,])
Ind.modes <- order(abs(Response[j]-Response.grid[Modes]))[1]
MODE[i,j] <- Response.grid[Modes][Ind.modes]
}
}
}

```

```

    CV[i,j] <- (Response[j]-MODE[i,j])^2
  }
  ind.knearest.opt[j] <- order(CV[,j])[1]
  q.opt[j]<-ceiling[ind.knearest.opt[j]/(klength^2)]
  ECrossvalidation[j] <- CV[ind.knearest.opt[j],j]
  Response.estimated[j] <- MODE[ind.knearest.opt[j],j]
}
  Mse.estimated <- sum(ECrossvalidation)/n
  if(!missing(PRED)){

    SMLEARN2NEW<-0
    Order.new<-0
    for (j in 1:nrow(PRED)){
      SSMLEARN2NEW <-semimetric.pca(CURVES, q.opt[j] , PRED)
      for ( i in 1:nrow(CURVES))
        { if (A[i]!=B[j]) SMLEARN2NEW[i]<- max (SSMLEARN2NEW)
          else SMLEARN2NEW[i]<-SSMLEARN2NEW[i,j]
        }
      Order.new[j] <- order(SMLEARN2NEW)[1]
    }

    Response.predicted <- Response.estimated[Order.new]
    return(list(Estimated.values=Response.estimated,
               Predicted.values=Response.predicted,
               Response.values=Response,
               Mse=Mse.estimated))
  }
  else{
    return(list(Estimated.values=Response.estimated,
               Response.values=Response,
               Mse=Mse.estimated))
  }
}

```



# Chapitre 4

## Commentaires et perspectives

L'objectif de ce chapitre est de commenter les résultats que nous avons obtenus dans cette thèse, avec comme souci majeur de mettre l'accent sur tous les problèmes rencontrés.

### 4.1 Quelques situations particulières

Pour mettre en valeur nos résultats nous allons revenir sur quelques situations particulières sur lesquelles nos méthodes peuvent être utilisées.

#### 4.1.1 Processus à temps continu

Nos résultats asymptotiques sont étroitement liés à un thème d'actualité en probabilité qui est celui des propriétés de concentration de la mesure de probabilité.

*Exemple 1 : Processus de Wiener.* K.L. Chung et J.L. Doob (1965) semblent être les premiers à avoir abordé le problème de la propriété de concentration sur les petites boules. Il a quantifié cette propriété de concentration sur le processus de Wiener  $W$ , en considérant la norme de la convergence uniforme, par

$$P(W \in B(0, r)) = \exp \left\{ -\frac{\pi^2}{8r^2}(1 + o(1)) \right\},$$

où  $B(0, r)$  est la boule centrée en zéro de rayon  $r$ . Si on garde les mêmes notations que lors des chapitres précédents et si on choisit les paramètres de

lissage  $h_K$  et  $h_H$  de type  $\eta/\log n$  et  $n^{-\epsilon}$ , avec  $\eta$  suffisamment grand et  $\epsilon > 0$ , on arrive à une vitesse de convergence de type  $O(\log n)^{\frac{-b_1}{2}}$ .

*Exemple 2 : Processus de Poisson.* A.A. Mogluskii (1974) a étudié la propriété de concentration sur le processus de Poisson, d'intensité  $\sigma$ , selon la norme de la convergence uniforme. Il a montré que la probabilité de la boule centrée de rayon  $r$  est :

$$\exp \left\{ -\frac{\pi^2 \sigma^2}{8r^2} (1 + o(1)) \right\}.$$

En utilisant les outils développés dans le chapitre 2 et pour un choix approprié de la fenêtre de lissage, nous arrivons à une vitesse de convergence  $O(\log n)^{\frac{-b_1}{2}}$ .

D'autres résultats sur la propriété de concentration de la mesure gaussienne ont été obtenus par Kulbs et *al.* (1993), Bogavhov (1999), Li et Shao (2001) (voir chapitre 2). Les vitesses de convergence des exemples précédents sont lentes. Ceci est dû au fait que toutes les métriques considérées actuellement en processus stochastique à temps continu ont une probabilité sur petites boules de type exponentielles. Cependant, pour améliorer ces vitesses de convergence il suffit d'utiliser des semi-métriques qui donnent une augmentation de la concentration de mesure. Une amorce de la solution est dans l'exemple suivant :

*Exemple 3 :* Soit  $X$  une variable fonctionnelle dans un espace de Hilbert  $\mathcal{F}$  muni du produit scalaire  $\langle \cdot, \cdot \rangle$ . En reprenant les outils de Ferraty et *al.* (2002) sur le modèle à indice fonctionnel, on peut définir une semi-métrique  $d$  par

$$\forall x, y \in \mathcal{F} \quad d(x, y) = | \langle x - y, \alpha \rangle |$$

où  $\alpha \in \mathcal{F}$ . C'est une semi-métrique qui peut s'appliquer à tous les processus stochastiques cités ci-dessus à condition que les trajectoires soient dans un espace de Hilbert ( $L^2$  par exemple). De plus, si on suppose que  $\langle X, \alpha \rangle$  admet une densité par rapport à la mesure de Lebesgue, alors la concentration de la mesure sur une boule de rayon  $h$  est de l'ordre de  $h$ . Ce qui permet d'arriver (après un choix convenable de  $h_K$  et de  $h_H$ ) à des vitesses de convergence  $O\left(\frac{n}{\log n}\right)^{\frac{-2b_1}{2b_1+1}}$ .



### 4.1.2 Cas fractal

Le cas fractal a été étudié par Ferraty et Vieu (2000). Il consiste à traiter les variables fonctionnelles qui possèdent la propriété suivante :

**Definition 4.1.1** Soit  $X$  une variable aléatoire à valeurs dans un espace semi-métrique noté  $(\mathcal{F}, d)$ , et soit  $B(x, r)$  la boule de centre  $x$  et de rayon  $r$ . On dit que la variable  $X$  est de dimension fractal  $\alpha \in \mathbb{R}^{*+}$  si et seulement si

$$\lim_{r \rightarrow 0^+} \frac{P(X \in B(x, r))}{r^\alpha} = C,$$

Dans cette situation, c'est-à-dire si on dispose d'une variable fonctionnelle de dimension fractale  $\alpha$ , la propriété de concentration est de l'ordre de  $h^\alpha$ . Donc les vitesses de convergence seront similaires à celles obtenues par Ferraty et Vieu (2000) pour le modèle de régression et qui sont :

$$O(h_K^{b_1}) + O(h_H^{b_2}) + O\left(\sqrt{\frac{\log n}{nh_K^\alpha}}\right),$$

pour la fonction de répartition, et

$$O(h_K^{b_1}) + O(h_H^{b_2}) + O\left(\sqrt{\frac{\log n}{nh_H^{2j+1}h_K^\alpha}}\right),$$

pour la dérivée d'ordre  $j$  de la densité conditionnelle.

### 4.1.3 Cas de dimension finie

Dans le cas où la dimension est finie, c'est-à-dire le cas où  $X \in \mathbb{R}^p$ , si  $X$  a une densité strictement positive, alors la concentration de sa mesure de probabilité sur la boule centrée en un point  $x$  et de rayon  $h$  est de type  $h^p$ . Ainsi, on retrouve les mêmes vitesses que celles obtenus par pour Youndjé (1996) qui valent :

$$O(h_K^{b_1}) + O(h_H^{b_2}) + O\left(\sqrt{\frac{\log n}{nh_H^{2j+1}h_K^p}}\right).$$

Par contre, nos conditions sont moins restrictives que celles de Youndjé. En effet, notre hypothèse de concentration dans le cas de dimension finie est une

hypothèse sur la fonction de répartition et non pas sur la densité. Autrement dit, on peut avoir une boule centrée en un point en lequel la densité est nulle bien que la probabilité de cette boule soit strictement positive. Par exemple, on peut citer la loi bien connue de  $\chi^2$  à un degré de liberté supérieur au égale à deux qui vérifiant ce type de propriété. Ainsi, nos estimateurs peuvent s'appliquer (même en dimension finie) en des points pour lesquels la stricte positivité de la densité n'est pas nécessaire.

## 4.2 Quelques questions pratiques

### 4.2.1 Choix de la semi-métrie

La question du choix optimal de la métrique a une importance capitale dans la pratique. A notre connaissance il n'existe pas encore de méthode pratique pour le choix optimal de la semi-métrie. Cependant, Ferraty et Vieu (2004), en considérant une famille de semi-métriques, définie par

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

où  $x^{(q)}$  est la dérivée d'ordre  $q$ , et en utilisant les données spectrométriques ont montré que la métrique qui minimise l'erreur de prédiction pour ces données est obtenue pour  $q = 2$ . D'autres familles de semi-métriques sont utilisées telles :

- *La semi-métrie de l'analyse en composantes principales fonctionnelles*. C'est une métrique introduite par (Besse et al. (1997)). Elle est définie par la topologie de la norme  $L^2$  dont le calcul est essentiellement basé sur l'analyse des composantes principales. Plus précisément, on cherche les  $q$  premiers vecteurs propres de l'opérateur de variance-covariance empirique, associés aux  $q$  plus grandes valeurs propres. La métrique est alors définie par :

$$\forall (x_1, x_2) \in \mathcal{F} \times \mathcal{F}, \quad d_q^{ACP}(x_1, x_2) = \int (\prod_q (x_1(t) - x_2(t)))^2 dt,$$

où  $\prod_q$  est le projecteur orthogonal sur l'espace  $L^2$  engendré par les  $q$  vecteurs propres associés aux  $q$  plus grandes valeur propres.

- *La semi-métrique de modèle d'indice fonctionnel.* Pour une variable fonctionnelle  $X$  dans un espace de Hilbert  $\mathcal{H}$  muni du produit scalaire  $\langle \cdot, \cdot \rangle$ , on définit la semi-métrique de modèle d'indice fonctionnel par

$$\forall \alpha, x, y \in \mathcal{H} \quad d_\alpha(x, y) = | \langle x - y, \alpha \rangle |.$$

Par des techniques de type validation croisée on pourrait estimer  $\hat{\alpha}$  qui donne la "meilleure" prédiction.

### 4.2.2 Choix du paramètre de lissage

Les expressions de nos résultats peuvent être divisées en deux parties : une partie biais et une partie dispersion. L'optimisation de la partie biais est obtenue en maximisant le paramètre de lissage  $h_n$ . Par contre, il faut minimiser ce paramètre pour optimiser la partie dispersion. Ainsi le choix optimal de  $h_n$  est celui qui consiste à faire un équilibre entre les deux. Dans nos applications, nous avons procédé en choisissant le paramètre de lissage par la méthode de la validation croisée sur le nombre de voisins les plus proches. Cette méthode a montré une compatibilité importante avec nos données. Mais, en général, la question du choix optimal du paramètre de lissage dans le cas des données fonctionnelles reste encore sans réponse. Dans le cas de la dimension finie ce problème a été largement étudié par Youndjé (1993). Cependant, pour le cas de dimension infinie, le problème reste ouvert, mais les idées de Youndjé (1993) devrait parvenir à être généralisées. En revanche, en régression fonctionnelle le problème est déjà en chantier (voir Rachdi et Vieu (2004)).

### 4.2.3 A propos de l'existence de la version régulière de la probabilité conditionnelle

Evidemment, dans toute la thèse on suppose que la version régulière de la probabilité conditionnelle existe dans l'espace semi-métrique. Cependant, ce problème est encore ouvert en probabilité et la littérature sur le sujet, à notre connaissance, n'a pas encore confirmé l'existence ou non de la version régulière dans cette structure topologique. Dans une structure plus restrictive tel qu'un espace métrique complet séparable, l'existence est assurée par le Théorème de Jirina (Tortrat, 1971, P.77).

Récemment, en employant les notions de complétude, compacité et séparabilité, Leão, Fragoso et Ruffino (1999,2004) ont élaborés des conditions nécessaires et suffisantes pour l'existence de cette version régulière sur des espaces moins restrictifs que ceux du Théorème de Jirina. Mais la question reste ouverte pour de espaces semi-métriques généraux. Ceci est un vrai défi pour les probabilistes qui, une fois relevé, augmentera bien plus les potentialités de l'application de notre méthode. Dans cette thèse, nous avons opté pour une présentation sous la structure semi-métrique en supposant l'existence de la version régulière pour deux raisons. La première est liée aux nombreuses applications offertes en considérant un espace semi-métrique au lieu de l'espace métrique ce qui permet d'avoir une grande souplesse dans le choix d'indices de proximités entre des objets fonctionnels. La deuxième vient du fait que les semi-métriques utilisées sur les variables initiales sont équivalentes à des métriques agissant sur les variables transformées, situations dans laquelle l'existence de la version régulière est vérifiée dès que l'espace métrique considéré est complet et séparable. A titre d'exemple, dans le cas de données spectrométriques nous avons considéré une famille de semi-métriques du type

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

où  $x^{(q)}$  est la dérivée d'ordre  $q$ . La semi-métrique optimale est donnée par  $q = 2$  (voir Ferraty et Vieu (2004)). Dans cet exemple, faire la prévision par  $X_i$  avec la semi-métrique  $d_2$  est équivalent à la prévision en utilisant comme donnée  $X_i^{(2)}$  et comme métrique celle induite par  $L^2$ , espace métrique complet et séparable où la version régulière existe. Mais comme il n'est pas possible de connaître a priori la meilleure semi-métrique, notre présentation (en considérant un espace semi-métrique avec les données originales) nous offre une multitude de choix concernant les semi-métriques. Finalement nous signalons que dans d'autres contextes statistiques, totalement différents de celui de la statistique non paramétrique fonctionnelle comme la théorie des problèmes inverses (voir Lehtinen et al. (1989)), la version régulière est implicitement supposée exister dans des espaces non nécessairement métriques.

### 4.3 Perspectives

Le travail développé dans cette thèse offre de nombreuses perspectives à court et long termes. Parmi les perspectives de court terme :

1. **La convergence en norme  $L^p$ .** En faisant une adaptation des outils développés par Niang et Rhomari (2003) on pourrait étudier la convergence en norme  $L^p$  des nos estimateur dans le cas dépendant (Un article traitant le cas i.i.d., a été déjà soumis avec Dabo-Niang)
2. **Le choix du paramètre de lissage.** On va tenter de généraliser le résultat de E.Youndjé (1993) sur le choix du paramètre de lissage pour l'estimation de la densité conditionnelle (Rachdi et Vieu (2004) ont déjà abordé ce problème dans le cas de la régression fonctionnelle). L'obtention des résultats de convergence en moyenne quadratique mentionnés plus haut devrait être très utile en ce sens.
3. **La normalité asymptotique.** La normalité asymptotique de nos estimateurs peut nous permettre de faire des tests et de construire des intervalles de confiance. E. Masry (2005) a obtenu des résultats sur ce problème en régression et il devrait être possible d'utiliser les mêmes idées en densité conditionnelle.
4. **Estimation robuste et données fonctionnelles.** Le modèle de la  $\psi$  régression introduit par Laib et Ould- Said (2000) est une autre façon d'estimer les quantiles conditionnels par la méthode du noyau. Nos outils devraient parvenir à s'adapter à cet estimateur.
5. **Choix de la semi-métrique.** Le choix de la semi-métrique est une question très importante. Ce choix a une influence dans tous les problèmes évoqués (voir paragraphe (4.1), (4.2.2) et (4.2.3)). En utilisant des familles de semi-métriques "adaptives" ( par exemple, celle qu'on pourrait déduire du modèle à indice fonctionnel), on peut espérer améliorer les performances de nos estimateurs en terme de prédiction.

D'autres thèmes peuvent être abordés à long terme tels le conditionnement par  $p$  variables fonctionnelles ou une combinaison linéaire de ces  $p$  variables fonctionnelles, la fonction de hasard conditionnelle, ...



# Bibliographie

- [1] Aneiros P.G., Cardot, H., Estévez, P.G., Vieu, P. (2004), *Maximum ozone concentration forecasting by functional non-parametric approaches*. *Environmetrics*, **15**, 675-685.
- [2] Ait Saidi, A., Ferraty, F., Kassa, R., (2005), *Single functional index model for time series*. *Rev. Roumaine Math. Pures Appl.*, in print.
- [3] Berlinet, A., Gannoun, A., Matzner-Løber, E., (1998), *Propriétés asymptotiques d'estimateurs convergents des quantiles conditionnels*. *C. R. Acad. Sci., Paris, Sér. I, Math.* **326**, No.5, 611-614.
- [4] Berlinet, A., Gannoun, A., Matzner-Løber, E., (1998), *Normalité asymptotique d'estimateurs convergents du mode conditionnel*. *Can. J. Stat.* **26**, No.2, 365-380.
- [5] Besse, P. Cardot, H., Ferraty, F., (1997), *Simultaneous nonparametric regressions of unbalanced longitudinal data*. *Comput. Statist. Data Anal.*, **24**, 255-270.
- [6] Besse, P., Cardot, H., Stephenson, D., (2000), *Autoregressive forecasting of some functional climatic variations*. *Scand. J. Stat.***27**, No.4, 673-687.
- [7] Boente, G., Fraiman, R., (1995), *Asymptotic distribution of smoothers based on local means and local medians under dependence*. *J. Multivariate Anal.* **54**, No.1, 77-90.
- [8] Bosq, D., Delecroix, M., (1985), *Nonparametric prediction of a Hilbert space valued random variable*. *Stochastic Processes Appl.* **19**, 271-280.
- [9] Bosq, D., (1991), *Modelization, non-parametric estimation and prediction for continuous time processes*. In Roussas, G. (Ed.), *Nonparametric Functional estimation and Related Topics*, NATO, ASI Series, pp.509-529.
- [10] Bosq, D., (2000), *Linear processes in function spaces. Theory and applications*. *Lecture Notes in Statistics*. New York.

- [11] Cardot, H., Ferraty, F., Sarda, P., (1999), *Functional linear model*. Stat. Probab. Lett. **45**, No.1, 11-22.
- [12] Cardot, H., Ferraty, F., Sarda, P., (2000), *Étude asymptotique d'un estimateur spline hybride pour le modèle linéaire fonctionnel*. C. R. Acad. Sci., Paris, Sér. I, Math. **330**, No.6, 501-504.
- [13] Cardot, H., Ferraty, F., Mas, A., Sarda, P., (2003), *Testing hypotheses in the functional linear model*. Scand. J. Stat. **30**, No.1, 241-255.
- [14] Cardot, H., Crambes, C., Sarda, P., (2004), *Spline estimation of conditional quantities for functional covariates*. C. R., Math., Acad. Sci. Paris **339**, No.2, 141-144.
- [15] Cadre, B., (2001), *Convergent estimators for the  $L_1$ -median of a Banach valued random variable*. Statistics **35**, No.4, 509-521.
- [16] Chen, G., (1997) *Berry-Esseen-type bounds for the kernel estimator of conditional distribution and conditional quantiles*. J. Stat. Plann. Inference **60**, No.2, 311-330.
- [17] Chung, K. L., Doob, J.L., (1965) *Fields, optionality and measurability*. Amer. J. Math. **87**, 397-424.
- [18] Collomb, G., Härdle, W., Hassani, S. (1987) *A note on prediction via estimation of the conditional mode function*. J. Stat. Plann. Inference **15**, 227-236.
- [19] Cuevas, A., Febrero, M., Fraiman, R., (2004), *An anova test for functional data*. Comput. Statist. Data Anal. **47**, no. 1, 111-122.
- [20] Dabo-Niang, S., (2002), *Estimation de la densité dans un espace de dimension infinie : Application aux diffusions*. C. R., Math., Acad. Sci. Paris **334**, No.3, 213-216.
- [21] 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**, No.1, 75-80.
- [22] Deville, J.C., (1974) *Méthodes statistiques et numériques de l'analyse harmonique*. Ann. Insee. **15**.
- [23] Doukhan, P. (1994) *Mixing : Properties and examples*. Lecture Notes in Statistics (Springer). 85. New York : Springer-Verlag.
- [24] Ferraty, F., Vieu, Ph., (2000), *Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés*. C. R. Acad. Sci., Paris, Sér. I, Math. **330**, No.2, 139-142.



- [25] Ferraty, F., Goia, A., Vieu, Ph., (2002), *Functional nonparametric model for time series : a fractal approach for dimension reduction*. Test **11**, No.2, 317-344.
- [26] Ferraty, F., Goia, A., Vieu, Ph., (2002), *Régression non-paramétrique pour des variables aléatoires fonctionnelles mélangeantes*. C. R., Math., Acad. Sci. Paris **334**, No.3, 217-220.
- [27] Ferraty, F., Vieu, Ph., (2002), *The functional nonparametric model and applications to spectrometric data*. Comput. Stat. **17**, No.4, 545-564.
- [28] Ferraty, F., Peuch, A., Vieu, Ph., (2003), *Modèle à indice fonctionnel simple*. C. R., Math., Acad. Sci. Paris **336**, No.12, 1025-1028.
- [29] Ferraty, F., Vieu, Ph., (2003), *Curves discrimination : a nonparametric functional approach*. Comput. Statist. Data Anal. **44** , no. 1-2, 161-173.
- [30] Ferraty, F., Vieu, Ph., (2004), *Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination*. J. Nonparametric Stat. **16**, No.1-2, 111-125.
- [31] Ferraty, F., Vieu, Ph., (2005), *Nonparametric functional data analysis*. Springer-Verlag (to appear).
- [32] Ferré, L., Villa, N. (2005), *Discrimination de courbes par régression inverse fonctionnelle*. Revue de Statistique Appliquée, **LIII**, (1), 39-57.
- [33] Gannoun, A., (1990), *Estimation non paramétrique de la médiane conditionnelle. Application à la prévision*. C. R. Acad. Sci., Paris, Sér. I **310**, No.5, 295-298.
- [34] Gannoun, A., (2002), *Sur quelques problèmes d'estimation fonctionnelle : Théorie, Méthode et Applications*. H.D.R. Université Montpellier II.
- [35] Gannoun, A., Saracco, J., Yu, K., (2003), *Nonparametric prediction by conditional median and quantiles*. J. Stat. Plann. Inference **117**, No.2, 207-223.
- [36] Gasser, T., Hall, P., Presnell, B., (1998), *Nonparametric estimation of the mode of a distribution of random curves*. J. R. Stat. Soc., Ser. B, Stat. Methodol. **60**, No.4, 681-691.
- [37] Geffroy, J., (1974), *Sur l'estimation d'une densité dans un espace métrique*. C. R. Acad. Sci., Paris, Sér. A **278**, 1449-1452.
- [38] Hall, P., Heckman, N. E., (2002), *Estimating and depicting the structure of a distribution of random functions*. Biometrika **89**, No.1, 145-158.

- [39] Hastie, T., Buja, A., Tibshirani, R., (1995), *Penalized discriminant analysis*. Ann. Stat. **23**, No.1, 73-102.
- [40] Holmstrom, I., (1963), *On a method for parametric representation of the state of the atmosphere*. Tellus, No.15, 127-149.
- [41] Kirkpatrick, M., Heckman, N. (1989) *A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters*. J. Math. Biol. **27**, No.4, 429-450.
- [42] Kuelbs, J., Li, W. V., (1993), *Metric entropy and the small ball problem for Gaussian measures*. J. Funct. Anal. **116**, No.1, 133-157.
- [43] Kuelbs, J., Li, W. V., (1993), *Small ball estimates for Brownian motion and the Brownian sheet*. J. Theor. Probab. **6**, No.3, 547-577.
- [44] Laïb, N., Ould-Saïd, E., (2000), *A robust nonparametric estimation of the autoregression function under an ergodic hypothesis*. Can. J. Stat. **28**, No.4, 817-828.
- [45] Laksaci, A., Yousfate, A. (2002), *Estimation fonctionnelle de la densité de l'opérateur de transition d'un processus de Markov à temps discret*. C. R., Math., Acad. Sci. Paris **334**, No.11, 1035-1038.
- [46] Lehtinen, M. S., Paivarinta, L., Somersalo, E., (1989), *Linear inverse problems for generalised random variables*, Inverse Probl. **5**, No.4, 599-612.
- [47] Leão, D., Jr., Fragoso, M. D., Ruffino, P. R. C., (1999), *Characterizations of Radon spaces*. Statist. Probab. Lett. **42**, no. 4, 409-413.
- [48] Leão, D., Jr., Fragoso, M., Ruffino, P. (2004) *Regular conditional probability, disintegration of probability and Radon spaces*. Proyecciones **23**, no. 1, 15-29.
- [49] Louani, D., Ould-Saïd, E., (1999), *Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis*. J. Nonparametric Stat. **11**, No.4, 413-442.
- [50] Li, Wenbo V., Shao, Q., (1999), *Small ball estimates for Gaussian processes under Sobolev type norms*. J. Theor. Probab. **12**, No.3, 699-720.
- [51] Mogulskii, A.A., (1974) *Small deviations in a space of trajectories*. Theory Probab. Appl. **19**, 726-736.
- [52] Masry, E. (2005), *Nonparametric regression estimation for dependent functional data : asymptotic normality*. Stochastic Process. Appl. **115**, No.1, 155-177.

- [53] Molenaar, P., Boomsma, D. (1987), *The genetic analysis of repeated measures : the karhunen-loeve expansion*. Behavior Genetics, **17**, 229-242.
- [54] Obhukov, A. (1960), *The statistically orthogonal expansion of empirical functions*. American Geophysical Union, 288-291.
- [55] Ould-Saïd, E., (1993), *Estimation non paramétrique du mode conditionnel. Application à la prévision*. C. R. Acad. Sci., Paris, Sér. I **316**, No.9, 943-947.
- [56] Ould-Saïd, E., (1997), *A note on ergodic processes prediction via estimation of the conditional mode function*. Scand. J. Stat.**24**, No.2, 231-239.
- [57] Ould-Saïd, E., (2003), *Sur l'estimation fonctionnelle pour des observations complètes ou censurées*. H.D.R. Université du Littoral Côte d'opale.
- [58] Parzen, E. (1962), *On estimation of a probability density function and mode*. Ann. Math. Stat. **33**, 1065-1076.
- [59] Quintela-del-Río, A., Vieu, Ph.(1997), *A nonparametric conditional mode estimate*. J. Nonparametric Stat. **8**, No.3, 253-266.
- [60] Rachdi, M., Vieu, Ph.(2004), *Nonparametric regression for functional data : automatic smoothing parameter selection*. Preprint.
- [61] Ramsay, J. O., Silverman, B. W., (1997) *Functional data analysis*. Springer Series in Statistics. New York.
- [62] Ramsay, J. O., Silverman, B. W., (2002), *Applied functional data analysis. Methods and case studies*. Springer Series in Statistics. Springer-Verlag, New York.
- [63] Roussas, G. G., (1968), *On some properties of nonparametric estimates of probability density functions*. Bull. Soc. Math. Grèce (N.S.) **9** fasc. 1, 29-43.
- [64] Samanta, M., (1989), *Non-parametric estimation of conditional quantiles*. Stat. Probab . Lett. **7**, No.5, 407-412.
- [65] Saracco, J., (2003), *Contribution à la modélisation statistique semi- et non-paramétrique*. H.D.R. Université Montpellier II.
- [66] Sarda, P., Vieu, Ph., (2000). *Kernel regression*. Schimek, Michael G. (ed.), Smoothing and regression. Approaches, computation and application. Chichester : Wiley. Wiley Series in Probability and Statistics. 43-70.
- [67] Schlee, W., (1982), *Estimation non paramétrique du alpha-quantile conditionnel et ses dérivées partielles*. Stat. Anal. Données **7**, No.1, 32-47.

- [68] Stone, C. J., (1977), *Consistent nonparametric regression. Discussion.* Ann. Stat. **5**, 595-645.
- [69] Tortrat, A. *Calcul des Probabilités et introduction aux processus aléatoires.* Masson, Paris (1971).
- [70] Youndjé, E., Sarda, P., Vieu, Ph., (1993), *Estimateur à noyau d'une densité conditionnelle : Choix de la fenêtre pour des observations dépendantes.* C. R. Acad. Sci., Paris, Sér. I **316**, No.9, 935-938.
- [71] Youndjé, E., (1993), *Estimation non paramétrique de la densité conditionnelle par la méthode du noyau.* Thèse de Doctarat, Université de Rouen.
- [72] Youndjé, E., Sarda, P., Vieu, Ph., (1994) *Validation croisée pour l'estimation non-paramétrique de la densité conditionnelle.* Publ. Inst. Stat. Univ. Paris **38**, No.1, 57-80.
- [73] Youndjé, E., (1996), *Propriétés de convergence de l'estimateur à noyau de la densité conditionnelle.* Rev. Roum. Math. Pures Appl. **41**, No.7-8, 535-566.
- [74] Youndjé, E., Sarda, P., Vieu, Ph., (1996), *Optimal smooth hazard estimates.* Test **5**, No.2, 379-394.