UNIVERSITÉ TOULOUSE III - PAUL SABATIER

U.F.R. MATHEMATIQUE INFORMATIQUE GESTION

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE TOULOUSE
délivré par l'Université Toulouse III - Paul Sabatier

en Mathématiques Appliquées

présentée par

Laurent Delsol

intitulée

Régression sur variable fonctionnelle : Estimation, Tests de structure et Applications.

Directeurs de thèse : Frédéric Ferraty et Philippe Vieu

Soutenue le 17 juin 2008 devant le jury composé de Mesdames et Messieurs :

Denis Bosq	Université Paris VI	Rapporteur
Frédéric Ferraty	Université Toulouse II	Directeur
Peter Hall	Australian National University	Rapporteur
Pascal Sarda	Université Toulouse II	Examinateur
Winfried Stute	University of Giessen	Examinateur
Ingrid Van Keilegom	Université Catholique de Louvain	Examinatrice
Philippe Vieu	Université Toulouse III	Directeur

Institut de Mathématiques de Toulouse
Unité mixte de recherche C.N.R.S. - U.M.R. 5219
Université Paul Sabatier Toulouse 3 - Bât 1R3, 31062 TOULOUSE cedex 9, France

TABLE DES MATIÈRES

Remerciements

Je tiens pour commencer à exprimer toute ma gratitude à Frédéric Ferraty et Philippe Vieu pour m'avoir fait confiance depuis l'année de D.E.A.. Ils m'ont permis au cours de ces quatre années de progresser dans la compréhension des problématiques liées à l'étude de variables fonctionnelles. Je tiens particulièrement à les remercier pour leur disponibilité et leur convivialité tout au long de ces années. Malgré toutes leurs occupations, ils ont su être à l'écoute de mes préoccupations et de mes questions, ce qui m'a sans cesse aidé à avancer. Leurs conseils et leur expérience m'ont permis de mener à bien ce travail de thèse. J'espère que nous aurons l'opportunité de collaborer dans les prochaines années. Je voudrais enfin les remercier pour m'avoir communiqué leur passion pour la recherche que ce soit en terme de résultats théoriques ou de leur mise en oeuvre concrète.

Je tiens bien entendu à exprimer ma sincère reconnaisance à Denis Bosq et Peter Hall qui ont accepté, malgré toutes leurs occupations, d'être les rapporteurs de ce manuscrit de thèse. Je leur sais gré de l'intérêt qu'ils ont porté à mon travail. Je souhaite les remercier pour leur lecture attentive de ce mémoire ainsi que pour leurs remarques et commentaires qui m'ont permis d'en améliorer le contenu. Enfin, j'ai pu trouver dans les questions qui m'ont été formulées des perspectives intéressantes à ce mémoire. Il sera important de les prendre en compte dans les années à venir.

Je suis très heureux qu'Ingrid Van Keilegom ait accepté de faire partie de mon jury. J'ai eu la chance de travailler avec elle sur la justification théorique des méthodes de rééchantillonnage présentées au Chapitre 11 de ce mémoire. Je tiens sincèrement à la remercier pour son accueil et sa disponibilité lors de mon séjour au sein de l'Institut de Statistique à Louvain-la-Neuve. Je suis également très flatté et heureux qu'elle me propose un postdoctorat à Louvain-La-Neuve pour l'année prochaine. Nous aurons ainsi l'occasion de continuer à travailler ensemble.

Je veux exprimer ma reconnaissance à Pascal Sarda qui a bien voulu faire partie de mon jury de thèse. J'ai eu l'occasion de discuter avec lui dans cadre du groupe de travail STAPH. Je tiens à le remercier pour l'intérêt qu'il a porté à mon travail, ses commentaires constructifs ainsi que sa disponibilité pour répondre à certaines de mes questions concernant l'estimation dans le modèle linéaire fonctionnel.

Je souhaite également remercier Winfried Stute pour avoir accepté de faire partie de mon jury ainsi que pour l'attention qu'il a portée à mon travail malgré son emploi du temps très chargé.

Je tiens à remercier de manière plus générale tous les membres de l'Institut de Mathématiques de Toulouse que j'ai eu l'opportunité de cotoyer au cours de ces trois années de thèse. Je pense dans un premier temps aux membres du groupe de travail STAPH : Alain Boudou, Yves Romain, Sylvie Viguier-Pla ainsi que Lubos Prchal et Emmanuel Cabral que je remercie pour leurs commentaires, leurs remarques et leur convivialité. Je veux également exprimer ma reconnaissance à Christophe Crambes et Ali Laksaci, avec qui j'ai eu le plaisir de travailler cette année. Je tiens aussi

à remercier Bernard Bercu, Serge Cohen et Gérard Letac pour la qualité de leurs enseignements et l'intérêt qu'ils ont porté à mon travail de recherche. Mes remerciements s'adressent également à Sébastien Déjean pour sa disponibilité pour résoudre des problèmes informatiques et sa convivialité. Je souhaite enfin exprimer ma reconnaissance à tous les enseignants avec qui j'ai eu l'occasion de travailler dans le cadre de mon monitorat et plus particulièrement à Patrick Cattiaux, Anne-Marie Mondot, Philippe Monnier et Claude André Roche pour leurs conseils.

Je souhaite aussi remercier les membres de la cellule informatique de l'Institut pour leur disponibilité et leur efficacité. Je veux également exprimer ma reconnaissance à Marie-Laure Ausset, Françoise Michel et Agnès Requis auprès de qui j'ai toujours pu trouver de l'aide pour résoudre les problèmes administratifs auxquels j'étais confronté. Je souhaite enfin remercier Jacqueline pour sa convivialité et sa bonne humeur.

Au cours de ces trois années j'ai eu la chance de cotoyer de nombreux doctorants. Je tiens tout d'abord à exprimer ma gratitude à ceux qui m'ont accueilli au début de ma thèse : Alexander, Agnès, Christophe, Delphine, Diana, Ignacio, Lionel, Myriam, Renaud et Solenn. J'ai eu la chance de partager mon bureau avec Renaud et Agnès pendant un temps (un an et plus de deux ans respectivement) au cours duquel j'ai pu apprécier leur humour, leur écoute et leur bonne humeur. Je tiens à remercier particulièrement Renaud, Christophe et Lionel pour leurs conseils mais aussi pour les temps de détente que nous avons pu partager à l'occasion du quizz à la cafétéria ou sur les terrains de foot de l'université.

Ces moments, j'ai également eu la chance de les partager avec les doctorants qui ont commencé leur thèse en même temps que moi : Amélie, Florent et Maxime. Cela fait bientôt quatre ans que je les connais et que j'apprécie leur dynamisme, leur convivialité et leur bonne humeur d'autant plus que j'ai la joie de partager mon bureau avec eux. Je tiens à leur dire que j'ai eu beaucoup de plaisir à passer ces années en leur compagnie. Je remercie Amélie pour ses conseils avisés, son sens de l'organisation et sa franchise. Je tiens à exprimer toute ma reconnaissance à Maxime pour sa disponibilité et son efficacité dans la résolutions de problèmes informatiques, son sens de l'humour ainsi que pour les discussions sportives que nous avons eues. Je suis très heureux d'avoir eu l'occasion de travailler avec lui dans le cadre de mes enseignements. Enfin, je partage le même bureau que Florent depuis le début de ma thèse. Je tiens à lui dire combien sa simplicité, son écoute ainsi que les discussions que nous avons eues m'ont aidé au cours de ces années. J'ai également apprécié son expérience des méthodes statistiques et espère que l'on aura l'occasion de travailler ensemble dans les années à venir.

Je veux remercier aussi deux autres doctorants Mylène et Aurélien que j'ai la chance de connaître depuis l'année de D.E.A.pour tous les bons moments que l'on a partagés. Je tiens également à remercier les doctorants arrivés cette année ou celle d'avant : Erwan, Jean-Paul, Julie, Michel, Mathieu, Maxime, Nolwen, qui ont apporté leur bonne humeur et leur dynamisme. Je remercie également Luís et Maryana avec qui je partage mon bureau depuis le mois de mars pour leur sympathie.

Mes remerciements s'adressent également à mes anciens camarades du lycée : Sandrine, Damien, Fabien, Sylvain, et de l'université : Nicolas et Sandrine qui m'ont accompagné et soutenu au cours de ces dernières années. Je pense également à mon professeur de lycée, Mr Dessennes, qui a su nous transmettre sa passion des mathématiques et de la recherche. Je souhaite également remercier Bruno pour son soutien.

Je veux maintenant dire aux membres de ma famille combien ils me sont chers et les remercier pour leur soutien constant. Mes premières pensées vont à mes parents qui m'ont sans aucun doute transmis la passion des mathématiques et de l'enseignement. Je leur suis infiniment redevable pour avoir eu confiance en moi, même lorsque l'avenir paraissait incertain. Je suis très heureux que ma mère soit là pour me voir exposer et je pense très fort à mon père qui est présent dans ma tête et dans mon coeur. Ma soeur Lise et son mari Jérôme m'ont sans cesse encouragé au cours de ces années. Je veux les remercier pour tout ce qu'ils m'ont apporté et qui m'a permis d'avancer même dans les temps difficiles. Je les remercie d'être présents et leur souhaite tout plein de réussite pour la fin de l'année. Je tiens à exprimer toute mon affection à mon frère Christophe qui m'a soutenu au cours de ces années et a su être à l'écoute de mes soucis et de mes difficultés. Après deux ans d'"exil" à Rodez, j'ai été heureux de le retrouver cette année sur Toulouse et de partager avec lui de très agréables moments. Je veux le remercier d'être là aujourd'hui. Je tiens enfin à remercier tous les membres de ma famille et toutes les personnes proches qui m'ont apporté leur soutien durant ces années. Je vous dédie ce mémoire de thèse.

Résumé

Au cours des dernières années, la branche de la statistique consacrée à l'étude de variables fonctionnelles a connu un réel essor tant en terme de développements théoriques que de diversification des domaines d'application. Nous nous intéressons plus particulièrement dans ce mémoire à des modèles de régression dans lesquels la variable réponse est réelle tandis que la variable explicative est fonctionnelle, c'est à dire à valeurs dans un espace de dimension infinie.

Les résultats que nous énonçons sont liés aux propriétés asymptotiques de l'estimateur à noyau généralisé au cas d'une variable explicative fonctionnelle. Nous supposons pour commencer que l'échantillon que nous étudions est constitué de variables α-mélangeantes et que le modèle de régression est de nature non-paramétrique. Nous établissons la normalité asymptotique de notre estimateur et donnons l'expression explicite des termes asymptotiquement dominants du biais et de la variance. Une conséquence directe de ce résultat est la construction d'intervalles de confiance asymptotiques ponctuels dont nous étudions les propriétés aux travers de simulations et que nous appliquons sur des données liées à l'étude du courant marin El Niño. On établit également à partir du résultat de normalité asymptotique et d'un résultat d'uniforme intégrabilité l'expression explicite des termes asymptotiquement dominants des moments centrés et des erreurs $\mathbb{L}^p$ de notre estimateur.

Nous considérons ensuite le problème des tests de structure en régression sur variable fonctionnelle et supposons maintenant que l'échantillon est composé de variables indépendantes. Nous construisons une statistique de test basée sur la comparaison de l'estimateur à noyau et d'un estimateur plus particulier dépendant de l'hypothèse nulle à tester. Nous obtenons la normalité asymptotique de notre statistique de test sous l'hypothèse nulle ainsi que sa divergence sous l'alternative. Les conditions générales sous lesquelles notre résultat est établi permettent l'utilisation de notre statistique pour construire des tests de structure innovants permettant de tester si l'opérateur de régression est de forme linéaire, à indice simple, ... Différentes procédures de rééchantillonnage sont proposées et comparées au travers de diverses simulations. Nos méthodes sont enfin appliquées dans le cadre de tests de non effet à deux jeux de données spectrométriques.

English summary

Functional data analysis is a typical issue in modern statistics. During the last years, many papers have been devoted to theoretical results or applied studies on models involving functional data. In this manuscript, we focus on regression models where a real response variable depends on a functional random variable taking its values in an infinite dimensional space.

We state various kinds of results linked with the asymptotic properties of the nonparametric kernel estimator generalized to the case of a functional explanatory variable. We firstly assume that the dataset under study is composed of strong-mixing variables and focus on a nonparametric regression model. A first result gives asymptotic normality of the kernel estimator with explicit expressions of the asymptotic dominant bias and variance terms. On one hand, we propose from this result a way to construct asymptotic pointwise confidence bands and study their properties on simulation studies and apply our method on a dataset dealing with El Niño phenomenon. On the other hand, we decline from both asymptotical normality and uniform integrability results the explicit expressions of the asymptotic dominant terms of centered moments and $\mathbb{L}^p$ errors of the kernel estimator.

We now assume that the dataset is composed of independent random variables and focus on structural testing procedures in regression on functional data. We construct a test statistic based on the comparison between the nonparametric kernel estimator and a particular one that must be chosen in accordance to the null hypothesis we want to test. We then state both asymptotic normality of our test statistic under the null hypothesis and its divergence under the alternative. Moreover, we prove our result under general conditions that enable to use our approach to construct innovative structural tests allowing to check if the regression operator is linear, single index, ... Different bootstrap procedures are proposed and compared through various simulation studies. Finally, we focus on no effect tests and apply our testing procedures on two spectrometric datasets

Written and oral communications

These works have led to the following publications :

1. Delsol, L. (2007a) CLT and $\mathbb{L}^q$ errors in nonparametric functional regression *C. R. Math. Acad. Sci.* **345** (7) 411-414.
2. Delsol, L. (2007b) Régression non-paramétrique fonctionnelle : Expressions asymptotiques des moments (in French) [Nonparametric functional regression : Asymptotic expressions of moments] *Annales de l'I.S.U.P.* **LI** (3) 43-67.
3. Delsol, L. (2008a) Tests de structure en régression sur variable fonctionnelle. (in French) [Structural tests in regression on functional variable] *C. R. Math. Acad. Sci. Paris* (In press)
4. Delsol, L. (2008b) Advances on asymptotic normality in nonparametric functional Time Series Analysis *Statistics* (In press)
5. Delsol, L., Ferraty, F., Vieu, P. (2008) Structural test in regression on functional variables (submitted)

They have also been presented through the following oral communications :
a) International and national conferences :

1. *Asymptotic normality in nonparametric time series analysis* **V° Workshop of the IAP research network** March, 16-17, 2006 ; UCL, Louvain la Neuve, BELGIQUE.
2. *Normalité asymptotique de la régression sur variable fonctionnelle avec application à la construction d'I.C.* (in French) [Asymptotic normality in regression on functional variable and application to the construction a confidence bands] **IV° Journées STAPH** June, 15-16, 2006 ; UPM F, Grenoble FRANCE.
3. *Régression non-paramétrique fonctionnelle : expressions asymptotiques des moments et des erreurs* $\mathbb{L}^q$ (in French) [Nonparametric functional regression : Asymptotic expressions of moments and $\mathbb{L}^q$ errors] **39° Journées de Statistique** June, 11-15, 2007 ; UCO, Angers FRANCE
4. *Régression non-paramétrique fonctionnelle, propriétés asymptotiques et tests.* (in French) [Nonparametric functional regresion, asymptotic properties and tests] **Deuxièmes Rencontres des Jeunes Statisticiens** September, 3-7, 2007 ; Aussois FRANCE.
5. *Nonparametric regression on functional variable and structural tests*, **Journées Franco-Tchèques**, May, 5, 2008 ; Toulouse FRANCE.

b) Local working groups and seminars :

– *Normalité asymptotique de l'estimateur à noyau pour des modèles nonparamétriques de régression avec données mélangantes* (in French) [Asymptotic normality of the kernel estimator for nonparametric regression models with dependent data] **Séminaire étudiant de Statistique et Probabilités** April, 11, 2006 ; Université Toulouse III.
– *Expressions asymptotiques des moments et erreurs* $\mathbb{L}^q$ *de l'estimateur à noyau dans un modèle de régression non-paramétrique* (in French) [Asymptotic expressions of moments and $\mathbb{L}^q$ errors of the kernel estimator in a nonparametric

regression model] **Groupe de Travail STAPH** November, 13, 2006 ; Université Toulouse III.
- *Régression non-paramétrique fonctionnelle : de la normalité asymptotique à l'expression des erreurs* $\mathbb{L}^p$ (in French) [Nonparametric functional regression : from asymptotic normality to $\mathbb{L}^p$ errors expressions] **Séminaire étudiant de Statistique et Probabilités** March, 13, 2007 ; Université Toulouse III.
- *Régression non-paramétrique fonctionnelle : Tests de structure* (in French) [Nonparametric functional regression : Structural tests] **Séminaire étudiant de Statistique et Probabilités** October, 2, 2007 ; Université Toulouse III.

Forthcoming communications :
- *Nonparametric regression on functional variable and structural tests*, **International Workshop on Functional and Operatorial Statistics**, June, 19-21, 2008 ; Toulouse FRANCE
- Invitation to write a chapter entitled *Nonparametric and α-mixing methods for functional random variables* in a **collective book dealing with the state of the art in functional statistics**.
- Invitation to have a talk (*Nonparametric regression on functional variable and structural tests*) at **Joint Statistical Meetings 2008**, August, 3-7, 2008 ; Denver USA.
- Invitation to have a talk (*Structural tests in regression on functional variable*) at **Joint Meeting of 4th World Conference of the IASC Computational Statistics and Data Analysis**, December, 5-8, 2008 ; Yokohama, Japan.

PARTIE I

INTRODUCTION GÉNÉRALE

The aim of this introductory part is to present the context in which the present dissertation takes place. Chapters 2, 3, and 4 are written in French. However, we propose in Chapter 1 a brief English introduction to make easier the international reading of this dissertation. The content of Chapter 1 is mainly an English summary of Chapter 4. In Chapter 2 we motivate and introduce the notion of functional data and their analysis through functional statistics. Then, Chapter 3 presents a bibliographical review on practical problems and theoretical results linked with the study of functional data. We discuss in Chapter 4 the model and the estimator we focus on and then we sum up the contribution of this dissertation. Finally, we present in Chapter 5 some prospects and open questions.

CHAPITRE 1

A SHORT ENGLISH INTRODUCTION

The present chapter provides a short English introduction to the topic of this dissertation and sums up its content. It corresponds to an English summary of Chapter 4. French readers are invited to read the more complete introduction given in Chapters 2, 3, and 4.

With recent technical advances on measuring apparatus efficiency, data are often collected on thinner and thinner discretization grid, what make them intrinsically functional. Because standard multivariate methods are often not relevant to deal with this kind of data, there is a need to propose alternative methods adapted to the study of functional random variables. This topic of modern statistics has encountered a strong infatuation in the beginning of the 80's (see notably Grenander, 1981, Dauxois *et al.*, 1982, and Ramsay, 1982) and becomes more and more popular in terms of applications and theoretical advances (see for instance the monographs of Ramsay and Silverman, 1997, 2002, 2005, Bosq, 2000, and Ferraty and Vieu, 2006a, or the special issues of Statistica Sinica, 2004, Computational Statistics & Data Analysis, 2006, Computational Statistics, 2007, and the forthcoming special issue of Journal of Multivariate Analysis, 2008). Complementary references are given in Chapters 2 and 3 in which are presented various important topics of functional statistics. In this chapter, we briefly present our contribution to the study of regression models with functional explanatory random variable. We firstly present the functional regression model we are interested in and recall the expression of the generalized kernel estimator. Then, we sum up the various results proposed in this dissertation. Some of them generalize former works to the case of dependent datasets (see Part II). The others provide a theoretical framework to construct innovative structural tests in regression on functional variable (see Part III). We present simulations studies and various applications.

1.1. Regression models for functional covariate

In the whole dissertation, we consider regression models of the form :

$$Y = r(X) + \epsilon, \tag{1.1}$$

where the response variable Y is real-valued while the explanatory variable X takes values in a infinite dimensional semimetric space $(E,\, d)$. We also assume that the variable ϵ (corresponding to the residual) fulfills $\mathbb{E}\left[\epsilon \mid X\right] = 0$ in such a way that $r\left(x\right) = \mathbb{E}\left[Y \mid X = x\right]$. The results given in this dissertation concern the estimation of this regression operator r and the issue of testing assumptions dealing with its structure. We consider a dataset of random variables $\left(X_i, Y_i\right)_{1 \leq i \leq n}$ identically distributed as (X, Y). In order to allow to consider the context of dependent datasets, and more precisely time series study through the method proposed by Bosq (1991), we establish some results for strong-mixing (i.e. α-mixing) datasets. Regression models of the form 1.1 have already been widely studied under various kinds of assumptions dealing with the structure of the operator and the nature of the dataset (see Chapter 3, Section 2.2.3.1).

1.1.1. Parametric and nonparametric functional regression models. — In this dissertation, we consider nonparametric functional regression models and propose structural testing procedures that may be used to check for some specific parametric functional regression models. Before going further, we have to define what we call a parametric (respectively nonparametric) model in regression on functional variable. The notion of parametric and nonparametric models may be ambiguous even in the case where both response and explanatory variables are real-valued. Indeed, one may define a model as parametric if :

(a) one assumes that the law of the pair (X, Y) belongs to a space indexed by a finite number of real parameters,

or alternatively if

(b) one assumes that the regression operator r belongs to a space indexed by a finite number of real parameters.

Let us for instance consider the three following regression models :

$$Y = aX + b + \epsilon,\ X \sim \mathcal{N}\left(\mu, \gamma^2\right) \text{ et } \epsilon \sim \mathcal{N}\left(0, \sigma^2\right) \tag{1.2}$$

$$Y = aX + b + \epsilon,\ \mathbb{E}\left[\epsilon \mid X\right] = 0. \tag{1.3}$$

$$Y = r\left(X\right) + \epsilon,\ r \in \mathcal{C}\left(\mathbb{R}\right) \text{ et } \mathbb{E}\left[\epsilon \mid X\right] = 0. \tag{1.4}$$

The model (1.2) (respectively (1.4)) is parametric (respectively nonparametric) according to both points of view. However, the model (1.3) is nonparametric according to (a) but parametric according to (b). Model (1.3) seems yet more similar to model (1.2) in terms of estimation issues. That is why we propose to adopt in this dissertation a definition based on the nature of the space we assume the regression operator belongs to. We have to take into account that the explanatory variable takes its values in an infinite dimensional space. We choose to follow the ideas introduced by Ferraty and Vieu (2006a, Definition 1.4) and propose to adapt definition (b) in the following way :

Definition 1.1.1. —
– *A model of regression on functional variable is a model of the form (1.1) in which one assumes that r belongs to a specific set $\mathcal{R}$ of operators.*
– *A model of regression on functional variable is called parametric if $\mathcal{R}$ may be indexed by a finite number of elements of E.*
– *A model of regression on functional variable is called nonparametric if $\mathcal{R}$ cannot be indexed by a finite number of elements of E.*

We propose the following examples to illustrate the previous definition :

1. The functional linear regression model is a model of regression on functional variable in which we assume the space E is an Hilbert space and r is a continuous linear operator (i.e. $\mathcal{R} = \mathcal{L}^C(E, \mathbb{R})$). Riesz representation theorem, implies that there exists a unique element α of E such that $r(\,.\,) =< .\,, \alpha >_E$. Consequently, any element of $\mathcal{R}$ is indexed by an element of E. According to the previous definition, the functional linear regression model is parametric and may be written :

$$Y =< X, \alpha >_E + \epsilon. \tag{1.5}$$

2. We now consider the situation where no assumption is made on the structure of the regression operator. It is yet necessary to make assumptions on its regularity. We assume that r is an Hölderian operator. The set $\mathcal{R}$ cannot be indexed by a finite number of elements of E. This kind of model is hence called nonparametric.

3. Let us finally focus on the functional single index model in which we assume $\mathcal{R} = \{\Phi(< .\,, \theta >_E), \theta \in E, \Phi \in Hol(\beta)\}$, where $Hol(\beta)$ stands for the set of Hölderian functions of order β defined on $\mathbb{R}$ and taking values in $\mathbb{R}$. This kind of model may be written :

$$Y = \Phi(< \theta, X >_E) + \epsilon. \tag{1.6}$$

According to the previous definition, this model is parametric if, and only if, the set $Hol(\beta)$ may be indexed by a finite number of elements of E. This highlights ambiguities that an explicit definition may lead to. By analogy with the multivariate case this model is called semi-parametric.

In this dissertation we mainly focus on nonparametric regression models like the one presented in the second example.

1.1.2. Dependency models. — In many situations we have to deal with dependent functional datasets. One of the most popular example comes from the study of a time series through the functional approach proposed by Bosq (1991). Instead of considering the time series observed on $[0, \tau N]$ through its discretized values, one uses a continuous time process modelization. Then the main idea is to cut the process into N successive curves $X_i(t) = \xi_{t+(i-1)\tau}, t \in [0, \tau[\,, i = 1, \ldots, N$. We assume that the initial process is stationary and aim to predict a value of the forthcoming

period from the last observed one. That leads us to consider the following functional regression model :

$$G(X_i) = r(X_{i-1}) + \epsilon_i, \tag{1.7}$$

where G is a known operator defined on E and taking its values in $\mathbb{R}$. Consequently, one has to study a regression model of the form (1.1) in the case of a dependent dataset. We implicitly assume $\mathbb{E}[G(X_i)|X_{i-1}]$ exists and is independent of i in order to be able to define the regression operator r. This condition is fulfilled when one makes the stronger assumption that the process ξ_t is strictly stationary. Strict stationarity is not necessary but additional assumptions may be considered to state some theoretical results.

There are various ways to modelize the dependence structure in a dataset. We focus in this dissertation on weakly dependent variables. Many modelizations of weak dependence have been proposed through the notions of α-mixing (Rosenblatt, 1956), β-mixing (Volkonskii et Rosanov, 1959), ϕ-mixing (Ibragimov, 1962), or others (see Doukhan, 1994, Doukhan et Louhichi, 1999, Bradley, 2005, Dedecker et Prieur, 2005). We will consider in this dissertation the notion of α-mixing variables that has been introduced by Rosenblatt (1956) in the following way :

Definition 1.1.2. —
- *Let $(\zeta_i)_{i\in\mathbb{Z}}$ be a sequence of random variables. One introduces the notations $\mathcal{F}_k = \sigma(X_i,\, i \le k)$ and $\mathcal{G}_l = \sigma(X_i,\, l \le i)$. The α-mixing coefficients of the sequence $(\zeta_i)_{i\in\mathbb{Z}}$ are defined by :*
$$\alpha(n) = \sup_{k\in\mathbb{Z}} \sup_{A\in\mathcal{F}_k,\, B\in\mathcal{G}_{n+k}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\,\mathbb{P}(B)|.$$
- *A sequence of random variables $(\zeta_i)_{i\in\mathbb{Z}}$ is called α-mixing if its α-mixing coefficients fulfill $\alpha(n) \overset{n\to+\infty}{\to} 0$.*

One also introduces the following definitions on the way the coefficients decrease to 0.

Definition 1.1.3. —
- *The α-mixing coefficients are called arithmetic of order $a > 0$ if there exists a positive constant C such that $\alpha(n) \le Cn^{-a}$.*
- *The α-mixing coefficients are called geometric if there exist a positive constant C and $0 \le \rho < 1$ such that $\alpha(n) \le C\rho^n$.*

Many results have been obtained for α-mixing sequences of random variables (see Doukhan, 1994, Bosq, 1996, 1998, Rio, 2000, and Yoshihara, 2004 and the references given in Chapter 4). They are useful to get the results proposed in Part II of this manuscript. The extension of our results to other dependence modelizations is discussed in Chapter 4.

1.2. Kernel estimation of the regression operator

1.2.1. Definition of the kernel estimator. — An historical and bibliographical introduction to kernel estimation in regression may be found in Chapter 4. More references on this topic are given by Collomb (1981, 1984, 1985), Györfi *et al.* (1989), Härdle (1990), Yoshihara (1994), Bosq (1996, 1998) and Sarda et Vieu (1999). In the whole dissertation we consider a generalization of the Nadaraya-Watson kernel estimator introduced by Ferraty et Vieu (2000) and study its asymptotic properties. For any element of E it is defined in the following way :

$$\hat{r}(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{d(X_i,x)}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{d(X_i,x)}{h_n}\right)}, \tag{1.8}$$

where d stands for a semimetric defined on the space E, h_n is the smoothing parameter and K is a kernel function that is nonnegative and nonincreasing on its support $[0,1]$. From a theoretical point of view it remains to define our estimator in the case where the denominator equals 0. In this case, the numerator equals 0 too (since K is nonnegative) and we choose (by convention) an usual way to define our estimator : $\hat{r}(x) = 0$.

Many asymptotic results have been obtained concerning almost sure convergence (see Ferraty and Vieu 2000, 2002 in the independent case and Ferraty *et al*, 2002a, 2002b, in the α-mixing case), asymptotic normality (see Ferraty *et al.*, 2007, in the independent case and Masry, 2005, in the α-mixing case), and convergence in $\mathbb{L}^p$ norm in the independent case (see Dabo-Niang and Rhomari, 2002). Recently some of these results have been extended to the case of long memory processes (see Hedli-Griche, 2008).

1.2.2. Semimetrics and small ball probabilities. — The curse of dimensionality is a well-known phenomenon in nonparametric regression on multivariate variable (see Stone, 1982). In multivariate nonparametric regression, convergence rates (for the dispersion part) are expressed in terms of h_n^d. In the functional case we adopt more general concentration notions called small ball probabilities and express our asymptotic results in function of these quantitities. Small ball probabilities are defined by :

$$F_x(h_n) := \mathbb{P}(d(X,x) \leq h_n).$$

The way they decrease to zero have a great influence on the convergence rate of the kernel estimator (1.8). One can find in many probability papers asymptotic equivalents for these small ball probabilities when d is a norm (see for instance Bogachev, 1998, Li et Shao, 2001, Dereich, 2003, Lifshits *et al.*, 2006, Shmileva, 2006, Gao et Li, 2007) or a specific seminorm (Lifshits et Simon, 2005, Aurzada et Simon, 2007). In the case of unsmooth processes (Brownian motion, Ornstein-Ullenbeck process, ...), these small ball probabilities have an exponential form (with regard to h_n) and hence the convergence rate is a power of $\log(n)$ (see Ferraty *et al.*, 2006, Section 5, Ferraty et Vieu, 2006a, Paragraph 13.3.2.). The choice of the

semimetric d has a direct influence on the topology and consequently on small ball probabilities. The diversity of semimetrics allows, in various situations, to find a topology that gives a relevant notion of proximity between curves (see Ferraty and Vieu, 2006a, Chapter 3). For instance, the choice of a projection semimetric (based on functional principal components, Fourier's decomposition, wavelets, splines, ...) constitutes an interesting alternative to the curse of dimensionality. When E is an separable Hilbert space, Ferraty et Vieu (2006a, Lemme 13.6, p.213), show that there exists a specific semimetric that leads to fractal small ball probabilities (i.e. $\exists C, \delta > 0$, $F_x(h) \overset{h\to 0}{\sim} C_x h^{\delta}$) hence the convergence rate of the kernel estimator is a power of n. In other situations, typically when the dataset is composed of smooth curves (like spectrometric curves presented in Chapter 3) it may be interesting to use semimetrics based on derivatives.

One may wonder how to choose the semimetric in practice. A first method has been proposed by Ferraty *et al.* (2002b). One firstly has to choose a family of semimetrcs from the information one gets on the data. Then, one determines, for instance by cross-validation, the semimetric (among this family) that is the most adapted to the data. Theoretical justification of the usefulness of a particular semimetric is still an open problem.

1.3. Contribution in kernel estimation of the regression operator

In this section, we sum up some of our results dealing with pointwise asymptotic convergence properties of the kernel estimator (1.8) and consider a dataset of α-mixing variables. In order to make the reading easier we do not explicit the assumptions that allow to get these results. Assumptions, comments and proofs are presented in detail in Chapters 6-9 in Part II.

One considers an element $x \in E$ and want to estimate $r(x)$. In order to present our results we have to introduce the following notations :

$$\begin{aligned}
&\sigma_\epsilon^2(X) := \mathbb{E}\left[\epsilon^2 \mid X\right] \text{ and } \sigma_\epsilon^2 := \sigma_\epsilon^2(x), \\
&\phi(s) = \mathbb{E}\left[\left(r(X) - r(x)\right) | d(X,x) = s\right], \\
&F(h) = \mathbb{P}\left(d(X,x) \leq h\right), \tau_h(s) = \frac{F(hs)}{F(h)}, s \in [0,1].
\end{aligned}$$

We assume that τ_h converges almost surely when h tends to 0 and we call the limit τ_0. We also introduce the following constants appearing in our results :

$$M_0 = K(1) - \int_0^1 \left(sK(s)\right)' \tau_0(s)\, ds,\ M_j = K^j(1) - \int_0^1 \left(K^j\right)'(s)\, \tau_0(s)\, ds,\ j = 1, 2.$$

1.3.1. Asymptotic normality. — The first result on asymptotic normality of the kernel estimator (1.8) comes from Masry (2005). He considers the case of a α-mixing dataset but does not give the explicit expression of the asymptotic dominant terms of bias and variance. Then, Ferraty *et al.* (2007) provide the explicit expression of the asymptotic law (i.e. of the dominant terms of bias and variance) in the case of an independent dataset. The results we state in the papers Delsol (2007a, 2008b) (see Chapters 6 and 8, Part II) make the link between these articles. They generalise the results of Ferraty *et al.* (2007) to the case of an α-mixing dataset. One gets the following result whose assumptions and proofs are given in the Chapter 6 (Part II) of this dissertation.

Theorem 1.3.1. — *Under various sets of assumptions, notably concerning α-mixing coefficients, one gets :*

$$\frac{M_1}{\sqrt{M_2\sigma_\epsilon^2}}\sqrt{n\hat{F}\left(h_n\right)}\left(\hat{r}\left(x\right)-r\left(x\right)-B_n\right)\rightarrow\mathcal{N}\left(0,1\right),$$

where $B_n=h_n\phi'\left(0\right)\frac{M_0}{M_1}$ *and* $\hat{F}\left(t\right)=\frac{1}{n}\sum_{i=1}^{n}1_{[d(X_i,x),+\infty[}\left(t\right)$.

From the explicit expressions of bias and variance asymptotic dominant terms, we propose a way to construct asymptotic pointwise confidence bands (see Chapter 6, Part II).

1.3.2. Asymptotic expressions of centered moments. — With both asymptotic normality and uniform integrability results, we get the explicit expressions of the asymptotic dominant terms of the centered moments of order q, smaller than a threshold ℓ, of our kernel estimator (see Delsol, 2007a, 2007b). We get the following result whose assumptions and proof are given in Chapter 7 (Part II) of this dissertation. We introduce a random variable $W\sim\mathcal{N}\left(0,1\right)$ to express our results.

Theorem 1.3.2. — *Under some technical assumptions, for all $q\in\mathbb{N}$ such that $0\leq q<\ell$, one gets :*

$$\mathbb{E}\left[\left(\hat{r}\left(x\right)-\mathbb{E}\left[\hat{r}\left(x\right)\right]\right)^q\right]=\left(\sqrt{\frac{M_2\sigma_\epsilon^2}{nF\left(h_n\right)M_1^2}}\right)^q\mathbb{E}\left[W^q\right]+o\left(\frac{1}{\left(nF\left(h_n\right)\right)^{\frac{q}{2}}}\right).$$

1.3.3. Asymptotic expressions of $\mathbb{L}^p$ errors. — With similar arguments we also get the explicit expressions of the asymptotic dominant terms of $\mathbb{L}^p$ errors ($p\leq\ell$) in the case of α-mixing variables. Dabo-Niang and Rhomari (2002) give $\mathbb{L}^p$ errors convergence rates in the case of independent variables but do not explicit the asymptotic dominant terms. We get the following result whose assumptions and proof are given in Chapter 7 (Part II) of this dissertation.

Theorem 1.3.3. — *Under some technical assumptions, one gets :*

(i) $\forall\ 0 \leq q < \ell$,

$$\mathbb{E}\left[\left|\hat{r}\left(x\right) - r\left(x\right)\right|^{q}\right] = \mathbb{E}\left[\left|h_n B + W\frac{V}{\sqrt{nF\left(h_n\right)}}\right|^{q}\right] + o\left(\frac{1}{\left(nF\left(h_n\right)\right)^{\frac{q}{2}}}\right),$$

(ii) $\forall m \in \mathbb{N},\ 2m < \ell$,

$$\mathbb{E}\left[\left|\hat{r}\left(x\right) - r\left(x\right)\right|^{2m}\right] = \sum_{k=0}^{m} \frac{V^{2k} B^{2(m-k)}\left(2m\right)!}{\left(2\left(m-k\right)\right)! k! 2^{k}} \frac{h_n^{2(m-k)}}{\left(nF\left(h_n\right)\right)^{k}} + o\left(\frac{1}{\left(nF\left(h_n\right)\right)^{m}}\right),$$

(iii) $\forall m \in \mathbb{N},\ 2m+1 < \ell$,

$$\mathbb{E}\left[\left|\hat{r}\left(x\right) - r\left(x\right)\right|^{2m+1}\right] = \frac{1}{\left(nF\left(h_n\right)\right)^{m+\frac{1}{2}}}\left(V^{2m+1}\psi_m\left(\frac{Bh_n\sqrt{nF\left(h_n\right)}}{V}\right) + o\left(1\right)\right),$$

where Ψ_m *is an explicit function,* $B = \phi'\left(0\right)\frac{M_0}{M_1}$, *and* $V = \sqrt{\frac{M_2\sigma_\epsilon^2}{M_1^2}}$. *(see Chapter 7, Part II).*

This result comes from Delsol (2007a, 2007b) (see Chapters 7 and 8, Part II). It is innovative in the functional case and seems also new in the multivariate case. Indeed, even if the $\mathbb{L}^2$ error has been widely studied, papers expliciting the expression of dominant terms of other $\mathbb{L}^p$ errors seem to be reduced to the paper of Wand (1990). He considers the case of a real fixed design and the $\mathbb{L}^1$ error. Moreover, our result opens interesting perspectives in terms of the choice of the smoothing parameter.

1.4. Contribution to structural testing procedures

The literature devoted to structural tests in regression on functional variable, seems to be reduced to four papers. Cardot *et al.* (2003) and Cardot *et al.* (2004) focus on structural tests in the functional linear model. Dembo and Gadiaga (2005) propose a no-effect test based on projection techniques while Chiou and Müller (2007) propose an heuristic goodness-of-fit test. It seems that no theoretical method has yet been proposed to test if the regression operator has a specific structure : linear, single index, ... We propose a theoretical framework to construct innovative structural tests. We adapt the method proposed by Härdle and Mammen (1993) to the case of a functional explanatory variable. We now assume that the dataset is composed of independent variables.

1.4.1. Testing if $r = r_0$. — We start with the simplest case in which we aim to test the null hypothesis

$$\mathcal{H}_0\ :\ \left\{\mathbb{P}\left(r\left(X\right) = r_0\left(X\right)\right) = 1\right\},$$

where r_0 is a known operator, against the local alternative

$$\mathcal{H}_{1,n} \ : \ \left\{ \|r - r_0\|_{\mathbb{L}^2(wdP_X)} \geq \eta_n \right\}.$$

The way η_n tends to 0 reflects the capacity of our test to detect smaller and smaller differences between r and r_0 when n grows. We take a leaf out of Härdle and Mammen (1993) and González-Manteiga *et al.* (2002) to construct the following test statistic :

$$T_n = \int \left(\sum_{i=1}^{n} \left(Y_i - r_0 \left(X_i \right) \right) K \left(\frac{d\left(x, X_i \right)}{h_n} \right) \right)^2 w\left(x \right) dP_X\left(x \right),$$

where w is a weight function. We study the asymptotic distribution of this test statistic. To express the asymptotic distribution of T_n under $\mathcal{H}_0$, we introduce $T_{1,n}$ and $T_{2,n}$ that provide bias and variance dominant terms.

$$\begin{aligned} T_{1,n} &= \int \sum_{i=1}^{n} K^2 \left(\frac{d\left(X_i, x \right)}{h_n} \right) \epsilon_i^2 w\left(x \right) dP_X\left(x \right), \\ T_{2,n} &= \int \sum_{1 \leq i \neq j \leq n} K \left(\frac{d\left(X_i, x \right)}{h_n} \right) K \left(\frac{d\left(X_j, x \right)}{h_n} \right) \epsilon_i \epsilon_j w\left(x \right) dP_X\left(x \right). \end{aligned}$$

In Delsol *et al.* (2008) (see also Delsol, 2008a) we state the following result whose assumptions and proof are given in the Chapter 10 (Part III) of this dissertation.

Theorem 1.4.1. — *Under general conditions one gets :*
– *under* $(\mathcal{H}_0)$, $\frac{1}{\sqrt{Var(T_{2,n})}} \left(T_n - \mathbb{E}\left[T_{1,n} \right] \right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 1 \right)$,
– *under* $(\mathcal{H}_{1,n})$, $\frac{1}{\sqrt{Var(T_{2,n})}} \left(T_n - \mathbb{E}\left[T_{1,n} \right] \right) \xrightarrow{P} +\infty$.

1.4.2. Testing if $r \in \mathcal{R}$. — In this paragraph, we use the ideas of the previous one to construct a more general test statistic allowing to check if the regression operator r belongs to a convex set $\mathcal{R}$ of operators. We propose to test the null hypothesis

$$\mathcal{H}_0 : \{ \exists r_0 \in \mathcal{R},\, P\left(r\left(X \right) = r_0\left(X \right) \right) = 1 \}$$

against the local alternative

$$\mathcal{H}_{1,n} : \left\{ \inf_{r_0 \in \mathcal{R}} \|r - r_0\|_{\mathbb{L}^2(wdP_X)} \geq \eta_n \right\}.$$

We assume we have two independent datasets D and D^* of respective lengths n and m_n. To test if $r \in \mathcal{R}$ we would like to use the projection r_0 of r on $\mathcal{R}$ but it is an unknown operator. We propose to estimate r_0 by an estimator r_0^* computed from the dataset D^* that has good convergence properties under $\mathcal{H}_0$ and we introduce the following test statistic :

$$T_n^* = \int \left(\sum_{i=1}^{n} \left(Y_i - r_0^*\left(X_i \right) \right) K \left(\frac{d\left(x, X_i \right)}{h_n} \right) \right)^2 w\left(x \right) dP_X\left(x \right).$$

If r_0^* fulfills general conditions, a result similar to Theorem 1.4.1 may be obtained. Details on assumptions and proof are given in the papers Delsol *et al.* (2008) and Delsol (2008a) (presented in Chapter 10, Part III).

Theorem 1.4.2. — *Under general conditions one gets :*
– *under* $(\mathcal{H}_0)$, $\frac{1}{\sqrt{Var(T_{2,n})}}\left(T_n^* - \mathbb{E}\left[T_{1,n}\right]\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,1\right)$,
– *under* $(\mathcal{H}_{1,n})$, $\frac{1}{\sqrt{Var(T_{2,n})}}\left(T_n^* - \mathbb{E}\left[T_{1,n}\right]\right) \xrightarrow{P} +\infty$.

This result is given under general assumptions that allow the use of our statistic to construct many innovative structural tests (see Chapter 10, Part III). For instance, it may be used to check if the regression operator is linear, single-index, if the effect of the explanatory functional variable may be reduced to the effect of vector composed of a finite number of features (extrema, inflexion points, integral, ...), ...

1.5. Some applications

We have introduced a method to construct asymptotic pointwise confidence bands and another to construct structural testing procedures. Simulation studies have been made to analyse the behaviour of these methods on finite samples before their application on real datasets.

The efficiency of our method to construct asymptotic pointwise confidence bands has been studied under various kinds of dependence (see Chapter 6, Part II). We have seen that, when n is fixed, dependence influences the performance of our estimated asymptotic confidence bands. However, the results obtained for α-mixing variables are fairly good. We have then applied our method on concrete datasets linked to the study of El Niño phenomenon and electricity consumption (see Chapter 6 and 9, Part II). We have also briefly compared, on the dataset linked to El Niño phenomenon, the performance of our kernel estimator with regard to estimators proposed respectively for the functional linear model and for an additive multivariate model. Our nonparametric functional approach seems to be adapted to the study of this phenomenon (see Chapter 6, Part II).

Finally, to apply our structural testing procedures, we need to estimate the critical value. The estimation of bias and variance terms seems difficult and it is often not relevant to estimate the critical value directly from the quantile of the asymptotic law. That is why we propose to compute empirical quantiles with a bootstrap procedure. We introduce various bootstrap procedures. We analyse their performances in the context of a no-effect test on simulated datasets (see Chapter 11, Part III).

We obtain satisfying results in terms of level and power. Then, no effect tests are applied on two spectrometric datasets. Simulations on linearity tests are in progress.

1.6. The contribution in curve discrimination

We now assume to be faced with a discrimination problem with k classes. In the dataset, any functional random variable X_i is linked to a random variable $T_i \in \{1, \ldots, k\}$ that stands for the class X_i belongs to. When we observe a new curve x, we aim to predict its corresponding class t (T stands for the corresponding random variable). The results given in this dissertation are expressed in the general context of regression on functional variable. They are yet linked to functional discrimination issues. Indeed, a classical way to predict t (see for instance Hall *et al.*, 2001, or Ferraty and Vieu, 2003) consists in estimating for each class j the probability

$$\pi_x^j := \mathbb{P}\left(T = j | X = x\right)$$

and then in predicting t by the value j for which $\hat{\pi_x^j}$ is the greatest. Now for each class j, the estimation of π_x^j corresponds to the estimation of the regression operator of the functional regression model associated to the variables

$$\left(X_i, Y_i^j\right), \text{ où } Y_i^j = 1_{\{T_i = j\}}.$$

Consequently, the results stated in this dissertation complete the literature on kernel estimation of π_x^j and open interesting prospects in terms of structural testing procedures.

1.7. Brief outline of the dissertation

The remaining of the manuscript is organized as follows. The first part goes on with three introductory chapters written in French. The first one is a general introduction to the aims and scopes of functional data analysis. In the second chapter we make a bibliographical review on various topics and models that have been considered. Finally, we present in the third chapter the model we are interested in and give a brief presentation of our contributions. To conclude Part I, we present and discuss in Chapter 5 some prospects and open questions linked with the content of this dissertation. In the second part, we focus more precisely on estimation issues in the case of a strong-mixing dataset. The fourth chapter is devoted to asymptotic normality and confidence bands. Chapter 7 is written in french and deals with explicit expressions of the asymptotic dominant terms involved in $\mathbb{L}^q$ errors. The reader may find in Chapter 8 an english note that summarizes the results of Chapters 6 and 7, while Chapter 9 gathers additional comments and applications. Finally, in Part III, we consider an independent dataset and focus on structural testing procedures in regression on functional variables. Chapter 10 presents some theoretical results allowing to construct general structural tests. In Chapter 11 we propose various bootstrap testing procedures and study their behaviour on simulated datasets.

CHAPITRE 2

INTRODUCTION À L'ANALYSE DE DONNÉES FONCTIONNELLES

L'objectif de ce chapitre est de motiver et d'introduire la notion de données fonctionnelles ainsi que l'approche statistique fonctionnelle au travers de laquelle nous allons les étudier.

2.1. Motivations

De très nombreux travaux ont été dédiés à l'étude de modèles impliquant des variables aléatoires multivariées et c'est un domaine de la statistique toujours très étudié. Cependant, les récentes innovations réalisées sur les appareils de mesure et les méthodes d'acquisition ainsi que l'utilisation intensive de moyens informatiques permettent souvent de récolter des données discrétisées sur des grilles de plus en plus fines, ce qui les rend intrinsèquement fonctionnelles. Les courbes de croissance, les enregistrements sonores, les images satellites, les séries chronologiques, les courbes spectrométriques ne sont que quelques exemples illustrant le grand nombre et la diversité des données de nature fonctionnelle auxquelles le statisticien peut être confronté. C'est une des raisons pour lesquelles un nouveau champ de la Statistique, dédié à l'étude de données fonctionnelles, a suscité un fort engouement au début des annés quatre-vingt, sous l'impulsion, notament, des travaux de Grenander (1981), Dauxois *et al.* (1982) et Ramsay (1982). Il a été popularisé par Ramsay et Silverman (1997), puis par les différents ouvrages de Bosq (2000), Ramsay et Silverman (2002), Ramsay et Silverman (2005), et Ferraty et Vieu (2006). C'est un domaine de la Statistique en plein essor comme en témoignent les hors-séries qui lui sont ou seront consacrés dans des revues reconnues comme Statistica Sinica (2004), Computational Statistics & Data Analysis (2006), Computational Statistics (2007), Journal of Multivariate Analysis (2008) (on pourra se rapporter au Chapitre 3 pour une bibliographie plus complète). De plus, même si les données dont dispose le statisticien ne sont pas de nature fonctionnelle, il peut être amené à étudier des variables fonctionnelles construites à partir de son échantillon initial. Un exemple classique est celui où l'on observe plusieurs échantillons de données réelles indépendantes et où l'on est ensuite amené à comparer les densités de ces différents échantillons ou bien à considérer des modèles où elles interviennent (on pourra se référer à Ramsay et Silverman, 2002, pour plus de détails). Dans le contexte particulier de l'étude de

séries temporelles l'approche introduite par Bosq (1991) fait apparaitre une suite de données fonctionnelles dépendantes qui modélisent la série chronologique observée. Elle consiste tout d'abord à voir le processus non plus au travers de sa forme discrétisée mais comme un processus à temps continu puis à le découper en un échantillon de courbes successives. Enfin, on trouve également dans la littérature de nombreux travaux portant sur des données dites longitudinales provenant de mesures répétées d'un même phénomène au cours du temps. Il s'agit en général de données discrétisées en seulement quelques points et les méthodes utilisées diffèrent souvent de celles utilisées lors de l'étude de variables aléatoires fonctionnelles (voir Davidian *et al.*, 2004, et Hall *et al.*, 2006, pour une comparaison plus approfondie). Cependant, certains outils de la statistique fonctionnelle peuvent s'adapter à ce type particulier de données fonctionnelles comme le montrent notamment les articles de James *et al.* (2000), James et Sugar (2003), Zhao *et al.* (2004), Müller (2005) et Hall *et al.* (2006).

2.2. Les défauts d'une approche multivariée classique

La manière la plus naïve de considérer ces données est de les voir comme réalisations d'une variable aléatoire multivariée de très grande dimension et d'y appliquer des méthodes multivariées classiques. Toutefois cette approche se heurte à un certain nombre d'écueils théoriques et pratiques. Tout d'abord, le nombre de points de discrétisation (c'est à dire la dimension de la variable à étudier) est souvent important par rapport à la taille de l'échantillon. De plus, les points de discrétisation ne sont pas toujours espacés régulièrement et il arrive fréquemment que la nature des points de discrétisation, ainsi que leur nombre, soit très différents d'une courbe à une autre. D'autre part, les différentes composantes de la variable multivariée étudiée sont clairement dépendantes car elles sont le fruit de la discrétisation d'une même fonction. Enfin, cette approche directe ne permet pas de prendre en compte ni de tirer profit de la nature fonctionnelle des données observées.

Afin d'illustrer les problèmes que l'on risque de rencontrer au travers d'une analyse multivariée de données fonctionnelles, nous allons nous intéresser à deux méthodes très couramment utilisées dans le contexte multivarié. Supposons que l'on dispose de données provenant de la discrétisation de n courbes aux mêmes p points $(t_1, \ldots, t_p)$, et que l'on cherche la relation entre ces courbes et d'autres variables réelles. Supposons tout d'abord que l'on veuille utiliser la méthode des moindres carrés ordinaires. Cette méthode, très "classique" pour des variables multivariées peut donner de mauvais résultats lorsque l'on étudie des variables fonctionnelles. En effet, si $x_{i,j}$ est la valeur de la courbe x_i au point t_j et $X = (x_{i,j})$ est la matrice qui regroupe ces valeurs, la méthode des moindres carrés ordinaires nécessite d'inverser la matrice ${}^tXX$ mais le fait que les colonnes de X soient dépendantes les unes des autres et que l'on puisse avoir un grand nombre p de points de discrétisation par rapport aux nombre de courbes n risque fort de rendre cela compliqué voir impossible.
Pour contourner ce problème on a envisagé d'autres méthodes telles que la "ridge

regression" (introduite initialement par Hoerl et Kennard, 1980), la regression sur composantes principales (Massy, 1965), ou la regression "partial least square" (voir Wold, 1975, Martens et Naes, 1989, Helland, 1990). Ces différentes méthodes ont été comparées par Frank et Friedman (1993) et dans leur discussion de cet article Hastie et Mallows (1993) ont insisté sur l'aspect réducteur de voir une courbe au travers du vecteur de ses valeurs discrétisées.

Supposons maintenant que l'on souhaite estimer, de manière non-paramétrique (c'est à dire sans supposer de forme particulière), l'effet de n courbes (discrétisées en p points) sur n variables réelles. D'après les résultats de type minimax obtenus par Stone (1980,1982,1983), la vitesse optimale de convergence des estimateurs non-paramétriques (de la fonction de régression mais aussi de la densité, des quantiles,...) se détériore très rapidement lorsque la dimension de la variable explicative augmente. Cela est dû à un phénomène bien connu de raréfaction des données contenues dans une boule de rayon fixé, lorsque la dimension augmente, que l'on appelle couramment fléau de la dimension. Par conséquent, lorsque le nombre p de points de discrétisation augmente, c'est à dire lorsque nous avons une meilleure connaissance de nos courbes, les estimateurs non-paramétriques construits à partir des valeurs discrétisées ont des vitesses de convergence moins bonnes. Cela semble un peu paradoxal car en ayant plus de connaissances des courbes on devrait savoir mieux faire ou tout au moins aussi bien. Comme de plus on peut s'attendre, grâce aux améliorations futures des systèmes d'acquisition et de stockage de données, à ce que les données soient discrétisées sur des grilles bien plus fines que ce qu'elles sont à présent, il est important de donner une méthode qui puisse s'adapter facilement à n'importe quelle finesse de discrétisation sans perdre de la qualité. Toutefois, dans le cas où l'on fait varier en même temps le pas de discrétisation et la longueur de l'intervalle sur lequel est observé le processus, une trop grande "finesse de discrétisation" peut être nuisible et faire "exploser" l'erreur d'estimation. On pourra voir à ce sujet les travaux de Blanke et Pumo (2003) ainsi que Bosq et Blanke (2007).
Les résultats donnés par Stone (1980, 1982, 1983) concernent la vitesse optimale de n'importe quel estimateur non-paramétrique, par conséquent le problème du fléau de la dimension n'est pas lié à l'estimateur non-paramétrique utilisé mais plutôt à la nature du modèle étudié. C'est à partir de ce constat qu'ont été faites des propositions de modèles à réduction de dimension. On peut notamment citer le modèle additif (Stone, 1985), le modèle additif généralisé (Hastie et Tibshirani, 1986, 1990), le modèle à transformations optimales (Breiman et Friedman, 1985), le modèle à projections révélatrices (Diaconnis et Shashahani, 1984) et le modèle à indices multiples (Li, 1991). Enfin, si ces modèles sont satisfaisants en statistique multivariée en terme de vitesse de convergence en ayant contourné le fléau de la dimension, leur application à des données fonctionnelles ne prend pas en compte la nature fonctionnelle des données et n'en tire pas profit.

2.3. Une approche fonctionnelle

Comme nous venons de le voir dans le paragraphe précédent, les méthodes classiques d'analyses de données multivariées ne semblent pas adaptées à l'étude de données provenant de variables de nature fonctionnelle. C'est pourquoi, plutôt que de considérer chacune de ces données de manière directe comme réalisation d'une variable aléatoire multivariée de grande dimension, on va choisir de prendre en compte sa nature fonctionnelle en la regardant comme discrétisation de la réalisation d'une variable aléatoire (dite fonctionnelle) à valeurs dans un espace $\mathcal{E}$ de dimension infinie.

De la même manière que dans le cas de l'analyse de données multivariées les atomes de base que l'on étudie sont des vecteurs aléatoires de dimension finie, les atomes de base de l'analyse de données fonctionnelles sont des variables aléatoires fonctionnelles. On a donc été amené à introduire de nouvelles notions, à construire de nouveaux outils théoriques et appliqués adaptés à l'étude de variables aléatoires fonctionnelles. Etant donné la relative jeunesse de la statistique fonctionnelle, il reste encore des pistes non explorées, des modèles non-étudiés, des questions ouvertes et sans doute la possibilité d'améliorer certains résultats. Cependant, les nombreux travaux consacrés à la statistique fonctionnelle (dont nous reparlerons plus en détails au Chapitre 3) ont permis de proposer des solutions à des problèmes très variés et donnent un panorama assez complet des situations que l'on peut rencontrer et des méthodes que l'on peut alors utiliser. Sous l'impulsion de l'article de Ramsay (1982), de nombreux travaux se sont attelés à essayer de proposer des variantes astucieuses de modèles et de méthodes déjà couramment utilisés pour des variables multivariées, qui permettent de prendre en compte et d'utiliser la nature fonctionnelle des variables et de contourner au mieux les problèmes que l'application directe de méthodes multivariées risquerait de provoquer. D'autre part, la nature fonctionnelle des variables étudiées a amené l'apparition de certaines notions et problématiques nouvelles : enregistrement de courbes, courbe médiane, choix de semi-métrique (voir Paragraphe 2.5), modélisation des courbes comme solutions d'une équation différentielle, ...

Le principal intérêt de considérer et de modéliser de cette manière les données fonctionnelles semble bien être la capacité à pouvoir prendre en compte des caractères typiquement fonctionnels comme par exemple les dérivées, les primitives, l'intégrale, les valeurs en des points non observés, la périodicité, la régularité, le taux de variation, le déphasage de deux courbes, ... D'autre part, on a remplacé un ensemble de p variables dépendantes par une seule variable fonctionnelle pour laquelle on a construit des modèles et des méthodes adéquats. Cette approche peut être vue comme une alternative au fléau de la dimension car les performances des méthodes qui ont été développées ne sont pas déteriorées lorsque le pas de discrétisation augmente. Par conséquent, la méthode restera la même si dans quelques années on dispose de données discrétisées sur des grilles plus fines. De plus, on n'est plus confronté au paradoxe qu'une meilleure connaissance des données (discrétisation sur une grille plus fine), entraîne une détérioration des vitesses de convergence de nos

estimateurs. Enfin, il est clair que les variables multivariées sont des variables fonctionnelles particulières, c'est pourquoi l'étude de variables aléatoires fonctionnelles peut être vue comme une généralisation de l'étude de variables aléatoires multivariées. On peut donc appliquer toutes les méthodes développées pour les variables fonctionnelles aux variables multivariées ce qui permet parfois de donner une variante ou une démonstration sous des hypothèses moins restrictives de résultats déjà bien connus.

2.4. Le prétraitement des données

En pratique, on ne dispose jamais, même si la grille de discrétisation est très fine, que d'un nombre fini de valeurs de la réalisation d'une variable fonctionnelle mais pas de cette réalisation fonctionnelle elle-même. De plus, il se peut qu'il y ait des valeurs manquantes ou bien que la discrétisation ne soit pas la même d'une donnée fonctionnelle à l'autre. D'autre part, il est assez fréquent d'être confronté à des courbes déphasées alors que l'on veut appliquer une méthode adaptée à des courbes alignées. On peut enfin supposer que les données récoltées contiennent des erreurs de mesure. Il est donc souvent nécessaire de commencer par effectuer un pré-traitement des données afin d'éliminer au maximum tous ces problèmes pratiques et de rendre les données adaptées aux méthodes de statistique fonctionnelle que l'on souhaite appliquer.

Par exemple, pour passer d'un ensemble de données discrétisées à une donnée fonctionnelle, l'approche la plus simple et la plus souvent utilisée consiste à lisser les données, c'est à dire à estimer de façon non-paramétrique chaque donnée fonctionnelle à partir des valeurs discrétisées. Parmi les nombreuses méthodes existantes, les plus couramment utilisées se servent de splines de lissage, d'ondelettes, de méthodes à noyaux ou de polynômes locaux. Si on a une idée plus précise de la forme de la fonction que l'on cherche à estimer, on pourra bien sûr utiliser d'autres méthodes mettant à profit ce supplément d'information. D'autre part, suivant le type discrétisaton, la régularité de la fonction que l'on cherche à estimer et la méthode statistique que l'on souhaite utiliser, il convient de choisir une méthode de lissage adaptée. Il faudra par exemple faire attention à ne pas trop sur-lisser des données provenant de la discrétisation d'une fonction irrégulière sur une grille fine, ni sous-lisser des données provenant de l'observation d'une fonction très régulière en peu de points, afin de ne pas perdre de l'information contenue dans les données initiales. En plus d'être essentielle pour passer de données discrétisées à une donnée fonctionnelle, l'étape de lissage des données permet aussi de contourner les problèmes posés par des données manquantes ou des différences de discrétisation. En effet, on peut malgré tout estimer les données fonctionnelles et poursuivre l'étude de celles-ci sans difficulté. D'autre part, certaines méthodes de statistique fonctionnelle demandent de disposer de données discrétisées sur une grille régulière. Si ce n'est pas le cas des données recueillies, on peut lisser celles-ci puis discrétiser de manière régulière la fonction estimée.

Revenons maintenant au cas où l'on suppose que des erreurs de mesure ont été commises, c'est à dire que l'on a mesuré la courbe x_i au point de discrétisation $t_{i,j}$ et que l'on a obtenu la valeur erronée $\tilde{x}_{i,j} := x_i(t_{i,j}) + \epsilon_{i,j}$ au lieu de la valeur exacte $x_i(t_{i,j})$. L'étape de lissage des données peut permettre d'atténuer l'effet de ces erreurs de mesure. En effet, supposons que la courbe x_i ait été mesurée de manière erronée et que nous lissions ces données erronées. On peut montrer que, sous de bonnes conditions, l'estimation réalisée à partir des données erronées converge vers la véritable fonction x_i (on pourra trouver des illustrations de l'intérêt de lisser des courbes bruitées dans différents types de situations en se rapportant à Hitchcock, Casella et Booth, 2006, Crambes, 2007, et Zhang et Chen, 2007). Enfin, si une courbe x_i a été mesurée (éventuellement de manière erronée) en p points, plus p est grand (plus on connait de valeurs de la courbe), meilleure est notre estimation de x_i. Le fait que le nombre de points de discrétisation augmente n'est donc plus vu comme un fléau ; au contraire cela apporte de l'information supplémentaire.

En dehors de l'étape de lissage dont nous venons de parler on peut envisager d'autres types de pré-traitements comme notamment l'utilisation de méthodes d'enregistrement. La nature fonctionnelle des données engendre des problématiques particulières. Par exemple, il se peut que les données recueillies pour différents individus présentent des décalages horizontaux ou des différences d'amplitudes qui sont suceptibles d'entraîner de mauvais résultats si on utilise les courbes telles qu'elles. Au cours d'une étude médicale, par exemple, on peut constater un déphasage en temps de nos observations qui peut être dû au fait que l'on n'a pas commencé nos mesures au même instant pour tous les individus ou que le temps de réaction au traitement est différent d'une personne à l'autre. Il se peut également qu'il y ait une différence au niveau de l'amplitude des courbes observées par exemple due aux caractéristiques propres à chaque individu. De manière plus générale, il peut arriver que les valeurs pour lesquelles on observe des points caractérisques (extrema, points d'inflexion,...) diffèrent suivant les courbes. Il se peut que les décalages observés contiennent des informations importantes par rapport au problème étudié. Mais si ce n'est pas le cas, on va chercher à supprimer ces déformations des courbes observées à l'aide de méthodes dites d'enregistrement. Si on note $X(t)$ la courbe observée et $X^*(t)$ la courbe non déformée, ces méthodes consistent principalement à trouver une fonction monotone h qui permette de faire un changement sur l'axe des abscisses de manière à enlever le décalage horizontal et une fonction a permettant de moduler l'amplitude de telle manière que $X^*(t) = a(t) X(h(t))$ (ou $X(t) = a(t) X^*(h(t))$ suivant les auteurs). Depuis l'article de Rao (1958) qui utilise des méthodes d'enregistrement pour transformer des courbes de croissance en droites, de nombreux auteurs se sont intéressés à ce problème important de la statistique fonctionnelle. Parmi les premiers travaux portant sur les méthodes d'enregistrement, on trouve l'article de Kneip et Gasser (1988) dans lequel les différences entre les courbes observées et une courbe de référence inconnue sont modélisées par des paramètres propres à chaque individu (modèle SEMOR). On s'intéresse également dans cet article à une modélisation plus simple par un modèle à forme invariante (SIM) dans lequel les courbes observées ont la même forme et diffèrent uniquement par des changements d'échelle ou des translations des différents axes. Différentes méthodes d'estimation de la forme commune et des paramètres

ont été proposés depuis notamment par Kneip et Engel (1995) pour les modèles SEMOR et SIM, puis pour le modèle SIM par Rønn (2001) grâce à des techniques basées sur des méthodes de maximum de vraisemblance, par Vimond (2005, 2006a) et Gamboa *et al.* (2007) au travers d'approches basées sur les décompositions de Fourier. D'autre part, Gasser et Kneip (1992) proposent une autre méthode basée sur l'alignement de points caractéristiques (extrema, points d'inflexion,...), Gasser et Kneip (1995) développent des outils pour identifier à partir d'un échantillon de courbes les points caractéristiques de la structure commune qu'il faut prendre en compte et enfin Wang et Gasser (1998) établissent des bornes de confiance pour la moyenne structurelle des données. Dans les travaux de Ramsay et Li (1998) et Kneip *et al.* (2000) on suppose connaître une forme commune et on cherche à estimer pour chacune des courbes la fonction h et la fonction a qui permettent de s'y ramener. Piccioni *et al.*(1998) considère le problème de données provenant de la déformation par un groupe de Lie d'une forme commune connue. D'autres articles comme ceux de Härdle et Marron (1990) ou Wang et Gasser (1997,1999) s'intéressent au cas ou l'on a observé deux courbes et que l'on cherche une déformation qui fasse passer de l'une à l'autre. Enfin, Liu et Müller (2004) s'intéressent aux problématiques d'enregistrement avec pour objectif de construire des sommes convexes de courbes. Des tests d'adéquation des données à des modèles particuliers de déformation ont été proposés notamment par Härdle et Marron (1990) et Vimond (2006b). Il existe de nombreuses méthodes d'enregistrement adaptées à des variables fonctionnelles autres que des courbes (images, surfaces,...) comme en témoignent les travaux de Dryden et Mardia (1998), Glasbey et Mardia (1998), Bigot *et al.* (2007) et les références qu'ils contiennent. Pour en savoir plus on pourra se rapporter aux monographies de Ramsay et Silverman (1997,2005) (Chapitre 7) qui donnent un point de vue plus général sur les différentes méthodes d'enregistrement de données fonctionnelles.

A partir de maintenant nous supposerons que le pré-traitement des données a été effectué et que nous considérons et étudions un échantillon de variables fonctionnelles.

2.5. Les espaces semi-métriques

Pour étudier des données on a souvent besoin d'avoir une notion de distance entre celles-ci. Il est bien connu qu'en dimension finie toutes les métriques sont équivalentes. Ce n'est plus le cas en dimension infinie, c'est pourquoi le choix de la métrique (et donc de la topologie associée) est un élément encore plus crucial pour l'étude de variables aléatoires fonctionnelles qu'il ne l'est en statistique multivariée. De nombreux auteurs définissent ou étudient les variables fonctionnelles comme variables aléatoires à valeurs dans $\mathbb{L}^2([0;1])$ (voir notament Crambes, Kneip et Sarda 2007), plus généralement dans un espace de Hilbert (voir par exemple Preda, 2007), de Banach (voir par exemple Cuevas et Fraiman, 2004) ou métrique (voir par exemple Dabo-Niang et Rhomari, 2003). D'autre part, Bosq (2000) considère des échantillons de variables aléatoires fonctionnelles dépendantes à valeur dans un

espace de Hilbert (ou de Banach) obtenues par découpage d'un même processus à temps continu. En plus des métriques disponibles il est assez souvent intéressant de considérer des semi-métriques permettant un éventail plus large de topologies possibles que l'on pourra choisir en fonction de la nature des données et du problème considéré. C'est pourquoi nous avons choisi dans ce mémoire de thèse de considérer et d'étudier des variables fonctionnelles définies comme variables aléatoires à valeurs dans un espace semi-métrique $(\mathcal{E}, d)$ de dimension infinie.

Un des avantages de cette définition générale des variables aléatoires fonctionnelles introduite par Ferraty et Vieu (2000) est qu'elle permet la modélisation de cas très divers comme par exemple des fonctions de carré intégrable ($\mathcal{E} = \mathbb{L}^2([0,1])$ muni de la norme usuelle ou bien d'une semi-métrique définie à partir des k premières composantes de la décomposition en base d'ondelettes), des images ($\mathcal{E} = \{f : [0,L] \times [0,\ell] \to [0,1]\}$ muni par exemple d'une semi-métrique invariante par rotations, translations et changements d'échelle), des opérateurs linéaires ($\mathcal{E} = \mathcal{L}(\mathbb{L}^2([0,1]), \mathbb{R})$ muni de sa norme usuelle ou d'une quelconque semi-métrique), ... Revenons par exemple au cas où nos données correspondent à des images. Au lieu de prendre une métrique qui détectera la moindre différence entre deux images, on peut décider de ne pas prendre en compte certaines déformations (comme par exemple les translations, rotations ou changements d'échelle) en utilisant des semi-métriques que ces déformations laissent invariantes. Dans d'autres cas de figure (comme les données spectrométriques présentées au chapitre suivant), les courbes sont très lisses et il peut être intéressant d'utiliser une semi-métrique basée sur les dérivées des données fonctionnelles.

En dehors de permettre la modélisation de phénomènes plus généraux, un autre intérêt d'utiliser une semi-métrique plutôt qu'une métrique est que cela peut constituer une alternative aux problèmes posés par la grande dimension des données. En effet, on peut prendre une semi-métrique définie à partir d'une projection de nos données fonctionnelles dans un espace de dimension plus petite que ce soit en réalisant une analyse en composantes principales fonctionnelles de nos données (Dauxois, Pousse et Romain, 1982, Besse et Ramsay, 1986, Hall et Hosseini-Nasab, 2006, Yao et Lee, 2006), ou bien en les projetant sur une base finie (ondelettes, splines, ...). Cela permet de réduire la dimension des données et augmente la vitesse de convergence des méthodes utilisées tout en préservant la nature fonctionnelle des données. On peut choisir la base sur laquelle on projette en fonction des connaissances que l'on a de la nature de la variable fonctionnelle. Par exemple, on pourra choisir la base de Fourier si on suppose que la variable fonctionnelle observée est périodique. On pourra se référer à Ramsay et Silverman (1997, 2005) ou Rossi, Delannay, Conan-Guez et Verleysen (2005) pour une discussion plus complète des différentes méthodes d'approximation par projection de données fonctionnelles. Une discussion plus approfondie de l'intérêt d'utiliser différents types de semi-métrique est faite dans Ferraty et Vieu (2006) (notamment au Paragraphe 3.4).

On peut retenir que le choix de la semi-métrique permet à la fois de prendre en compte des situations plus variées et de pouvoir contourner le fléau de la dimension. Ce choix ne doit cependant pas être fait à la légère mais en prenant en compte la nature des données et du problème étudié.

2.6. Plan du mémoire

Le reste de ce mémoire est organisé comme suit. La première partie se poursuit au travers de deux autres chapitres introductifs. Le Chapitre 3 propose un survol bibliographique des problématiques concrètes, des méthodes et des modèles que l'on peut rencontrer en statistique fonctionnelle. Le Chapitre 4 est quant à lui consacré à la présentation rapide des différents apports de ce mémoire. Enfin, on présente et discute dans le Chapitre 5 quelques perspectives rattachées au contenu de ce mémoire. Ensuite, une deuxième partie est consacrée à de nouveaux résultats asymptotiques concernant l'estimation de l'opérateur de régression. Le Chapitre 6 porte sur la normalité asymptotique tandis que le Chapitre 7 concerne les erreurs $\mathbb{L}^p$. Une note résumant ces deux chapitres constitue le Chapitre 8 tandis que le Chapitre 9 est consacré à des commentaires et applications supplémentaires. Enfin, une troisième partie est consacrée à un résultat important de ce mémoire qui donne une approche générale (Chapitre 10) permettant de construire des tests de structure très variés en régression sur variable fonctionnelle pour lesquels aucune statistique de test n'avait encore été proposée. Le Chapitre 11 présente l'utilisation de méthodes de rééchantillonnage dans le cas particulier de tests de non-effet à travers de nombreuses simulations.

CHAPITRE 3

PROBLÉMATIQUES CONCRÈTES ET ÉTAT DE L'ART EN STATISTIQUE POUR VARIABLES FONCTIONNELLES

L'objectif de ce chapitre n'est bien entendu pas de dresser un panorama complet de tous les résultats obtenus dans le domaine de la statistique fonctionnelle. On souhaite plutôt présenter quelques exemples concrets d'étude de variables fonctionnelles, les questions qui leur sont rattachées ainsi que différentes solutions et modélisations proposées dans la littérature.

3.1. Quelques situations particulières et concrètes

Ce paragraphe a le double objectif d'illustrer, à partir de jeux de données concrets, différentes problématiques auxquelles on peut être confronté et de mettre en relief le large spectre de domaines où la statistique fonctionnelle a été utilisée.

3.1.1. Différents types de questions que l'on peut se poser. — Nous présentons dans ce paragraphe trois jeux de données fonctionnelles particuliers provenant de domaines assez variés (linguistique, climatologie et chimiométrie) pour illuster quels sont les avantages et les enjeux liés à leur étude par une approche fonctionnelle.

3.1.1.1. Reconnaissance vocale. — Dans le domaine de la linguistique, le problème de la reconnaissance vocale est un sujet d'actualité. L'objectif est de pouvoir retranscrire phonétiquement des mots et des phrases prononcés par un individu. Pour cela, il est auparavant nécessaire de procéder à une étape d'étalonnage qui consiste à recueillir différents enregistrements sonores pour chaque phonème. Les données obtenues sont intrinsèquement de nature fonctionnelle. Elles résultent de la discrétisation, sur une grille très fine, des signaux sonores enregistrés au cours du temps. On n'étudie pas directement ces signaux mais plutôt les log-périodogrammes correspondants. Le jeu de données particulier auquel nous nous intéressons provient de la collaboration entre Andrea Buja, Werner Stuetzle et Martin Maechler et a été étudié dans Hastie *et al.* (1995) (les données et leur description sont disponibles sur le site de Hastie *et al.* (2001) : http :www-stat.stanford.edu/ElemStatLearn) ainsi que dans Ferraty et

Vieu (2003, 2006). L'étude s'intéresse en particulier à cinq groupes d'enregistrements correspondant aux phonèmes suivants :

- "sh" comme dans le mot anglais "she",
- "iy" comme la voyelle dans le mot anglais "she",
- "dcl" comme dans le mot anglais "dark",
- "aa" come la voyelle dans le mot anglais "dark",
- "ao" comme la première voyelle dans le mot anglais "water".

On dispose pour chacun de ces phonèmes de plus de cinq cents log-périodogrammes obtenus à partir d'enregistrements répétés de cinquante voix différentes d'hommes. On dispose d'un nombre de valeurs suffisant pour bien refléter la nature fonctionnelle du phénomène (voir Fig. 3.1). Il y a des décalages en temps et des variations d'amplitude entre les différents périodogrammes obtenus, on peut les prendre en compte au travers de l'approche fonctionnelle en utilisant des techniques de recalages de courbes avant d'aller plus loin.

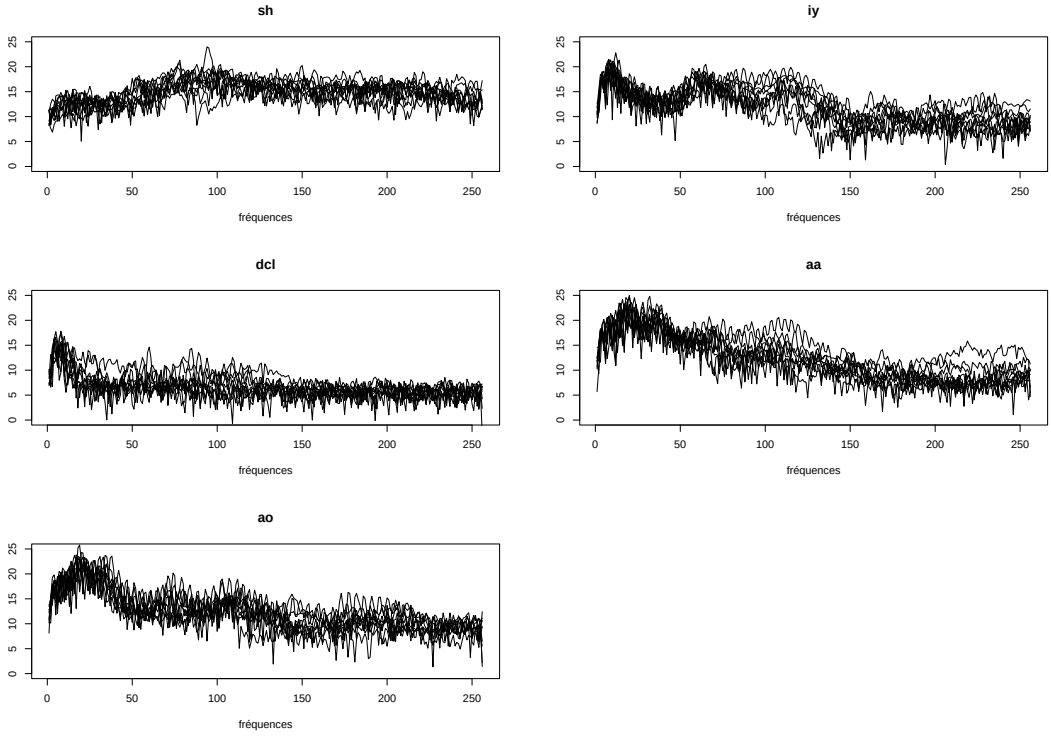

FIGURE 3.1. Echantillons de 10 log-périodogrammes pour chacun des cinq phonèmes

Dans l'optique de faire de la reconnaissance vocale, on souhaite pouvoir attribuer à un nouvel enregistrement vocal (en fait le log-périodogramme correspondant) le ou les phonèmes qui lui correspondent. Pour cela on pourra se servir des échantillons de données recueillis pour chaque phonème pendant la phase d'étalonnage. On est amené à résoudre un problème de discrimination de variables aléatoires fonctionnelles. On pourrait aussi se poser d'autres types de questions : peut-on détecter des différences entre la prononciation de différents individus ? quel est le phonème dont

les différentes prononciations sont les plus semblables ?, ... Enfin, la prise en compte de la nature fonctionnelle des données est un atout supplémentaire dans la comparaison de ces courbes puisqu'elle permet de dégager à la fois la forme générale du log-périodogramme et les variations d'amplitudes et de fréquences de ses oscillations plus fines. Les apports de cette thèse ne sont pas directement liés aux problématiques de classification. Néanmoins, le problème de discrimination peut être vu comme un problème de régression particulier (voir Ferraty et Vieu, 2003). Il est donc assez lié au contenu de ce mémoire.

3.1.1.2. Etude du phénomène El Niño. — Nous nous intéressons à présent à un jeu de données provenant de l'étude d'un phénomène climatologique assez important couramment appelé El Niño. C'est un grand courant marin qui survient de manière exceptionnelle (en moyenne une ou deux fois par décennie) le long des côtes péruviennes à la fin de l'hiver. Ce courant provoque des dérèglements climatiques à l'échelle de la planète. Le jeu de données dont nous disposons est constitué de relevés de température mensuels de la surface océanique effectués depuis 1950 dans une zone située au large du nord du Pérou (aux coordonnées 0-10° Sud, 80-90° Ouest) où peut apparaître le courant marin El Niño. Ces données et leur description sont disponibles sur le site du centre de prévision du climat américain : http ://www.cpc.ncep.noaa.gov/data/indices/. L'évolution des températures au cours du temps est réellement un phénomène continu. Le nombre de mesures dont nous disposons en rend assez bien compte (voir Figure 3.2) et permet de prendre en considération la nature fonctionnelle des données.

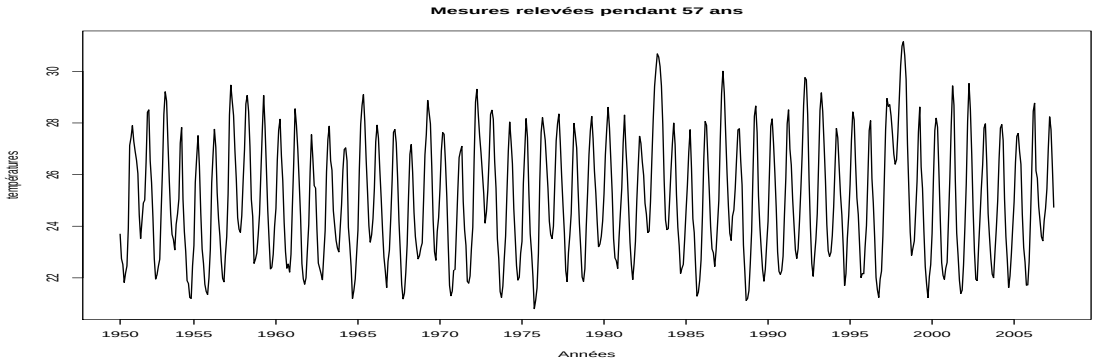

FIGURE 3.2. Mesures mensuelles de la température à la surface autour du courant marin El Niño depuis 1950.

A partir de ces données on peut s'intéresser à essayer de prédire l'évolution du phénomène à partir des données recueillies lors des années précédentes. Plusieurs approches on été introduites pour tenter de répondre à ce problème. Les premiers travaux cherchent à prédire la température du mois suivant à partir des p températures mensuelles précédentes (voir Katz, 2002, pour plus de références). Toutefois cette modélisation ne permet pas de prendre en compte la nature fonctionnelle du phénomène étudié ni d'en tirer profit. Plus récemment, suite à l'approche

introduite par Bosq (1991), on a choisi de considérer le processus non plus au travers de sa version discrétisée mais bien comme un processus continu que l'on décompose en courbes successives de 12 mois (entre juin d'une année et mai de l'année suivante, voir Fig. 3.3).

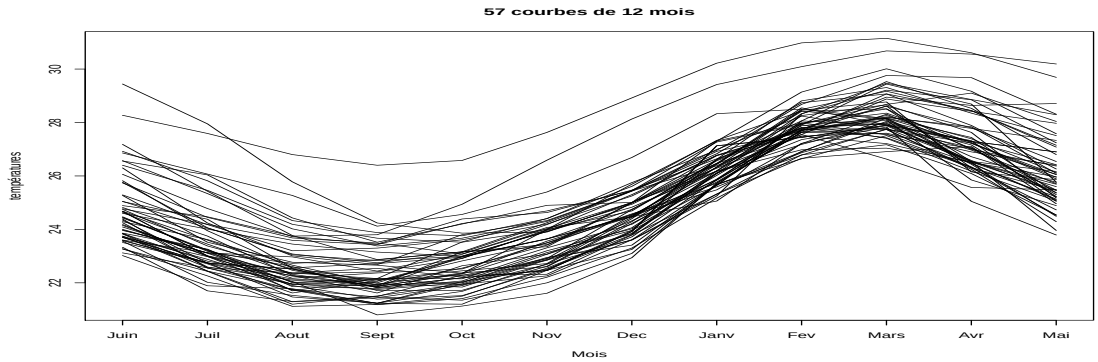

FIGURE 3.3. 57 courbes représentant la température à la surface autour du courant marin El Niño par tranches de 12 mois depuis juin 1950.

On peut ensuite chercher à prédire la courbe des douze mois suivants (ou certaines valeurs particulieres de celle-ci) à partir d'une ou de plusieurs courbes précédentes. Cette approche fonctionnelle a l'avantage de rester adaptée même lorsque le nombre de points de discrétisation augmente et permet de prendre en compte, pour chaque période de douze mois, la courbe des températures dans sa globalité. Les travaux de Besse *et al.* (2000), Valderama *et al.* (2002), Antoniadis et Sapatinas (2003) ou Ferraty *et al.* (2005) illustrent différentes manières de répondre à ce problème et de modéliser la dépendance des courbes successives. Les résultats donnés dans la première partie de ce mémoire sont établis pour des variables fonctionnelles faiblement dépendantes et peuvent eux aussi être appliqués pour répondre à des problématiques de prédiction de séries temporelles.

3.1.1.3. Taux de graisse et courbes spectrométriques. — Nous considérons enfin un jeu de données provenant de la chimie quantitative (branche de la chimie utilisant des outils mathématiques, également appelée chimiométrie). Les données ont été récoltées en vue de répondre à un problème de contrôle qualité dans l'industrie agro-alimentaire. Lors du conditionnement d'émincés de viandes, il est obligatoire de mettre sur l'emballage le taux de graisse. Une analyse chimique permet de donner de façon exacte le taux de graisse contenu dans un morceau de viande. Cependant cette méthode prend du temps, détériore en partie le morceau étudié et coûte assez cher, c'est pourquoi on s'intéresse à d'autres méthodes plus rentables. C'est ainsi que l'on a envisagé de prédire le taux de graisse à partir de courbes spectrométriques dont l'obtention est moins coûteuse (en temps et en argent) et ne nécessite pas de détériorer en partie la viande étudiée. Les données spectrométriques sont obtenues en mesurant pour chaque morceau de viande l'absorbance de lumières de différentes longueurs d'onde. Ce sont des données de nature intrinsèquement fonctionnelle comme

le souligne Leurgans *et al.* (1993) : "the spectra observed are to all intents and purposes functional observations". On peut donc les résumer par des courbes (appelées spectrométriques) représentant l'absorbance en fonction de la longueur d'onde. Nous nous intéressons à un jeu de données provenant de l'étude par analyse chimique et par spectrométrie de 215 morceaux de viande. On dispose ainsi de 215 courbes spectrométriques (voir Figure 3.4) ainsi que des taux de graisse correspondants (obtenus par analyse chimique). Ces données et leur description précise sont disponibles sur le site de StatLib (http ://lib.stat.cmu.edu/datasets/tecator).

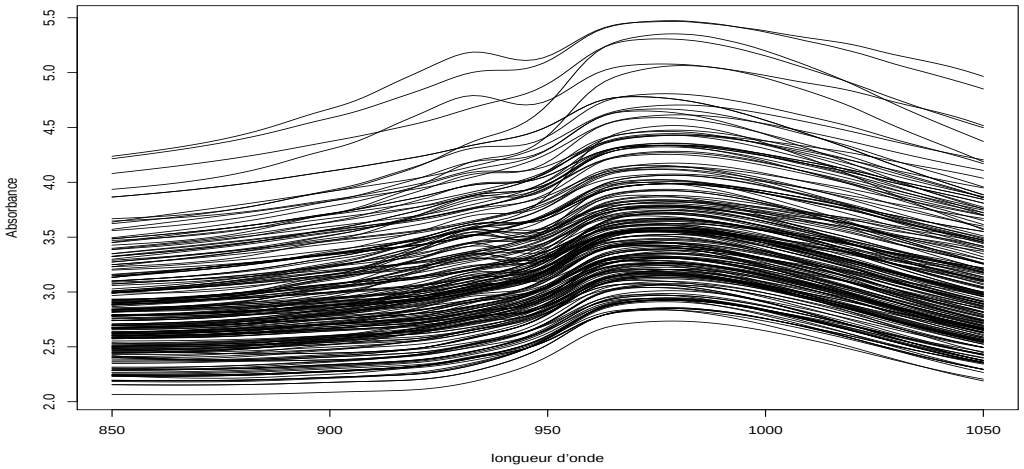

FIGURE 3.4. Courbes spectrométriques obtenues à partir des 215 morceaux de viande étudiés

On peut voir que les courbes spectrométriques sont très régulières et de formes semblables en dehors d'une translation verticale. On peut se demander si ce shift contient une information importante pour prédire le taux de graisse. Si c'est le cas il faudrait en tenir compte dans le choix de la semi-métrique. Dans le cas contraire, comme les courbes sont très lisses et de formes semblables, il peut être intéressant d'utiliser les courbes dérivées plutôt que les courbes elles-mêmes. Nous nous intéressons à prédire le taux de graisse d'un nouveau morceau de viande à partir de la courbe spectrométrique qui lui est associée. Les premiers travaux qui se sont intéressés à ce problème sont dus à Borggaard et Thodberg (1992) et utilsent des méthodes de réseaux de neurones. Depuis d'autres articles parmi lesquels Ferraty et Vieu (2002), Ferré et Yao (2005), Ferraty *et al.* (2006), Ferraty et Vieu (2006), Aneiros-Pérez et Vieu (2006), Ferraty, Mas et Vieu (2007) et Mas et Pumo (2007) ont proposé et appliqué d'autres méthodes pour répondre à cette problématique. D'autre part, les résultats donnés dans la première partie de ce mémoire complètent la littérature dédiée à la prédiction d'une variable réelle à partir d'une variable fonctionnelle et peuvent donc être utilisés pour prédire le taux de graisse à partir de la courbe spectrométrique. Pour finir, signalons que ces données peuvent donner lieu à d'autres problématiques que la régression. Par exemple, Dabo-Niang *et al.*

(2006) et Ferraty *et al.* (2007) se sont intéressés à ces données dans des optiques de classification et de détection de structure commune des courbes spectrométriques.

Enfin, avant d'essayer de prédire le taux de graisse en fonction de la courbe spectrométrique on peut se demander si il y a un lien entre ces deux variables. De plus, parmi les articles précédemment cités, plusieurs font une hypothèse a priori sur la nature du lien entre le taux de graisse et la courbe spectrométrique. On peut se demander si cette hypothèse est raisonnable. Les résultats donnés dans la seconde partie de ce mémoire donnent des méthodes pour tester si le lien entre une variable fonctionnelle et une variable réelle existe ou s'il est d'une nature particulière et permettrons donc de répondre à ce genre de question.

3.1.2. De nombreux autres champs d'application. — Les trois exemples de jeux de données que nous venons de présenter proviennent de domaines assez différents et reflètent des problématiques variées. Afin d'illustrer la grande variété des domaines où l'on peut être confronté à des données de nature fonctionnelle, nous proposons dans ce paragraphe un rapide tour d'horizon de ces champs d'application. Etant donné l'essor de la statistique fonctionnelle et le très grand nombre de cas où elle est utilisée, nous ne présentons qu'un panorama incomplet de ses domaines d'application. Nous avons choisi de limiter nos références à des publications portant sur l'étude de données correspondant à des courbes au travers de méthodes de statistique fonctionnelle, en privilégiant dans la littérature existante les travaux précurseurs ou les plus récents. Cependant, de même que tous les résultats de statistique nonparamétrique fonctionnelle, les résultats de cette thèse s'appliquent directement à des échantillons de données fonctionnelles de nature différente telles que les images, les surfaces, ...

- Le premier travail dans lequel on considère et prend en compte l'aspect fonctionnel des données est celui de Rao (1958) dans lequel sont étudiées et comparées des courbes de croissance. L'auteur envisage l'analyse en composantes principales et l'analyse factorielle de données fonctionnelles. Il suppose que l'on peut faire une transformation de l'axe du temps de telle sorte que les courbes de croissances transformées soient presque linéaires. Il propose ensuite de construire des tests non pas à partir de l'ensemble des points mesurés mais en les résumant par deux données : le point initial et une estimation de la pente. D'autres travaux ont été consacrés à l'étude fonctionnelle de courbes de croissance parmi lesquels on peut citer Ramsay *et al.* (1995), Gasser *et al.* (1998), James *et al.* (2000), James et Sugar (2003) et Ramsay et Silverman (2005).
- De manière plus générale, dans le domaine de la médecine on peut observer l'utilisation de la statistique fonctionnelle au travers d'études de phénomènes différents tels que l'évolution de certains cancers (voir Ramsay et Silverman, 2005, Cao et Ramsay, 2007, et Erbas *et al.*, 2007), l'effet placébo (voir Tarpey *et al.*, 2003), l'activité cardiaque (voir Clot, 2002, Ratcliffe *et al.*, 2002a, 2002b et Harezlak *et al.*, 2007), les mouvements du genou pendant l'effort sous contrainte (voir Abramovich et Angelini, 2006, et Antoniadis et Sapatinas, 2007) ou certaines déformations de la cornée (voir Locantore *et al.*, 1999).

– Depuis quelques années le secteur de la génétique est en plein essor. Grâce aux progrès effectués au niveau des appareils et des méthodes de mesure, les biologistes parviennent à faire plusieures mesures de l'expression des gènes au cours du temps. Ces mesures ont pour objectif de permettre une meilleure compréhension de la fonction des gènes et des intéractions entre certains de leurs effets (par exemple les phénomènes de régularisation d'une substance par une autre). On s'intéresse aussi à l'identification de groupes de gènes responsables de l'évolution d'un phénomène biologique complexe observé au cours du temps. Récemment, plusieurs méthodes de statistique fonctionnelle ont été appliquées aux profils d'expression des gènes au cours du temps à travers les travaux de Araki *et al.* (2004), Leng et Müller (2006), Opgen-Rhein et Strimmer (2006) et Song *et al.* (2007).
– Dans le domaine de la biologie animale, on a utilisé la statistique fonctionnelle pour étudier l'évolution de certains phénomènes au cours du temps. On peut par exemple évoquer les différentes études portant sur l'évolution de la ponte de mouches effectuées par Chiou, Müller, Wang et Carey (2003), Chiou, Müller et Wang (2003, 2004), Müller et Stadtmüller (2005), Cardot (2006) et Chiou et Müller (2007).
– L'étude de phénomènes liés à l'environnement et de leur évolution est très souvent liée à l'étude de données fonctionnelles qui peuvent correspondre à l'évolution d'un phénomène au cours du temps ou en fonction d'un autre paramètre (altitude, température, ...). En dehors de l'étude du courant marin El Niño évoquée précédemment, on trouve des travaux portant sur la prédiction de pics d'ozone (voir Damon et Guillas, 2002, Aneiros-Perez *et al*, 2004, Cardot *et al.* 2004, 2006), l'étude de l'évolution du taux d'ozone à différentes altitudes (Meiring, 2005), de la pollution par certains gaz à effet de serre (Febrero *et al.*, 2007), de la qualité de l'eau (Henderson, 2006, Nerini et Ghattas, 2007) ou des tests concernant des relevés de radioactivité à différentes altitudes au cours du temps (Cardot *et al.*, 2007).
– Les mesures et notamment les images recueillies par des satellites sont également des données dont l'étude peut être faite à partir de méthodes de statistique fonctionnelle. On peut par exemple citer les travaux de Vidakovic (2001) dans le domaine de la météorologie ou ceux de Dabo-Niang *et al.* (2004b, 2007) dans le domaine de la géophysique. Ils s'intéressent à classer des courbes recueillies par le satellite en différents endroits de l'Amazonie pour identifier la nature du sol. On peut également évoquer Cardot *et al.* (2003) et Cardot et Sarda (2006) qui étudient l'évolution de la végétation à partir de données satellites.
– Dans le domaine de l'économétrie on est confronté à de nombreux phénomènes que l'on peut modéliser par des variables fonctionnelles. On peut citer par exemple des travaux étudiant la dynamique de l'index mensuel de production de denrées périssables (Ramsay et Ramsay, 2002), la prédiction de la consommation électrique (Ferraty *et al.*, 2002), la volatilité de marchés financiers (Müller *et al.*, 2007), le rendement d'une entreprise (Kawasaki et Ando, 2004), l'évolution du prix d'un objet au cours d'enchères (Reddy et Das, 2006, Wang *et al.*, 2007), le commerce électronique (Jank et Shmueli, 2006) ou l'intensité de transactions financières (Laukaitis et Rackauskas, 2002 et Laukatis, 2006). On peut se reporter à Kneip et Utikal (2001), Benko (2006) et Benko *et al.* (2006) pour des références supplémentaires.

– En graphologie, on étudie la façon dont sont manuscrites des lettres ou des mots. La manière dont sont écrites ces lettres peut être décrite à la fois par le tracé final mais aussi par les mouvements (vitesse, accélération, rotation) du stylo au cours de l'écriture. L'analyse de ce type de données est typiquement liée à des problèmes de reconnaissance de signature. L'analyse fonctionnelle est bien adaptée à la nature de ces données et a été appliquée à diverses reprises notamment par Hastie *et al.* (1995), Ramsay (2000a, 2000b), ainsi que Ramsay et Silverman (2002).
– Enfin, on peut être amené à étudier des variables aléatoires fonctionnelles même si l'on dispose de données initiales réelles ou multivariées indépendantes. C'est ainsi le cas lorsque l'on souhaite comparer ou étudier des fonctions que l'on peut estimer à partir des données. Parmi les exemples typiques de ce type de situation on peut évoquer la comparaison de différentes fonctions de densité (voir Kneip et Utikal, 2001, Ramsay et Silverman, 2002, Delicado, 2007 et Nerini et Ghattas, 2007), de fonctions de régressions (Härdle et Marron, 1990, Heckman et Zamar, 2000), l'étude de la fonction représentant la probabilité qu'un individu a de répondre correctement à un test en fonction de ses "qualités" (Rossi *et al.*, 2002, Ramsay et Silverman, 2002), ...

Il existe d'autres domaines où la statistique fonctionnelle a été employée comme par exemple l'étude de certains mouvements (voir par exemple Ramsay *et al.*,1996, Ramsay et Silverman, 2002, Ormoneit *et al.*, 2005) , le traitement de signaux sonores (Lucero, 1999) ou enregistrés par un radar (Hall *et al*, 2001), la météorologie (voir Ramsay et Silverman, 2005, Hall et Vial, 2006b), les études démographiques (voir Hyndman et Ullah, 2007) , la géologie (voir Manté *et al*, 2007) , ... On pourra compléter ce rapide tour d'horizon en se référant au livre de Ramsay et Silverman (2002) dans lequel on trouve des applications dans des domaines aussi variés que la criminologie (comment modéliser et comparer l'évolution de la criminalité d'un individu au cours du temps ?), la paléopathologie (peut-on dire si un individu souffrait d'arthrite à partir de la forme de son fémur ?), l'étude de résultats à des tests scolaires, ... On peut imaginer que dans les années à venir l'utilisation de méthodes de statistique fonctionnelle sera étendue à d'autres domaines. Bien qu'incomplet, ce rapide tour d'horizon illustre le fait que l'utilisation de méthodes adaptées à l'étude de variables aléatoires fonctionnelles est en plein essor à la fois en ce qui concerne le nombre d'articles qui leurs sont consacrés et la diversité des phénomènes étudiés. Cet essor va de paire avec le nombre croissant de résultats et méthodes théoriques nouveaux concernant l'étude de variables aléatoires fonctionnelles. Le paragraphe suivant a pour objectif d'en donner un aperçu au travers d'un survol bibliographique.

3.2. Quelques méthodes d'analyse de données fonctionnelles

L'essor que connaît la statistique fonctionnelle au travers de ses divers champs d'application se retrouve au niveau des nombreuses approches théoriques développées pour l'étude de variables aléatoires fonctionnelles. Bien qu'incomplet, l'aperçu que nous en proposons dans ce paragraphe a pour objectif de rendre compte de leur

diversité. Nous avons choisi de limiter notre revue bibliographique aux publications portant sur l'étude de variables fonctionnelles correspondant à des courbes, en privilégiant les travaux précurseurs ou les plus récents. Cependant, les résultats présentés dans ce mémoire de thèse s'appliquent directement à des échantillons de variables fonctionnelles de nature différente (images, surfaces, ...) comme tous les résultats de statistique non-paramétrique fonctionnelle. Parmi ce survol bibliographique, on trouve à la fois des adaptations de méthodes bien connues en statistique multivariée (sous l'impulsion de Ramsay, 1982) mais aussi des résultats répondant à des problèmes nouveaux liés à la nature fonctionnelle des données. Le lecteur pourra se rapporter aux monographies de Ramsay et Silverman (1997, 2002, 2005), Bosq (2000) et Ferraty et Vieu (2006) pour avoir un aperçu plus complet de la statistique fonctionnelle.

3.2.1. Analyses factorielles. — Nous évoquons dans ce paragraphe sous le nom d'analyses factorielles les méthodes d'analyse de données basées sur l'étude d'éléments propres de différents opérateurs. Dès les premiers travaux prenant en compte la nature fonctionnelle des variables étudiées, Rao (1958) et Tucker (1958), l'application de méthodes d'analyse factorielle telles que l'analyse en composantes principales et l'analyse factorielle "pure" a été envisagée. Voici un rapide tour d'horizon des diverses méthodes d'analyse factorielle développées pour l'étude de variables fonctionnelles.

3.2.1.1. A. Analyse en composantes principales fonctionnelles. — Les variables aléatoires fonctionnelles sont par définition à valeurs dans un espace de dimension infinie. Cependant certaines d'entre elles peuvent être convenablement approchées par leur projection sur un espace fonctionnel de dimension finie. On peut par exemple envisager d'approcher des variable aléatoires fonctionnelles par leurs projections sur une base finie de fonctions trigonométriques, d'ondelettes, de polynômes, de splines, ... Toutefois, le choix de la base de fonctions sur laquelle on projette est important et nécessite d'avoir un a priori sur la nature des variables que l'on étudie. Au lieu de cela, on peut essayer de construire cette base de fonctions à partir des variables dont nous disposons. Cette approche est appelée analyse en composantes principales fonctionnelles. Elle consiste à construire, à partir de l'étude des valeurs propres et des fonctions propres de l'opérateur de covariance de nos variables, une base de composantes fonctionnelles orthogonales qui maximisent (à dimension fixée) la proportion de variance totale expliquée. C'est une adaptation au cas fonctionnel de la méthode d'analyse en composantes principales très couramment utilisée en statistique multivariée. L'analyse en composantes principales fonctionnelles est l'une des méthodes les plus utilisées en statistique fonctionnelle notamment parce qu'elle a l'intérêt de permettre dans certaines situations de se ramener à un problème de statistique multivariée. Elle est envisagée par Rao (1958) et elle a été depuis étudiée à travers de nombreux travaux parmi lesquels on peut citer Deville (1974), Dauxois *et al.* (1982) dans le cas de fonctions de carré intégrables, Ramsay (1982) pour des variables à valeur dans un espace de Hilbert, Besse et Ramsay (1986) dans le cas où on a des hypothèses de régularité supplémentaires sur certaines dérivées, Bosq (1991) et Aguilera *et al.* (1997, 1999) dans le cas de l'étude des processus. Par la

suite on trouve Ramsay et Dalzell (1991) qui proposent d'utiliser des L-splines pour traiter séparément la partie structurelle et la partie résiduelle des courbes, Pezzuli et Silverman (1993) donnent des résultats de consistance pour certains processus gaussiens et Silverman (1995) étudie le cas où la variable fonctionnelle peut dépendre de paramètres. On trouve ensuite les variantes proposées par Silverman (1996) qui construit des composantes orthogonales par rapport à d'autres produits scalaires que celui de $\mathbb{L}^2$, Locantore *et al.* (1999) proposent une approche robuste, Ocaña *et al* (1999) s'intéressent à l'importance du produit scalaire utilisé, Boente et Fraiman (2000) utilisent des méthodes à noyau, Cardot (2000) utilise des B-splines, James *et al* (2000) et Yao *et al.* (2005a) proposent une méthode adaptée à des données longitudinales. Citons également les travaux de Ramsay et Silverman (1997, 2005, Chapitre 8) ainsi que Spitzner *et al.* (2003) qui proposent des méthodes d'analyse en composantes principales de données fonctionnelles bi-variées. Les monographies de Ramsay et Silverman (1997, 2005) donnent un aperçu général de l'état de l'art de l'analyse en composantes principales fonctionnelles. Plus récemment Hall *et al.* (2006) étudient l'influence de la discrétisation sur l'estimation des valeurs propres et des vecteurs propres, Hall et Hosseini-Nasab (2006) s'intéressent aux propriétés d'approximation d'une analyse en composante principale en fonction de l'espacement des valeurs propres, Cardot (2007) propose une analyse en composantes principales fonctionnelles conditionelle et Ocaña *et al.* (2007) proposent une méthode algorithmique générale.

3.2.1.2. B. D'autres analyses factorielles. — L'analyse en composantes fonctionnelles principales est la méthode d'analyse factorielle la plus couramment utilisée dans l'étude de variables fonctionnelles. Un autre type d'analyse factorielle permettant de réduire la dimension d'une variable aléatoire est évoquée dans la littérature sous le nom d'analyse factorielle "pure". Elle est assez semblable à l'analyse en composantes principales. On suppose que la variable étudiée X peut être décomposée en une variable de dimension plus petite U et un terme d'erreur ϵ. L'idée est de construire une base de fonctions ("facteurs") qui maximisent, à dimension fixée, la proportion de la variance de U expliquée. Cela se fait à partir des valeurs et fonctions propres d'une version modifiée de l'opérateur de covariance prenant en compte l'erreur. En dehors des articles de Rao (1958) et Tucker (1958) dans lesquels elle a été envisagée, on trouve assez peu de travaux s'intéressant à l'analyse factorielle "pure" de données fonctionnelles. On peut cependant évoquer l'approche théorique proposée par Pousse et Téchené (1997, 1998).

En dehors des deux analyses factorielles que nous venons d'évoquer, on peut envisager d'adapter les méthodes d'analyse canonique des corrélations et d'analyse discriminante au cas de variables fonctionnelles. De même que dans le cas multivarié, l'analyse canonique des corrélations de deux variables fonctionnelles X et Y a pour objectif de mettre en évidence les intéractions entre ces deux variables aléatoires. Pour cela, si par exemple X et Y sont des variables aléatoires à valeurs dans $\mathbb{L}^2(\mu)$, on recherche les fonctions (a, b) de $\mathbb{L}^2(\mu)$ pour lesquelles les variables $U := \int aXd\mu$ et $V := \int aXd\mu$ sont le plus corrélées possible (par rapport à un critère de corrélation pénalisé). Le (premier) couple de variables fonctionnelles

canoniques (aX, bY) est une première approximation du lien existant entre les variables étudiées. Celle-ci peut être affinée en construisant de manière itérative un nombre plus grand k de couples de fonctions canoniques. Lorsque l'on a déjà construit les $j-1$ premiers couples de variables canoniques, on contruit le couple numéro j en recherchant les fonctions (a_j, b_j) de $\mathbb{L}^2(\mu)$ pour lesquelles les variables $U_j := \int a_j X d\mu$ et $V_j := \int a_j X d\mu$ sont le plus corrélées possible (par rapport à un critère de corrélation pénalisé) et telles que U_j (respectivement V_j) soit non corrélée par rapport aux autres fonctions canoniques U_s, $1 \leq s \leq j-1$ (respectivement V_s, $1 \leq s \leq j-1$). Dans la littérature on trouve une approche théorique de l'analyse canonique dans les espaces de Hilbert, basée sur la théorie des opérateurs, proposée notamment par Dauxois et Pousse (1975), Dauxois et Nkiet (2002) et Dauxois *et al.* (2004). D'autres articles s'intéressent à la manière dont peut être généralisée l'approche multivariée au cas de variables fonctionnelles et proposent des méthodes d'estimation. On peut par exemple citer les travaux de Leurgans *et al.* (1993) qui mettent en évidence la nécessité de pénaliser le critère de corrélation par rapport au cas multivarié. Par la suite, les travaux de He *et al.* (2000, 2003) montrent que certaines propriétés de l'analyse canonique multivariée restent vraies dans le cas de l'analyse canonique de deux processus. Un algorithme robuste basé sur les réseaux de neurones est proposé par Gou et Fyfe (2004). Enfin, He *et al.* (2004) proposent et comparent quatre méthodes d'estimation des variables et des fonctions de poids canoniques. On pourra aussi se référer aux commentaires de Ramsay (1982) et aux monographies de Ramsay et Silverman (1997, 2005, Chapitre 11).

L'analyse discriminante linéaire a pour objectif d'attribuer une classe à une nouvelle donnée à partir d'un échantillon de données dont on connaît la classe d'appartenance. On note X la variable aléatoire fonctionnelle ($\in \mathbb{L}^2(\mu)$) étudiée et $Y \in \mathbb{R}^k$ le vecteur aléatoire correspondant à sa classe (si la classe de X est c on pose $Y_j := \delta_{c,j}$, $1 \leq j \leq k$). L'analyse discriminante linéaire est étroitement liée à l'analyse canonique puisque l'on cherche une fonction $a \in \mathbb{L}^2(\mu)$ et un vecteur b de $\mathbb{R}^k$ tels que les variables $\int aX d\mu$ et ${}^t bY$ soient le plus corrélées possibles (au sens d'un certain critère de corrélation pénalisé). Il semble que peu de travaux aient été consacrés à l'analyse discriminante fonctionnelle. De même que dans le cas de l'analyse canonique fonctionnelle il est nécessaire d'uiliser un critère de corrélation pénalisé comme cela est expliqué par Hastie *et al.* (1995) et repris par Ramsay et Silverman (1997, 2005). Plus récemment, James et Hastie (2001) proposent une nouvelle approche basée sur une réduction préalable de la dimension des données en les projetant sur une base de fonctions splines. Enfin, Preda *et al.* (2007) relient la notion d'analyse discriminante linéaire à celle de régression linéaire "partial least square" et proposent d'utiliser une méthode de régression "partial least square" fonctionnelle.

On peut s'attendre dans les années qui viennnent à ce que soient généralisées au cadre fonctionnel d'autres méthodes d'analyse factorielle multivariée comme par exemple celle de l'analyse en composantes indépendantes (voir Hyvärinen *et al.*, 2001).

3.2.2. Analyse exploratoire d'un échantillon de variables fonctionnelles. — En dehors des analyses factorielles évoquées dans le paragraphe précédent, l'analyse exploratoire d'un échantillon de données fonctionnelles se fait au travers de paramètres de centralité tels que la moyenne, la médiane, et le mode. On peut également penser à étudier d'autres quantités liés à la distribution des données ou à s'intéresser à la notion de profondeur d'une courbe au sein de l'échantillon. Ces notions pourront être utiles pour répondre à des problématiques de classification non supervisée.

3.2.2.1. A. Paramètres de centralité et distribution. — Le paramètre de centralité le plus simple est celui de la moyenne. Il est important d'avoir recalé les données (voir Chapitre 2) avant d'envisager de calculer la courbe moyenne. La manière la plus naive de construire la moyenne d'un échantillon de courbes est de considérer la fonction qui vaut en chaque point la moyenne des valeurs que prennent les autres courbes au même point (voir Rice et Silverman, 1991, Ramsay et Silverman 1997, 2005). On pourra aussi se réferer à Ferraty et Vieu (2006) pour une discussion des inconvénients de cette construction de la moyenne. Cependant d'autres méthodes d'estimation basées sur des méthodes de lissage ont été proposées afin de mieux prendre en compte la régularité de la courbe moyenne que l'on cherche à estimer. On peut par exemple évoquer Rice et Silverman (1991) qui estiment la moyenne par l'argument minimum d'un critère quadratique pénalisé et Gervini (2006) qui utilise des splines à noeuds et donne des intervalles de confiance. Cependant il est bien connu que la moyenne n'a pas de bonnes propriétés de robustesse dans le cas multivarié. Il en est de même dans le cas fonctionnel. C'est pourquoi de nombreux auteurs ont proposé la notion de "trimmed mean" qui consiste à calculer la moyenne seulement à partir des valeurs centrales de l'échantillon en écartant les extrêmes. On s'appuie sur des notions de profondeur (voir par exemple Fraiman et Muniz, 2001, Lopez-Pintado et Romo, 2005, 2006a, 2007b et Cuevas *et al*, 2007) d'une courbe par rapport à l'échantillon étudié pour retirer les données les plus "abhérentes" avant de calculer la moyenne. C'est un domaine assez récent pour lequel il n'existe encore que quelques travaux parmi lesquels on peut citer le travail initial de Fraiman et Muniz (2001), ainsi que les travaux complémentaires menés par Lopez-Pintado et Romo (2005, 2006a, 2007b), Cuevas *et al.* (2006, 2007) et Febrero *et al.* (2007). Une approche différente appelée "impartial trimming" permettant de supprimer les courbes abhérentes directement à partir de l'échantillon et ne reposant plus sur la notion de profondeur a été récemment introduite par Cuesta-Albertos et Fraiman (2007).

Pour répondre aux problèmes de robustesse bien connus de la moyenne on peut également s'intéresser à d'autres paramètres de centralité. La notion de courbe médiane peut être définie de différentes manières. On peut généraliser la notion de médiane à des espaces de Hilbert par la notion de $\mathbb{L}^1$-médiane étudiée notamment par Kemperman (1987) et Cadre (2001) (voir aussi les références qu'il contient). On peut également définir la courbe médiane comme la courbe qui possède la plus grande profondeur dans l'échantillon (voir Fraiman et Muniz, 2001). On obtient ainsi

différentes courbes médianes correspondant aux différentes notions de profondeur (voir par exemple Fraiman et Muniz, 2001, Lopez-Pintado et Romo, 2005, 2006a, 2007b, ainsi que Cuevas *et al.*, 2006, 2007). Enfin, Ferraty et Vieu (2006) proposent une définition de la médiane liée à une semi-métrique.

Un des autres paramètres de centralité robuste est le mode. Cependant, généraliser la notion de mode aux variables fonctionnelles n'est pas simple pour deux raisons. Qui dit mode, dit densité par rapport à une mesure. Or, il n'existe pas de mesure de référence dans les espaces de dimension infinie comme peut l'être la mesure de Lebesgue dans le cas multivarié. D'autre part, même si on suppose que nos variables ont une distribution absolument continue par rapport à une mesure particulière, il reste difficile d'estimer la densité (voir Dabo-Niang, 2002, 2004a et 2004b). Notons qu'il existe d'autres travaux concernant des problématiques d'estimation de la densité de mesures définies sur des espaces généraux (voir par exemple Jacob et Oliveira, 1995). Le fait de rechercher la courbe modale va de paire avec une estimation de la distribution. Parmi les premières approches on trouve celle proposée par Gasser *et al.* (1998) qui consiste à ce ramener au cas multivarié en projetant les variables fonctionnelles sur une base finie de fonctions. Une autre approche est développée par Hall et Heckman (2002) qui proposent une méthode de "lignes ascendantes" construites à partir d'une relation d'ordre particulière entre les courbes basée sur la manière dont les probabilités de petites boules centrées en chacune des courbes tendent vers 0. Plus récemment enfin, une autre méthode basée sur des pseudo-estimateurs non-paramétriques de la densité par des méthodes à noyau a été proposée par Dabo-Niang *et al.* (2004a, 2006) dans le cas de variables fonctionnelles à valeurs dans un espace semi-métrique. On retrouve cette dernière méthode dans l'article de Cuevas *et al.* (2006).

En statistique univariée on s'intéresse très souvent aux quantiles. Il n'est pas aisé de généraliser la notion de quantile au cadre fonctionnel car il n'y a plus de notion d'ordre immédiate. Cependant, à partir des notions de profondeur que nous avons évoquées précédemment (voir Fraiman et Muniz, 2001, Lopez-Pintado et Romo, 2005, 2006a, 2007b et Cuevas *et al*, 2007) il est possible de définir des ensembles de fonctions dans lesquels on peut s'attendre à avoir au moins un pourcentage donné de courbes (voir Lopez-Pintado et Romo, 2007a). Ces ensembles sont les équivalents pour le cadre fonctionnel de ce que sont des intervalles $]-\infty; q_\alpha]$ pour le cadre univarié.

En dehors des paramètres de la distribution évoqués précédemment, on trouve aussi des articles qui s'intéressent à l'estimation de l'opérateur de covariance comme par exemple Rice et Silverman (1991), Lee *et al.* (2002) et Gervini (2006). Comme cela est très lié à la problématique de l'analyse en composantes principales on pourra se rapporter à l'étude bibliographique qui y est consacrée dans le paragraphe précédent pour obtenir plus de références. On trouve également un article de Hall et Vial (2006a) qui s'intéresse à l'étude des maxima et minima locaux des vraies composantes principales fonctionnelles au travers de ceux de leur version empirique.

3.2.2.2. B. Classification non supervisée. — Les différents indices de centralité et la notion de profondeur évoqués précédemment peuvent permettre de répondre à des problématiques de classification non supervisée. L'approche hiérarchique descendante générale décrite dans Ferraty et Vieu (2006a) permet de réaliser une classification non supervisée des données à partir de chacun des différents indices de centralité que nous venons d'évoquer. Le cas plus particulier de classification à partir du mode est traité par Dabo-Niang *et al.* (2004b, 2006, 2007). Différentes versions de la méthode "k-means" ont été adaptées à l'étude des variables fonctionelles. Abraham *et al.* (2003) se ramènent au cas multivarié en projetant les courbes sur une base de splines, Mizuto (2004) propose une méthode à deux temps basée sur l'application de la méthode "k-means" en chacun des points de la discrétisation, Tarpey et Kinateder (2003) donnent des résultats dans le cas de processus gaussiens et Cuesta-Albertos et Fraiman (2007) s'intéressent à un cadre fonctionnel plus général. On peut également évoquer l'approche particulière développée par James et Sugar (2003) pour la classification de données fonctionnelles observées en peu de points. L'article de Mizuto (2004) présente différentes méthodes de classification, Rossi *et al.* (2004) proposent de projeter les courbes sur une base de dimension finie puis de classer les coefficients à l'aide d'un algorithme (SOM) utilisant les réseaux de neurones, tandis que Nerini et Ghattas (2007) utilisent des arbres de régression pour classer des densités. Enfin, Preda et Saporta (2004, 2005b, 2007) utilisent des méthodes basées sur la régression sur composantes principales fonctionnelles et sur la régression "partial least square" pour classer des couples (X_i, Y_i) (où la variable explicative fonctionnelle X_i, et la variable réponse réelle Y_i sont reliées par un modèle linéaire fonctionnel) en fonction des coéfficients du modèle de régression linéaire qui les relie.

Les notions de profondeur peuvent aussi être utilisées dans le cadre de discrimination de courbes comme on peut notamment le voir au travers des travaux de Lopez-Pintado et Romo (2006) et Cuevas *et al.* (2007). Ces méthodes de discrimination complètent les autres approches basées sur l'analyse discriminante (voir paragraphe précédent) ou la régression (voir paragraphe suivant).

3.2.3. Régression pour variables fonctionnelles. — L'étude de modèles de régression adaptés à des données fonctionnelles est un domaine important de la statistique fonctionnelle. On y retrouve des situations très différentes suivant que la variable explicative, la variable réponse ou les deux variables sont de nature fonctionnelle. Nous verrons à travers cette étude bibliographique que c'est un domaine en plein essor, que certaines situations ont été plus étudiées que d'autres et qu'il reste encore de nombreuses questions ouvertes. Nous verrons enfin que l'étude des séries temporelles ou la discrimination de courbes peuvent être directement reliées à des problèmes de régression faisant intervenir des variables fonctionnelles. Les références seront regroupées et présentées suivant la nature de la variable explicative X et de la variable réponse Y.

3.2.3.1. A. Cas où Y est réelle et X fonctionnelle.— On suppose pour commencer que l'on s'intéresse au cas de la régression d'une variable réelle sur une variable fonctionnelle. C'est la forme de régression qui a été le plus étudiée dans la littérature. Comme dans le cas multivarié, différentes méthodes ont été proposées selon la nature de l'opérateur de régression. Le premier modèle de régression considéré est le modèle linéaire fonctionnel, introduit et étudié par Ramsay et Dalzell (1991) puis Hastie et Mallows (1993), dans lequel on suppose que l'opérateur de régression est linéaire. On a vu dans le Chapitre 2 que les méthodes de régression "partial least square", de régression sur composantes principales et de "ridge" régression sont bien adaptées à la régression sur variable multivariée à composantes dépendantes. C'est pourquoi de nombreux auteurs ont proposé de généraliser ces méthodes au cas d'une variable explicative fonctionnelle afin de permettre la prise en compte de la nature fonctionnelle de la variable explicative. Ainsi Preda (1999), puis Preda et Saporta (2002, 2004, 2005a, 2005b) proposent des méthodes basées sur une généralisation de la méthode "partial least square" au cas fonctionnel. Cardot *et al.* (1999), dans le cas d'une variable explicative non bruitée, puis Crambes (2007), dans le cas d'une variable explicative bruitée, adoptent une méthode de régression sur composantes principales fonctionnelles (introduite par Bosq, 1991). Cardot *et al.* (2000, 2003) utilisent une approche de type "ridge" régression, tandis que Cardot *et al.* (2007) adaptent la méthode "total least square" pour étudier le cas où la variable explicative est entachée d'erreur. D'autres travaux se sont intéressés à une étude plus fine de la vitesse de convergence $\mathbb{L}^2$ de l'estimateur par régression "ridge". On peut notamment citer Hall et Cai (2006) qui donnent deux types de vitesses de convergence selon que l'on considère le problème de l'estimation ou celui de la prédiction. Crambes *et al.* (2007) ainsi que Hall et Horowitz (2007) obtiennent les vitesses de convergence optimales en prenant en compte la régularité de la variable explicative. Un article récent de Cardot *et al.* (2007) établit des résultats de convergence en loi de type "théorème de la limite centrale" pour l'estimation et la prédiction dans le modèle linéaire fonctionnel. On pourra se référer à Ramsay et Silverman (1997, 2005, Chapitre 14), Besse *et al.* (2005), Cardot et Sarda (2006), Cardot *et al.* (2006) et Crambes (2006) pour des compléments d'information concernant l'estimation de l'opérateur de régression linéaire.

Au cours des années, des modèles fonctionnels plus généraux ont été proposés. De nombreux travaux se sont intéressés au modèle général dans lequel on ne fait pas d'hypothèse sur la forme de l'opérateur de régression mais seulement sur sa régularité. Les premiers résultats ont été obtenus à partir de l'étude d'un estimateur généralisé introduit par Ferraty et Vieu (2000). On pourra se référer à Ferraty et Vieu (2004, 2006a, 2006b) pour un panorama plus complet de l'utilisation de méthodes à noyau en statistique fonctionnelle. Les résultats présentés dans ce mémoire de thèse s'inscrivent dans ce cadre et sont présentés succintement dans le chapitre suivant, puis détaillés dans les Parties II et III de ce mémoire. On trouve dans la littérature différents travaux qui ont permis d'établir des propriétés asymptotiques de cet estimateur. Les vitesses de convergence presque complète ont été obtenues dans le cas d'un échantillon dépendant par Ferraty et Vieu (2000, 2002) ainsi que par Dabo-Niang et Rhomari (2003) et dans le cas dépendant par Ferraty *et al.* (2002a, 2002b). D'autre part Masry (2005) établit un résultat de

normalité asymptotique dans le cas d'un échantillon de variables dépendantes. L'étude plus approfondie menée par Ferraty *et al* (2007) permet de donner les expressions explicites des termes dominants du biais et de la variance dans le cas d'un échantillon indépendant. Cela leur permet de donner l'expression de l'erreur $\mathbb{L}^2$ et de construire des intervalles de confiance asymptotiques. Récemment, Aspirot *et al.* (2008) obtiennent un résultat de normalité asymptotique dans le cas d'échantillons provenant de processus non-stationnaires. Enfin, la convergence en norme $\mathbb{L}^p$ a été obtenue par Dabo-Niang et Rhomari (2003) dans le cas indépendant. L'utilisation d'un estimateur à noyau amène souvent la question du choix du paramètre de lissage. On pourra consulter à ce propos les articles de Rachdi et Vieu (2007) pour le choix du paramètre de lissage global par validation croisée et Benhenni *et al.* (2007) en ce qui concerne le choix du paramètre de lissage local. Dans un article récent, Burba *et al.* (2008) s'intéressent à la méthode des k plus proches voisins dans le cadre de l'estimation à noyau de l'opérateur de régression. D'autres méthodes d'estimation ont été proposées. Rossi et Conan-Guez (2005b,2005c, 2006) ainsi que Rossi *et al.* (2005) proposent des méthodes de réseaux de neurones adaptées à l'étude de données fonctionnelles. D'autre part, Preda (2007) propose une méthode d'estimation utilisant les propriétés des espaces autoreproduisants et des méthodes de choix de modèle.

Les résultats présentés dans ce mémoire sont liés à l'étude des propriétés asymptotiques de l'estimateur à noyau introduit par Ferraty et Vieu (2000). Ceux que nous énonçons dans la Partie II complètent les résultats de convergence évoqués précédemment. Nous nous intéressons tout d'abord à la généralisation des résultats de Ferraty *et al* (2007) au cas de variables dépendantes au travers des articles Delsol (2007a, 2008b) (voir Partie II, Chapitre 6 pour plus de détails). Nous évoquons ensuite les articles Delsol (2007a, 2007b) dans lesquels la convergence en norme $\mathbb{L}^p$ obtenue par Dabo-Niang et Rhomari (2003) est généralisée à des échantillons dépendants, tout en donnant les termes asymptotiquement dominants des erreurs $\mathbb{L}^p$ (se reporter au Chapitre 7, Partie II pour des détails supplémentaires).

Au fil des années d'autres modèles de régression sur variable fonctionnelle ont été proposés. On trouve ainsi le modèle linéaire généralisé fonctionnel introduit et étudié par James (2002) qui propose une approche bien adaptée dans le cas de variables fonctionnelles observées en peu de points, Ratcliffe *et al.* (2002a) ainsi que Escabias *et al.* (2004, 2005) dans le cas particulier du modèle logit, Besse *et al* (2005) dans celui du modèle multilogit, Cardot et Sarda (2005) au travers de méthodes de vraisemblance pénalisée ainsi que Müller et Stadtmüller (2005) qui utilisent la décomposition de Karhunen-Loeve et des méthodes de quasi-vraisemblance. Une généralisation des modèles de type "generalized projection pousuit" au cas de variables explicatives fonctionnelles est proposée par James et Silverman (2005) sous le nom de "functional adaptative model estimation" (FAME). D'autre part, une généralisation du modèle (SIR) a été proposée par Dauxois, Ferré *et al.* (2001) puis étudié par Ferré et Yao (2003, 2005) ainsi que par Ferré et Villa (2005, 2006). D'autres travaux comme ceux de Ferraty *et al.* (2003), Ait-Saïdi *et al.* (2005) et Ait-Saïdi *et al* (2008) ont obtenu différents résultats de convergene pour le modèle à

indice simple fonctionnel. Plus récement, un modèle semi-fonctionnel partiellement linéaire a été proposé et étudié par Aneiros-Perez et Vieu (2006). Enfin on peut aussi évoquer les travaux de Ratcliffe *et al.* (2002b) sur un modèle linéaire pour variable explicative observée plusieurs fois et ceux de Rossi et Conan-Guez (2005a) dans le cas d'un modèle non linéaire dépendant de paramètres vectoriels.

Toutes les méthodes que nous venons d'évoquer concernent l'estimation de l'espérance conditionnelle. Cependant, la régression sur variable fonctionnelle peut aussi être réalisée à partir de l'estimation d'autres quantités liées à la distribution conditionnelle de Y sachant X. Un des premiers exemples auxquels on peut penser est celui de la régression par quantiles conditionnels pour lesquels des méthodes d'estimation linéaire ont été proposées et étudiées par Cardot *et al.* (2004a, 2004b, 2005). Une méthode d'estimation des quantiles conditionnels à partir de l'estimation à noyau de la fonction de répartition conditionelle a également été proposée et étudiée par Ferraty, Rabhi *et al.* (2005), Ferraty *et al.* (2006), Ferraty et Vieu (2006a) et Ezzahrioui (2007). En ce qui concerne la régression par le mode conditionnel, la méthode proposée et étudiée par Ferraty, Laksaci *et al.* (2005), Ferraty et Vieu (2006a), Ferraty *et al.* (2006), Dabo-Niang et Laksaci (2006) et Ezzahrioui (2007) est basée sur l'estimation de la densité conditionnelle par des estimateurs à noyau. Les méthodes à noyaux développées pour estimer la fonction de répartition et la fonction de densité conditionnelles (voir Laksaci, 2007) permettent de considérer le problème de l'estimation de la fonction de hasard conditionnelle (voir Ezzahrioui, 2007, et Ferraty *et al.*, 2008). On pourra regarder également les récents travaux de Attouch *et al.* (2007) sur les estimateurs robustes en régression sur variable fonctionnelle.

Quelques liens entre discrimination et régression :

Revenons enfin au problème de la discrimination de courbes. Nous avons évoqué dans les paragraphes précédents les approches basées sur l'analyse discriminante et la notion de profondeur. Une autre manière de proposer une solution à ce type de problèmes est de constater qu'ils peuvent être reliés à des problèmes de régression sur variable fonctionnelle particuliers. En effet, lorsque l'on est confronté à un problème de discrimination de courbes on dispose d'un échantillon de courbes auxquelles sont associées des classes. L'objectif est d'utiliser l'information donnée par les courbes déjà classées afin de classer d'autres courbes dont on ne connaît pas la classe d'appartenance. On peut modéliser la relation entre une courbe (on note X la variable fonctionnelle correpondant à cette courbe) et sa classe d'appartenance (on note Y la variable à valeurs dans $\{1, 2, \ldots, k\}$ représentant la classe associée à X) au travers d'un modèle de régression sur variable fonctionnelle. Une première approche consiste à utiliser un modèle de régression dans lequel la variable réponse est supposée catégorielle comme par exemple les modèles linéaires généralisés (multi)logit (voir les références données précédemment). Une autre approche, utilisée notamment par Hall *et al.*(2001) ainsi que Ferraty et Vieu (2003), consiste dans un premier temps à estimer, pour toute nouvelle courbe x à classer, la

quantité $P(Y = c|X = x)$ pour chaque classe c. Cela revient à estimer la valeur en x de l'opérateur de régression par moyenne conditionnelle reliant la variable $1_{Y=c}$ à la variable fonctionnelle X. Ensuite on attribue à la courbe x la classe c pour laquelle la valeur estimée de $P(Y = c|X = x)$ est la plus grande. Hall *et al.* (2001) et Ferraty et Vieu (2003) proposent des méthodes d'estimation non-paramétriques de ces probabilités conditionnelles basées respectivement sur une approche par projection et sur des méthodes à noyau. Plus récemment, Abraham *et al.* (2006) ont étudié la consistance (suivant la nature de l'espace fonctionnel et la loi des variables fonctionnelles étudiées) d'une version particulière de ces classificateurs, appelée classificateur de la fenêtre mobile dans le cas de la discrimination en deux classes. Des résultats de consistance pour une variante appelée classificateur des k plus proches voisins ont été établis par Biau *et al.* (2005) ainsi que Cerou et Guyader (2006). On trouve d'autre part des méthodes de discrimination de courbes construites à partir du modèle SIR fonctionnel comme dans l'article de Ferré et Villa (2005). Une autre méthode de discrimination basée sur l'utilisation de SVM généralisée au cadre fonctionnel a été proposée et étudiée par Rossi et Villa (2006) puis par Villa et Rossi (2006a, 2006b) qui établissent des résultats de consistance. Enfin, Costanzo *et al.* (2006) puis Preda *et al.* (2007) proposent l'utilisation de méthodes de régression "partial least square" fonctionnelle qu'ils relient à l'analyse dicriminante fonctionnelle.

Les résultats présentés dans cette thèse portent sur l'estimation de l'opérateur de régression sur variable fonctionnelle à partir de méthodes à noyau. Ils ne traitent pas de manière directe des problèmes de discrimination fonctionnelle cependant, ils sont liés aux problèmes de discrimination au travers de l'approche utilisée par Ferraty et Vieu (2003).

3.2.3.2. B. Autres situations. — Parmi les différents modèles de régression où interviennent des variables fonctionnelles, le cas le plus étudié est celui que nous venons d'évoquer où la variable réponse Y est réelle tandis que la variable explicative X est fonctionnelle. Nous donnons tour à tour un aperçu de la littérature consacrée aux situations où Y est fonctionnelle et X vectorielle, X et Y sont fonctionnelles, puis où X et Y sont fonctionnelles et proviennent de l'étude d'un même processus à temps continu.

On se place pour commencer dans le cas où la variable réponse est fonctionnelle et la variable explicative vectorielle. Il semble que les premiers travaux consacrés à ce type de situation soient ceux de Faraway (1997), Ramsay et Silverman (1997, 2005, Chapitre 13) ainsi que Brumback et Rice (1998) dans le cas d'une fonction de régression linéaire. Des généralisations du modèle linéaire ont été proposées et étudiées par la suite. On peut notamment citer Guo (2002) qui étudie une version générale du modèle à effet mixte fonctionnel. D'autre part Chiou *et al.* (2003) s'intéressent à un modèle de quasi-vraisemblance fonctionnel, Chiou Müller Wang et Carey (2003) considèrent un modèle à effets multiplicatifs et enfin Chiou *et al.* (2004) étudient un modèle qui généralise ces deux modèles. De plus, Chiou, Müller,

Wang et Carey (2003) étudient aussi le modèle le plus général dans lequel aucune hypothèse de forme n'est faite sur la fonction de régression. On peut également évoquer l'approche particulière proposée par Li *et al.* (2003) dans laquelle sont proposées des techniques de réduction de la dimension de la variable réponse fonctionnelle et des variables explicatives. Plus récemment, Hlubinka et Prchal (2007) proposent et comparent différents types d'estimateurs. On peut également évoquer les méthodes d'arbres de régression fonctionnels proposées et étudiées par Yu et Lambert (1999) puis Nerini et Ghattas (2007).

On se place à présent dans le cas où à la fois la variable réponse et la variable explicative sont fonctionnelles. On suppose que l'on dispose d'un échantillon formé de n paires indépendantes de variables fonctionnelles. Ce type de modèle de régression est déjà évoqué dans les travaux de Ramsay et Dalzell (1991) ainsi que ceux de Hastie et Tibshirani (1993). Le plus simple de ces modèles est celui où la valeur de la variable réponse au temps t dépend de la variable explicative uniquement au travers de la valeur qu'elle prend au même instant. Ce type de modèle apparaît dans la littérature sous le nom de "functional concurrent model". D'assez nombreuses études ont été menées sur ce genre de modèles dans le cas où l'opérateur est de type linéaire et à coeficients variables. On peut par exemple citer à ce sujet les travaux de Hastie et Tibshirani (1993), Ramsay et Silverman (1997, 2005, Chapitre 14), Hoover *et al.* (1998), Wu *et al.* (1998), Fan et Zhang (2000), Eubank *et al.* (2004) ainsi que les références qu'ils contiennnent. Dans d'autres modèles, on cherche à prédire la valeur de la variable réponse au temps t à partir de la totalité de la variable explicative fonctionnelle. Il semble qu'il existe assez peu de travaux portant sur ce modèle. Ramsay et Silverman (1997, 2005 Chapitre 16) proposent une approche basée sur la projection de chacune des variables fonctionnelles sur une base finie. Cuevas *et al.* (2002) s'intéressent au problème de la régression linéaire d'une variable fonctionnelle sur un ensemble de fonctions non aléatoires ("fixed functional design"). D'autre part, Yao *et al.* (2005b) considèrent le problème de la régression linéaire entre deux variables longitudinales. Enfin, Müller *et al.* (2008) ainsi que Prchal et Sarda (2008) obtiennent des résultats de convergence dans le cas où X et Y sont aléatoires. Des généralisations du modèle linéaire ont également été proposées par Kneip *et al.* (2004) qui ont étudié des modèles à effets mixtes. Finalement, on peut faire le constat suivant : si les deux variables fonctionnelles sont mesurées simultanément, on ne peut prédire la valeur de Y au temps t à venir qu'à partir des valeurs de X dont on dispose avant cet instant $(X(s),\, s \leq t)$. C'est une des raisons pour lesquelles Malfait et Ramsay (2003) ont introduit la notion de modèle linéaire fonctionnel historique qui n'utilise que les valeurs $X(s),\, s \leq t$ pour prédire $Y(t)$. Ce constat est également pris en compte dans les travaux de Müller et Zhang (2005) qui étudient un modèle de régression particulier qui peut être vu comme un modèle linéaire historique généralisé dans lequel la fonction de lien varie au cours du temps.

3.2.3.3. C. Etude et prévision de processus à temps continu aux travers de modèles de régression fonctionnelle. — Lorsque l'on étudie des séries temporelles, il peut

être intéressant de ne plus voir le processus que l'on observe au travers de sa forme discrétisée mais comme un processus à temps continu observé sur l'intervalle $[0, \tau N]$. L'idée novatrice proposée par Bosq (1991) est de découper ce processus en N trajectoires successives observées sur l'intervalle $[0, \tau]$ puis de prédire certaines caractéristiques d'une trajectoire ou la trajectoire elle-même à partir de celles qui lui sont antérieures. Le travail précurseur de Bosq et Delecroix (1985) s'intéresse à ce type de problématique dans le cas de processus de Markov à valeur dans des espaces mesurables et propose des méthodes de pédiction d'une variable aléatoire Hilbertienne liée à une valeur future du processus.

Considérons tout d'abord le cas particulier dans lequel on ne souhaite pas prédire la trajectoire mais plutôt une quantité réelle liée à celle-ci. Cette quantité peut par exemple correspondre à la valeur de la courbe en un point donné, à la valeur maximale ou minimale atteinte, à la valeur moyenne, ... On est dans un modèle de régression où la variable réponse est réelle et la variable explicative fonctionnelle et notre échantillon est composé de variables dépendantes. Une grande partie des travaux utilisant des méthodes à noyau que nous avons évoqués dans la section A sont adaptés (ou ont été généralisés) à des échantillons faiblement dépendants. On trouve ainsi des résultats de convergence presque sûre (voir Ferraty *et al.*, 2002a, 2002b) et de normalité asymptotique (voir Masry, 2005 et Aspirot *et al.*, 2008). Il existe également des résultats de convergence presque complète et de normalité asymptotique pour les estimateurs à noyau des quantiles et du mode conditionnels (voir Ferraty, Rabhi *et al.*, 2005, Ferraty Laksaci *et al.*, 2005, Ferraty et Vieu, 2006a, Ezzahrioui, 2007). Ait Saidi *et al.* (2005) considèrent le cas de variables faiblement dépendantes dans le modèle à indice simple fonctionnel et enfin Ferraty et Vieu (2006a) proposent une méthode de discrimination de courbes adaptée à des courbes faiblement dépendantes. Quelques résultats ont même été obtenus dans le cas de forte dépendance par Hedli-Griche (2008) pour l'espérance conditionnelle.

On souhaite à présent prédire une trajectoire à partir des précédentes. On est alors confronté à un problème de régression où la variable explicative et la variable réponse sont fonctionnelles mais, à la différence des modèles présentés dans l'un des paragraphes précédents, l'échantillon que l'on étudie est constitué de variables fonctionnelles dépendantes provenant du découpage d'un même processus. Une façon de modéliser une telle situation est de considérer les processus autorégressifs linéaires hilbertiens (ARH) comme le propose Bosq (1989, 1990, 1991). On pourra se référer à Bosq (2000) pour un point de vue complet sur ce type de processus et de manière générale sur les processus linéaires hilbertiens. Différents travaux se sont intéressés aux problématiques de prédiction dans les modèles ARH comme par exemple Bosq (1991) qui utilise des techniques d'analyse en composantes principales fonctionnelles, Besse et Cardot (1996) puis Cardot (1998) utilisent des splines, Antoniadis et Sapatinas (2003) puis Laukaitis (2007) proposent des méthodes d'estimation par ondelettes. Bosq (1999) et Mas (2000, 2004a) établissent des résultats asymptotiques concernant l'estimateur empirique de l'opérateur de covariance, Mas (1999, 2004b) et Guillas (2001) étudient les propriétés asymptotiques de l'estimateur de l'opérateur d'autocorrélation. D'autre part, Mas et Menneteau

(2003) obtiennent des résultats de grande déviation ou de déviation modérée pour les estimateurs de la moyenne et de l'opérateur de covariance tandis que Menneteau (2005) obtient la loi du logarithme itéré pour l'estimateur de la covariance. Mas (2007b) propose quant à lui des résultats de convergence en loi concernant l'estimateur de l'opérateur d'autocorrélation. Enfin, Ruiz-Medina *et al.* (2007) étudient des données fonctionnelles spatiales à l'aide d'un modèle autorégressif hilbertien. D'autres auteurs proposent de généraliser ces modèles au cas où les variables aléatoires fonctionnelles sont à valeurs dans un espace de Banach. Il semble que les premiers travaux s'intéressant à des modèles autorégressifs linéaires banachiques (ARB) soient ceux de Pumo (1992). Les problématiques de prédiction dans le cas particulier des processus autorégressifs linéaires à valeur dans l'espace $\mathbb{C}([0;1])$ (ARC) ont été étudiés par Pumo (1995, 1998) puis Mokhtari et Mourid (2002). Par la suite, les articles de Bosq (1993a, 1996a, 2002) et Mourid (1993, 2002) démontrent différentes formes de convergence (loi des grands nombres, loi du logarithme itéré, TCL) de la moyenne et de l'opérateur de covariance empiriques dans le cadre de l'étude de processus ARB(1) et ARB(p). Les travaux de Llabas et Mourid (2003) étendent certains résultats concernant l'estimation de l'opérateur d'autocorrélation et la prédiction pour les ARH ou les ARC au cadre de processus ARB. Enfin, Rachedi et Mourid (2003) ainsi que Rachedi (2004, 2005) étudient un estimateur crible de l'opérateur d'autocorrélation. On trouve également les travaux de Marion et Pumo (2004) ainsi que Mas et Pumo (2007) qui s'intéressent aux modèles ARHD qui prennent en compte les dérivées du processus hilbertien. D'autre part, Guillas (2000) puis Damon et Guillas (2005) proposent un modèle autorégressif hilbertien dans lequel est introduite une variable exogène. D'autre part Guillas (2002) étudie un processus hilbertien doublement stochastique, qui s'écrit sous la forme d'un processus hilbertien autorégressif dans lequel l'opérateur de corrélation dépend d'un processus réel. Une autre famille particulière de processus linéaires hilbertiens est celle des processus hilbertiens à moyenne mobile. On pourra se référer à Bosq (2003b, 2004, 2007) pour une caractérisation de ces processus ainsi que des résultats asymptotiques et Turbillon (2007) pour des résultats récents concernant l'estimation et la prédiction dans ces modèles. Des résultats de convergence ont également été obtenus dans le cadre général de processus linéaires fonctionnels notamment par Merlevède (1996, 1997), Merlevède *et al.* (1997), Bosq (2000, 2003a) et Mas (2000, 2002). Enfin, Bosq (2007) propose une définition plus générale de différents types de processus linéaires hilbertiens. Dans le cas de processus autorégressifs non-linéaires fonctionnels, des méthodes d'estimation ont été proposées par Besse *et al.* (2000) qui introduisent des méthodes d'estimation à noyau fonctionnelle, puis par Antoniadis *et al.* (2006) qui utilisent des ondelettes. D'autres processus stochastiques peuvent être étudiés à l'aide de méthodes de statistique fonctionnelle comme en témoignent par exemple les travaux de Valderama *et al.* (2002), Ortega-Moreno *et al.* (2002), Valderama *et al.* (2003) et ceux de Aguilera *et al.* (2002) et Bouzas *et al.* (2006a, 2006b) qui étudient l'intensité et la moyenne d'un processus de Cox.

Notre apport dans ce domaine concerne la prédiction d'une caractéristique réelle d'une trajectoire à partir des trajectoires précedentes. Les résultats donnés dans la Partie II de ce mémoire complètent les travaux de Ferraty *et al.* (2002a, 2002b) et Masry (2005) en proposant des résultats concernant la normalité asymptotique

(voir Delsol, 2007a, 2008b) et les erreurs $\mathbb{L}^p$ (voir Delsol, 2007a, 2007b) d'un estimateur à noyau de l'opérateur de régression dans le cas d'un échantillon de variables dépendantes.

3.3. Tests

On trouve un nombre assez restreint de tests dans la littérature consacrée à la statistique fonctionnelle. Voici un aperçu de ce qui a été proposé.

On trouve des tests permettant de comparer des familles de courbes. On peut citer à ce propos les travaux de Cuevas *et al.* (2004) dans lesquels sont proposés deux tests de type ANOVA qui consistent à tester si l'espérance des variables fonctionnelles est la même dans tous les groupes. Ferraty *et al.* (2007) introduisent un test heuristique de comparaison de différents groupes de données fonctionnelles basé sur l'analyse des facteurs. Ce test considère des hypothèses portant sur les covariances dans les différents groupes. D'autre part, Delicado (2007) propose différentes approches permettant de tester si la distribution qui a engendré les données dans chacun des groupes est la même. Dans le cas particulier de la comparaison de différents groupes de fonctions de régression sur variable réelles ou vectorielles, la nature fonctionnelle des données peut être contournée en s'intéressant plutôt aux résidus. Nous renvoyons à Pardo Fernández (2005, 2007) pour plus de références sur ce sujet.

D'autres procédures de test s'intéressent à des hypothèses concernant la structure des variables fonctionnelles. James et Sood (2006) ainsi que Mas (2007a) proposent des tests portant sur la forme de l'espérance de la variable aléatoire fonctionnelle étudiée. Viele (2001) propose un test permettant de valider ou d'invalider une modélisation stochastique de la loi de probabilité qui a engendré les données. Enfin Hall et Vial (2006b) proposent un test concernant une possible réduction de la dimension de la variable fonctionnelle étudiée.

Evoquons à présent les tests portant sur des modèles de régression où interviennent des variables fonctionnelles. On se place pour commencer dans le cas où la variable réponse est réelle et la variable explicative fonctionnelle. Les tests proposés dans la littérature pour ce genre de modèle ne sont pas très nombreux. Gadiaga et Ignaccolo (2005) introduisent un test de non-effet de la variable explicative basé sur des méthodes de projection. Cardot *et al.* (2003) puis Cardot *et al.* (2004) considèrent des tests d'hyothèses dans le modèle linéaire fonctionnel. Plus récemment, Chiou et Müller (2007) proposent un test d'adéquation heuristique basé sur la décomposition en composantes principales fonctionnelles de la variable explicative. La littérature existante semble être restreinte à ces quelques travaux. Dans le cas où la variable réponse est fonctionnelle et la variable explicative vectorielle, Shen et Faraway (2004) introduisent une procédure pour tester l'influence d'une variable explicative dans le modèle linéaire, Cardot *et al* (2007) proposent

des tests de structure basés sur des méthodes de permutation. On trouve également dans Mas (2000) un test de nullité de l'opérateur d'autocorrélation d'un ARH(1). Enfin, des tests ont été proposés dans le cas général où les deux variables sont fonctionnelles notamment par Antoniadis et Sapatinas (2007) dans le modèle à effets mixtes fonctionnel.

Il n'existe pas de test de structure général permettant de tester si un modèle régression d'une variable réelle sur une variable fonctionnelle est linéaire, à indice simple, si l'effet de la variable explicative fonctionnelle peut être résumé par l'effet de quelques valeurs caractéristiques (extrema, points d'inflexion, moyenne,...) de celle-ci, ... L'approche développée dans Delsol *et al.* (2008) et détaillée dans la Partie III de ce mémoire comble en partie ce manque en proposant une manière générale de construire des tests de structure dans les modèles de régression sur variable fonctionnelle. On établit la convergence en loi de la statistique de test étudiée sous l'hypothèse nulle ainsi que sa divergence sous l'alternative. Ce résultat est obtenu sous des conditions générales qui permettent d'utiliser la statistique proposée pour réaliser des tests de structure innovants et très variés tels que les tests de non-effet, de linéarité, de modèle à indice simple, ...

CHAPITRE 4

LA CONTRIBUTION DE CE MÉMOIRE EN RÉGRESSION NON-PARAMÉTRIQUE FONCTIONNELLE

Dans ce chapitre nous donnons un rapide aperçu de la contribution de cette thèse à la statistique fonctionnelle. Nous présentons dans un premier temps le modèle de régression sur variable fonctionnelle que nous étudions. Nous rappelons ensuite l'expression de l'estimateur à noyau et nous proposons un bref résumé des résultats de cette thèse. Certains d'entre eux consistent en la généralisation de résultats obtenus dans le cas indépendant (voir Partie II). Les autres sont tout à fait novateurs et introduisent une approche générale pour construire des tests de structure en régression sur variable fonctionnelle (voir Partie III). Nous évoquerons ensuite l'étude de ces différents résultats au travers de simulations et d'applications à des données réelles. Enfin, nous expliquerons comment les résultats de ce mémoire sont peuvent être utilisés dasn le cadre de la discrimination de courbes.

4.1. Modèles de régression sur variable fonctionnelle

Tout au long de ce mémoire on considère les modèles de régression pouvant s'écrire

$$Y = r(X) + \epsilon, \tag{4.1}$$

où la variable réponse Y est à valeurs réelle tandis que la variable explicative X est à valeur dans un espace semi-métrique $(E,\, d)$ de dimension infinie. On suppose d'autre part que la variable correspondant au résidu ϵ vérifie $\mathbb{E}[\epsilon \mid X] = 0$ de telle sorte que $r(x) = \mathbb{E}[Y \mid X = x]$. Les résultats énoncés dans ce mémoire concernent l'estimation de cet opérateur de régression r ainsi que des tests de structure qui lui sont attachés. Nous considérons un échantillon de variables aléatoires $(X_i, Y_i)_{1\leq i\leq n}$ de même loi que (X, Y). Afin notamment de pouvoir considérer des problèmes de prédiction de séries temporelles au travers de l'approche fonctionnelle introduite par Bosq (1991), certains de nos résultats sont établis pour des échantillons de variables faiblement dépendantes. Comme nous l'avons vu dans le chapitre précédent, ce type de modèle de régression a été étudié sous des formes très variées que ce soit par rapport aux hypothèses faites sur la forme de l'opérateur de régression r ou sur la dépendance des échantillons étudiés.

4.1.1. Modèles paramétriques et non-paramétriques de régression sur variable fonctionnelle. — Nous nous intéressons principalement dans ce mémoire à une approche exploratoire des modèles de régression sur variable fonctionnelle au travers de l'étude d'un estimateur non-paramétrique de l'opérateur de régression. Cet estimateur, dont l'expression est donné dans la Section 3 de ce chapitre, est conçu pour l'étude de modèles de régression non-paramétrique (sur variable) fonctionnelle.

Avant d'aller plus loin, il est important de définir clairement ce qu'est, pour nous, un modèle de régression non-paramétrique fonctionnelle. La frontière entre les modèles paramétriques et non-paramétriques n'est pas clairement définie dans la littérature. En effet, même dans le cas le plus simple d'un modèle de régression où X et Y sont deux variables aléatoires réelles il peut y avoir ambiguité suivant que l'on considère l'une ou l'autre des définitions suivantes :

(a) un modèle est paramétrique si la loi du couple (X, Y) est supposée appartenir à un espace indexé par un nombre fini de paramètres réels.

(b) un modèle est paramétrique si l'opérateur de régression r est supposé appartenir à un espace indexé par un nombre fini de paramètres réels.

Considérons par exemple les trois modèles de régression suivants :

$$Y = aX + b + \epsilon,\ X \sim \mathcal{N}\left(\mu, \gamma^2\right) \text{ et } \epsilon \sim \mathcal{N}\left(0, \sigma^2\right) \tag{4.2}$$

$$Y = aX + b + \epsilon,\ \mathbb{E}\left[\epsilon \mid X\right] = 0. \tag{4.3}$$

$$Y = r\left(X\right) + \epsilon,\ r \in \mathcal{C}\left(\mathbb{R}\right) \text{ et } \mathbb{E}\left[\epsilon \mid X\right] = 0. \tag{4.4}$$

Le modèle (4.2) (respectivement (4.4)) est paramétrique (respectivement non-paramétrique) selon les deux points de vue (a) et (b). Cependant, le modèle (4.3) est non-paramétrique au sens de (a) mais paramétrique au sens de (b) car la fonction de régression est paramétrée par le vecteur (a, b). Il apparaît toutefois qu'en ce qui concerne les problématiques d'estimation, le modèle (4.3) est plus proche du modèle (4.2) que du modèle (4.4). C'est une des raisons pour lesquelles il nous paraît préférable d'adopter dans ce mémoire une définition basée sur la nature de l'ensemble dans lequel on suppose que l'opérateur de régression r se trouve. Cela permet de mettre d'un côté les modèles où l'on ne fait que des hypothèses de forme concernant l'opérateur de régression (comme par exemple le modèle linéaire) et de l'autre ceux pour lesquels on fait des hypothèses concernant le régularité de r.

On cherche une manière de définir les modèles paramétriques en régression sur variable fonctionnelle. Utiliser telle quelle la définition donnée dans le cas multivarié n'est pas satisfaisant puisque par exemple le modèle linéaire fonctionnel serait considéré comme non-paramétrique. Ce problème provient du fait qu'en statistique fonctionnelle, on étudie des échantillons de variables qui ne sont pas multivariées mais à valeurs dans un espace de dimension infinie E. C'est une des raisons pour lesquelles Ferraty et Vieu (2006a) proposent d'adapter la définition du cas multivarié au cas fonctionnel en définissant un modèle comme paramétrique s'il ne dépend que d'un ensemble fini d'éléments de E. Cette définition a l'avantage de faire la différence entre des modèles dans lesquels on ne fait que des hypothèses sur

la forme (linéaire par exemple) de l'objet que l'on cherche à estimer (dans notre cas l'opérateur de régression r) et des modèles plus généraux où on fait des hypothèses sur sa régularité (continuité, de type Hölder, ...). Nous adoptons dans ce mémoire cette façon de définir les modèles paramétriques et non-paramétriques fonctionnels :

Definition 4.1.1. —
- *Un modèle de régression sur variable fonctionnelle est un modèle de la forme (4.1) dans lequel on suppose que r appartient à un ensemble particulier $\mathcal{R}$ d'opérateurs.*
- *Un modèle de régression sur variable fonctionnelle est dit paramétrique si $\mathcal{R}$ est indexé par un nombre fini d'éléments de E.*
- *Un modèle de régression sur variable fonctionnelle est dit non-paramétrique si $\mathcal{R}$ ne peut pas être indexé par un nombre fini d'éléments de E.*

Cette définition présente clairement la frontière entre les modèles paramétriques et non-paramétriques de régression sur variable fonctionnelle. Nous allons voir au travers des deux premiers exemples suivants que l'on retrouve comme dans le cas multivarié la distinction entre des modèles paramétriques comme le modèle linéaire fonctionnel et des modèles plus généraux, dans lesquels on ne fait que des hypothèses de régularité sur r. Toutefois, le dernier exemple que nous évoquons montre que la définition que nous avons choisie doit être utilisée avec précaution.

Quelques exemples de modèles étudiés dans la littérature :

1. Le modèle linéaire fonctionnel est un modèle de régression sur variable fonctionnelle dans lequel on suppose que E est un espace de Hilbert et que l'opérateur de régression r est linéaire et continu c'est-à-dire que $\mathcal{R} = \mathcal{L}^C(E, \mathbb{R})$. On sait par le théorème de représentation de Riesz qu'il existe un unique élément $\alpha \in E$ tel que, si $< ., . >_E$ représente le produit scalaire sur E, on ait $r(\,.\,) =< . , \alpha >_E$. Par conséquent chaque élément de $\mathcal{R}$ est indexé par un élément de E. C'est pourquoi, selon la définition que nous venons de donner, le modèle linéaire fonctionnel est un modèle de régression paramétrique fonctionnelle et peut s'écrire :

$$Y =< X, \alpha >_E + \epsilon. \tag{4.5}$$

2. On considère à présent la situation où l'on ne souhaite pas faire d'hypothèse à priori sur la forme de l'opérateur de régression. Il est nécessaire cependant de faire des hypothèses concernant sa régularité. On suppose que r est Hölderien d'ordre β de E dans $\mathbb{R}$. L'ensemble $\mathcal{R}$ ne peut pas être indexé par une famille finie d'éléments de E. On dira par conséquent que ce type de modèle de régression est non-paramétrique.

3. Evoquons maintenant le modèle à indice simple fonctionnel dans lequel on suppose que $\mathcal{R} = \{\Phi(< ., \theta >_E), \theta \in E, \Phi \in Hol(\beta)\}$, où on note $Hol(\beta)$ l'ensemble des fonctions Hölderiennes (d'ordre β) de $\mathbb{R}$ dans $\mathbb{R}$. Ce type de modèle peut s'écrire

$$Y = \Phi(< \theta, X >_E) + \epsilon. \tag{4.6}$$

Suivant la nature de l'espace semi-métrique E, ce type de modèle peut être défini comme paramétrique ou non-paramétrique. En effet, si $Hol(\beta)$ peut être indexé par un nombre fini d'éléments de E, le modèle est paramétrique. Par exemple, si $Hol(\beta) \subset E$ le modèle est indexé par le couple d'éléments de E : (θ, Φ) et est donc paramétrique. Cependant, si $Hol(\beta)$ ne peut pas être indexé par un nombre fini d'éléments de E, alors on dira que ce modèle est non-paramétrique. Ce type de situation illustre bien les ambiguités que peut produire une définition précise des modèles paramétriques. Par analogie avec le cas multivarié, ce type de modèles est qualifié de semi-paramétrique.

On s'intéresse principalement dans ce mémoire à des modèles de régression non-paramétrique sur variable aléatoire fonctionnelle du même type que celui évoqué dans le deuxième exemple.

4.1.2. Modèles de dépendance. — De nombreux résultats de statistique fonctionnelle ont été établis en considérant des échantillons indépendants. Cependant, il est parfois intéressant de considérer et d'étudier des échantillons dépendants afin de pouvoir répondre à des situations où les données ne sont pas indépendantes.

C'est par exemple le cas lorsque l'on s'intéresse à l'étude de séries temporelles au travers de l'approche fonctionnelle introduite par Bosq (1991). On ne regarde plus la série temporelle au travers de sa forme discrétisée mais comme la réalisation d'un processus à temps continu ξ_t observé sur un intervalle de temps $[0; N\tau]$. Ensuite, l'idée est de découper le processus en N courbes successives $X_i(t) = \xi_{t+(i-1)\tau}, t \in [0, \tau[\,, i = 1, \ldots, N$. Supposons que le processus que l'on étudie est stationnaire et que l'on cherche à prédire à partir de la dernière période observée une valeur associée à la période suivante. On est amené à étudier le modèle de régression sur variable fonctionnelle suivant :

$$G(X_i) = r(X_{i-1}) + \epsilon_i, \tag{4.7}$$

où G est un opérateur de E dans $\mathbb{R}$ connu. On suppose implicitement en écrivant ce modèle de régression que $\mathbb{E}[G(X_i)\,|X_{i-1}]$ existe et ne dépend pas de i afin de pouvoir définir l'opérateur de régression r. Cette condition est notamment vérifiée lorsque l'on fait l'hypothèse plus forte que le processus ξ_t est strictement stationnaire. La stricte stationnarité n'est pas nécessaire mais on pourra être amené à faire des hypothèses supplémentaires pour obtenir certains résultats théoriques. On peut penser à différents choix possibles de G comme par exemple l'opérateur qui associe à une fonction sa valeur en un point particulier, la valeur maximale ou minimale qu'elle atteint, la valeur du taux de variation entre deux points,... On est donc ramené à étudier un modèle de régression du type de (4.1) mais à partir d'un échantillon de données dépendantes.

Il y a plusieurs types de modélisation de la dépendance au sein d'un échantillon. Nous nous intéresserons dans ce mémoire à des phénomènes de dépendance faible. Il existe diverses modélisations de la dépendance faible à l'aide, entre autres (voir

Bradley, 2005), des notions de variables α-mélangeantes (Rosenblatt, 1956), β-mélangeantes (Volkonskii et Rozanov, 1959) ou ϕ-mélangeantes (Ibragimov, 1962). Les travaux de Doukhan (1994) et Bradley (2005) proposent un point de vue global sur ces différentes formes de dépendance faible et les comparent. Il apparaît notamment que la notion de variables α-mélangeantes est la plus générale. Ces auteurs évoquent également diverses situations où l'on peut démontrer qu'une famille de variables est faiblement mélangeante, complétant ainsi les nombreux travaux qui étudient la nature faiblement dépendante de chaînes de Markov et de processus autorégressifs (voir par exemple Davydov, 1973, Chanda, 1974, Withers, 1981, ou Mokkadem, 1990). Dans le cadre des processus fonctionnels on peut évoquer l'article de Allam et Mourid (2002) qui étudient la nature β-mélangeante de processus ARB. Récemment, des généralisations de ces notions de mélange ont été proposées par Dedecker et Prieur (2005). D'autre part, Doukhan et Louhichi (1999) ont introduit une nouvelle notion de dépendance faible qui peut être vérifiée pour des processus plus généraux. On pourra également se référer à Dedecker *et al.* (2007) pour de plus amples explications et références. Comme ces notions de dépendance faible sont plus récentes que les autres on dispose d'un plus petit nombre de résultats. Nous avons choisi, dans ce mémoire, de modéliser la dépendance au sein de l'échantillon étudié en considérant des variables α-mélangeantes. Ce choix a été motivé par le fait que cette modélisation est assez générale (voir par exemple Doukhan, 1994 et Bradley, 2005) et que l'on dispose de nombreux résultats permettant d'étudier ce type de variables. Il est toutefois envisageable de généraliser nos résultats aux types de dépendance proposés par Doukhan et Louhichi (1999) et Dedecker et Prieur (2005) puisque certains des résultats que nous utilisons ont récemment été généralisés à ces formes de dépendance (voir par exemple Coulon-Prieur et Doukhan, 2000, Dedecker et Prieur, 2005, Doukhan et Neumann, 2007 ou Dedecker *et al.*, 2007). On peut également envisager l'étude d'autres formes de dépendance comme par exemple celle de dépendance forte. Les phénomènes de dépendance forte sont souvent modélisés à l'aide de processus à longue mémoire (voir par exemple Mandelbrot et Van Ness, 1968, Beran, 1994 ou Palma, 2007). Les résultats pour ce type d'échantillons sont souvent de nature différente de ceux obtenus dans le cas indépendant (voir par exemple Hedli-Griche, 2008, pour de premiers résultats dans le cas d'une variable explicative fonctionnelle).

On rappelle maintenant la définition de suite de variables α-mélangeantes introduite par Rosenblatt (1956).

Definition 4.1.2. —
- *Soit* $(\zeta_i)_{i\in\mathbb{Z}}$ *une suite de variables aléatoires. On note* $\mathcal{F}_k = \sigma(X_i,\, i \leq k)$ *et* $\mathcal{G}_l = \sigma(X_i,\, l \leq i)$. *On définit les coefficients de* α*-mélange associés à la suite* $(\zeta_i)_{i\in\mathbb{Z}}$ *par :*
$$\alpha(n) = \sup_{k\in\mathbb{Z}} \sup_{A\in\mathcal{F}_k,\, B\in\mathcal{G}_{n+k}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\,\mathbb{P}(B)|\,.$$
- *Une suite de variables* $(\zeta_i)_{i\in\mathbb{Z}}$ *est dite* α*-mélangeante si ses coefficients de* α*-mélange vérifient* $\alpha(n) \overset{n\to+\infty}{\to} 0$.

Si la suite de variables que l'on étudie est de la forme $(\zeta_i)_{i\in\mathcal{I}}$ avec $\mathcal{I} \subsetneq \mathbb{Z}$, alors la convention veut que l'on complète celle-ci en posant $\zeta_i = 0$, $i \notin \mathcal{I}$. La suite initiale sera dite α-mélangeante si la suite complétée l'est. Parmi les différents types de décroissance des coefficients de α-mélange on distinguera les cas de décroissance arithmétique et géométrique.

Definition 4.1.3. —
- *Les coefficients de α mélange sont dits arithmétiques d'ordre $a > 0$ s'il existe une constante positive C telle que $\alpha(n) \leq Cn^{-a}$.*
- *Les coefficients de α-mélange sont dits géométriques s'il existe une constante positive C et $0 \leq \rho < 1$ tels que $\alpha(n) \leq C\rho^n$.*

On peut remarquer facilement que les coefficients de α-mélange géométriques sont arithmétiques d'ordre a pour tout a. On pourra donc appliquer les résultats donnés dans la Partie II de ce mémoire pour des coefficients de α-mélange arithmétiques à des variables géométriquement α-mélangeantes. On trouve dans la littérature un nombre conséquent de travaux consacrés à l'étude de variables α-mélangeantes réelles dont les monographies de Doukhan (1994), Bosq (1996, 1998), Rio (2000) et Yoshihara (2004) donnent un point de vue global. On peut y trouver, par exemple, des résultats de type théorème de la limite centrale (voir également, par exemple, Doukhan *et al.*, 1994, Rio, 1995, Liebscher 1996, 2001), des inégalités exponentielles (voir également, par exemple, Bosq, 1975, 1993b, Carbon, 1983 et Rhomari, 2002) et des inégalités concernant les covariances. Nous renvoyons aux différents livres évoqués précédemment pour compléter cette bibliographie en ce qui concerne d'autres résultats concernant par exemple la loi du logarithme itéré, la loi des grands nombres, ... Enfin, il semble qu'il existe assez peu de résultats concernant des propriétés de suites α-mélangeantes à valeur dans des espaces de dimension infinie. Cependant, les travaux de Rhomari (2002), proposent des inégalités exponentielles pour des suites β-mélangeantes banachiques.

Tous ces résultats constituent des outils primordiaux dans l'étude de variables α-mélangeantes et permettent de généraliser certains résultats obtenus avec des échantillons indépendants. L'utilisation de ces outils nous permet d'établir des résultats asymptotiques (normalité asymptotique, expression asymptotique des moments et des erreurs $\mathbb{L}^p$) concernant l'estimation de l'opérateur de régression pour des échantillons α-mélangeants fonctionnels (voir Partie II). Les résultats concernant les problématiques de tests de structure (voir Partie III) sont établis seulement dans le cas d'échantillons indépendants car il manque, pour l'instant, un résultat donnant la normalité asymptotique de U-statistiques construites à partir d'échantillons α-mélangeants. Cependant, il est intéressant de noter qu'il existe des résultats de type théorème de la limite centrale pour des U-statistiques construites à partir de variables β-mélangeantes (voir par exemple Kashimov, 1993, Yoshihara, 1976, Harel et Puri, 1994, Fan et Li, 1999, Borovkova *et al,*. 2001 ou Harel et Elharfaoui, 2003). On peut par conséquent envisager de généraliser certains des

résultats énoncés dans la Partie III de ce mémoire à des échantillons de variables β-mélangeantes.

4.2. L'estimateur à noyau de l'opérateur de régression

Tout au long de cette thèse nous nous intéressons à une approche exploratoire des données en ce sens que nous essayons de faire le moins possible d'hypothèses concernant la forme de l'opérateur que l'on cherche à estimer. Ceci nous amène à considérer un modèle non-paramétrique de régression.

4.2.1. Bref retour au cas multivarié. — Dans le cas particulier de variables explicatives multivariées, l'utilisation de méthodes à noyau pour estimer la fonction de régression fait partie des méthodes non-paramétriques classiques. L'idée consiste à estimer la valeur de r associée à un élément x de $\mathbb{R}^p$ par une moyenne pondérée des valeurs de la variable réponse dont on dispose :

$$\hat{r}(x) = \frac{\sum_{i=1}^{n} Y_i W_{i,n}(x)}{\sum_{i=1}^{n} W_{i,n}(x)}.$$

Les poids ont pour objectif de donner plus d'importance aux valeurs Y_i pour lesquelles X_i est "proche" de x. En 1961, Tukey introduit un estimateur appelé régressogramme pour lequel les poids sont choisis de la forme particulière $W_{i,n} = \sum_{j=1}^{K} 1_{B_j}(x)\, 1_{B_j}(X_i)$, où B_j, $1 \leq j \leq K$ est une partition choisie par l'utilisateur. Bosq et Delecroix (1985) étudient cet estimateur dans un contexte plus général. Cet estimateur revient à estimer $r(x)$ par une fonction constante sur chaque B_j dont la valeur correspond à la moyenne des Y_i pour lesquels les X_i appartiennent à B_j. Il a comme principaux inconvénients de dépendre du choix de la partition utilisée et de manquer de régularité. Pour résoudre le premier de ces problèmes, un estimateur de type fenêtre mobile a été proposé. Il consiste à ne prendre en compte, pour estimer $r(x)$, que les valeurs Y_i pour lesquelles la distance entre X_i et x est inférieure à une valeur seuil h_n appelée paramètre de lissage. Cela revient à considérer des poids de la forme $W_{i,n} = 1_{|X_i - x| \leq h_n}$. Cette méthode ne résoud pas le problème de la régularité puisque l'estimateur est toujours constant par morceaux. En plus de cela, on attribue le même poids aux valeurs Y_i pour lesquelles X_i est très proche de x qu'à celles pour lesquelles X_i et x sont un peu plus éloignés. Pour résoudre ces deux types de problèmes Nadaraya (1964) et Watson (1964) proposent d'utiliser une fonction K, appelée noyau, pour construire des poids de la forme $W_{i,n} = K\left(\frac{X_i - x}{h_n}\right)$. Les noyaux usuels sont supposés à support dans $[-1, 1]$, symétriques autour de l'origine et décroissants sur $[0; 1]$ mais dans certains cas il peut être intéressant de considérer des noyaux plus généraux comme par exemple ceux proposés par Collomb (1976), Gasser et Müller (1979) et Gasser *et al.* (1985). Cet estimateur dépend encore d'un paramètre h_n que l'utilisateur doit choisir avec précaution afin d'éviter des problèmes de sur-lissage ou de sous-lissage. L'expression explicite et relativement simple de cet estimateur ainsi que la facilité avec laquelle il peut être mis en oeuvre sur des données réelles ont suscité beaucoup d'intérêt. C'est pourquoi il n'est pas possible de donner un point de vue exhaustif des résultats

obtenus dans le cas multivarié. Nous renvoyons le lecteur aux travaux de Collomb (1981, 1985), Härdle (1990) et Sarda et Vieu (1999) ainsi qu'aux références qu'ils contiennent pour tout complément en ce qui concerne l'estimation à noyau de la fonction de régression à partir d'un échantillon de données indépendantes. Des résultats ont également été obtenus pour l'estimateur à noyau dans le cas d'un échantillon de variables dépendantes. Parmi les nombreux travaux consacrés au cas de variables faiblement mélangeantes on peut citer le papier précurseur de Collomb (1984) puis les ouvrages généraux de Györfi *et al.* (1989), Yoshihara (1994) et Bosq (1996, 1998).

4.2.2. Présentation de l'estimateur. — L'estimateur que nous étudions dans ce mémoire est une généralisation des estimateurs de type Nadaraya-Watson au cas de variables fonctionnelles introduite par Ferraty et Vieu (2000). Pour tout élément x de E il s'écrit de la manière suivante

$$\hat{r}\,(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{d(X_i,x)}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{d(X_i,x)}{h_n}\right)}, \tag{4.8}$$

où d est la semi-métrique associée à E, h_n le paramètre de lissage et K un noyau qui est positif et décroissant sur son support $[0,1]$. Pour être complet, il nous faut définir la valeur de notre estimateur lorsque le dénominateur est nul. Le fait que K soit positif assure que si le dénominateur est nul alors il en est de même pour le numérateur et on choisira (par convention) de définir notre estimateur de manière usuelle : $\hat{r}\,(x) = 0$.

Différents résultats ont déjà été obtenus pour cet estimateur en ce qui concerne la convergence presque sûre dans le cas d'un échantillon indépendant (voir Ferraty et Vieu, 2000, 2002) ou α-mélangeant (voir Ferraty *et al.*, 2002a, 2002b), la normalité asymptotique dans le cas indépendant (voir Ferraty *et al*, 2007) ou dans le cas α-mélangeant (voir Masry, 2005) et la convergence en norme $\mathbb{L}^p$ dans le cas indépendant (voir Dabo-Niang et Rhomari, 2002). Récemment, certains de ces résultats ont été étendus au cas de longue mémoire (voir Hedli-Griche, 2008).

4.2.3. Probabilités de petites boules et semi-métriques. — Le problème du fléau de la dimension est un phénomène bien connu dans le cas du modèle non-paramétrique de régression multivariée. Il provoque une décroissance exponentielle des vitesses de convergences des estimateurs nonparamétriques en fonction de la dimension (voir Stone, 1982). Par conséquent, il est légitime de penser que les méthodes non-paramétriques dans des modèles de régression sur variable fonctionnelle risquent d'avoir une vitesse de convergence très lente. Dans le cas où la variable explicative est multivariée (i.e. à valeurs dans un espace de dimension finie d), les vitesses de convergence de l'estimateur à noyau sont exprimées en fonction d'un terme de la forme h_n^d, provenant de la probabilité que la variable explicative appartienne à la boule de centre x et de rayon h_n. Dans le cas d'une variable explicative fonctionnelle (i.e. à valeur dans un espace semi-métrique de dimension infinie (E,d)), les

résultats asymptotiques sont exprimés à partir de quantités plus générales appelées probabilités de petites boules et définies par :

$$F_x(h_n) := \mathbb{P}(d(X,x) \leq h_n).$$

Au travers des différents résultats de convergence concernant l'estimateur (4.8) (voir les références données dans le paragraphe précédent ou les Parties II et III de ce mémoire), on observe que la vitesse de convergence est fonction de la manière dont décroissent ces probabilités de petites boules. Il existe dans la littérature un nombre assez important de résultats probabilistes qui étudient la manière dont ces probabilités de petites boules tendent vers 0 dans le cas où d est une norme (voir par exemple Bogachev, 1998, Li et Shao, 2001, Lifshits *et al.*, 2006, Shmileva, 2006, Gao et Li, 2007) ou un type particulier de semi-norme (voir les travaux récents de Lifshits et Simon, 2005, Aurzada et Simon, 2007). On pourra également lire les travaux de Dereich (2003, Chapitre 7) consacrés au comportement des probabilités de petites boules dont les centres sont aléatoires. Au travers de ces travaux on peut voir par exemple que dans le cas de processus non-lisses tels que le mouvement brownien ou le processus d'Ornstein-Ullenbeck, ces probabilités de petites boules sont de forme exponentielle (par rapport à h_n) et que par conséquent la vitesse de nos estimateurs est en puissance de $\log(n)$ (voir par exemple Ferraty *et al.*, 2006, Paragraphe 5, Ferraty et Vieu, 2006a, Paragraphe 13.3.2 pour une discussion plus approfondie sur ce sujet). On retrouve le type de résultat auquel on pouvait s'attendre à cause du fléau de la dimension.

Comme on vient de le voir, la nature des probabilités de petites boules joue un rôle prépondérant en ce qui concerne les vitesses de convergence de notre estimateur (4.8). Celle-ci est à la fois liée à la loi de la variable explicative et à la topologie que l'on considère. On n'a pas d'emprise sur la loi de la variable explicative. Par contre, considérer une semi-métrique d plutôt que de se restreindre à une norme constitue un atout majeur pour construire des topologies très diverses adaptées aux différentes situations auxquelles on peut être confronté (voir à ce propos Ferraty et Vieu, 2006a, Chapitre 3). On a vu précédemment qu'avec des topologies usuelles induites par des normes, le fléau de la dimension rend les vitesses de convergence très faibles. Une manière de tenter de remédier à cela est de chercher à remplacer ces topologies par une topologie qui restitue de façon pertinente les proximités entre les données. Cela peut par exemple être fait à l'aide d'une semi-métrique de projection basée sur les composantes principales fonctionnelles, les décompositions en base de Fourier, d'ondelettes, de splines, ... Lorsque la variable explicative est à valeur dans un espace de Hilbert séparable, un résultat proposé par Ferraty et Vieu (2006a, Lemme 13.6, p.213) montre que l'on peut définir de manière générale une semi-métrique de projection qui permette de se ramener à des probabilités de petites boules de type fractal (i.e. $\exists C, \delta > 0$, $F_x(h) \overset{h \to 0}{\sim} C_x h^\delta$). On condense ainsi les données en réduisant leur dimension et on contourne ainsi le fléau de la dimension. En effet, on revient à des vitesses de convergence en puissance de n. Dans d'autres situations, on peut être confronté à des données très lisses (comme les courbes spectrométriques présentées au Chapitre 3). Il peut alors être intéressant d'utiliser plutôt des semi-métriques basées sur les dérivées (voir Ferraty et Vieu, 2006a). Ces semi-métiques peuvent également être utiles lorsque les données présentent un shift

vertical artificiel (i.e. non informatif vis-à-vis des réponses). Elles ont alors pour effet d'éliminer ces décalages verticaux qui nuisent à la qualité de prédiction. Enfin, on peut envisager d'autres types de semi-métriques permettant de prendre en compte d'autres types de phénomènes comme par exemple des décalages horizontaux (voir Dabo-Niang *et al.*, 2006), ou bien dans le cas du traitement d'image évoqué au Chapitre 2.

Face à la grande diversité des semi-métriques, on peut se demander comment choisir celle qui sera la mieux adaptée. Une première méthode est proposée par Ferraty *et al.* (2002b). Elle consiste dans un premier temps à choisir a priori une famille de semi-métriques candidates à partir des informations dont on dispose sur les données (forme, régularité, nature du phénomène étudié, ...). On détermine ensuite, par exemple par validation croisée, quelle est parmi ces semi-métriques celle qui est le mieux adaptée aux données. La justification théorique de l'adaptation d'une semi-métrique particulière par rapport aux données reste un problème ouvert qu'il serait intéressant d'étudier.

4.3. L'apport de ce mémoire en termes d'estimation de l'opérateur de régression

Les résultats que nous résumons maintenant sont établis pour des échantillons de variables α-mélangeantes. Ils ont trait à des propriétés de convergence ponctuelle de l'estimateur à noyau (4.8). Par souci de clarté nous ne donnons pas pour l'instant les hypothèses et renvoyons le lecteur à la Partie II de ce mémoire dans laquelle il trouvera une version détaillée des résultats que nous évoquons ici.

On considère un élément $x \in (E, d)$ auquel on souhaite estimer l'opérateur de régression. Afin d'énoncer nos résultats, nous avons besoin d'introduire les notations suivantes :

$$\begin{aligned}
&\sigma_\epsilon^2(X) := \mathbb{E}\left[\epsilon^2 \mid X\right] \text{ et } \sigma_\epsilon^2 := \sigma_\epsilon^2(x) \\
&\phi(s) = \mathbb{E}\left[(r(X) - r(x)) \,|\, d(X, x) = s\right] \\
&F(h) = \mathbb{P}(d(X, x) \leq h),\ \tau_h(s) = \frac{F(hs)}{F(h)}, s \in [0, 1]
\end{aligned}$$

On suppose que τ_h converge presque sûrement lorsque h tend vers 0 et on note la limite τ_0. Les résultats de cette partie sont exprimés à l'aide des constantes suivantes :

$$M_0 = K(1) - \int_0^1 (sK(s))' \tau_0(s)\, ds, M_j = K^j(1) - \int_0^1 \left(K^j\right)'(s)\, \tau_0(s)\, ds,\ j = 1, 2.$$

Remarque : Quelques cas particuliers où l'expression des constantes M_i, $i = 1, 2, 3$ peut être simplifiée.

– Si le processus et de type fractal (i.e. s'il existe $\delta_x, C_x > 0$ tel que l'on ait $F_x(h_n) \overset{n\to+\infty}{\sim} C_x h_n^{\delta_x}$), alors on a $\tau_0 \equiv s^{\delta_x} 1_{[0,1]}(s)$, $M_0 = \delta_x \int_0^1 s^{\delta_x} K(s)\, ds$ et $M_j = \delta_x \int_0^1 s^{\delta_x - 1} K^j(s)\, ds$, $j = 1, 2$.
– Si le processus est de type exponentiel (i.e. s'il existe $C_{1,x}, C_{2,x}, b_x > 0$ tels que $F_x(h_n) \overset{n\to+\infty}{\sim} C_{1,x} \exp -\frac{C_{2,x}}{h_n^{b_x}}$), alors on a $\tau_0 \equiv 1_{\{1\}}$, $M_0 = M_1 = K(1)$ et $M_2 = K^2(1)$.
– Si le noyau K vérifie $K \equiv 1_{[0,1]}$, alors on a $M_0 = 1 - \int_0^1 \tau_0(s)\, ds$ et $M_1 = M_2 = 1$.

4.3.1. Normalité asymptotique. — Les premiers travaux s'intéressant à la normalité asymptotique de l'estimateur à noyau (4.8) sont dus à Masry (2005). Il considère le cas d'un échantillon constitué de variables α-mélangeantes mais ne donne pas l'expression des termes asymptotiquement dominants du biais et de la variance. Parallèlement, Ferraty *et al.* (2007) ont obtenu l'expression explicite de la loi asymptotique (i.e. des termes dominants du biais et de la variance) dans le cas d'un échantillon de variables indépendantes. Les résultats que nous énonçons dans les articles Delsol (2007a, 2008b) (voir Chapitres 6 et 8 de la Partie II de ce mémoire) font le lien entre ces deux articles en généralisant les résultats de Ferraty *et al.* (2007) au cas de variables dépendantes. On obtient le résultat suivant dont les hypothèses et démonstrations sont détaillées dans la Partie II, au Chapitre 6, de cette thèse.

Théorème 4.3.1. — *Sous différents jeux d'hypothèses, notamment en ce qui concerne les coefficients de α-mélange, on montre que :*

$$\frac{M_1}{\sqrt{M_2 \sigma_\epsilon^2}} \sqrt{n \hat{F}(h_n)} \left(\hat{r}(x) - r(x) - B_n\right) \to \mathcal{N}(0,1),$$

où $B_n = h_n \phi'(0) \frac{M_0}{M_1}$ *et* $\hat{F}(t) = \frac{1}{n} \sum_{i=1}^n 1_{[d(X_i,x),+\infty[}(t)$.

Le fait d'avoir explicité les termes dominants du bias et de la variance de la loi asymptotique nous permet notamment de construire des intervalles de confiance asymptotiques ponctuels et de donner l'expression de l'erreur $\mathbb{L}^2$ (voir Chapitre 6, Partie II).

4.3.2. Expressions asymptotiques des moments centrés. — L'obtention des expressions explicites des termes dominants des moments centrés est une autre conséquence envisageable des résultats de normalité asymptotique précédents. On peut obtenir des résultats d'uniforme intégrabilité qui, combinés aux résultats de convergence en loi, permettent d'obtenir la convergence des moments centrés jusqu'à un certain seuil (que l'on note ℓ). On obtient ainsi le résultat suivant dont les hypothèses et démonstrations sont détaillées dans le Chapitre 7 de la Partie II ce mémoire. On introduit, pour exprimer nos résultat, une variable aléatoire $W \sim \mathcal{N}(0,1)$. Ce résultat est issu des articles Delsol (2007a, 2007b).

Théorème 4.3.2. — *Sous certaines hypothèses techniques, pour tout $q \in \mathbb{N}$ tel que $0 \leq q < \ell$, on a :*

$$\mathbb{E}\left[\left(\hat{r}\left(x\right) - \mathbb{E}\left[\hat{r}\left(x\right)\right]\right)^{q}\right] = \left(\sqrt{\frac{M_2\sigma_\epsilon^2}{nF\left(h_n\right)M_1^2}}\right)^{q}\mathbb{E}\left[W^q\right] + o\left(\frac{1}{\left(nF\left(h_n\right)\right)^{\frac{q}{2}}}\right).$$

4.3.3. Expressions asymptotiques des erreurs $\mathbb{L}^p$. — Les travaux de Dabo-Niang et Rhomari (2002) ont donné les vitesses de convergence en norme $\mathbb{L}^p$ dans le cas de variables indépendantes mais n'ont pas donné l'expression explicite des termes dominants de ces erreurs. Avec des arguments similaires à ceux utilisés pour obtenir les expressions explicites des moments, on peut également obtenir les expressions explicites des erreurs $\mathbb{L}^p$, $p \leq \ell$ de notre estimateur dans le cas d'un échantillon de variables α-mélangeantes. On obtient le résultat suivant dont les hypothèses et démonstrations sont détaillées dans le Chapitre 7 de la Partie II de ce mémoire.

Théorème 4.3.3. — *Sous certaines hypothèses techniques, on a :*

(i) $\forall\ 0 \leq q < \ell$,

$$\mathbb{E}\left[\left|\hat{r}\left(x\right) - r\left(x\right)\right|^q\right] = \mathbb{E}\left[\left|h_nB + W\frac{V}{\sqrt{nF\left(h_n\right)}}\right|^q\right] + o\left(\frac{1}{\left(nF\left(h_n\right)\right)^{\frac{q}{2}}}\right),$$

(ii) $\forall m \in \mathbb{N}$, $2m < \ell$,

$$\mathbb{E}\left[\left|\hat{r}\left(x\right) - r\left(x\right)\right|^{2m}\right] = \sum_{k=0}^{m}\frac{V^{2k}B^{2(m-k)}\left(2m\right)!}{\left(2\left(m-k\right)\right)!k!2^k}\frac{h_n^{2(m-k)}}{\left(nF\left(h_n\right)\right)^k} + o\left(\frac{1}{\left(nF\left(h_n\right)\right)^m}\right),$$

(iii) $\forall m \in \mathbb{N}$, $2m+1 < \ell$,

$$\mathbb{E}\left[\left|\hat{r}\left(x\right) - r\left(x\right)\right|^{2m+1}\right] = \frac{1}{\left(nF\left(h_n\right)\right)^{m+\frac{1}{2}}}\left(V^{2m+1}\psi_m\left(\frac{Bh_n\sqrt{nF\left(h_n\right)}}{V}\right) + o\left(1\right)\right),$$

où $B = \phi'\left(0\right)\frac{M_0}{M_1}$, $V = \sqrt{\frac{M_2\sigma_\epsilon^2}{M_1^2}}$ *et où* Ψ_m *est une fonction dont on donne l'expression explicite (voir Chapitre 7, Partie II).*

Ce résultat, énoncé dans Delsol (2007a, 2007b) (voir Chapitres 7 et 8, Partie II), est innovant dans le cadre fonctionnel et paraît également l'être dans le cas multivarié. En effet, l'erreur $\mathbb{L}^2$ a été assez abondamment étudiée que ce soit dans le cas multivarié ou fonctionnel, mais il semble que les seuls travaux donnant les termes asymptotiquement dominants de l'erreur $\mathbb{L}^1$ soient ceux de Wand (1990) dans le cas d'un modèle de type "fixed design" réel. Le fait de donner l'expression explicite des termes dominants des erreurs $\mathbb{L}^p$ ouvre des perspectives intéressantes en ce qui concerne le choix asymptotiquement optimal du paramètre de lissage.

4.4. L'apport de ce mémoire en termes de tests de structure

Nous avons vu au cours de la revue bibliographique qu'il existe très peu de résultats concernant les tests de structure en régression sur variable fonctionnelle. En effet on ne trouve que les articles de Cardot *et al.* (2003) et Cardot *et al.* (2004) qui s'intéressent à des tests de non-effet dans le cas du modèle linéaire, celui de Dembo et Gadiaga (2005) qui propose un test de non-effet basé sur des techniques de projection, ainsi qu'un test d'adéquation heuristique proposé par Chiou et Müller (2007). Il n'existe pas d'approche théorique générale permettant de tester si un modèle possède une structure particulière : linéaire, à indice simple, ... Nous proposons une approche permettant de construire des tests de structure très généraux en adaptant la méthode proposée par Härdle et Mammen (1993) au cadre d'une variable explicative fonctionnelle. Les résultats que nous énonçons à présent sont établis pour des échantillons de variables indépendantes.

4.4.1. Tester si $r = r_0$. — Commençons par le cas le plus simple où l'on souhaite tester l'hypothèse nulle

$$\mathcal{H}_0 \ : \ \left\{ \mathbb{P}\left(r\left(X \right) = r_0\left(X \right) \right) = 1 \right\},$$

où r_0 est un opérateur connu, contre l'alternative locale

$$\mathcal{H}_{1,n} \ : \ \left\{ \left\| r - r_0 \right\|_{\mathbb{L}^2(wdP_X)} \geq \eta_n \right\}.$$

Cela revient à faire un test de non effet sur l'échantillon $\left(Y_i - r_0\left(X_i \right), X_i \right)_{1 \leq i \leq n}$. La façon dont η_n tend vers 0 reflète la capacité de notre test à détecter des différences de plus en plus petites entre r et r_0 lorsque n croît. On s'inspire des statistiques de test proposées par Härdle et Mammen (1993) pour construire la statistique de test suivante :

$$T_n = \int \left(\sum_{i=1}^{n} \left(Y_i - r_0\left(X_i \right) \right) K\left(\frac{d\left(x, X_i \right)}{h_n} \right) \right)^2 w\left(x \right) dP_X\left(x \right),$$

où w est une fonction de poids. On étudie ensuite la loi asymptotique de cette statistique de test. Pour exprimer la loi asymptotique de T_n sous $\mathcal{H}_0$, on introduit $T_{1,n}$ et $T_{2,n}$ qui permettent d'écrire les termes de biais et de variance.

$$\begin{aligned} T_{1,n} &= \int \sum_{i=1}^{n} K^2\left(\frac{d\left(X_i, x \right)}{h_n} \right) \epsilon_i^2 w\left(x \right) dP_X\left(x \right), \\ T_{2,n} &= \int \sum_{1 \leq i \neq j \leq n} K\left(\frac{d\left(X_i, x \right)}{h_n} \right) K\left(\frac{d\left(X_j, x \right)}{h_n} \right) \epsilon_i \epsilon_j w\left(x \right) dP_X\left(x \right). \end{aligned}$$

On obtient le résultat suivant dont les hypothèses et la démonstration se trouvent dans l'article Delsol *et al.* (2008) que nous reprenons au Chapitre 10 de la Partie III de ce mémoire (voir aussi Delsol, 2008a).

Théorème 4.4.1. — *On démontre sous des hypothèses générales que l'on a :*
– *Sous* $(\mathcal{H}_0)$, $\frac{1}{\sqrt{Var(T_{2,n})}}\left(T_n - \mathbb{E}\left[T_{1,n}\right]\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,1\right)$,
– *Sous* $(\mathcal{H}_{1,n})$, $\frac{1}{\sqrt{Var(T_{2,n})}}\left(T_n - \mathbb{E}\left[T_{1,n}\right]\right) \xrightarrow{P} +\infty$.

Ce résultat complète les résultats donnés par Cardot *et al.* (2003), Cardot *et al.* (2004), puis Dembo et Gadiaga (2005) concernant les tests de non effet.

4.4.2. Tester si $r \in \mathcal{R}$. — Le résultat précédent nous sert de préliminaire pour construire une statistique de test plus générale permettant de tester si r fait partie d'une famille convexe $\mathcal{R}$ d'opérateurs. On propose de tester l'hypothèse nulle

$$\mathcal{H}_0 : \left\{\exists r_0 \in \mathcal{R},\ P\left(r\left(X\right) = r_0\left(X\right)\right) = 1\right\}$$

contre l'alternative

$$\mathcal{H}_{1,n} : \left\{ \inf_{r_0 \in \mathcal{R}} \left\| r - r_0 \right\|_{\mathbb{L}^2(wdP_X)} \geq \eta_n \right\}.$$

On suppose qu'en plus de l'échantillon de taille n on dispose également d'un autre échantillon D^* de taille m_n indépendant du premier. On veut tester si $r \in \mathcal{R}$. On aurait envie pour cela d'utiliser la projection r_0 de r sur $\mathcal{R}$ mais c'est un opérateur inconnu. On propose de l'estimer à partir du second échantillon à l'aide d'un estimateur r_0^* qui a de bonnes propriétés de convergence sous $\mathcal{H}_0$. C'est pourquoi on introduit la statistique de test suivante :

$$T_n^* = \int \left(\sum_{i=1}^{n} \left(Y_i - r_0^*\left(X_i\right)\right) K\left(\frac{d\left(x, X_i\right)}{h_n}\right) \right)^2 w\left(x\right) dP_X\left(x\right).$$

Si l'estimateur r_0^* vérifie des conditions générales, on peut obtenir le même résultat que précédemment. Les hypothèses et la démonstration sont détaillés dans Delsol *et al.* (2008) et Delsol (2008a) (voir Chapitre 10, Partie III).

Théorème 4.4.2. — *Sous les hypothèses générales on a :*
– *Sous* $(\mathcal{H}_0)$, $\frac{1}{\sqrt{Var(T_{2,n})}}\left(T_n^* - \mathbb{E}\left[T_{1,n}\right]\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,1\right)$,
– *Sous* $(\mathcal{H}_{1,n})$, $\frac{1}{\sqrt{Var(T_{2,n})}}\left(T_n^* - \mathbb{E}\left[T_{1,n}\right]\right) \xrightarrow{P} +\infty$.

Ce théorème permet d'envisager l'utilisation de la statistique de test T_n^* pour construire des procédures innovantes de tests de structure. En effet, les différents jeux de conditions proposées dans l'article permettent l'utilisation de notre statistique dans des situations très variées. (voir Chapitre 10, Partie III). En particulier, cela permet de tester si un modèle de régression sur variable fonctionnelle est linéaire, à indice simple, si l'effet de la variable explicative fonctionnelle peut se résumer à l'effet d'un vecteur de taille finie de valeurs particulières qui lui sont associées (extrema, points d'inflexion, intégrale, ...)....

4.5. Quelques Applications

Le comportement, à taille d'échantillon fixée, des résultats que nous venons d'énoncer a été étudié au travers de simulations. A la suite de ces études nous avons appliqué nos méthodes sur des données réelles. Tout d'abord, comme nous l'avons remarqué précedemment, nous pouvons construire des intervalles de confiance ponctuels à partir du résultat de normalité asymptotique que nous obtenons dans lequel les termes dominants du biais et de la variance sont explicités. On a étudié la normalité asymptotique et les performances de nos intervalles de confiance au travers de simulations en faisant varier le type de dépendance de notre échantillon (voir Chapitre 6, Partie II). On a ainsi vu qu'à taille d'échantillon fixée, la dépendance des données a une influence sur la performance des intervalles de confiance mais que les résultats dans le cas de variables α-mélangeantes sont assez bons. On a ensuite appliqué notre méthode de construction d'intervalles de confiance ponctuels à des données concrètes liées à l'étude du courant marin El Niño ou de consommation électrique (voir Chapitres 6 et 9, Partie II). Nous avons également brièvement comparé les performances en terme de prédiction de notre estimateur à noyau par rapport à l'estimateur linéaire fonctionnel et un estimateur additif multivarié sur les données liées à l'étude du courant marin El Niño. Il semble que notre approche non-paramétrique et fonctionnelle soit bien adaptée à l'étude de ce phénomène (voir Chapitre 6, Partie II).

Enfin, la mise en oeuvre de notre statistique de test nécessite de trouver une valeur seuil au delà de laquelle on rejette l'hypothèse nulle. Les termes de biais et de variance sont difficiles à estimer et, de plus, utiliser directement la loi asymptotique pour obtenir des quantiles donne souvent d'assez mauvais résultats. C'est pourquoi nous proposons plutôt de construire des quantiles empiriques par diverses méthodes de rééchantillonnage. Diverses méthodes, élaborées à partir de formes de rééchantillonnage différentes, ont été abondamment étudiées au travers de simulations dans le cas particulier du test de non-effet (voir Chapitre 11, Partie III). Il en ressort des résultats satisfaisants en termes de puissance et de niveau. On présente ensuite l'utilisation de ces tests de non effet sur deux jeux de données spectrométriques. Des simulations sont en cours dans le cas plus général du test de linéarité.

4.6. L'apport de ce mémoire en discrimination fonctionnelle

On suppose que l'on est confronté à un problème de discrimination en k classes. A chacune des variables aléatoires fonctionnelles X_i de notre échantillon on associe une variable aléatoire $T_i \in \{1, \ldots, k\}$ correspondant à la classe à laquelle elle appartient. Supposons que l'on observe une nouvelle courbe x dont on veut prédire la classe d'appartenance t (on note T la variable aléatoire correspondante). Les résultats de ce mémoire sont énoncés dans le contexte général des modèles de régression sur variable fonctionnelle. Ils peuvent toutefois être reliés à la problématique de la discrimination

fonctionnelle. En effet, une méthode classique de discrimination (voir par exemple Hall *et al.*, 2001, et Ferraty et Vieu, 2003) consiste à estimer pour chaque classe j la probabilité

$$\pi_x^j := \mathbb{P}\left(T = j | X = x\right)$$

puis à attribuer à la courbe x la classe j pour laquelle la valeur de π_x^j est la plus grande. Or, pour chaque classe j, estimer la probabilité conditionnelle π_x^j revient à estimer l'opérateur de régression de l'échantillon constitué des couples

$$\left(X_i, Y_i^j\right), \text{ où } Y_i^j = 1_{\{T_i = j\}}.$$

Les résultats que nous énonçons dans ce mémoire complètent donc les résultats de la littérature concernant la convergence de l'estimateur à noyau de π_x^j et ouvrent d'intéressantes perspectives en terme de tests de structure.

CHAPITRE 5

SOME PROSPECTS AND OPEN QUESTIONS

In the present chapter we discuss some interesting prospects of this dissertation. Although some of them appear as direct generalizations or consequences of the content of this dissertation, others are linked with open questions and need further works and developments.

Kernel methods
We have studied in this dissertation a regression model in which there is a real response variable and a functional explanatory variable. It seems that some results obtained in this dissertation for the kernel estimator of the regression operator might be generalized to other kernel estimators.

(K.1) The results of this dissertation have been expressed in the regression context. They can be used in a straightforward way in curves discrimination (see Chapters 1 and 4). Structural testing procedures introduced in Part III open interesting prospects. For instance, they allow to check if the explanatory variable has any effect on the probability to belong to a given class. It would be interesting to study how our assumptions and results may be improved in the context of curve discrimination. Moreover, it would be worth providing a practical point of view of the use of our results in this specific context.

(K.2) Recently, Azzedine *et al.* (2006) and Attouch *et al.* (2007) have considered and introduced robust kernel estimators. With similar ideas as those used in Chapter 7, it seems possible to provide the explicit expression of the asymptotic dominant terms of $\mathbb{L}^q$ errors for these estimators.

(K.3) Asymptotic normality results have been obtained for other kernel estimators. Consequently, it seems possible to get the expression of $\mathbb{L}^q$ errors for other kernel estimators as for instance the kernel conditional cumulative distribution function estimator, the kernel conditional density estimator, the kernel estimator of the conditional hazard function, ...

(K.4) In this dissertation, we have proposed a no effect test based on the conditional mean. It would be interesting to propose other no effect tests based on other conditional quantities (density, cumulative distribution function, mode, quantiles, hazard function, ...).

(K.5) Moreover, it would be interesting to propose a test to check if the conditional law or the conditional hazard function have a specific nature.

(K.6) Finally, there exist few results on the estimation of the density function of functional random variables with regard to a known measure. Consequently, one actually needs further theoretical advances to consider the interesting question of testing the nature of the density function.

(K.7) Finally, it would be interesting to adapt our results to the case of a functional random response. There exist few results in this context. Developing kernel methods for such regression models is an important challenge for the future.

Dependence assumptions

We have chosen in Part II of this dissertation to state our results under α-mixing assumptions. However, the results given in Part III have been obtained for independent datasets.

(D.1) In Part III of this manuscript, we use a central limit theorem for U-statistics. There exist central limit theorems for U-statistics constructed from β-mixing variables (see for instance Kashimov, 1993, Borovkova *et al.*, 2001, or Harel and Elharfaoui, 2003). Consequently, it seems that the results given in Part III may be generalized to the case of a β-mixing dataset.

(D.2) New mixing coefficients have been proposed by Doukhan and Louhichi (1999) and Dedecker and Prieur (2005). Some of the results we have used in the context of α-mixing sequences have been generalized to these new mixing coefficients. Even if these mixing assumptions, defined for real random variables, seem not directly usable, one can imagine to generalize our results to such mixing assumptions adapted to the functional case.

(D.3) Finally, there exist few results (see Hedli-Griche, 2008) adapted to the case of long memory dependence. In general the results are of different nature. However, the question of asymptotic results and structural testing procedures in the case of long memory dependence are interesting challenges for the futur.

Bootstrap

In Chapter 11, we propose and study from a practical point of view various bootstrap procedures for estimating the critical value of the test. Bootstrap methods may also be useful in other situations.

(B.1) We introduce in Chapter 6 a method to estimate asymptotic pointwise confidence bands. We propose to assume that the bias is negligible and estimate variance terms with kernel estimators. However, a more accuracy estimation method, for instance by bootstrap methods, may lead to better results.

(B.2) It would be interesting to provide a theoretical justification of the use of bootstrap (to estimate the critical value) in the context of structural testing procedures.

(B.3) Finally, to generalize the structural testing procedures to the case of a β-mixing dataset, we will be faced with the problem of bootstrapping dependent

variables. Some results have been proposed in this context (see for instance Politis and Romano, 1994).

Choice of h_n and d

In practice, one has to choose the value of the smoothing parameter h_n and the semimetric d. For applications presented in this dissertation we have chosen the smoothing parameter by cross-validation. We have observed that leads to fairly good results. Concerning the choice of the semimetric, we have chosen the semimetric from the knowledge we have on the data. However some of the results given in this dissertation opens interesting prospects in terms of the choice of the smoothing parameter and the semimetric.

(C.1) We now consider two semimetrics d_0 and d such that $Hol_{d_0} \subset Hol_d$ (i.e. a function is Hölderian for d_0 if it is Hölderian for d). We may imagine to use structural testing procedures presented in Part III to test the regularity of the regression operator r with regard to d_0 against its regularity with regard to d.

(C.2) We provide in Chapter 7 the asymptotic explicit expression of the $\mathbb{L}^q$ errors. This opens interesting propects in terms of the asymptotic optimal choice of the smoothing parameter. Bootstrap methods might be used to estimate the constants involved in the expression of $\mathbb{L}^q$ errors. One can also imagine to use plug-in methods.

Tests

We introduce in Chapter 10 a theoretical framework to construct general testing procedures. We then propose in Chapter 11 various bootstrap procedures and compare their empirical level and power in the case of a no effect test.

(T.1) To complete our simulation studies, a deep simulation study of the empirical level and power of other testing procedures constructed from the ideas of Part III should be done.

(T.2) Let us now consider the problem of testing if the effect of two functional random variables (X_1, X_2) is indeed reduced to the effect of one of them. Theorem 10.3.2 given in Chapter 10 allows to make this kind of test but may require in some cases to split the original dataset into samples of various sizes. It would be interesting to study the empirical power and level of such tests on a simulated dataset.

(T.3) Recently, Lavergne and Patilea (2007) have proposed a new way to test for no effect or dimension reduction in the multivariate case. Their approach consists in a modification of standard test statistics improving the power of the corresponding test. It would be interesting to try to adapt their method to our testing procedures in the functional context.

(T.4) We have introduced in Part III the test statistic T_n^*. It would be interesting to introduce other test statistics (see for instance Härdle and Mammen, 1993, Zhang et Dette, 2004) and compare their properties.

(T.5) Finally, in Part III we split the dataset into independent datasets. We can imagine to study the possibility of using the whole dataset to compute the

test statistic T_n^*, to estimate r_0^* and to approximate the integral.

Other modelizations and applications

We have studied in this dissertation a regression model in which there is a real response variable and a functional explanatory variable. In our simulation studies and applications we have considered functional data corresponding to curves. Let us discuss other situations.

(O.1) The results given in this dissertation, and more generally most functional kernel methods, are adapted to the study of functional variables taking values in an infinite dimensional semimetric (or semi-normed) space (curves, images, surfaces, ...). They form a theoretical background allowing to study samples of images (or more complex functional data). However, there is a deep lack of practical developments, in particular concerning the construction and implementation of specific semimetrics. An important challenge for the futur is to construct such semimetrics in order to be able to apply our results in the context of images study.

(O.2) In the same way, one can imagine to use the results given in Part II to study a sequence of dependent images.

Some of the previous propects ((K.1), (K.2), (D.1), (B.1), (C.1), and (T.1)) may be obtained fairly easily under some technical assumptions and considerations. Other prospects ((K.3)-(K.5), (D.2), (B.2), (C.2), and (T.2)-(T.4)) seem reachable in the forthcoming years but require more attention and work. Finally, prospects ((K.6), (K.7), (D.3), (B.3), (T.5), (O.1), and (O.2)) need important theoretical or practical advances and constitute actual challenges for the futur.

PARTIE II

ESTIMATION DE L'OPÉRATEUR DE RÉGRESSION

In this part we present results dealing with asymptotic pointwise convergence properties of the kernel estimator of the regression operator. We firstly state asymptotic normality results in Chapter 6. Then, in Chapter 7 we get from these results explicit expressions of the asymptotic dominant terms of centered moments and $\mathbb{L}^q$ errors. Chapter 7 is written in French but an English version of its main result is given in Chapter 8. Chapter 8 sums up the main results of Chapters 6 and 7. Finally, additional comments on the assumptions of Chapters 6 may be found in Chapter 9 in which an other application example is also presented.

CHAPITRE 6

ADVANCES ON ASYMPTOTIC NORMALITY IN NONPARAMETRIC FUNCTIONAL TIME SERIES ANALYSIS

The content of the present chapter is the one of the paper Delsol (2008b) accepted for publication in Statistics. Some of the results presented here have also been stated in the short note, Delsol (2007a) (see Chapter 8, Part II). Moreover, some of the results given in the present chapter are used in the next one.

Abstract : We consider a stationary process and desire to predict future values from previous ones. Instead of considering the process through its discretized form, we choose to see it as a sample of dependent curves. Then, we cut the process into N successive curves. Obviously, the N curves are not independent. The prediction issue can be translated into a nonparametric functional regression problem from dependent functional variables. This paper aims to revisit and complete two recent works on this topic. This article extends recent literature and provides asymptotic law with explicit constants under α-mixing assumptions. Then we establish pointwise confidence bands for the regression function. To conclude, we present how our results behave on a simulation and on a real time series.

6.1. Introduction

Since the monograph by Ramsay and Silverman (2002) and (2005), statistics for functional data know a great infatuation. Statistical modelling for functional variables and functional data analysis are multi-facetted questions as attested by the wide scope of methods discussed in Davidian *et al.* (2004) or Gonzalez-Manteiga and Vieu (2006). Recently, the monograph by Ferraty and Vieu (2006) presents the state of the art in nonparametric statistics for functional data. This paper aims to give advances in nonparametric functional time series analysis. Typically, the goal of such issues is to forecast time series future values from the past ones. In other words, one considers a real stationary process $(Z_t)_{t\in T}$ and wants to estimate Z_{t+u}, $u > 0$ from the value of $Z_s, s < t$. For instance, one considers the sea temperature during fifty four years (see Figure 6.1), and aims to predict temperature for the next year.

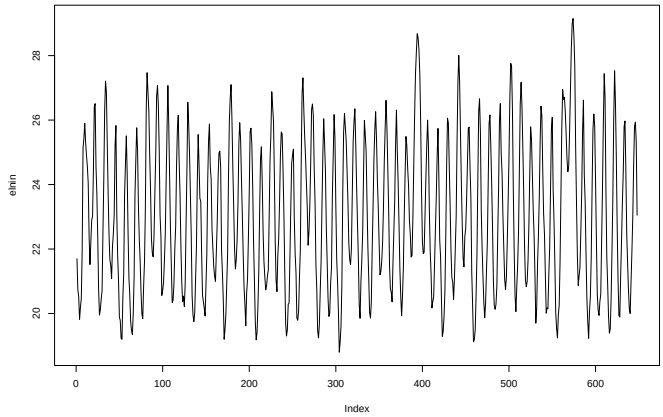

FIGURE 6.1. Sea temperature near EL NINO

For many years, statisticians have studied the following discretized model :

$$Z_i = r\left(Z_{i-1}, Z_{i-2}, \ldots, Z_{i-p}\right) + \epsilon_i,$$

where r cannot be indexed by a finite number of real parameters. However, such nonparametric regression models do not reflect reality in the sense that the thinner discretization is or the more past values we take into account (i.e. p increases), the less results are relevant. Indeed Stone (1982) showed that the 'curse of dimensionality' deteriorates optimal convergence rates of nonparametric estimators of r.

To avoid this phenomenon, we introduce a functional approach. We propose to consider Z_t no longer through its discretized version $(Z_i)_{i\in\mathbb{Z}}$ but as a continuous process $(Z_t)_{t\in[0,\tau N]}$. Then, we cut the process into N functional dependent variables defined on $[0,\tau]$ by $\xi_i(x) := Z_{i\tau+x}$. This idea was initially introduced by Bosq (1991) (see the literature therein for more details). Going on with our first example, we obtain fifty four yearly curves gathered in Figure 6.2.

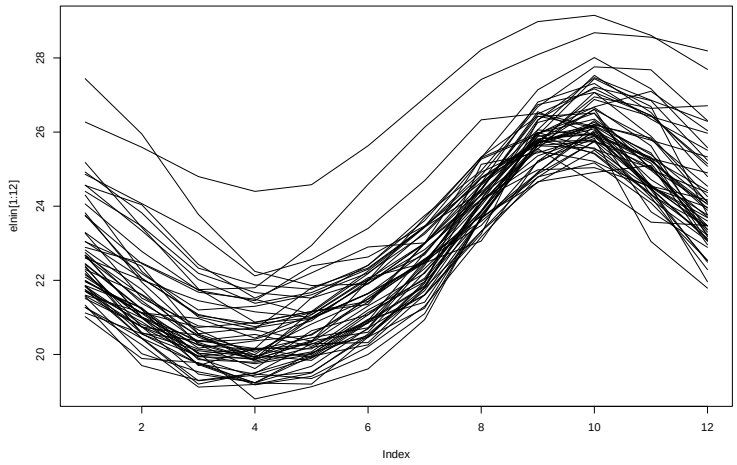

FIGURE 6.2. Yearly sea temperature near EL NINO

In this context, a regression model taking into account a whole trajectory in the past can be written as :

$$G(\xi_i) = r(\xi_{i-1}) + \epsilon_i,$$

where G is a known real-valued operator defined on $C^0([0,\tau])$. Consequently, time series analysis may be viewed as a functional regression problem with dependent data. In addition, we will show that taking into account the time series as functional variable leads us to a method insensitive to any change with respect to the frequency of the observation (i.e. the grid of the discretization). In this sense, one can say that functional methods avoid the so-called 'curse of dimensionality'.

This kind of model has been widely studied when r is assumed linear (see Bosq, 2000, for larger references). Now, when general regularity constraints are required for r (i.e. the nonparametric setting), first results can be found in Ferraty *et al.* (2002b). More recently, Masry (2005) showed the asymptotic normality of the kernel estimator with dependent data. However, these results are not easy to apply since the constants involved are not explicit. At the same time, Ferraty *et al.* (2007) have given exact expressions of variance and expectation for kernel estimators, which led them to a central limit theorem with explicit constants in the independent case.

To take into account the dependence in our results, we suppose the data are α-mixing. In other words, we assume that when $|i-j|$ tends to infinity, ξ_i and ξ_j become roughly independent. This most general case of weakly dependent variables has been widely studied notably by Rio (1995 and 2000), Bosq (1996b) and Liebscher (2001). In addition, their advances are useful to extend results given for independent data to the dependent context.

The aim of this paper is to explicit the constants involved in the asymptotic law of $\hat{r}(x)$. This will make possible real data application. To do that, we revisit and complete previous results on asymptotic normality and $\mathbb{L}^2$ error given by Masry (2005) and Ferraty *et al.* (2007). The first section gathers main results dealing with asymptotic normality of the kernel estimator. The first one is given under general α-mixing assumptions whereas a corollary clarifies assumptions in algebraic and geometric cases. The second part shows applications of the first one. We highlight the dependence influence on confidence bands efficiency through a simulation and conclude with an application to the previous sea temperature example. Finally, in order to make the reading easier, we have gathered all the proofs of this article in the last section.

6.2. Nonparametric regression for dependent functional data

6.2.1. Regression model for dependent functional data. — We consider n pairs $(X_i, Y_i)_{i=1..n}$ identically distributed as (X,Y) such that the sequence $(X_i, Y_i)_i$ is α-mixing. We also assume X takes values on a semi-metric space $(\mathbf{E}, d)$ whereas Y is a real random variable. In all Section 2, we focus on the following model :

$$Y_i = r(X_i) + \epsilon_i,$$

in which ϵ_i is a real random variable satisfying $\mathbb{E}\left[\epsilon_i^2/X_i\right] = \sigma_\epsilon^2(X_i) < +\infty$ and $\mathbb{E}[\epsilon_i/X_i] = 0$. We will give results for the functional kernel estimator introduced by Ferraty and Vieu (2006) :

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

This estimator is an extension of the Nadaraya-Watson estimator when the explanatory random variable is a functional one, in which K is a kernel function and h_n a real smoothing parameter.

6.2.2. Notations and Main Assumptions. — We suppose that x is a fixed element of the functional space E and we define some notations :

$$\phi(s) = \mathbb{E}\left[(r(X) - r(x))\,|d(X,x) = s\right],$$
$$F(h) = \mathbb{P}(d(X,x) \leq h) \text{ and } \tau_h(s) = \frac{F(hs)}{F(h)}.$$

We start with some assumptions on the model.

(6.1) r is bounded on a neighbourhood of x,

(6.2) $F(0) = 0$, $\phi(0) = 0$ and $\phi'(0)$ exists,

(6.3) $\sigma_\epsilon^2(.)$ is continuous on a neighbourhood of x and $\sigma_\epsilon^2 := \sigma_\epsilon^2(x) > 0$,

(6.4) $\forall s \in [0;1]$, $\lim_{n\to+\infty} \tau_{h_n}(s) = \tau_0(s)$ with $\tau_0(s) \neq 1_{[0;1]}(s)$.

Let us make some remarks on previous assumptions. Firstly, no assumption is made on the density of the variable X. All the hypothesis concerning the distribution of X are expressed in terms of small ball probabilities through the function F. Then, (6.1)-(6.3) are very standard (see Ferraty *et al.*, 2007, or Masry, 2005) while (6.4) was introduced by Ferraty *et al.* (2007) in the independent case to explicit constants. Besides, (6.4) holds when X corresponds to standard processes (see again Ferraty *et al.*, 2007).
Now, to control the covariances we propose the following assumptions on conditional moments :

(6.5) $\exists p > 2$, $\exists M > 0$, $\mathbb{E}\left[|\epsilon|^p|X\right] \leq M \quad a.s.$,

(6.6) $max\left(\mathbb{E}\left[|Y_iY_j||X_i, X_j\right], \mathbb{E}\left[|Y_i||X_i, X_j\right]\right) \leq M$ $a.s.$ $\forall i, j \in \mathbb{Z}$.

However, assumption (6.5) is always used to obtain that

$$\mathbb{E}\left[\left|Y_i K\left(\frac{d(x,X_i)}{h_n}\right)\right|^s\right] \leq M\mathbb{E}\left[\left|K\left(\frac{d(x,X_i)}{h_n}\right)\right|^s\right],\ \forall s \leq p.$$

As we have $K\left(\frac{d(x,X_i)}{h_n}\right)$ in factor and because this term is null if $d(x,X_i) > h_n$ we only need local assumption on the conditional moment. Consequently, all the results can be established if one of the next hypothesis held instead of (6.5) :

– $\exists p > 2,\ \exists M > 0,\ \mathbb{E}\left[|\epsilon|^p 1_{d(X,x)\leq h_n}|X\right] \leq M$ a.s.,
– $\exists p > 2,\ u(x) := \mathbb{E}\left[|\epsilon|^p|X = x\right]$ is bounded on a neighbourhood of x.

Finally, we make assumptions on the kernel functional estimator :

$$h_n = O\left(\frac{1}{\sqrt{nF(h_n)}}\right),\ \lim_{n\to+\infty} nF(h_n) = +\infty, \tag{6.7}$$

$$K \text{ has a compact support, is } C^1 \text{ and unincreasing on }]0;1[\ ,\ K(1) > 0. \tag{6.8}$$

It is worth noting that assumption (6.7) is fulfilled when the smoothing parameters sequence is asymptotically optimal with regard to the $\mathbb{L}^2$ error. Indeed, optimal smoothing parameters have to balance the bias and variance terms. Hence they have the following properties : $h_n \sim C\sqrt{nF(h_n)}^{-1}$ and $h_n \to 0$. Our results will be expressed using the following constants :

$$\begin{aligned}
M_0 &= \left(K(1) - \int_0^1 (sK(s))'\tau_0(s)\,ds\right),\\
M_1 &= \left(K(1) - \int_0^1 K'(s)\tau_0(s)\,ds\right),\\
M_2 &= \left(K^2(1) - \int_0^1 \left(K^2\right)'(s)\tau_0(s)\,ds\right).
\end{aligned}$$

What differs from the independent case studied by Ferraty *et al.* (2007) is that we need additional assumptions to control the dependence between variables. We assume that the n pairs (X_i, Y_i) are α-mixing with the usual definition of α-mixing coefficients introduced by Rosenblatt (1956) and given by Rio (1995, p37, Definition 2, Notation 1.2). We introduce some notations :

$$\begin{aligned}
&\Theta(s) := \max\left(\max_{i\neq j} P\left(d(X_i,x)\leq s,\ d(X_j,x)\leq s\right), F^2(s)\right),\\
&\Gamma_i := Y_i K\left(\frac{d(X_i,x)}{h_n}\right),\ \Delta_i := K\left(\frac{d(X_i,x)}{h_n}\right), U_{i,n} = \frac{\Gamma_i\mathbb{E}[\Delta_i] - \Delta_i\mathbb{E}[\Gamma_i]}{F(h_n)\sqrt{nF(h_n)}},
\end{aligned}$$

and the general technical assumptions (H1) and (H2) below. Because it is the key of our results, we firstly choose to express them in a very general way :

$$\exists (u_n)_{n\in\mathbb{N}} \in \mathbb{N}^{\mathbb{N}},\ O\left(\frac{n\left[\alpha(u_n)\right]^{\frac{p-2}{p}}}{F(h_n)^{\frac{p-2}{p}}}\right) + O\left(u_n\frac{\Theta(h_n)}{F(h_n)}\right) \overset{n\to+\infty}{\to} 0, \tag{H1}$$

and

$$I_n = n\int_0^1 \alpha^{-1}\left(\frac{x}{2}\right) Q^2_{U_{1,n}}(x)\, min\left(\frac{3M_1\sqrt{\sigma_\epsilon^2 M_2}}{2}, \alpha^{-1}\left(\frac{x}{2}\right) Q_{U_{1,n}}(x)\right) dx \to 0, \text{(H2)}$$

where $Q_{U_{i,n}}(x) = \inf\{t,\, \mathbb{P}(|U_{i,n}| > t) \le x\}$, $\alpha^{-1}\left(\frac{x}{2}\right) = \inf\left\{t,\, \alpha([t]) \le \frac{x}{2}\right\}$.

To fix ideas, some special cases of processes for which they are fulfilled will be presented later in Section 2.4. Assumption (H1) makes the covariance term negligible with respect to the variance terms. That enables us to compute exact expression of dominant terms of the variance. Assumption (H2) is a version of Lindeberg's condition for α-mixing random variables and allows us to state the asymptotic normality of our estimator. In the independent case both previous assumptions vanish since they are always fulfilled (indeed, in this case $\alpha(i) \equiv 0$ for all $i > 0$, so (H2) is fairely easy to obtain and (H1) is fulfilled with $u_n = 1$ since $\Theta(h_n) = F^2(h_n)$).

6.2.3. The main result : Asymptotic Normality. —

Theorem 6.2.1. — *Asymptotic Normality.*
Under assumptions (6.1)-(6.8), (H1) and (H2), we have

$$\frac{M_1}{\sqrt{M_2\sigma_\epsilon^2}}\sqrt{n\hat{F}(h_n)}\left(\hat{r}(x) - r(x) - B_n\right) \to \mathcal{N}(0,1)$$

where $B_n = h_n\phi'(0)\frac{M_0}{M_1}$ *and* $\hat{F}(t) = \frac{1}{n}\sum_{i=1}^n 1_{[d(X_i,x),+\infty[}(t)$.

Démonstration. — Let us first give the main lemmas which present the frame of the proof. To do that, we introduce the notations :

$$\hat{f}(x) := \frac{1}{nF(h_n)}\sum_{i=1}^n \Delta_i \text{ and } \hat{g}(x) := \frac{1}{nF(h_n)}\sum_{i=1}^n \Gamma_i.$$

The following lemmas will be proved in the appendix.

Lemma 6.2.1. — *Under assumptions (6.2),(6.5) and (6.8) we have :*

$$\frac{\mathbb{E}[\hat{g}(x)]}{\mathbb{E}\left[\hat{f}(x)\right]} - r(x) = h_n\phi'(0)\frac{M_0}{M_1} + o(h_n).$$

Lemma 6.2.2. — *Under assumptions (6.1)-(6.8) and (H1) we have :*

$$\sum_{1\le i\ne j\le n} |Cov(\Gamma_i, \Gamma_j)| = o\left(\sum_{i=1}^n Var(\Gamma_i)\right).$$

Lemma 6.2.3. — *Under assumptions (6.1)-(6.8) and (H1) we have :*

$$\sum_{1\le i\ne j\le n} |Cov(\Delta_i, \Delta_j)| = o\left(\sum_{i=1}^n Var(\Delta_i)\right)$$

$$and \sum_{1\leq i\neq j\leq n} |Cov\left(\Gamma_i, \Delta_j\right)| = o\left(\frac{1}{nF\left(h_n\right)}\right).$$

Lemma 6.2.4. — *If (6.1)-(6.8) and* $(H1)$ *hold then we have :*

$$\begin{aligned} Var\left(\hat{g}\left(x\right)\right) &= \left(\sigma_\epsilon^2 + r^2\left(x\right)\right)\frac{M_2}{nF\left(h_n\right)}\left(1+o\left(1\right)\right), \\ Var\left(\hat{f}\left(x\right)\right) &= \frac{M_2}{nF\left(h_n\right)}\left(1+o\left(1\right)\right), \\ Cov\left(\hat{f}\left(x\right), \hat{g}\left(x\right)\right) &= \frac{M_2}{nF\left(h_n\right)}\left(r\left(x\right)+o\left(1\right)\right). \end{aligned}$$

Lemma 6.2.5. — *Under assumptions (6.1)-(6.8),* $(H1)$ *and* $(H2)$ *we have the following results :*
$\hat{f}\left(x\right) \rightharpoonup M_1$, $\mathbb{E}\left[\hat{f}\left(x\right)\right] \to M_1$, $\frac{\hat{F}(h_n)}{F(h_n)} \rightharpoonup 1$, *where* $\rightharpoonup$ *stands for convergence in law.*

From the lemmas and using the following decomposition :

$$\begin{aligned} &\sqrt{nF\left(h_n\right)}\left(\hat{r}\left(x\right) - r\left(x\right) - B_n\right) \\ &= \sqrt{nF\left(h_n\right)}\left(\frac{\left(\hat{g}\left(x\right) - \mathbb{E}\left[\hat{g}\left(x\right)\right]\right)\mathbb{E}\left[\hat{f}\left(x\right)\right] - \left(\hat{f}\left(x\right) - \mathbb{E}\left[\hat{f}\left(x\right)\right]\right)\mathbb{E}\left[\hat{g}\left(x\right)\right]}{\hat{f}\left(x\right)\mathbb{E}\left[\hat{f}\left(x\right)\right]} + o\left(h_n\right)\right) \\ &= \frac{\sum_{i=1}^n U_{i,n}}{\hat{f}\left(x\right)\mathbb{E}\left[\hat{f}\left(x\right)\right]} + o\left(1\right), \end{aligned}$$

we get directly the proof of Theorem 6.2.1 with the Corollary 1 of Rio (1995). Lemmas 6.2.1-6.2.4 ensure the assumptions of Corollary 1 hold for $U_{i,n}$ while Lemma 6.2.5 allows to treat the denominator.

□

Remark. Assumption (H2) is intricate but (6.5) gives a necessary condition. Indeed, if (6.5) holds then $Q_{i,n}\left(x\right)$ may be bounded in the following way :

$$Q_{U_{i,n}}\left(x\right) \leq \frac{CF^{\frac{1}{q}}\left(h_n\right)}{x^{\frac{1}{q}}\sqrt{nF\left(h_n\right)}}, \forall\ 1 \leq q \leq p.$$

Furthermore α^{-1} may be written as follows :

$$\alpha^{-1}\left(x\right) := \sum_{i=0}^{n-1} 1_{]-\infty;\alpha(i)[}\left(x\right) = \sum_{i=0}^{n-1}\left(i+1\right)1_{[\alpha(i+1);\alpha(i)[}\left(x\right)$$

Consequently, with Fubini-Tonelli, (H2) holds as soon as there exists a sequence $\left(u_n\right)_n$, $1 \leq q_1 \leq p$ and $2 < q_2 \leq p$, such that

$$I_{1,n} \to 0, I_{2,n} \to 0,$$

where

$$I_{1,n} := \frac{F^{\frac{3-q_1}{q_1}}(h_n)}{\sqrt{nF(h_n)}} \sum_{i=0}^{u_n} (i+1)^2 \int_{2\alpha(i+1)}^{2\alpha(i)} x^{-\frac{3}{q_1}} dx,$$

and

$$I_{2,n} := F^{\frac{2-q_2}{q_2}}(h_n) \sum_{i=u_n+1}^{n-1} (i+1) \int_{2\alpha(i+1)}^{2\alpha(i)} x^{-\frac{2}{q_2}} dx.$$

The arithmetic case

The result given here for arithmetic α-mixing variables is not a trivial consequence of Theorem 6.2.1. Some assumptions may be precised and proofs are based on thinner arguments. Let us specify $(H1)$ and $(H2)$ in the case of arithmetic α-mixing coefficients, that is when $\alpha(n) \leq Cn^{-a}$. In this particular case, we can write a simplified version of Theorem 6.2.1 by introducing the assumptions :

$$\exists \nu > 0,\ \Theta(h_n) = O\left(F(h_n)^{1+\nu}\right), \text{ with } a > \frac{(1+\nu)p - 2}{\nu(p-2)}, \tag{6.9}$$

$$\exists \gamma > 0,\ nF^{1+\gamma}(h_n) \to +\infty \text{ and } a > max\left(\frac{4}{\gamma}, \frac{p}{p-2} + \frac{2(p-1)}{\gamma(p-2)}\right). \tag{6.10}$$

Theorem 6.2.2. — *Asymptotic Normality for arithmetic mixing process. Under assumptions (6.1)-(6.10), we have :*

$$\frac{M_1}{\sqrt{M_2\sigma_\epsilon^2}} \sqrt{n\hat{F}(h_n)} \left(\hat{r}(x) - r(x) - B_n\right) \to \mathcal{N}(0,1).$$

Démonstration. — The proof of Theorem 6.2.2 is based on the both following lemmas (their proofs being deferred in the appendix).

Lemma 6.2.6. — *If (6.9) holds then* $\sum_{1\leq i\neq j\leq n} |Cov(\Gamma_i, \Gamma_j)| = o\left(\sum_{i=1}^n Var(\Gamma_i)\right)$.

Lemma 6.2.7. — *Under assumption (6.5) and (6.10), (H2) holds.*

Note that the proof of Theorem 6.2.2 differs from the one of Theorem 6.2.1 only in the fact that Lemma 6.2.6 enables to control covariances without (H1).

□

Remark : Using a result given by Liebscher (2001), we may have a more general assumption in place of (6.10) :

$$\exists (v_n)_n \in \mathbb{N}^{\mathbb{N}}\ v_n = o\left(\sqrt{nF(h_n)}\right) \text{ and } (n/F(h_n))^{\frac{1}{2}} v_n^{-a} \to 0. \tag{6.11}$$

Moreover, in the arithmetic α-mixing case one may choose $v_n = \sqrt{nF(h_n)}(\log(n))^{-1}$. Then assumption (6.11) holds as soon as there exists $\gamma > 0$ such that $n(F(h_n))^{1+\gamma} \to +\infty$ and $a > \frac{2}{\gamma} + 1$.

Remarks :

– It is worth noting geometrical α-mixing coefficients, that is when $\alpha(n) \leq C\rho^n$, are arithmetic ones for any a. Indeed, for any a, ρ and for n large enough, we have $\rho^n \leq n^{-a}$. Thus, in the geometric case, assumptions (6.9) and (6.10) are always fulfilled as soon as there exist ν, $\gamma > 0$ such that $nF^{1+\gamma}(h_n) \to +\infty$ and $\Theta(h_n) = O\left(F^{1+\nu}(h_n)\right)$.

– A usual application of asymptotic normality is for providing confidence bands. In order to avoid the difficult estimation of the bias term we may assume that $h_n\sqrt{nF(h_n)} \to 0$ and derive asymptotic pointwise confidence bands from the following kernel estimators of the constants M_1, M_2 :

$$\begin{aligned}\hat{M}_2(x) &:= \frac{1}{n\hat{F}(h_n)}\sum_{i=1}^{n} K^2\left(\frac{d(X_i,x)}{h_n}\right),\\ \hat{M}_1(x) &:= \frac{1}{n\hat{F}(h_n)}\sum_{i=1}^{n} K\left(\frac{d(X_i,x)}{h_n}\right),\end{aligned}$$

and any estimator $\hat{\sigma_\epsilon^2}$ that converges in law to σ_ϵ^2 as for instance :

$$\hat{\sigma_\epsilon^2} := \frac{\sum_{i=1}^n Y_i^2 K\left(\frac{d(X_i,x)}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{d(X_i,x)}{h_n}\right)} - \left(\hat{r}(x)\right)^2.$$

– Under additional assumptions :

(6.12) $\exists M,\ \max_i \mathbb{E}\left[\epsilon_i^2 |\ X_1,\ldots,X_n\right] \leq M\ a.s.,$

$$p > 4,\ \exists \nu_3,\nu_4 > 0,\ \Theta_k(s) = O\left(F^{1+\nu_k}(s)\right),\ k = 3,4,\ \nu_4 \leq \nu_3 + 1 \leq \nu + 2$$

(6.13) and $a > \max\left(3\frac{p}{p-4} + \frac{3}{\nu_4}, 2\frac{p}{p-4} + \frac{2}{\nu_3}, \frac{p}{p-4} + \frac{1}{\nu}\right),$

where Θ_k is defined by

$$\Theta_k(s) := \max\left(\max_{1\leq i_1<\cdots<i_k\leq n} P\left(d\left(X_{i_j},x\right) \leq s,\ 1 \leq j \leq k\right), F^k(s)\right),$$

one can obtain the explicit expression of the dominant terms in $\mathbb{L}^2$ error of the kernel estimator :

$$\mathbb{E}\left[\left(\hat{r}(x) - r(x)\right)^2\right] = \left(\phi'(0)\frac{M_0}{M_1}\right)^2 h_n^2 + \frac{1}{nF(h_n)}\frac{M_2\sigma_\epsilon^2}{M_1^2} + o\left(\frac{1}{nF(h_n)}\right).$$

Moreover, uniform integrability arguments may be used to give from the explicit expression of the asymptotic law the asymptotic expressions of $\mathbb{L}^q$ errors for $q \in \mathbb{N}$. (see Delsol, 2007a, for more details)

6.3. Application to Time Series Forecasting

As explained in the introduction, many issues on Time Series may be viewed as nonparametric regression problems with dependent data. One can find more explanations of what such an approach enables to do in Ferraty *et al.* (2002b) or Ferraty

and Vieu (2006). In this part we will apply the results of the previous section on a simulation study and the El Nino data presented before.

6.3.1. Some simulations. — We will show how our estimator behaves in independent, m-dependent and α-mixing contexts through asymptotic normality and confidence bands. To do that we simulate 100 variables $w_{0,i}$ uniformly distributed on $\left[0, \frac{\pi}{4}\right]$. Then we create for every $w_{0,i}$ a vector $\left(X_{0,i}\left(t_j\right)\right)_{j\in\{1,\ldots,100\}}$ (see Figure) that is the discretized version of the function $t \mapsto cos(w_{0,i} + \pi\left(2t-1\right))$ on $[0,1]$:

$$X_0[i,j] = cos(w_{0,i} + \pi\left(\frac{2j}{100} - 1\right)).$$

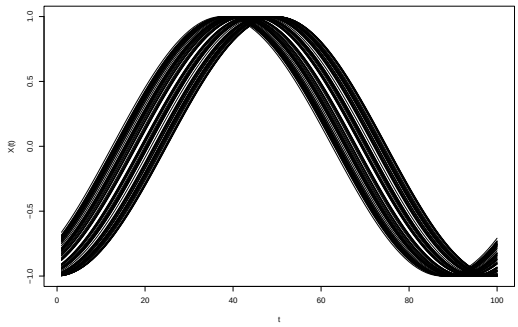

FIGURE 6.3. Set of simulated random curves

In this simulation r is defined from $C^1\left(\left[0, \frac{\pi}{2}\right]\right)$ into $\mathbb{R}$ by

$$r\left(f\right) := \frac{1}{2\pi}\int_{\frac{1}{2}}^{\frac{3}{4}} \left(f'\left(t\right)\right)^2 dt \ ,$$

that is why we take the semi-metric d based on the first order of differentiation. Then we repeat 100 times the following sequence :

1. We simulate 400 w_i uniformly distributed on $[0, \pi/4]$ and their corresponding $X\left[i,j\right]$,

2. We simulate 400 independent, m-dependent or α-mixing centered gaussian errors with variance 0.01. To make 9-dependent errors e_i from independent gaussian variables η_i, we take $e_i = \frac{1}{\sqrt{10}}\sum_{i=1}^{10}\eta_i$. For α-mixing errors we define recursively $e_i = \frac{1}{\sqrt{2}}\left(e_{i-1} + \eta_i\right)$,

3. For each curve $X_{0,i},\ 1 \leq i \leq 100$ we take $x = X_{0,i}$ and compute

$$\frac{\hat{M}_1}{\sqrt{\hat{\sigma}_\epsilon^2 \hat{M}_2}}\sqrt{n\hat{F}\left(h_n\right)}\left(\hat{r}\left(x\right) - r\left(x\right)\right),$$

using h_n given by cross validation (the algorithm used is the one given by Ferraty and Vieu (2006)).

When the loop ends, for each curve $X_{0,i}$, we have 100 estimations, loaded in the vector E_i. To observe the asymptotic normality of our estimator, we plot with R, for each i, the estimated density of the vector E_i. We then obtain the Figure 6.4 in which appears that the more dependence is strong, the less the density obtained is a standard gaussian one. Indeed, it appears clearly that variance grows with the dependency. We also compute the percentage P of points falling into asymptotic confidence bands (with h_n given by cross validation). As a consequence of the loss of normality when dependency grows, the number of points in asymptotic confidence bands decreases.

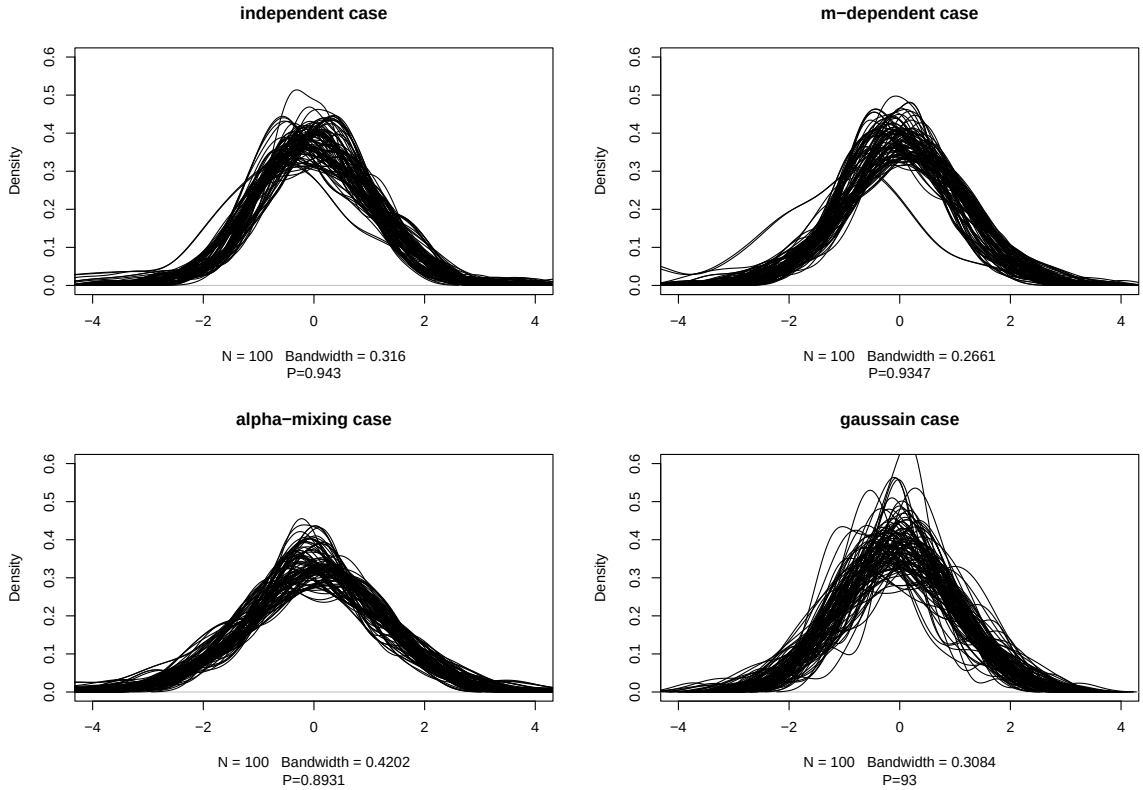

FIGURE 6.4. Normality in independent, m-dependent and alpha-mixing case
and number of points in the confidence bands of level 0.95

However, we can have better results with h_n different from the one provided by cross-validation (that is what the figure 6.5 displays). This is concordant with the way we construct asymptotic confidence bands. We assume that the smoothing parameter makes the bias term asymptotically negligible with regard to the variance one. Hence we suppose the smoothing parameter is asymptotically smaller than the one given by cross validation criterion. We call *hopt* the better choice of h_n from the family $\{0.05, 0.1, ..., 2\}$ (that is to say the value of the smoothing parameter for which the percentage of points in confidence bands is the nearest from 0.95) and *hcv*

the one given by the cross validation criterion. The solid and dotted lines represent the value obtained with *hopt* and *hcv* respectively.

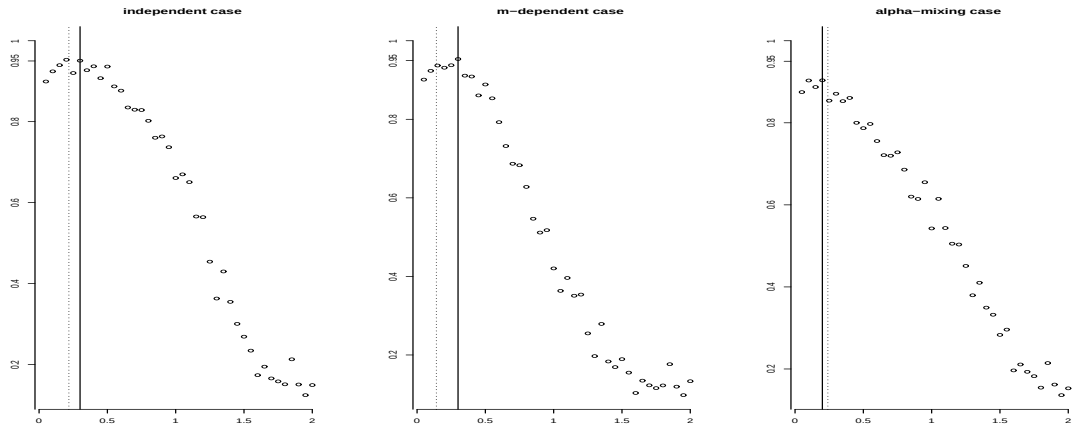

FIGURE 6.5. Percentages of points in asymptotic confidence bands relatively to h_n in independent, m-dependent and alpha-mixing case

Moreover, Table 1 gathers some important values of the study. The values given in the last columns correspond to the best percentage P_{opt} of points in the confidence bands obtained with *hopt* and the percentage obtained with *hcv*.

Table 1

	hopt	*hcv*	P_{opt}	P_{cv}
Independent case	0.3	0.21	0.9507	0.9359
m-dependent case	0.3	0.1417	0.9534	0.9286
α-mixing case	0.2	0.2401	0.9035	0.8832

These figures show that the more dependence is strong the less percentages stay near 0.95 for h_n small. For α-mixing case, percentages always decrease whereas in other cases they stay constant before decreasing. However, the cross-validation criterion provides fairly good results. Because of the well behaviour of the method for a bandwidth selected by cross-validation and since we do not have at hand other criteria, we will use this cross-validation's procedure in the next study concerning El Nino.

We have also studied how confidence bands behave when the quotient

$$Q := \frac{Var\,(error)}{Var\,(r\,(X))}$$

changes. It appeared that interval's length grows with Q. Besides, we have also observed that performances of our confidence bands do not seem to depend on Q. Indeed, after many studies, no common structure appeared. It could be related to the fact that error variance appears in confidence bands expression. The fact that

intervals length grows with Q permit to limit the loss of accuracy coming from increasing errors.

6.3.2. Application to El Nino. — The aim of this part is to apply our estimator on real functional data. Clearly, we are not able to show if our confidence bands provide good results since r remains unknown. We will give some graphics obtained with our routines implemented with R in order to give an idea on how our results behave from a practical point of view. We will study the El Nino dataset which is available on-line on the web site $http://kdd.ics.uci.edu/databases/el\_nino/elnino.data.html$. Previous studies of these data in a functional nonparametric prediction setting using conditional features can be found in Ferraty and Vieu (2006).

Our dataset is composed of sea temperature curves. We have monthly measures covering 54 years. In order to study this time series, one cuts data as in Figure 6.2 and obtains 54 curves. Each curve corresponds to the temperature evolution during one year. In order to study the 54^{th} year, we have a dataset of 53 other curves. We split them into a learning set composed of the 52 first years and a test set containing the 53^{th} year. We estimate each of the 12 values of the 54^{th} year and show our results in the next figure. Because the curves are not smooth, we take the PCA metric. The kernel is the quadratic one whereas the smoothing parameter is the one obtained by cross-validation criterion. A first survey consists in plotting the predicted and the real 54^{th} year (see Figure 6.6). The dotted line stands for the real data where the solid one is the prediction.

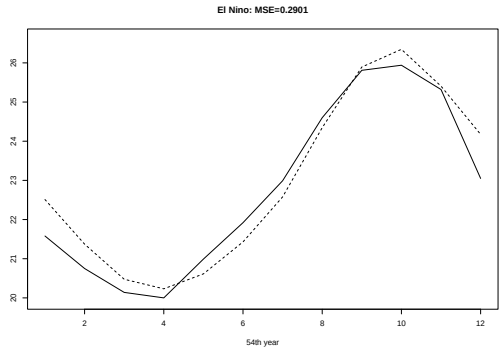

FIGURE 6.6. Prediction of the 54^{th} year

However, our results enable to go deeper and give confidence bands for $r(x)$. Figure 6.7 presents the prediction and two confidence bands of level 0.2 and 0.05 respectively. It is important to note that these are pointwise confidence bands.

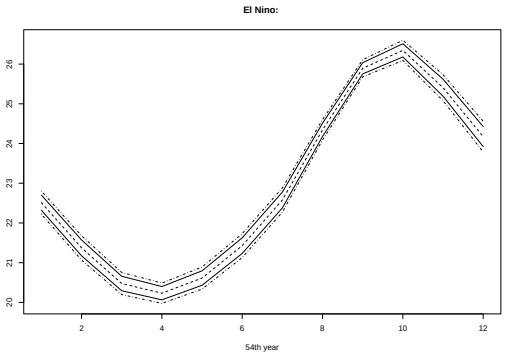

FIGURE 6.7. Confidence bands for sea temperature during the 54^{th} year

In order to highlight both nonparametric and functional interests of the proposed method, here is a short comparison study with functional parametric and multivariate methods. More precisely, we are interested in comparing the considered functional nonparametric estimators with functional linear (functional and parametric), or multivariate additive (multivariate and "nonparametric") ones. By simplicity we use the same notations as in the introduction part, denote $(\xi_i)_{1\leq i\leq 54}$ the consecutive annual curves and $(Z_i)_{1\leq i\leq 648}$ monthly temperature measures. For each month k we note G_k the operator that associates to any annual curve its value during the month k. Then we are interested in estimators constructed for the following functional nonparametric model :

$$G_k(\xi_i) = r(\xi_{i-1}) + \epsilon_i,$$

a version of the functional linear model introduced by Ramsay and Silverman (1991) :

$$G_k(\xi_i) =< \xi_{i-1}, \alpha > +\epsilon_i,$$

and a version of the multivariate generalised additive model proposed by Hastie and Tibshirani (1986) :

$$Z_{12i+k} = \sum_{j=1}^{12} r_j\left(Z_{12(i-1)+j}\right) + \epsilon_i.$$

The first two models are functional that means that they are adapted to functional variables. The functional linear model is parametric in the sense that it depends on a parameter function α. The generalised additive model (g.a.m.) has been introduced to avoid the curse of dimensionality for multavariate random variables. Each function r_j has a nonparametric structure that is why g.a.m. may be viewed as a "nonparametric" model. For each method we compute the estimator from the first fifty three yearly curves $(\xi_i)_{1\leq i\leq 53}$ (or the corresponding monthly measures $(Z_i)_{1\leq i\leq 636}$) and try to predict the twelve values of the fifty fourth year from the data collected during the previous year. We use the algorithms proposed by Ferraty and Vieu (2006) for nonparametric functional model, presented by Cardot *et al* (2006) for the functional linear model, and the package "gam", written by Hastie in 2006, for the multivariate additive model. These methods are compared through their mean squared error (gathered in Table 2).

Table 2

	Func. Nonparametric	Func. Linear	Multiv. Additive
M.S.E.	0.2901	0.6742	0.8932

This short comparison study reflects the interest of functional and nonparametric methods for time series analysis. However, if the link between two consecutive years is linear, one may expect the functional linear estimator to have better properties.

6.3.3. How to choose the semi-metric, the banwidth and the kernel in practice. — The nonparametric functional approach presented in this paper depends on three parameters : the semimetric d, the bandwidth h_n and the kernel function K. In this paragraph we discuss the way to choose them in practice. See the monograph by Ferraty and Vieu (2006) for a deeper discussion on this topic.

The metric choice is important in the multivariate case where all metrics are equivalent. In the functional case that is no longer true hence the metric choice is more crucial. Moreover in addition to the usual functional metrics it is worth considering semimetrics because it may be an alternative to the curse of dimensionality, it enables to take into account more general situations and it may be more relevant when observed curves are very smooth. See the monograph by Ferraty and Vieu (2006) for a deeper discussion on some arguments to choose the semimetric from the curves nature. For instance, when the observed curves are smooth, it may be interesting to use semimetrics based on derivatives (see for instance the spectrometric dataset studied in Ferraty and Vieu, 2006, p.106). In other cases, when the curves are not smooth, it may be useful to consider projection semimetrics (based for instance on functional principal components, first coefficients in Fourrier's decomposition, ...) to avoid the curse of dimensionality. Despite these considerations, it is allways possible to consider a family of semimetrics and choose the one that is the best adapted to the considered dataset by cross-validation. Theoretical justification of the usefulness of a particular semimetric is still an open problem.

The most popular way to choose the smoothing parameter from the dataset is to take the one given by cross-validation criterion : h_{CV}. The choice of the optimal smoothing parameter in functional regression by cross-validation criterion has been addressed in a theoretical way in the recent paper by Rachdi and Vieu (2007). To avoid the estimation of the bias term we have assumed the smoothing parameter to be small enough to make the bias term negligible with regard to the variance one. That is why we obtain in simulations that the best smoothing parameter seems to be smaller than h_{CV}. However, simulations made in Section 3.1 show that the results obtained with h_{CV} are fairly good. Moreover, there is no automatic method to choose the optimal bandwidth in practice that is why we use h_{CV}.

As in the real case there exists various kinds of kernel functions. The main difference is that in the functional case we consider kernel functions with compact

support $[0;1]$ such that $K(1) > 0$. Most standard kernel functions are the restriction of classical indicator, triangular, quadratic or gaussian kernels to the set $[0;1]$. The choice of the kernel function is linked we the smoothness of the operator we want to estimate. On the other hand, in the particular case of the indicator function of $[0;1]$ the constants M_1 and M_2 are equal to one and hence we do not have to estimate them.

6.3.4. Conclusion and prospects. — The new approach presented in this article enables to complete the results obtained in Masry (2005) in the sense that we explicit the various constants involved in the asymptotic law of the kernel estimator. As a consequence, we are able to give and estimate confidence bands. The previous simulation highlights that dependence deteriorates confidence bands results.
In order to motivate further researches, we would like to emphasize some interesting open questions. One can imagine to use other methods to estimate the various unknown constants : bootstrap, other estimators, ... Then, even if the cross-validation criterion gives good results, one could try to provide other ways to choose h_n. One may finally imagine to obtain the same results for the integrated error and construct some specification testing procedures.

6.4. Appendix : Proofs

In all our proofs C stands for any arbitrary positive constant.

Proof of Lemma 6.2.1 :
This lemma is the same as the one obtained in Ferraty *et al.* (2007) in the independent case. Indeed, the proof lies on the asymptotic expressions of $\mathbb{E}[\Delta_i]$ and $\mathbb{E}[\Gamma_i]$ with the constants M_0, M_1 and M_2 and these quantities are not affected by the dependence structure.

Proof of Lemma 6.2.2 :
To bound the sum of covariances, one introduces a real sequence (u_n) and split the sum as follows :

$$\sum_{i \neq j} |Cov(\Gamma_i, \Gamma_j)| = A + B, \tag{6.14}$$

where $A = \sum_{1 \leq |i-j| \leq u_n} |Cov(\Gamma_i, \Gamma_j)|$ and $B = \sum_{u_n < |i-j| \leq n} |Cov(\Gamma_i, \Gamma_j)|$.
Concerning A :

$$A \leq \sum_{1 \leq |i-j| \leq u_n} \left(\mathbb{E}[|\Gamma_i \Gamma_j|] + \mathbb{E}[\Gamma_i]^2 \right).$$

On the one hand, taking conditional expectation, it comes :

$$\mathbb{E}[|\Gamma_i \Gamma_j|] = \mathbb{E}\left[\mathbb{E}[|Y_i Y_j| \, | X_i, X_j] \, K\left(\frac{d(X_i, x)}{h_n} \right) K\left(\frac{d(X_j, x)}{h_n} \right) \right],$$

and (6.6) gives

$$\begin{aligned}\mathbb{E}\left[|\Gamma_i\Gamma_j|\right] &\leq C\mathbb{E}\left[K\left(\frac{d\left(X_i,x\right)}{h_n}\right)K\left(\frac{d\left(X_j,x\right)}{h_n}\right)\right]\\ &\leq CK^2\left(0\right)P\left(d\left(X_i,x\right)\leq h_n,\, d\left(X_j,x\right)\leq h_n\right).\end{aligned}$$

On the other hand, conditioning by X_i allows us to get :

$$\begin{aligned}\mathbb{E}\left[\Gamma_i\right] &= \mathbb{E}\left[Y_iK\left(\frac{d\left(X_i,x\right)}{h_n}\right)\right]\\ &= \mathbb{E}\left[r\left(X_i\right)K\left(\frac{d\left(X_i,x\right)}{h_n}\right)\right].\end{aligned}$$

Furthermore, (6.1) and (6.8) give $\left|\mathbb{E}\left[\Gamma_i\right]\right|\leq CF\left(h\right)$ where C is a positive constant. Thus, we obtain the following bound for A :

$$\begin{aligned}A &\leq Cnu_n\left(F^2\left(h_n\right)+max_{1\leq|i-j|\leq u_n}P\left(d\left(X_i,x\right)\leq h_n,\, d\left(X_j,x\right)\leq h_n\right)\right)\\ &\leq Cnu_n\Theta\left(h_n\right). \qquad (6.15)\end{aligned}$$

Now, let us establish an upper bound of B. Since (6.8) ensures K to be continuous on $]0;1[$ and have a compact support, K belongs to $L^\infty\left[0,1\right]$. Moreover, conditionally on X_i, ϵ_i belongs to L^p thus Y_i and Γ_i does by (6.1). An application of an inequality of Rio (2000) on α-mixing variables Γ_i gives :

$$\left|Cov\left(\Gamma_i,\Gamma_j\right)\right|\leq C\left[\alpha\left(\left|i-j\right|\right)\right]^{\frac{p-2}{p}}F\left(h_n\right)^{\frac{2}{p}},$$

and it comes

$$\begin{aligned}B &= \sum_{u_n<|i-j|\leq n}\left|Cov\left(\Gamma_i,\Gamma_j\right)\right|\\ &\leq Cn\left(n-u_n\right)\left[\alpha\left(u_n\right)\right]^{\frac{p-2}{p}}F\left(h_n\right)^{\frac{2}{p}}\\ &\leq Cn^2\left[\alpha\left(u_n\right)\right]^{\frac{p-2}{p}}F\left(h_n\right)^{\frac{2}{p}}. \qquad (6.16)\end{aligned}$$

Then, we get from (6.14), (6.15) and (6.16) that :

$$\frac{1}{n^2F^2\left(h_n\right)}\sum_{i\neq j}\left|Cov\left(\Gamma_i,\Gamma_j\right)\right|\leq C\left[\frac{\left[\alpha\left(u_n\right)\right]^{\frac{p-2}{p}}F\left(h_n\right)^{\frac{2}{p}}}{F^2\left(h_n\right)}+\frac{u_n}{n}\frac{\Theta\left(h_n\right)}{F^2\left(h_n\right)}\right].$$

Since $\sum_{i=1}^n Var\left(\Gamma_i\right)=nF\left(h_n\right)M_2\left(\sigma_\epsilon^2+r^2\left(x\right)\right)\left(1+o\left(1\right)\right)$ (see Ferraty *et al.*, 2007, for more details), assumption $(H1)$ ends the proof.

□

Proof of Lemma 6.2.3 :
The first part of this lemma is an application of the previous one with $Y_i\equiv 1$.To prove the second statement of the lemma, we bound $\mathbb{E}\left[|\Gamma_i\Delta_j|\right]$ using (6.6),

$$\mathbb{E}\left[|Y_i|\,K\left(\frac{d\left(X_i,x\right)}{h_n}\right)K\left(\frac{d\left(X_j,x\right)}{h_n}\right)\right]\leq C\mathbb{E}\left[K\left(\frac{d\left(X_i,x\right)}{h_n}\right)K\left(\frac{d\left(X_j,x\right)}{h_n}\right)\right],$$

and conclude with the same arguments as in the proof of Lemma 6.2.2.

□

Proof of Lemma 6.2.4 :
This proof comes directly from Lemmas 6.2.2, 6.2.3 and expressions of variances given in the independent case by Ferraty *et al.* (2007).

□

Proof of Lemma 6.2.5 :
Lemma 6.2.4 gives $Var\left(\hat{f}(x)\right) \to 0$ and hence $\hat{f}(x) - \mathbb{E}\left[\hat{f}(x)\right] \to 0$ in probability. Besides, Ferraty *et al.* (2007) provide an asymptotic expression of $\mathbb{E}\left[\hat{f}(x)\right]$:

$$\mathbb{E}\left[\hat{f}(x)\right] = M_1\left(1 + o(1)\right).$$

This is enough to prove the first part of this lemma. The second part is just a special case of the first one when the kernel K is taken to be uniform (i.e. $K \equiv 1_{[0;1]}$).

□

Proof of Lemma 6.2.6 :
For all $\eta > 0$, we have :

$$\begin{aligned} B &\leq CnF(h_n)^{\frac{2}{p}} u_n^{-\left(a\frac{p-2}{p}-1-\eta\right)} \sum_{u_n<k\leq n} k^{a\frac{p-2}{p}-1-\eta} k^{-a\frac{p-2}{p}} \\ &\leq C\frac{nF(h_n)^{\frac{2}{p}}}{u_n^{\left(a\frac{p-2}{p}-1-\eta\right)}}. \end{aligned}$$

Consequently, $\sum_{1\leq i\neq j\leq n} |Cov(\Gamma_i, \Gamma_j)| = o\left(\sum_{i=1}^n Var(\Gamma_i)\right)$ as soon as

$$\frac{F(h_n)^{\frac{2-p}{p}}}{u_n^{\left(a\frac{p-2}{p}-1-\eta\right)}} + u_n\frac{\Theta(h_n)}{F(h_n)} \to 0.$$

One takes

$$u_n = \left[\frac{F(h_n)^{\frac{2}{a(p-2)-\eta p}}}{\Theta(h_n)^{\frac{p}{a(p-2)-\eta p}}}\right],$$

where [ ] denotes the function integer part. If we replace u_n by its expression, we obtain

$$u_n\frac{\Theta(h_n)}{F(h_n)} \leq CF(h_n)^{\frac{2}{a(p-2)-\eta p}-1}\Theta(h_n)^{1-\frac{p}{a(p-2)-\eta p}} \leq O(1)F(h_n)^{\frac{\nu(a(p-2)-p-\eta p)-(p-2)}{a(p-2)-\eta p}}.$$

To conclude, one uses assumption (6.9) and takes η small enough to make the exponent positive.

□

Proof of Lemma 6.2.7 :
Without loss of generality one may assume that $\alpha(i) = Ci^{-a}$. Moreover, one can obtain an upper bound for $Q_{X_{i,n}}(x)$. Indeed, (6.1), (6.8) and (6.5) ensure $\sqrt{nF(h_n)}U_{i,n}$ is in $\mathbb{L}^p$. Consequently, with Bienaime-Tchebitchef's inequality, (6.5), and (6.8) it comes :

$$Q_{U_{i,n}}(x) \leq \frac{CF^{\frac{1}{q}}(h_n)}{x^{\frac{1}{q}}\sqrt{nF(h_n)}}, \forall\ 1 \leq q \leq p.$$

Consequently, since $\alpha^{-1}\left(\frac{x}{2}\right) = \sum_{i=0}^{n-1}(i+1)\,1_{[2\alpha(i+1);2\alpha(i)[}(x)$, I_n may be bounded with

$$n\int_0^1 \sum_{i=0}^{n-1} 1_{[2\alpha(i+1);2\alpha(i)[}(x)\,(i+1)\left(\frac{CF^{\frac{1}{q_i}}(h_n)}{x^{\frac{1}{q_i}}\sqrt{nF(h_n)}}\right)^2 \inf\left(1,(i+1)\frac{CF^{\frac{1}{q_i}}(h_n)}{x^{\frac{1}{q_i}}\sqrt{nF(h_n)}}\right)dx$$
$$= C\sum_{i=0}^{n-1}(i+1)\,F^{\frac{2-q_i}{q_i}}(h_n)\int_{2\alpha(i+1)}^{2\alpha(i)} x^{-\frac{2}{q_i}}\inf\left(1,(i+1)\frac{CF^{\frac{1}{q_i}}(h_n)}{x^{\frac{1}{q_i}}\sqrt{nF(h_n)}}\right)dx,$$

by applying Fubini-Tonelli's theorem. The idea of the proof is to bound I_n in the following way, taking $q_i = q_1$, $i \in \{0,...,u_n\}$ and $q_i = q_2$, $i \in \{u_n+1,...,n-1\}$, we obtain :

$$\begin{aligned} I_n &\leq C\Big(F^{\frac{2-q_2}{q_2}}(h_n)\sum_{i=u_n+1}^{n-1}(i+1)\int_{2\alpha(i+1)}^{2\alpha(i)} x^{-\frac{2}{q_2}}dx \\ &\qquad + \frac{F^{\frac{3-q_1}{q_1}}(h_n)}{\sqrt{nF(h_n)}}\sum_{i=0}^{u_n}(i+1)^2\int_{2\alpha(i+1)}^{2\alpha(i)} x^{-\frac{3}{q_1}}dx\Big) \\ &\leq C\left(I_{1,n}+I_{2,n}\right), \end{aligned}$$

and then to choose q_1, q_2 and u_n such that $I_{j,n} \to 0$, $j = 1,2$. Let us first choose $q_2 = p$. Since $a(p-2) > p$ by (6.9), we have :

$$\begin{aligned} I_{1,n} &= CF^{\frac{2-p}{p}}(h_n)\sum_{i=u_n+1}^{n-1}(i+1)\left[x^{\frac{p-2}{p}}\right]_{2\alpha(i+1)}^{2\alpha(i)} \\ &= CF^{\frac{2-p}{p}}(h_n)\left(\sum_{i=u_n+1}^{n-1}\left[2\alpha(i)\right]^{\frac{p-2}{p}} + (u_n+2)\left[2\alpha(u_n+1)\right]^{\frac{p-2}{p}}\right) \\ &\leq CF^{\frac{2-p}{p}}(h_n)\,u_n^{1+\eta-\frac{a(p-2)}{p}}, \end{aligned}$$

where $\eta > 0$ such that $1+\eta-\frac{a(p-2)}{p} < 0$. The other term may be bounded as follows :

$$\begin{aligned} I_{2,n} &\leq C\frac{F^{\frac{3-q_1}{q_1}}(h_n)}{\sqrt{nF(h_n)}} sign(q_1-3)\sum_{i=0}^{u_n}(i+1)^2\left((\alpha(i))^{\frac{q_1-3}{q_1}}-(\alpha(i+1))^{\frac{q_1-3}{q_1}}\right) \\ &\leq C\frac{F^{\frac{3-q_1}{q_1}}(h_n)}{\sqrt{nF(h_n)}} sign(q_1-3)\Big(\sum_{i=1}^{u_n}(2i+1)(\alpha(i))^{\frac{q_1-3}{q_1}}+(\alpha(0))^{\frac{q_1-3}{q_1}} \\ &- (u_n+1)^2(\alpha(u_n+1))^{\frac{q_1-3}{q_1}}\Big) \\ &\leq C\frac{F^{\frac{3-q_1}{q_1}}(h_n)}{\sqrt{nF(h_n)}} u_n^{\left(2+\beta-\frac{a(q_1-3)}{q_1}\right)_+}, \end{aligned}$$

with $\beta > 0$, $1 \leq q_1 \leq p$, $q_1 \neq 3$, and where $(\ .\ )_+$ denotes the positive part. It is possible to obtain the same bound if $q_1 = 3$ because $\lim_{x\to+\infty}\frac{ln(x)}{x^\beta} = 0$, for all $\beta > 0$. Now it remains to choose q_1 and u_n. We have to consider two different cases :

– $p < 3$ or $p > 3$ and $a \leq \frac{2p}{p-3}$: taking $q_1 = p$ and

$$u_n = \left[(F(h_n))^{-\frac{p-2}{a(p-2)-(1+\eta)p}}\left[n(F(h_n))^{1+\frac{2(p(1+\beta-\eta)-1+3\eta-2\beta)}{a(p-2)-(1+\eta)p}}\right]^{\frac{1}{4\left(2+\beta-\frac{a(q_1-3)}{q_1}\right)}}\right],$$

it comes

$$\begin{aligned} I_{1,n} &\leq C\left[n(F(h_n))^{1+\frac{2(p(1+\beta-\eta)-1+3\eta-2\beta)}{a(p-2)-(1+\eta)p}}\right]^{\frac{1+\eta-\frac{a(p-2)}{p}}{4\left(2+\beta-\frac{a(p-3)}{p}\right)}} \to 0, \\ I_{2,n} &\leq C\frac{F^{\frac{3-p}{p}}(h_n)}{\sqrt{nF(h_n)}} u_n^{\left(2+\beta-\frac{a(p-3)}{p}\right)} \leq C\left[n(F(h_n))^{1+\frac{2(p(1+\beta-\eta)-1+3\eta-2\beta)}{a(p-2)-(1+\eta)p}}\right]^{-\frac{1}{4}} \to 0, \end{aligned}$$

for η and β small enough since $\gamma > \frac{2(p-1)}{a(p-2)-p}$.

– $p > 3$ and $a > \frac{2p}{p-3}$: taking $q_1 = \frac{3a}{a-2}$, and

$$u_n = \left[(F(h_n))^{-\frac{p-2}{a(p-2)-(1+\eta)p}}\frac{1}{o(1)}\right],$$

it comes

$$\begin{aligned} I_{1,n} &\leq o(1), \\ I_{2,n} &\leq C\frac{1}{\sqrt{nF^{1+\frac{4}{a}}(h_n)}} \to 0. \end{aligned}$$

□

ANNEX : Proof of Corollary 6.2.2 under assumption (6.11) :
When the α-mixing coefficients are arithmetic one may get a more general proof of the Corollary 6.2.2 using the Theorem 2.1. of Liebscher (2001) to the array of α-mixing variables $(U_{i,n})_{i=1,...,n}$. Firstly, using assumptions (6.5) and (6.8) and the fact that the variables are identically distributed, one gets

$$\begin{aligned} W_n &= \max_{i=1,...,n} \mathbb{E}\left[U_{i,n}^2\right] \\ &= \mathbb{E}\left[U_{i,n}^2\right] \\ &= \frac{1}{nF(h_n)}\mathbb{E}\left[\left(\Gamma_i \frac{\mathbb{E}[\Delta_i]}{F(h_n)} - \Delta_i \frac{\mathbb{E}[\Gamma_i]}{F(h_n)}\right)^2\right] \\ &\leq C\frac{1}{nF(h_n)}\mathbb{E}\left[\Gamma_i^2 + \Delta_i^2 + 2|\Gamma_i\Delta_i|\right] \\ &\leq C\frac{1}{nF(h_n)}\mathbb{E}\left[K^2\left(\frac{d(X,x)}{h_n}\right)\right] \\ &\leq C\frac{1}{n}. \end{aligned}$$

Then one gets with the same arguments and assumption (6.6) :

$$\begin{aligned} \gamma_n &= \max_{j,k,\,1\leq j,k\leq n,\,j\neq k} \left(\mathbb{E}[|U_{j,n}U_{k,n}|] + \mathbb{E}[|U_{j,n}|]\,\mathbb{E}[|U_{k,n}|]\right) \\ &\leq C \max_{j,k,\,1\leq j,k\leq n,\,j\neq k} \Big(\frac{\mathbb{E}[|\Gamma_{j,n}\Gamma_{k,n}|] + \mathbb{E}[|\Gamma_{j,n}\Delta_{k,n}|] + \mathbb{E}[|\Delta_{j,n}\Gamma_{k,n}|] + \mathbb{E}[|\Delta_{j,n}\Delta_{k,n}|]}{nF(h_n)} \\ &\quad + \frac{\left(\mathbb{E}[|\Gamma_{j,n}|] + \mathbb{E}[|\Delta_{j,n}|]\right)^2}{nF(h_n)}\Big) \\ &\leq C\frac{1}{nF(h_n)} \max_{j,k,\,1\leq j,k\leq n,\,j\neq k} \left(\mathbb{E}[|\Delta_{j,n}\Delta_{k,n}|] + \left(\mathbb{E}[|\Delta_{j,n}|]\right)^2\right) \\ &\leq C\frac{1}{nF(h_n)} \max_{j,k,\,1\leq j,k\leq n,\,j\neq k} \left(\mathbb{P}\left(d(X_j,x,)\leq h_n, d(X_k,x)\leq h_n\right) + F^2(h_n)\right) \\ &\leq C\frac{\Theta(h_n)}{nF(h_n)} \\ &\leq C\frac{F(h_n)^{\nu}}{n}. \end{aligned}$$

Besides it is easy to see that assumption (6.9) implies that $a > \frac{p}{p-2}$ and hence we get :

$$\sum_{k=1}^{\infty} k^{\frac{2}{p-2}}\alpha(k) < +\infty.$$

Then assumptions (6.5) and (6.8) enable to give an upper bound for the $\mathbb{L}^p$ norm of $U_{i,n}$:

$$\begin{aligned}
\left(\mathbb{E}\left[|U_{i,n}|^p\right]\right)^{\frac{1}{p}} &\leq \frac{1}{\sqrt{nF\left(h_n\right)}}\left(\mathbb{E}\left[\sum_{k=0}^{p} C_p^k \left|\Gamma_i \frac{\mathbb{E}\left[\Delta_i\right]}{F\left(h_n\right)}\right|^k \left|\Delta_i \frac{\mathbb{E}\left[\Gamma_i\right]}{F\left(h_n\right)}\right|^{p-k}\right]\right)^{\frac{1}{p}} \\
&\leq \frac{C}{\sqrt{nF\left(h_n\right)}}\left(\sum_{k=0}^{p} C_p^k \mathbb{E}\left[\left|\Gamma_i\right|^k \left|\Delta_i\right|^{p-k}\right]\right)^{\frac{1}{p}} \\
&\leq \frac{C}{\sqrt{nF\left(h_n\right)}}\left(\sum_{k=0}^{p} C_p^k \mathbb{E}\left[\left|\Delta_i\right|^p\right]\right)^{\frac{1}{p}} \\
&\leq \frac{C}{\sqrt{nF\left(h_n\right)}}\left(F\left(h_n\right)\right)^{\frac{1}{p}}
\end{aligned}$$

One takes, for $\eta > 0$, $m_n = \left[\left(F\left(h_n\right)\right)^{-\frac{\nu}{2}-\frac{p-2}{2(a(p-2)-\eta(p-2)-p)}}\right]$ where $[\,.\,]$ represents the function interger part. It is easy to see that $m_n \to +\infty$ since $F\left(h_n\right) \to 0$. Then one gets assumption (2.1) of Theorem 2.1 of Liebscher :

$$\begin{aligned}
nm_n\gamma_n &\leq Cn\left(F\left(h_n\right)\right)^{-\frac{\nu}{2}-\frac{p-2}{2(a(p-2)-\eta(p-2)-p)}} \frac{F\left(h_n\right)^{\nu}}{n} \\
&\leq C\left(F\left(h_n\right)\right)^{+\frac{\nu}{2}-\frac{p-2}{2(a(p-2)-\eta(p-2)-p)}} \\
&\leq C\left(F\left(h_n\right)\right)^{\frac{(p-2)\left(\nu\left(a-\eta-\frac{p}{p-2}\right)-1\right)}{2(a(p-2)-\eta(p-2)-p)}} \\
&\leq o\left(1\right).
\end{aligned}$$

The last line comes from assumption (6.9) for η small enough.

Moreover, one also gets assumption (2.2) of Theorem 2.1 of Liebscher :

$$\begin{aligned}
&\left(\sum_{j=m_n+1}^{+\infty} j^{\frac{2}{p-2}}\alpha\left(j\right)\right)^{1-\frac{2}{p}} \sum_{i=1}^{n}\left(\mathbb{E}\left[|U_{i,n}|^p\right]\right)^{\frac{2}{p}} \\
\leq\ & C\left(\sum_{j=m_n+1}^{+\infty} j^{-1-\eta} j^{\frac{p}{p-2}+\eta-a}\right)^{1-\frac{2}{p}} n\frac{\left(F\left(h_n\right)\right)^{\frac{2}{p}}}{nF\left(h_n\right)} \\
\leq\ & C\left(m_n^{\frac{p}{p-2}+\eta-a}\right)^{1-\frac{2}{p}}\left(F\left(h_n\right)\right)^{\frac{2}{p}-1} \\
\leq\ & C\left(\left(F\left(h_n\right)\right)^{\left(\frac{p}{p-2}+\eta-a\right)\left(-\frac{\nu}{2}-\frac{p-2}{2(a(p-2)-\eta(p-2)-p)}\right)-1}\right)^{1-\frac{2}{p}} \\
\leq\ & C\left(\left(F\left(h_n\right)\right)^{-\frac{1}{2}\left(1+\left(\frac{p}{p-2}+\eta-a\right)\nu\right)}\right)^{1-\frac{2}{p}} \\
\leq\ & o\left(1\right).
\end{aligned}$$

The last line comes from assumption (6.9) for η small enough. Consequently the array $\left(u_{i,n}\right)_{1\leq i\leq n}$ fulfills condition $C\left(p\right)$ of Theorem 2.1 of Liebscher.

Afterwards one may see that using Hölder and Markov inequalities it comes :

$$\begin{aligned}\sum_{i=1}^{n}\mathbb{E}\left[U_{i,n}^{2}1_{|U_{i,n}|>\frac{1}{\tau_n}}\right] &\leq n\left\|U_{1,n}\right\|_{p}^{2}\left(\mathbb{P}\left(|U_{i,n}|>\frac{1}{\tau_n}\right)\right)^{1-\frac{2}{p}}\\ &\leq C\frac{n\left(F\left(h_n\right)\right)^{\frac{2}{p}}}{\left(\sqrt{nF\left(h_n\right)}\right)^{2}}\left(\frac{\tau_n^{p}F\left(h_n\right)}{\sqrt{nF\left(h_n\right)}^{p}}\right)^{1-\frac{2}{p}}\\ &\leq C\left(\frac{\tau_n}{\sqrt{nF\left(h_n\right)}}\right)^{p-2}.\end{aligned}$$

Besides if (6.11) holds, that is to say if there exists a sequence of positive integers v_n such that $v_n = o\left(\sqrt{nF\left(h_n\right)}\right)$ and $\left(n/F\left(h_n\right)\right)^{\frac{1}{2}}v_n^{-a}\to 0$, and if we now take $\tau_n = v_n^{\frac{a}{a+1}}\sqrt{nF\left(h_n\right)}^{\frac{1}{a+1}}$ then we get assumption (2.3) of Theorem 2.1 of Liebscher :

$$\left(\frac{\tau_n}{\sqrt{nF\left(h_n\right)}}\right)^{p-2} = o\left(1\right).$$

Furthermore, since $\alpha\left(i\right)\leq Ci^{-a}$ one gets assumption (2.4) of Theorem 2.1 of Liebscher :

$$\begin{aligned}\forall\epsilon_0>0,\ \frac{n}{\tau_n}\alpha\left(\left[\epsilon_0\tau_n\right]\right) &= C\epsilon_0^{-a}\left(v_n\right)^{-a}\frac{n}{\sqrt{nF\left(h_n\right)}}\\ &\leq C\epsilon_0^{-a}\sqrt{\frac{n}{F\left(h_n\right)}}v_n^{-a}\\ &\leq o\left(1\right),\end{aligned}$$

what notably implies that τ_n tends to infinity. We may also note that $\tau_n = o\left(\sqrt{n}\right)$ since we know that $v_n = o\left(\sqrt{nF\left(h_n\right)}\right)$.

With the result of Ferraty *et al.* (2007) we conclude easily that assumption (2.5) of the Theorem 2.1 of Liebscher holds with $\sigma^2 = M_1^2M_2\sigma_\epsilon^2$.

We finish our proof applying the Theorem 2.1 of Liebscher to the variables $\left(U_{i,n}\right)_{1\leq i\leq n}$.

CHAPITRE 7

RÉGRESSION NON-PARAMÉTRIQUE FONCTIONNELLE : EXPRESSIONS ASYMPTOTIQUES DES MOMENTS CENTRÉS ET DES ERREURS $\mathbb{L}^q$

The content of this chapter is the one of the paper Delsol (2007b) that has been published in the "Annales de l'ISUP" (Annals of the Statistical Institute of the University of Paris). This chapter is written in French. However, some of the results presented in this chapter have also been published in the short English note Delsol (2007a) (see Chapter 8, Part II). Moreover, some of the results given in the present chapter use results given in Chapter 6.

Abstract : Recent works on functional nonparametric regression explicit the asymptotic law of the kernel estimator of the regression function. This article aims to get from these results asymptotic expressions of the moments, which lead to asymptotic expressions of $\mathbb{L}^q$ errors of this estimator. The results provided are available for both independent and α-mixing functional datasets and may be easily applied to the classical multivariate case.

Résumé : Dans la littérature récente on trouve des résultats concernant la loi asymptotique explicite de l'estimateur à noyau de la fonction de régression. Cet article a pour objectif d'établir à l'aide de ceux-ci l'expression asymptotique des moments de cet estimateur. Ces résultats sont établis dans le cas où la variable explicative est fonctionnelle mais peuvent tout aussi bien s'appliquer au cas où elle est vectorielle. Les résultats contenus dans ce papier pourront être utilisés dans l'optique du choix du paramètre de lissage optimal pour l'erreur $\mathbb{L}^q$.

7.1. Introduction

Avec les progrès des appareils de mesure, le statisticien dispose souvent de données mesurées sur des grilles de plus en plus fines les rendant intrinsèquement fonctionnelles. C'est pourquoi une nouvelle branche des statistiques traitant de données fonctionnelles a vu le jour. Cet article s'inscrit dans cette démarche en s'intéressant plus particulièrement au problème de la régression par rapport à une variable explicative fonctionnelle. Ce sujet a déjà été largement étudié dans le cas linéaire (voir [366] en

ce qui concerne le modèle linéaire fonctionnel et [46] par rapport à l'étude de processus linéaires pour plus de références). Dans le cas non-paramétrique (c'est à dire lorsque l'on fait seulement des hypothèses de régularité sur la fonction de régression) les premiers résultats proviennent de [164]. Plus récemment, [314] puis [167] et [134] ont montré la normalité asymptotique de l'estimateur à noyau de la fonction de régression ; les deux derniers donnant une expression explicite des termes de biais et de variance. Cet article a pour objectif de compléter l'éventail de ces propriétés asymptotiques en exprimant les termes asymptotiquement dominants des moments centrés et des erreurs $\mathbb{L}^q$ de l'estimateur à noyau. Ces résultats innovants dans le cas fonctionnel le sont également dans le cas vectoriel puisque si à ma connaissance le risque $\mathbb{L}^2$ a été abondamment étudié (notamment par [265] et [209]), les seuls travaux explicitant le risque $\mathbb{L}^1$ sont établis par [421] dans le cas réel et "fixed design". En plus de leur utilisation comme simples outils de calcul, ils peuvent se révéler particulièrement intéressants dans l'optique du choix optimal du paramètre de lissage par rapport aux erreurs $\mathbb{L}^q$.

Nous allons étudier tout au long de cet article le modèle de régression usuel :

$$Y = r(X) + \epsilon, \tag{7.1}$$

où Y est une variable aléatoire réelle et X une variable aléatoire à valeurs dans l'espace fonctionnel semi-métrique (E, d). On ne fait aucune hypothèse sur la forme de l'opérateur r si ce n'est des hypothèses de régularité, c'est pourquoi notre modèle est dit non-paramétrique. Enfin, on considère un échantillon constitué de n paires (X_i, Y_i) qui suivent la même loi que (X, Y). Si dans un premier temps nous supposerons que ces paires sont indépendantes, nous nous attacherons par la suite à généraliser notre approche au cas α-mélangeant. Notre motivation à généraliser nos résultats à ce type de dépendance provient notamment de la volonté de pouvoir les utiliser dans l'étude de séries temporelles à travers l'approche introduite par [39].

On considère un élément x de l'espace E et on estime $r(x)$ par l'estimateur à noyau suivant :

$$\hat{r}(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{d(X_i,x)}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{d(X_i,x)}{h_n}\right)},$$

où h_n et K seront précisés ultérieurement. Cet estimateur a été introduit et abondamment étudié dans [176] comme une généralisation de celui de Nadaraya-Watson au cas de variables explicatives fonctionnelles.

Le paragraphe 2 rassemble des résultats généraux concernant la convergence de certains moments de cet estimateur obtenue à partir de résultats d'uniforme intégrabilité et de convergence en loi. De ces résultats on pourra déduire dans le paragraphe 3 des résultats plus importants en terme de retombée statistique, en particulier le Théorème 7.3.2 qui donne les expressions asymptotiques des erreurs $\mathbb{L}^q$ de l'estimateur à noyau étudié. On explicitera les termes dominants de ces erreurs lorsque q est entier. Enfin on étudiera succintement le cas vectoriel standard pour lequel certaines simplifications sont à noter. Par souci de lisibilité et de clarté,

les démonstrations des résultats des deux sections précédentes seront données en annexe.

7.2. Résultats généraux de convergence des moments

Dans un premier temps, par souci de clarté, nous allons nous restreindre au cas où les paires (X_i, Y_i) sont indépendantes. Nous verrons ensuite qu'en rajoutant certaines hypothèses il est possible de généraliser les résultats au cas où elles sont α-mélangeantes. Cela peut être utile pour faire de la prévision dans le cadre de séries temporelles (voir Partie IV de [176]).

7.2.1. Le cas indépendant. —

7.2.1.1. Notations et principales hypothèses. — Dans ce paragraphe nous reprenons les notations et les hypothèses introduites dans [167] qui nous seront utiles dans notre étude. Pour commencer, il nous faut introduire quelques fonctions qui joueront un rôle clef dans nos résultats.

$$\phi(s) = \mathbb{E}\left[(r(X) - r(x)) \,|\, d(X,x) = s\right],$$
$$F(h) = \mathbb{P}\left(d(X,x) \leq h\right),\ \tau_h(s) = \frac{F(hs)}{F(h)},$$

Nos premières hypothèses portent sur le modèle :

(7.2) r est borné sur un voisinage de x que l'on notera $\mathcal{V}_x$,

(7.3) $F(0) = 0$, $\phi(0) = 0$ et $\phi'(0)$ existe,

(7.4) $\sigma_\epsilon^2(x) = \mathbb{E}\left[\epsilon^2 \mid X = x\right]$ est continue sur $\mathcal{V}_x$ et $\sigma_\epsilon^2 := \sigma_\epsilon^2(x) > 0$,

(7.5) $\forall s \in [0;1]$, $\lim_{n\to+\infty} \tau_{h_n}(s) = \tau_0(s)$ avec $\tau_0(s) \neq 1_{[0;1]}(s)$.

L'hypothèse (7.3) permet de donner l'expression exacte des constantes là où une hypothèse standard de type Lipschitz ne donnerait que des ordres de grandeur. On pourra aussi remarquer que les hypothèses faites sur la loi de X ne se font qu'à travers des probabilités de petites boules et que la condition (7.5) est vérifiée par de nombreux processus standards (voir [167]). La deuxième partie de l'hypothèse assure que la constante M_0 définie plus loin est strictement positive. Ensuite, on introduit une hypothèse portant sur le moment conditionnel d'ordre p de ϵ sachant X. Par souci de simplicité on fait l'hypothèse :

(7.6) $$\exists p > 2,\ \exists M > 0,\ \mathbb{E}\left[|\epsilon|^p / X\right] \leq M \ p.s..$$

On peut remarquer que si ϵ est bornée p.s. ou si elle est indépendante de X et a un moment d'ordre p alors cette hypothèse est vérifiée. Enfin il nous faut faire quelques

hypothèses concernant notre estimateur à noyau :

$$h_n = O\left(\frac{1}{\sqrt{nF(h_n)}}\right), \lim_{n\to+\infty} nF(h_n) = +\infty, \tag{7.7}$$

$$K \text{ a un support compact, est } C^1 \text{ et décroissante sur }]0;1[\ , K(1) > 0. \tag{7.8}$$

Il semble intéressant de faire quelques commentaires sur l' hypothèse (7.6) :
Remarque : On utilise seulement l'hypothèse (7.6) pour montrer que :

$$\forall s \leq p,\ \forall l \geq 0,\ \mathbb{E}\left[|\epsilon|^s \left|K\left(\frac{d(x,X)}{h_n}\right)\right|^l\right] = O\left(\mathbb{E}\left[\left|K\left(\frac{d(x,X)}{h_n}\right)\right|^l\right]\right).$$

Comme K est à support compact dans $[0;1]$, on peut démontrer les résultats de cet article sous l'hypothèse moins restrictive

$$\exists p > 2,\ \exists \mu > 0,\ \exists M > 0,\ \mathbb{E}\left[|\epsilon|^p \mid X\right] 1_{\{d(x,X)\leq\mu\}} \leq M \ p.s..$$

Nos résultats seront exprimés à l'aide des constantes suivantes :

$$M_0 = K(1) - \int_0^1 (sK(s))' \tau_0(s)\, ds,\ M_j = K^j(1) - \int_0^1 \left(K^j\right)'(s)\, \tau_0(s)\, ds,\ j = 1,2\ ,$$

et de la variable aléatoire

$$Z_n := \frac{M_1}{\sqrt{M_2\sigma_\epsilon^2}} \sqrt{nF(h_n)}\left(\hat{r}(x) - r(x) - B_n\right)$$

où $B_n = h_n\phi'(0)\frac{M_0}{M_1}$.
Pour toute la suite de cet article on introduit une variable aléatoire gaussienne centrée réduite indépendante de $(X_i, Y_i)_{1\leq i,j\leq n}$ que l'on notera W. On notera G sa fonction de répartition et g sa densité.

7.2.1.2. Convergence des moments. — A partir de la convergence en loi et de l'uniforme intégrabilité de $|Z_n|^l$ pour $0 \leq l < p$, on obtient la convergence des moments présentée dans le théorème suivant.

Théorème 7.2.1. — *On suppose que les couples (X_i, Y_i) sont indépendants, que les hypothèses (7.2)-(7.8) sont vérifiées, alors pour tout $0 \leq q < p$*

$$\mathbb{E}\left[|Z_n|^q\right] \overset{n\to+\infty}{\to} \mathbb{E}\left[|W|^q\right].$$

De plus si $q \in \mathbb{N}$ alors on a aussi $\mathbb{E}\left[Z_n^q\right] \overset{n\to+\infty}{\to} \mathbb{E}\left[W^q\right]$.

Démonstration. — La preuve repose sur les lemmes suivants :

Lemme 7.2.2. — *(Voir [167] Théorème 1)*
Si les hypothèses (7.2)-(7.5) et (7.7)-(7.8) sont vérifiées et si les paires (X_i, Y_i) sont indépendantes alors

$$Z_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Lemme 7.2.3. — *Si les couples (X_i, Y_i) sont indépendants, si (7.2), (7.3) et (7.5)-(7.8) sont vérifiées, si $0 \leq q < p$, alors $|Z_n|^q$ est uniformément intégrable.*

Le Lemme 7.2.2 implique que Z_n converge en loi vers W. Par conséquent puisque $x \mapsto |x|^q$ est continue sur $\mathbb{R}$, on a également la convergence en loi de $|Z_n|^q$ vers $|W|^q$. De plus le Lemme 7.2.3 implique que $|Z_n|^q$ est uniformément intégrable. On obtient donc :

$$\mathbb{E}\left[|Z_n|^q\right] \overset{n\to+\infty}{\to} \mathbb{E}\left[|W|^q\right].$$

Pour obtenir la seconde partie du théorème il suffit de remarquer que lorsque q est un entier naturel Z_n^q est uniformément intégrable puisque $|Z_n|^q$ l'est et converge en loi vers W^q puisque $x \mapsto x^q$ est continue sur $\mathbb{R}$. □

7.2.2. Le cas α-mélangeant. — Lorsque l'on s'intéresse à certains problèmes, l'hypothèse d'indépendance des variables ne peut être conservée. C'est par exemple le cas lorsque l'on étudie une série temporelle. Typiquement, on veut prévoir certaines des valeurs futures à partir des valeurs dont on dispose. On considère la série temporelle comme un processus a temps continu $(\zeta_t)_{t\in[0,\tau N]}$. On découpe ensuite ce processus en N variables fonctionnelles dépendantes définies sur $[0,\tau]$ par $\xi_i(x) := \zeta_{i\tau+x}$. Cette idée a été initialement introduite dans [39]. Dans ce contexte, on propose le modèle de régression suivant prenant en compte la totalité de la trajectoire passée :

$$H(\xi_i) = r(\xi_{i-1}) + \epsilon_i,$$

où H est un opérateur connu à valeurs réelles défini sur $C^0([0,\tau])$. Un exemple standard d'utilisation de ce modèle est le cas où $H(\xi) = \xi(t_0)$. On retrouve le modèle de régression (7.1) en remplaçant (X_i, Y_i) par $(\xi_{i-1}, H(\xi_i))$. Par conséquent, l'étude de séries temporelles peut être vue comme un problème de régression avec des données dépendantes. Pour pouvoir utiliser nos résultats dans ce contexte il faut prendre en compte la dépendance des variables fonctionnelles. On supposera qu'elles sont α-mélangeantes. Ce cas le plus général de dépendance faible entre les variables a été largement étudié notamment dans [43], [376] et [377].

7.2.2.1. Notations et hypothèses supplémentaires. — Dans le cas où les variables sont dépendantes il nous faut pouvoir contrôler la somme des covariances ; on utilise pour cela l'hypothèse :

$$(7.9)\quad \exists u \le p,\ \exists M,\ \forall q_j \in \mathbb{N},\ \sum_{j=1}^{u} q_j \le u,\ \max_{1\le i_j\le n} \mathbb{E}\left[\left|\prod_{j=1}^{u} \epsilon_{i_j}^{q_j}\right| \mid X_1, \ldots, X_n\right] \le M \ p.s..$$

On peut faire une remarque semblable à celle sur l'hypothèse (7.6) par rapport à l'hypothèse (7.9) et l'affaiblir un peu. Il nous faut également des hypothèses supplémentaires concernant la dépendance de nos variables. On suppose que les n paires (X_i, Y_i) sont α-mélangeantes (voir [376], p37, Definition 2, Notation 1.2) avec des coefficents de mélange arithmétiques d'ordre a. Il nous faut encore introduire quelques notations :

$$\forall k \ge 2,\ \Theta_k(s) := \max\left(\max_{1\le i_1<\cdots<i_k\le n} P\left(d\left(X_{i_j}, x\right) \le s,\ 1 \le j \le k\right), F^k(s)\right)$$

S'ajoutent également des hypothèses portant sur les fonctions Θ_k et l'ordre a des coefficients de mélange.

$$\exists t < p,\ \forall k \leq t,\ \exists \nu_k > 0, \Theta_k(s) = O\left(F(s)^{1+\nu_k}\right),\ \text{avec } \nu_k \leq \nu_{k-1} + 1$$
$$\text{et } \ a > \max_{2\leq k\leq t} (k-1)\frac{(1+\nu_k)p - t}{\nu_k (p-t)} \tag{7.10}$$

$$\exists \gamma > 0,\ nF^{1+\gamma}(h_n) \to +\infty \text{ et } a > \frac{2}{\gamma} + 1. \tag{7.11}$$

On peut remarquer en ce qui concerne l'hypothèse (7.10) que, par définition de Θ_k, $0 \leq \nu_k \leq k-1$ et que par conséquent la deuxième partie implique que $a > 2$.

Il est intéressant de noter que les coefficients géométriques sont arithmétiques d'ordre a quelque soit a. On peut donc supprimer, si les coefficients de mélange sont géométriques, les conditions portant sur a des hypothèses précédentes.

7.2.2.2. Convergence des moments. —

Théorème 7.2.4. — *On suppose que les couples (X_i, Y_i) sont arithmétiquement α-mélangeants d'odre a et que les hypothèses (7.2)-(7.11) sont vérifiées . On a alors, pour tout $0 \leq q < 2\left[\frac{min(t,u)}{2}\right]$*

$$\mathbb{E}\left[|Z_n|^q\right] \overset{n\to+\infty}{\to} \mathbb{E}\left[|W|^q\right].$$

De plus si $q \in \mathbb{N}$ alors on a aussi $\mathbb{E}\left[Z_n^q\right] \overset{n\to+\infty}{\to} \mathbb{E}\left[W^q\right]$.

Démonstration. — Le schéma de la preuve est le même que celui la démonstration du Théorème 7.2.1, en remplaçant les Lemmes 7.2.2 et 7.2.3 par les lemmes suivants.

Lemme 7.2.5. — *(Voir [134] Théorème 2.2) Si les hypothèses (7.2)-(7.11) sont vérifiées, alors*

$$Z_n \overset{\mathcal{L}}{\to} \mathcal{N}(0,1).$$

Lemme 7.2.6. — *Si les couples (X_i, Y_i) sont arithmétiquement α-mélangeants d'ordre a, si (7.2), (7.3), (7.5)-(7.11) sont vérifiées et si $0 < q < 2\left[\frac{min(t,u)}{2}\right]$, alors $|Z_n|^q$ est uniformément intégrable.*

□

7.3. Expression asymptotique des moments et erreurs $\mathbb{L}^q$

7.3.1. Résultats principaux. — A partir des résultats généraux de la partie précédente, il est possible de donner l'expression asymptotique des moments centrés de l'estimateur à noyau mais aussi celle des moments de $\hat{r}(x) - r(x)$ et donc celle des

erreurs $\mathbb{L}^q$. Intéressons nous tout d'abord à l'expression asymptotique des moments centrés de l'estimateur à noyau. Les lemmes dont nous avons eu besoin pour montrer les résultats précédents permettent d'en exprimer les termes dominants.

Théorème 7.3.1. — *Sous les hypothèses du Théorème 7.2.1 avec $\ell = p$ ou celles du Théorème 7.2.4 avec $\ell = 2\left[\frac{\min(t,u)}{2}\right]$, pour tout $q \in \mathbb{N}$ tel que $0 \leq q < \ell$, on a :*

$$\mathbb{E}\left[\left(\hat{r}\left(x\right) - \mathbb{E}\left[\hat{r}\left(x\right)\right]\right)^q\right] = \left(\sqrt{\frac{M_2\sigma_\epsilon^2}{nF\left(h_n\right)M_1^2}}\right)^q \mathbb{E}\left[W^q\right] + o\left(\frac{1}{\left(nF\left(h_n\right)\right)^{\frac{q}{2}}}\right).$$

Remarque : Le résultat donné dans le Théorème 7.3.1 provient d'un résultat plus général qui permettrait d'énoncer un théorème équivalent pour les moments absolus d'ordres non nécessairement entiers (voir la preuve du Théorème 7.3.1 en annexe).

D'un point de vue statistique dans une perspective de choix de fenêtre, on s'intéresse plus souvent aux erreurs $\mathbb{L}^q$ c'est à dire aux moments absolus de $\hat{r}\left(x\right) - r\left(x\right)$. Dans cette optique on établit les expressions asymptotiques des moments de $\hat{r}\left(x\right) - r\left(x\right)$. C'est l'objectif du corollaire suivant dans lequel nous préciserons les constantes selon la parité de l'exposant. En particulier, le cas impair nous amène à introduire les deux suites de polynômes suivantes : $P_{2m+1}\left(u\right) = \sum_{l=0}^{m} a_{m,l}u^{2l+1}$ et $Q_{2m+1}\left(u\right) = \sum_{l=0}^{m} b_{m,l}u^{2l}$, où

$$\begin{aligned} a_{m,l} &= \frac{(2m+1)!}{(2l+1)!2^{(m-l)}\left(m-l\right)!} \\ \text{et } b_{m,l} &= \sum_{j=m-l+1}^{m}\left[C_{2m+1}^{2j+1}\frac{2^j j!}{2^{j+l-m}\left(j+l-m\right)!} - C_{2m+1}^{2j}\frac{(2j)!2^{j+l-m}\left(j+l-m\right)!}{2^j j!\left(2\left(j+l-m\right)\right)!}\right] \\ &\quad + C_{2m+1}^{2(m-l)+1}2^{m-l}\left(m-l\right)!, \end{aligned}$$

ainsi que la fonction $\psi_m\left(u\right) = \left(2G\left(u\right)-1\right)P_{2m+1}\left(u\right) + 2g\left(u\right)Q_{2m+1}\left(u\right)$ (avec la convention qu'une somme sur un ensemble vide d'indices est nulle).

Théorème 7.3.2. — *Sous les hypothèses du Théorème 7.2.1 avec $\ell = p$ ou celles du Théorème 7.2.4 avec $\ell = 2\left[\frac{\min(t,u)}{2}\right]$, on a :*

(i) $\forall\ 0 \leq q < \ell, \mathbb{E}\left[\left|\hat{r}\left(x\right) - r\left(x\right)\right|^q\right] = \mathbb{E}\left[\left|h_n B + W\frac{V}{\sqrt{nF(h_n)}}\right|^q\right] + o\left(\frac{1}{\left(nF(h_n)\right)^{\frac{q}{2}}}\right),$

(ii) $\forall m \in \mathbb{N},\ 2m < \ell,$

$$\mathbb{E}\left[\left|\hat{r}\left(x\right) - r\left(x\right)\right|^{2m}\right] = \sum_{k=0}^{m}\frac{V^{2k}B^{2(m-k)}\left(2m\right)!}{\left(2\left(m-k\right)\right)!k!2^k}\frac{h_n^{2(m-k)}}{\left(nF\left(h_n\right)\right)^k} + o\left(\frac{1}{\left(nF\left(h_n\right)\right)^m}\right),$$

(iii) $\forall m \in \mathbb{N},\ 2m+1 < \ell,$

$$\mathbb{E}\left[\left|\hat{r}\left(x\right) - r\left(x\right)\right|^{2m+1}\right] = \frac{1}{\left(nF\left(h_n\right)\right)^{m+\frac{1}{2}}}\left(V^{2m+1}\psi_m\left(\frac{Bh_n\sqrt{nF\left(h_n\right)}}{V}\right) + o\left(1\right)\right),$$

où $B = \phi'\left(0\right)\frac{M_0}{M_1}$ *et* $V = \sqrt{\frac{M_2\sigma_\epsilon^2}{M_1^2}}$.

Remarque : le résultat (i) du Théorème 7.3.2 provient d'un résultat plus général qui implique également le même résultat sans les valeurs absolues de part et d'autre de l'égalité lorsque $q \in \mathbb{N}$ (voir la preuve du (i) en annexe pour plus de détails). De là, on peut obtenir l'expression des moments d'ordre entier de $\hat{r}(x) - r(x)$ (voir la preuve de (ii) en annexe pour plus de détails).

7.3.2. Cas particulier : erreurs $\mathbb{L}^1$ et $\mathbb{L}^2$. — Pour illustrer les résultats (ii) et (iii), on peut donner les expressions asymptotiques des erreurs $\mathbb{L}^1$ et $\mathbb{L}^2$:

Corollaire 7.3.3. — *Si les hypothèses du Théorème 7.3.2 sont vérifiées, alors on a :*

$$\mathbb{E}\left[|\hat{r}(x) - r(x)|\right] = \sqrt{\frac{M_2\sigma_\epsilon^2}{M_1^2 nF(h_n)}}\psi_0\left(\frac{h_n\phi'(0)\,M_0\sqrt{nF(h_n)}}{\sqrt{M_2\sigma_\epsilon^2}}\right) + o\left(\frac{1}{\sqrt{nF(h_n)}}\right),$$

et

$$\mathbb{E}\left[|\hat{r}(x) - r(x)|^2\right] = \left(\frac{M_0}{M_1}\phi'(0)\,h_n\right)^2 + \frac{M_2\sigma_\epsilon^2}{M_1^2 nF(h_n)} + o\left(\frac{1}{nF(h_n)}\right).$$

On peut noter l'expression relativement simple de l'erreur $\mathbb{L}^1$ puisque dans ce cas, $\psi_0(u) = 2uG(u) + 2g(u) - u$. D'un point de vue statistique, on remarque que pour minimiser ces erreurs il va falloir choisir h_n de façon à équilibrer des termes croissants et d'autres décroissants en h_n.

7.3.3. Cas particulier : $E = \mathbb{R}^d$. — Les résultats précédents sont donnés dans le cas général où la variable explicative est fonctionnelle. Ils s'appliquent donc bien entendu au cas où elle est à valeurs dans $\mathbb{R}^d$. Voyons à présent ce que deviennent nos hypothèses et nos résultats dans ce cas particulier. Si on suppose qu'il existe $t < p$ tel que pour tout $k \leq t$, X_i et $(X_{i_1}, \ldots, X_{i_k})$ admettent des densités f et $c_k(i_j, 1 \leq j \leq k)$ continues au voisinage de x et telles que $f(x) > 0$, alors on montre facilement que

$$F(h_n) = f(x)\,h_n^d\,(1 + o(1)) \text{ et } \Theta_k(h_n) = O\left(h_n^{kd}\right).$$

Par conséquent l'hypothèse (7.5) portant sur τ_{h_n} est toujours vérifiée avec $\tau_0(s) = s^d$. De plus la première partie de l'hypothèse (7.10) est vérifiée avec $\nu_k = k - 1$. Enfin, on peut remplacer $F(h_n)$ par $f(x)\,h_n^d$ dans les hypothèses et les résultats, ce qui donne les hypothèses :

$$h_n^{2+d} = O\left(\frac{1}{n}\right), \lim_{n\to+\infty} nh_n^d = +\infty, \tag{7.12}$$

$$\exists t > 0, a > \frac{t(p-1)}{(p-t)}, \tag{7.13}$$

$$\exists \gamma > 0,\ nh_n^{d(1+\gamma)}(h_n) \to +\infty \text{ et } a > \frac{2}{\gamma} + 1. \tag{7.14}$$

On remarque que les constantes ne dépendent plus de la loi de X :

$$M_0 = d\int_0^1 s^d K(s)\,ds \quad , \quad M_j = d\int_0^1 s^{d-1}K^j(s)\,ds,\ j = 1, 2\ .$$

Le Théorème 7.3.2 devient donc :

Corollaire 7.3.4. — *Sous les hypothèses (7.2)-(7.6), (7.8) et (7.12) dans le cas indépendant avec $\ell = p$ ou en y ajoutant (7.9), (7.13) et (7.14) dans le cas arithmétiquement α-mélangeant avec $\ell = 2\left[\frac{\min(t,u)}{2}\right]$, on a :*

(i) $\forall\ 0 \leq s < 2\left[\frac{min(t,u)}{2}\right]$,

$$\mathbb{E}\left[|\hat{r}(x) - r(x)|^s\right] = \mathbb{E}\left[\left|h_n\phi'(0)\frac{M_0}{M_1} + W\sqrt{\frac{M_2\sigma_\epsilon^2}{nf(x)h_n^d M_1^2}}\right|^s\right] + o\left(\frac{1}{(nh_n^d)^{\frac{s}{2}}}\right),$$

(ii) $\forall m \in \mathbb{N},\ 0 \leq 2m < 2\left[\frac{min(t,u)}{2}\right]$,

$$\begin{aligned}\mathbb{E}\left[|\hat{r}(x) - r(x)|^{2m}\right] &= \sum_{k=0}^{m} \frac{\left(\frac{M_2\sigma_\epsilon^2}{M_1^2}\right)^k \left(\frac{M_0}{M_1}\phi'(0)\right)^{2(m-k)} (2m)!}{(2(m-k))!k!2^k (f(x))^k} h_n^{2(m-k)-dk} n^{-k} \\ &+ o\left(\frac{1}{(nh_n^d)^m}\right),\end{aligned}$$

(iii) $\forall m \in \mathbb{N},\ 2m+1 < 2\left[\frac{min(t,u)}{2}\right]$,

$$\begin{aligned}\mathbb{E}\left[|\hat{r}(x) - r(x)|^{2m+1}\right] &= \left(\frac{M_2\sigma_\epsilon^2}{M_1^2 nf(x)h_n^d}\right)^{m+\frac{1}{2}} \psi_m\left(\frac{h_n^{1+\frac{d}{2}}\phi'(0)M_0\sqrt{nf(x)}}{\sqrt{M_2\sigma_\epsilon^2}}\right) \\ &+ o\left(\frac{1}{(nh_n^d)^{m+\frac{1}{2}}}\right).\end{aligned}$$

Remarque : ces résultats généralisent à des moments d'ordre quelconque ceux donnés par [209] pour l'erreur $\mathbb{L}^2$ et par [421] pour l'erreur $\mathbb{L}^1$.

7.4. Conclusion et perspectives

Grâce à l'expression explicite de la loi asymptotique de l'estimateur à noyau, on a pu obtenir les expressions asymptotiques des moments. Ces résultats sont très motivants car ils ouvrent d'intéressantes perspectives, en particulier vis-à-vis du choix de fenêtre au travers de l'erreur $\mathbb{L}^q$. Il serait notamment intéressant d'étudier comment l'estimateur se comporte sur des données réelles lorsque l'on choisit d'utiliser d'autres critères que l'erreur $\mathbb{L}^1$ ou $\mathbb{L}^2$. De plus les résultats établis sont innovants dans le cadre fonctionnel général mais aussi dans le cadre vectoriel où ils s'appliquent sous des hypothèses plus simples. Il serait enfin intéressant de compléter ces résultats en donnant des résultats concernant les erreurs intégrées.

7.5. Annexe

Dans les preuves suivantes nous noterons C les constantes positives quelconques par souci de simplicité. On sera amené à utiliser plusieurs fois l'inégalité suivante portant sur la covariance de variables α-mélangentes établie dans [377] ((1.12b) p 10). Si $X_i \in \mathbb{L}^p$ et $X_j \in \mathbb{L}^q$ avec $1 < p, q \leq \infty$, alors en définissant r par $\frac{1}{r}+\frac{1}{p}+\frac{1}{q} = 1$ on a :

$$|Cov(X_i, X_j)| \leq B_0 (\alpha(|i-j|))^{\frac{1}{r}} \|X_i\|_p \|X_j\|_q, \tag{7.15}$$

où B_0 est une constante positive. On introduit les variables

$$\Gamma_{i,n} := Y_i K\left(\frac{d(X_i,x)}{h_n}\right), \Delta_{i,n} := K\left(\frac{d(X_i,x)}{h_n}\right), U_{i,n} = \frac{\Gamma_{i,n}\mathbb{E}[\Delta_{i,n}] - \Delta_{i,n}\mathbb{E}[\Gamma_{i,n}]}{F(h_n)\sqrt{nF(h_n)}}.$$

Il est établi dans [167] que sous (7.3) et (7.5) on a :

$$\frac{\mathbb{E}[\Gamma_{1,n}]}{\mathbb{E}[\Delta_{1,n}]} - r(x) = B_n + o(h_n).$$

On en déduit facilement (voir [134] pour les détails), grâce à (7.7), que

$$Z_n = \frac{M_1}{\sqrt{M_2\sigma_\epsilon^2}} \frac{\sum_{i=1}^n U_{i,n}}{\hat{f}(x)\,\mathbb{E}\left[\hat{f}(x)\right]} + o(1).$$

On introduit les variables

$$W_{i,h} = Y_i K\left(\frac{d(X_i,x)}{h}\right) \frac{\mathbb{E}\left[K\left(\frac{d(X_i,x)}{h}\right)\right]}{F(h)} - K\left(\frac{d(X_i,x)}{h}\right) \frac{\mathbb{E}\left[Y_i K\left(\frac{d(X_i,x)}{h}\right)\right]}{F(h)}.$$

On utilisera par la suite pour plus de clarté la notation :

$$\hat{f}(x) = \frac{1}{nF(h_n)} \sum_{i=1}^n \Delta_{i,n}.$$

Remarque : Attention, on utilise ici la même notation que dans le cas multivarié pour le dénominateur de l'estimateur à noyau mais ici $\hat{f}$ n'estime pas la densité. On a alors :

$$U_{i,n} = \frac{W_{i,h_n}}{\sqrt{nF(h_n)}} \text{ et } Z_n = \frac{M_1}{\sqrt{nF(h_n)\,M_2\sigma_\epsilon^2}} \frac{\sum_{i=1}^n W_{i,h_n}}{\hat{f}(x)\,\mathbb{E}\left[\hat{f}(x)\right]} + o(1).$$

Preuve du Lemme 7.2.3 ∴ — Nous allons montrer que Z_n a tous ses moments d'ordre $q \leq p$ bornés, ce qui impliquera que Z_n^q est uniformément intégrable pour

tout $q < p$. On va utiliser la décomposition suivante :

$$\begin{aligned}
\mathbb{E}\left[|Z_n|^q\right] &= \mathbb{E}\left[|Z_n|^q 1_{\left|\hat{f}(x)-\mathbb{E}\left[\hat{f}(x)\right]\right| \leq \frac{K(1)}{2}}\right] + \mathbb{E}\left[|Z_n|^q 1_{\left|\hat{f}(x)-\mathbb{E}\left[\hat{f}(x)\right]\right| > \frac{K(1)}{2}}\right] \\
&\leq C + \mathbb{E}\left[\left|\frac{M_1}{\sqrt{M_2\sigma_\epsilon^2}} \frac{\sum_{i=1}^n U_{i,n}}{\hat{f}(x)\,\mathbb{E}\left[\hat{f}(x)\right]}\right|^q 1_{\left|\hat{f}(x)-\mathbb{E}\left[\hat{f}(x)\right]\right| \leq \frac{K(1)}{2}}\right] \\
&+ \mathbb{E}\left[\left|\frac{M_1}{\sqrt{M_2\sigma_\epsilon^2}} \frac{\sum_{i=1}^n U_{i,n}}{\hat{f}(x)\,\mathbb{E}\left[\hat{f}(x)\right]}\right|^q 1_{\left|\hat{f}(x)-\mathbb{E}\left[\hat{f}(x)\right]\right| > \frac{K(1)}{2}}\right] \\
&= C + A_{1,n} + A_{2,n} \qquad (7.16)
\end{aligned}$$

Il nous faut à présent montrer que les suites $A_{1,n}$ et $A_{2,n}$ sont bornées. On s'intéresse dans un premier temps à la majoration de $A_{1,n}$.

Première étape : majoration des moments du numérateur. Prouvons tout d'abord que pour tout $0 \leq q \leq p$,

$$\mathbb{E}\left[\left|\sum_{i=1}^n W_{i,h_n}\right|^q\right] = O\left(\left[nF(h_n)\right]^{\frac{q}{2}}\right). \qquad (7.17)$$

En fait tout repose sur le lemme suivant :

Lemme 7.5.1. — *Si les couples (X_i, Y_i) sont indépendants, sous les hypothèses (7.2), (7.6), (7.7) et (7.8), pour tout $0 \leq q \leq p$ il existe des constantes K_1 et K_2 telles que pour tout $l \in \mathbb{N}$ et h assez petit (tel que $B(x,h) \subset \mathcal{V}_x$) on a :*

$$\mathbb{E}\left[\left|\sum_{i=1}^l W_{i,h}\right|^q\right] \leq K_1(q)\, lF(h) + K_2(q)\,(lF(h))^{\frac{q}{2}}.$$

Preuve du Lemme 7.5.1 :. — On peut remarquer qu'en utilisant les hypothèses (7.2), (7.6) et (7.8) et en conditionnant par X, on a pour tout $q \leq p$:

$$\begin{aligned}
\mathbb{E}\left[|W_{i,h}|^q\right] &\leq C\mathbb{E}\left[K^q\left(\frac{d(X,x)}{h}\right)\right] \\
&\leq CF(h).
\end{aligned}$$

On va décomposer la preuve suivant les valeurs de q.

Preuve du lemme pour $q = 2$:. — On peut montrer assez aisément que lemme est vrai pour $q = 2$. En effet, puisque les $W_{i,h}$ sont centrées et indépendantes et de

même loi, on a :

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{i=1}^{l} W_{i,h}\right)^2\right] &= l\,\mathbb{E}\left[(W_{1,h})^2\right] \\
&= l\,\mathbb{E}\Big[\Big(Y_i K\left(\frac{d(X_i,x)}{h}\right)\frac{1}{F(h)}\mathbb{E}\left[K\left(\frac{d(X_i,x)}{h}\right)\right] \\
&\quad -K\left(\frac{d(X_i,x)}{h}\right)\frac{1}{F(h)}\mathbb{E}\left[Y_i K\left(\frac{d(X_i,x)}{h}\right)\right]\Big)^2\Big] \\
&\leq C\,l\,\Big\{\mathbb{E}\left[(r(X_i)+\epsilon_i)^2 K^2\left(\frac{d(X_i,x)}{h}\right)\right]+\mathbb{E}\left[K^2\left(\frac{d(X_i,x)}{h}\right)\right] \\
&\quad -2\mathbb{E}\left[(r(X_i)+\epsilon_i)K^2\left(\frac{d(X_i,x)}{h}\right)\right]\Big\} \\
&\leq C\,l\,\mathbb{E}\left[K^2\left(\frac{d(X_i,x)}{h}\right)\right] \\
&\leq ClF(h).
\end{aligned}$$

On obtient les dernières lignes en utilisant les hypothèses (7.2), (7.6) et (7.8) qui impliquent qu'en conditionnant par X_i, pour $u \in \{0,1,2\}$ et $s>0$, on a :

$$\mathbb{E}\left[Y_i^u K^s\left(\frac{d(X_i,x)}{h}\right)\right] \leq C\,F(h).$$

□

Preuve du lemme pour $0 \leq q < 2$ **:**. — On démontre ensuite le lemme pour $q<2$ en utilisant l'inégalité de Hölder :

$$\mathbb{E}\left[\left|\sum_{i=1}^{l} W_{i,h}\right|^q\right] \leq \left(\mathbb{E}\left[\left|\sum_{i=1}^{l} W_{i,h}\right|^2\right]\right)^{\frac{q}{2}} \leq K_3(q)\left(lF(h)\right)^{\frac{q}{2}}.$$

□

Preuve du lemme pour $q>2$ **:**. — On va se servir des résultats L^2 pour obtenir ceux d'ordre $q<p$ par récurrence. En s'inspirant d'une preuve de [264] et de [139], on va montrer que si l'inégalité précédente est vérifiée pour $q=m\geq 2$ et $m\in\mathbb{N}$, alors elle l'est pour $q=m+\nu$, $0<\nu\leq 1$. On définit $C_{l,h}$ de la façon suivante :

$$C_{l,h} = \mathbb{E}\left[\left|\sum_{i=1}^{l} W_{i,h}\right|^{m+\nu}\right]$$

ainsi que

$$S_{l,h} = \sum_{i=1}^{l} W_{i,h},\ \hat{S}_{l,h} = \sum_{i=l+1}^{2l} W_{i,h}.$$

On veut montrer que pour tout m et ν il existe des constantes K_1 et K_2 telles que

$$C_{l,h} \leq K_1 (m+\nu) lF(h) + K_2 (m+\nu) (lF(h))^{\frac{m+\nu}{2}}.$$

Montrons dans un premier temps qu'il existe des constantes A_1 et A_2 positives et indépendantes de l et de h telles que

$$\mathbb{E}\left[\left|S_{l,h} + \hat{S}_{l,h}\right|^{m+\nu}\right] \leq 2C_{l,h} + A_1 (lF(h))^{\frac{m+\nu}{m}} + A_2 (lF(h))^{\frac{m+\nu}{2}}.$$

$$\begin{aligned}
& \mathbb{E}\left[\left|S_{l,h} + \hat{S}_{l,h}\right|^{m+\nu}\right] \\
\leq\ & \mathbb{E}\left[\left|S_{l,h} + \hat{S}_{l,h}\right|^{m}\left(|S_{l,h}|^{\nu} + \left|\hat{S}_{l,h}\right|^{\nu}\right)\right] \\
\leq\ & 2C_{l,h} + \mathbb{E}\left[\sum_{j=0}^{m-1} C_m^j |S_{l,h}|^{j+\nu}\left|\hat{S}_{l,h}\right|^{m-j} + \sum_{j=1}^{m} C_m^j |S_{l,h}|^{j}\left|\hat{S}_{l,h}\right|^{m-j+\nu}\right]
\end{aligned}$$

Dans ce qui suit $B_1, \ldots, B_8$ désignent des constantes indépendantes de l et de h. En utilisant l'indépendance de nos variables, une inégalité de Hölder puis l'hypothèse de récurrence au rang m on obtient :

$$\begin{aligned}
\mathbb{E}\left[|S_{l,h}|^{j}\left|\hat{S}_{l,h}\right|^{m+\nu-j}\right] &= \mathbb{E}\left[|S_{l,h}|^{j}\right]\mathbb{E}\left[\left|\hat{S}_{l,h}\right|^{m+\nu-j}\right] \\
&\leq \left[\mathbb{E}\left[|S_{l,h}|^{m}\right]\right]^{\frac{j}{m}}\left[\mathbb{E}\left[|S_{l,h}|^{m}\right]\right]^{\frac{m+\nu-j}{m}} \\
&\leq \left(K_1(m) lF(h) + K_2(m)(lF(h))^{\frac{m}{2}}\right)^{\frac{m+\nu}{m}} \\
&\leq B_1 (lF(h))^{\frac{m+\nu}{m}} + B_2 (lF(h))^{\frac{m+\nu}{2}}.
\end{aligned}$$

Avec un raisonnement identique on obtient, pour $0 \leq j \leq m-1$:

$$\mathbb{E}\left[|S_{l,h}|^{j+\nu}\left|\hat{S}_{l,h}\right|^{m-j}\right] \leq B_1 (lF(h))^{\frac{m+\nu}{m}} + B_2 (lF(h))^{\frac{m+\nu}{2}}.$$

On a donc

$$\mathbb{E}\left[\left|S_{l,h} + \hat{S}_{l,h}\right|^{m+\nu}\right] \leq 2C_{l,h} + B_3 (lF(h))^{\frac{m+\nu}{m}} + B_4 (lF(h))^{\frac{m+\nu}{2}}.$$

On peut maintenant majorer le terme $C_{2l,h}$.

$$\begin{aligned}
C_{2l,h} &= \mathbb{E}\left[\left|S_{l,h} + \hat{S}_{l,h}\right|^{m+\nu}\right] \\
&\leq 2C_{l,h} + A_1 (lF(h))^{\frac{m+\nu}{m}} + A_2 (lF(h))^{\frac{m+\nu}{2}}.
\end{aligned}$$

Par conséquent on obtient en remplaçant $2l$ par 2^r

$$\begin{aligned}
& C_{2^r,h} \\
\leq\ & 2C_{2^{r-1},h} + A_1 \left(2^{r-1}F(h)\right)^{\frac{m+\nu}{m}} + A_2 \left(2^{r-1}F(h)\right)^{\frac{m+\nu}{2}} \\
\leq\ & 4C_{2^{r-2},h} + A_1 \left(2^{r-1}F(h)\right)^{\frac{m+\nu}{m}} \left(1 + \frac{2}{2^{\frac{m+\nu}{m}}}\right) + A_2 \left(2^{r-1}F(h)\right)^{\frac{m+\nu}{2}} \left(1 + \frac{2}{2^{\frac{m+\nu}{2}}}\right) \\
\leq\ & 2^r C_{1,h} + A_1 \left(2^{r-1}F(h)\right)^{\frac{m+\nu}{m}} \left(\sum_{k=0}^{r-1} \left(\frac{2}{2^{\frac{m+\nu}{m}}}\right)^k\right) \\
& + A_2 \left(2^{r-1}F(h)\right)^{\frac{m+\nu}{2}} \left(\sum_{k=0}^{r-1} \left(\frac{2}{2^{\frac{m+\nu}{2}}}\right)^k\right) \\
\leq\ & 2^r C_{1,h} + \frac{A_1 \left(2^{r-1}F(h)\right)^{\frac{m+\nu}{m}}}{1 - \frac{2}{2^{\frac{m+\nu}{m}}}} + \frac{A_2 \left(2^{r-1}F(h)\right)^{\frac{m+\nu}{2}}}{1 - \frac{2}{2^{\frac{m+\nu}{2}}}} \\
\leq\ & B_5 2^r F(h) + B_6 \left(2^r F(h)\right)^{\frac{m+\nu}{2}}
\end{aligned}$$

car $1 < \frac{m+\nu}{m} \leq \frac{m+\nu}{2}$ et $C_{1,h} \leq CF(h)$. On écrit maintenant l en base 2 :

$$l = 2^r + a_1 2^{r-1} + \ \ldots\ + a_r \leq 2^r + 2^{r-1} + \ \ldots\ + 1 < 2^{r+1},$$

où les a_i valent 0 ou 1. On peut alors majorer $C_{l,h}$ en décomposant la somme de l termes en une somme de 2^r termes, une autre de $a_1 2^{r-1}$ termes et ainsi de suite. On utilise ensuite l'inégalité de Minkowski pour obtenir :

$$\begin{aligned}
C_{l,h} &\leq \left[\left(\mathbb{E}\left[\left|S_{2^r,h}\right|^{m+\nu}\right]\right)^{\frac{1}{m+\nu}} + \left(\mathbb{E}\left[\left|S_{2^{r-1},h}\right|^{m+\nu}\right]\right)^{\frac{1}{m+\nu}} + \cdots + \left(\mathbb{E}\left[\left|S_{1,h}\right|^{m+\nu}\right]\right)^{\frac{1}{m+\nu}}\right]^{m+\nu} \\
&\leq \left[\sum_{j=0}^{r} \left(B_5 2^{r-j} F(h) + B_6 \left(2^{r-j}F(h)\right)^{\frac{m+\nu}{2}}\right)^{\frac{1}{m+\nu}}\right]^{m+\nu} \\
&\leq \left[\sum_{j=0}^{r} \left(B_5 2^{r-j} F(h)\right)^{\frac{1}{m+\nu}} + \sum_{j=0}^{r} \left(B_6 \left(2^{r-j}F(h)\right)^{\frac{m+\nu}{2}}\right)^{\frac{1}{m+\nu}}\right]^{m+\nu} \\
&\leq B_7 2^r F(h) + B_8 2^{\frac{r(m+\nu)}{2}} F(h)^{\frac{m+\nu}{2}} \\
&\leq K_1 (m+\nu)\, l F(h) + K_2 (m+\nu) \left(l F(h)\right)^{\frac{m+\nu}{2}}.
\end{aligned}$$

On passe de la deuxième ligne à la troisième par l'inégalité $|a+b|^\nu \leq |a|^\nu + |b|^\nu$ qui est valable pour $0 \leq \nu \leq 1$. Ensuite on obtient la quatrième ligne en utilisant l'inégalité $(a+b)^l \leq 2^l a^l + 2^l b^l$ pour tout a et b positifs. Les constantes B_7 et B_8 dépendent donc clairement de ν. La propriété étudiée est héréditaire, donc vraie pour tout exposant $0 \leq q \leq p$. □

La preuve est ainsi achevée. □

En utilisant le lemme précédent avec $l = n$ et $h = h_n$ on obtient qu'il existe K_1 et K_2 telles que

$$\begin{aligned}\mathbb{E}\left[\left|\sum_{i=1}^{n} W_{i,h_n}\right|^q\right] &\leq K_1(q)\, nF(h_n) + K_2(q)\,(nF(h_n))^{\frac{q}{2}} \\ &\leq K_3\,(nF(h_n))^{\frac{q}{2}}.\end{aligned}$$

La dernière ligne provient du fait que si $q < 2$, $K_1 = 0$ (voir la preuve du lemme 7.5.1) et sinon $nF(h_n) \to +\infty$ donc le premier terme est négligeable devant le deuxième. On peut en déduire que pour tout $0 \leq q \leq p$ il existe une constante $C(q)$ telle que

$$\sup_n \mathbb{E}\left[\left|\sum_{i=1}^{n} U_{i,n}\right|^q\right] = \sup_n \mathbb{E}\left[\left|\sum_{i=1}^{n} \frac{1}{\sqrt{nF(h_n)}} W_{i,h_n}\right|^q\right] \leq C(q).$$

Deuxième étape : bornes pour $A_{1,n}$. On peut montrer sous l'hypothèse (7.8), que l'on a $\mathbb{E}\left[\hat{f}(x)\right] \geq K(1)$. Par conséquent, on en déduit qu'on a

$$\begin{aligned}\left|\hat{f}(x)\right|^{-1} 1_{\left|\hat{f}(x)-\mathbb{E}\left[\hat{f}(x)\right]\right| \leq \frac{K(1)}{2}} &= \left|\hat{f}(x) - \mathbb{E}\left[\hat{f}(x)\right] + \mathbb{E}\left[\hat{f}(x)\right]\right|^{-1} 1_{\left|\hat{f}(x)-\mathbb{E}\left[\hat{f}(x)\right]\right| \leq \frac{K(1)}{2}} \\ &\leq \frac{2}{K(1)}\ p.s.\end{aligned}$$

On déduit facilement de l'étape précédente et de ce qui précède que pour tout $0 \leq q \leq p$ il existe une constante $C(q)$ strictement positive telle que

$$\sup_n A_{1,n} \leq C(q). \tag{7.18}$$

On doit à présent majorer $A_{2,n}$, pour cela on utilise la majoration suivante :

$$\begin{aligned}A_{2,n} &= O\left(\mathbb{E}\left[\left|\frac{\sqrt{nF(h_n)}\hat{r}(x)}{\mathbb{E}\left[\hat{f}(x)\right]}\right|^q 1_{\left|\hat{f}(x)-\mathbb{E}\left[\hat{f}(x)\right]\right| > \frac{K(1)}{2}}\right]\right) \\ &\quad + O\left(\mathbb{E}\left[\left|\frac{\sqrt{nF(h_n)}}{\mathbb{E}\left[\hat{f}(x)\right]}\right|^q 1_{\left|\hat{f}(x)-\mathbb{E}\left[\hat{f}(x)\right]\right| > \frac{K(1)}{2}}\right]\right) \\ &= \sqrt{nF(h_n)}^q \Big(O\left(\mathbb{E}\left[|\hat{r}(x)|^q 1_{\left|\hat{f}(x)-\mathbb{E}\left[\hat{f}(x)\right]\right| > \frac{K(1)}{2}}\right]\right) \\ &\quad + O\left(\mathbb{P}\left(\left|\hat{f}(x) - \mathbb{E}\left[\hat{f}(x)\right]\right| > \frac{K(1)}{2}\right)\right)\Big)\end{aligned} \tag{7.19}$$

On obtient la dernière majoration en se servant de la minoration de $\mathbb{E}\left[\hat{f}(x)\right]$ par $K(1)$. L'idée est maintenant d'utiliser l'hypothèse (7.6) qui permet d'établir que

$$\mathbb{E}\left[|\hat{r}(x)|^q \,|X_1, \ldots, X_n\right] \leq M p.s..$$

On en déduit à partir de (7.19) la majoration suivante.

$$\begin{aligned} A_{2,n} &\leq C\sqrt{nF(h_n)}^q \mathbb{P}\left(\left|\hat{f}(x) - \mathbb{E}\left[\hat{f}(x)\right]\right| > \frac{K(1)}{2}\right) \\ &\leq C\sqrt{nF(h_n)}^q \frac{\mathbb{E}\left[\left|\hat{f}(x) - \mathbb{E}\left[\hat{f}(x)\right]\right|^q\right] 2^q}{K^q(1)} \\ &\leq C\mathbb{E}\left[\left|\sqrt{nF(h_n)}\left(\hat{f}(x) - \mathbb{E}\left[\hat{f}(x)\right]\right)\right|^q\right]. \end{aligned}$$

Pour conclure on remarque que la preuve du Lemme 7.5.1 peut etre facilement adaptée afin de prouver qu'il existe une constante $C_2(q)$ telle que

$$\sup_n \mathbb{E}\left[\left|\sqrt{nF(h_n)}\left(\hat{f}(x) - \mathbb{E}\left[\hat{f}(x)\right]\right)\right|^q\right] \leq C_2(q).$$

On en déduit aisément qu'il existe une constante $C_0(q)$ telle que

$$\sup_n A_{2,n} \leq C_0(q). \tag{7.20}$$

On déduit aisément de (7.18) et (7.20) que les moments d'ordre $\frac{p}{q} > 1$ de $|Z_n|^q$ sont bornés.

Etape finale : uniforme intégrabilité de Z_n. Ce qui précède va nous permettre de montrer que pour tout $0 \leq q < p$, $|Z_n|^q$ est uniformément intégrable.

$$\begin{aligned} \lim_{M\to+\infty} \limsup_{n\to+\infty} \mathbb{E}\left[|Z_n|^q 1_{\{|Z_n|^q > M\}}\right] &\leq \lim_{M\to+\infty} \limsup_{n\to+\infty} \left(\mathbb{E}\left[|Z_n|^p\right]\right)^{\frac{q}{p}} \left(\mathbb{P}\left(|Z_n|^q > M\right)\right)^{1-\frac{q}{p}} \\ &\leq \lim_{M\to+\infty} \limsup_{n\to+\infty} \left(C(p)\right)^{\frac{q}{p}} \left(\frac{\mathbb{E}\left[|Z_n|^p\right]}{M^{\frac{p}{q}}}\right)^{1-\frac{q}{p}} \\ &\leq \lim_{M\to+\infty} C(p) M^{-\frac{p-q}{q}} \\ &= 0. \end{aligned}$$

Par conséquent $\lim_{M\to+\infty} \limsup_{n\to+\infty} \mathbb{E}\left[|Z_n|^q 1_{\{|Z_n|^q > M\}}\right] = 0$ et donc $|Z_n|^q$ est uniformément intégrable.

□

Preuve du Lemme 7.2.6 *:.* — La preuve est semblable à celle du Lemme 7.2.3 et repose sur le lemme suivant :

Lemme 7.5.2. — *Si les couples (X_i, Y_i) sont arithmétiquement α-mélangeants d'ordre a, sous les hypothèses (7.2), (7.6)-(7.10) et si il existe $\gamma > 0$ tel que $nF^{1+\gamma}(h_n) \to +\infty$, alors pour tout $0 \leq q \leq 2\left[\frac{min(t,u)}{2}\right]$, il existe une constante K telle que pour tout $l \in \mathbb{N}$ et $h \in \mathbb{R}$ on a :*

$$\mathbb{E}\left[\left|\sum_{i=1}^n W_{i,h_n}\right|^q\right] \leq K(q)\left(nF(h_n)\right)^{\frac{q}{2}}.$$

Preuve du Lemme 7.5.2 *:.* — On s'inspire d'une preuve de [434]. On va décomposer la preuve suivant les valeurs de q. On démontrera le lemme par récurrence pour des puissances paires : $q = 2m \leq 2\left[\frac{min(t,u)}{2}\right]$, si (7.9) et (7.10) sont vérifiées.
A partir de la majoration pour les exposants pairs on déduit celle pour les exposants inférieurs grâce à une inégalité de Hölder.

Preuve du lemme pour $q = 2m$ **:.** — Pour simplifier les notations on définit pour tout $i \in \mathbb{N}^*$, $W_{n+i,h_n} \equiv 0$ p.s.. Commençons tout d'abord par remarquer que dans notre cas les valeurs absolues ne sont plus nécessaires et développons. On notera dans nos calculs $C_{2m}^{q_1,q_2,...,q_k} = \frac{(2m)!}{q_1!q_2!...q_k!}$ les coefficients du multinome.

$$\begin{aligned}\mathbb{E}\left[\left|\sum_{i=1}^{n} W_{i,h_n}\right|^{2m}\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{n} W_{i,h_n}\right)^{2m}\right] \\ &= \mathbb{E}\left[\sum_{q_j,\sum_{j=1}^{n} q_j=2m} C_{2m}^{q_1,q_2,...,q_n} \prod_{i=1}^{n} W_{i,h_n}^{q_i}\right]\end{aligned}$$

On fait la décomposition suivante :

$$\begin{aligned}\mathbb{E}\left[\left|\sum_{i=1}^{n} W_{i,h_n}\right|^{2m}\right] &= \mathbb{E}\left[\sum_{k=1}^{2m} \sum_{\substack{q_j,\, q_j \geq 1 \\ \sum_{j=1}^{k} q_j = 2m}} C_{2m}^{q_1,q_2,...,q_k} \sum_{1\leq i_1<\cdots<i_k\leq n} \prod_{j=1}^{k} W_{i_j,h}^{q_j}\right] \\ &= \sum_{k=1}^{2m} \sum_{\substack{q_j,\, q_j \geq 1 \\ \sum_{j=1}^{k} q_j = 2m}} C_{2m}^{q_1,q_2,...,q_k} \sum_{1\leq i_1<\cdots<i_k\leq n} \mathbb{E}\left[\prod_{j=1}^{k} W_{i_j,h_n}^{q_j}\right]. \qquad (7.21)\end{aligned}$$

On continue ensuite par l'étude de la partie finale de l'expression précédente.

$$\begin{aligned}\sum_{1\leq i_1<\cdots<i_k\leq n} \mathbb{E}\left[\prod_{j=1}^{k} W_{i_j,h_n}^{q_j}\right] &\leq \sum_{1\leq i_1<\cdots<i_k\leq n} \left|\mathbb{E}\left[\prod_{j=1}^{k} W_{i_j,h_n}^{q_j}\right]\right| \\ &\leq \sum_{i_1=1}^{n} \sum_{1\leq e_1,e_2,...,e_{k-1}\leq n} \left|\mathbb{E}\left[\prod_{j=1}^{k} W_{i_1+\sum_{v=1}^{j-1} e_v,h_n}^{q_j}\right]\right|.\end{aligned}$$

On va montrer par récurrence sur m que cette dernière majoration est bornée quelque soient k et $q_j, 1 \leq j \leq k$ de la manière suivante :

$$(H_m): \qquad \forall k \leq 2m,\ \forall q_j \in \mathbb{N}^*,\ \sum_{i=1}^{k} q_j = q \leq 2m, \exists C_1, C_2 > 0,\ \forall n \in \mathbb{N},$$
$$\sum_{i_1=1}^{n} \sum_{1 \leq e_1, e_2, \ldots, e_{k-1} \leq n} \left| \mathbb{E}\left[\prod_{j=1}^{k} W_{i_1+\sum_{l=1}^{j-1} e_l, h}^{q_j} \right] \right| \leq C_1 \left(nF\left(h_n\right)\right)^{\frac{q}{2}}.$$

On va dans un premier temps montrer que dans le cas où $m = 1$ l'hypothèse de récurrence est vraie.

Preuve de (H_1) **:.** — Dans le cas où $m = 1$ on peut avoir $q = 1$ ou $q = 2$. Dans un premier temps notons pour tout $m \in \mathbb{N}$, si $q = 1$ alors la majoration est vérifiée puisque les W_{i,h_n} sont centrés. En effet $q = 1$ implique $k = 1$ on a donc :

$$n \left|\mathbb{E}\left[W_{1,h_n}\right]\right| = 0 \leq C_1 \left(nF\left(h_n\right)\right)^{\frac{q}{2}}.$$

On peut montrer assez aisément que (H_1) est vérifiée lorque $q = 2$ en utilisant notamment les hypothèses (7.9) et (7.10) qui permettent de contrôler les covariances. Dans la preuve du Lemme 7.5.1 on a montré que

$$\sum_{i=1}^{n} \mathbb{E}\left[W_{i,h_n}^2\right] \leq C\ nF\left(h_n\right).$$

Il nous suffit donc de montrer que

$$\sum_{i=1}^{n} \sum_{1 \leq j \leq n} \left|\mathbb{E}\left[W_{i,h_n} W_{i+j,h_n}\right]\right| \leq C\ nF\left(h_n\right).$$

Pour ce faire, pour toute suite u_n, on peut faire la décomposition suivante :

$$\sum_{i=1}^{n} \sum_{1 \leq j \leq n} \left|\mathbb{E}\left[W_{i,h_n} W_{i+j,h_n}\right]\right| \leq A + B, \tag{7.22}$$

où $A = \sum_{i=1}^{n} \sum_{1 \leq j \leq u_n} \left|\mathbb{E}\left[W_{i,h_n} W_{i+j,h_n}\right]\right|$ et $B = \sum_{i=1}^{n} \sum_{u_n < j \leq n-1} \left|\mathbb{E}\left[W_{i,h_n} W_{i+j,h_n}\right]\right|$.

Etudions tout d'abord A :

$$\begin{aligned}
A &\leq C\sum_{i=1}^{n}\sum_{1\leq j\leq u_n}\mathbb{E}\left[|Y_iY_{i+j}|\,K\left(\frac{d\left(X_i,x\right)}{h_n}\right)K\left(\frac{d\left(X_{i+j},x\right)}{h_n}\right)\right]\\
&+ C\sum_{i=1}^{n}\sum_{1\leq j\leq u_n}\mathbb{E}\left[K\left(\frac{d\left(X_i,x\right)}{h_n}\right)K\left(\frac{d\left(X_{i+j},x\right)}{h_n}\right)\right]\\
&+ C\sum_{i=1}^{n}\sum_{1\leq j\leq u_n}\mathbb{E}\left[|Y_i|\,K\left(\frac{d\left(X_i,x\right)}{h_n}\right)K\left(\frac{d\left(X_{i+j},x\right)}{h_n}\right)\right]\\
&+ C\sum_{i=1}^{n}\sum_{1\leq j\leq u_n}\mathbb{E}\left[|Y_{i+j}|\,K\left(\frac{d\left(X_i,x\right)}{h_n}\right)K\left(\frac{d\left(X_{i+j},x\right)}{h_n}\right)\right]\\
&\leq C\sum_{i=1}^{n}\sum_{1\leq j\leq u_n}\mathbb{E}\left[K\left(\frac{d\left(X_i,x\right)}{h_n}\right)K\left(\frac{d\left(X_{i+j},x\right)}{h_n}\right)\right]\\
&\leq C\sum_{i=1}^{n}\sum_{1\leq j\leq u_n}\Theta_2\left(h_n\right)\\
&\leq Cnu_n\Theta_2\left(h_n\right). \qquad (7.23)
\end{aligned}$$

On obtient la quatrième ligne avec les hypothèses (7.2), (7.6), (7.9) et (7.8) en conditionnant par (X_i, X_j).

Obtenons à présent une majoration pour B. Une application de l'inégalité (7.15) pour des variables α-mélangeantes qui ont des moments d'odre p donne

$$\left|\mathbb{E}\left[W_{i,h_n}W_{j,h_n}\right]\right| \leq C\left[\alpha\left(|i-j|\right)\right]^{\frac{p-2}{p}}F\left(h_n\right)^{\frac{2}{p}}.$$

Par conséquent on peut majorer B. On a pour tout η suffisamment petit :

$$\begin{aligned}
B &= \sum_{i=1}^{n}\sum_{u_n<j\leq n-1}\left|\mathbb{E}\left[W_{i,h_n}W_{i+j,h_n}\right]\right|\\
&\leq C\sum_{i=1}^{n}\sum_{u_n<j\leq n-1}\left[\alpha\left(j\right)\right]^{\frac{p-2}{p}}F\left(h_n\right)^{\frac{2}{p}}\\
&\leq CF\left(h_n\right)^{\frac{2}{p}}n\sum_{u_n<k\leq n-1}k^{-\frac{a(p-2)}{p}}\\
&\leq CF\left(h_n\right)^{\frac{2}{p}}nu_n^{-\frac{a(p-2)}{p}+1+\eta}\sum_{u_n<k\leq n-1}k^{-1-\eta}\\
&\leq CF\left(h_n\right)^{\frac{2}{p}}nu_n^{-\frac{a(p-2)}{p}+1+\eta}, \qquad (7.24)
\end{aligned}$$

puisque $-\frac{a(p-2)}{p}+1+\eta<0$ pour $\eta>0$ assez petit d'après (7.10).

En combinant (7.22), (7.23) et (7.24) on obtient

$$\sum_{i=1}^{n}\sum_{1\leq j\leq n}\left|\mathbb{E}\left[W_{i,h_n}W_{i+j,h_n}\right]\right| \leq Cn\left(u_n\Theta_2\left(h_n\right)+F\left(h_n\right)^{\frac{2}{p}}u_n^{-\frac{a(p-2)}{p}+1+\eta}\right)$$

On prend u_n de façon à équilibrer les deux termes de droite :

$$u_n=\frac{\left(F\left(h_n\right)\right)^{\frac{2}{a(p-2)-\eta p}}}{\left(\Theta_2\left(h_n\right)\right)^{\frac{p}{a(p-2)-\eta p}}}.$$

On a donc

$$\begin{aligned}\sum_{i=1}^{n}\sum_{1\leq j\leq n}\left|\mathbb{E}\left[W_{i,h_n}W_{i+j,h_n}\right]\right| &\leq Cn\left(F\left(h_n\right)\right)^{\frac{2}{a(p-2)-\eta p}}\Theta_2\left(h_n\right)^{\frac{a(p-2)-\eta p-p}{a(p-2)-\eta p}}\\ &\leq Cn\left(F\left(h_n\right)\right)^{\frac{2+(1+\nu)(a(p-2)-\eta p-p)}{a(p-2)-\eta p}}\\ &\leq Cn\left(F\left(h_n\right)\right)^{1+\frac{2-p+\nu(a(p-2)-\eta p-p)}{a(p-2)-\eta p}}\\ &\leq CnF\left(h_n\right)\end{aligned}$$

On finit la preuve en remarquant que $F\left(h\right)\leq 1$ et que pour $\eta>0$ assez petit on a $2-p+\nu\left(a\left(p-2\right)-\eta p-p\right)>0$ d'après l'hypothèse (7.10).

□

On vient de montrer que l'hypothèse de récurrence est vraie au rang $m=1$. La prochaine étape est de montrer qu'elle est héréditaire.

Preuve de l'héréditarité de (H_m) *:.* — On va maintenant supposer que l'hypothèse de récur-rence est vérifiée au rang $m\geq 1$ et on va démontrer qu'alors elle est vraie au rang $m+1$. L'idée est de faire le même type de raisonnement que dans la preuve de (H_1). On va utiliser des inégalités sur les covariances de variables aléatoires α-mélangeantes en prenant bien soin de choisir les indices les plus éloignés possibles. On se fixe $k\leq 2m+2$ et le vecteur d'exposants $q_j\in\mathbb{N}^*, 1\leq j\leq k$ tels que $q=\sum_{j=1}^{k}q_j\leq 2m+2$. A cause d'une précédente remarque on peut aussi supposer que $q\geq 2$ puisque si $q=1$ la majoration à démontrer est toujours vérifiée.

$$\begin{aligned}&\sum_{i_1=1}^{n}\sum_{1\leq e_1,e_2,\ldots,e_{k-1}\leq n}\left|\mathbb{E}\left[\prod_{j=1}^{k}W_{i_1+\sum_{l=1}^{j-1}e_l,h_n}^{q_j}\right]\right|\\ \leq\ &\sum_{i_1=1}^{n}\sum_{s=1}^{k-1}\sum_{e_s=1}^{n}\sum_{1\leq e_j\leq e_s,j\neq s}\left|\mathbb{E}\left[\prod_{j=1}^{k}W_{i_1+\sum_{l=1}^{j-1}e_l,h_n}^{q_j}\right]\right|. \qquad (7.25)\end{aligned}$$

Pour plus de clarté, on s'intéresse dans un premier temps à la majoration de la partie finale du membre de droite de l'inégalité précédente.

$$\begin{aligned}
&\sum_{i_1=1}^{n}\sum_{e_s=1}^{n}\sum_{1\le e_j\le e_s, j\neq s}\left|\mathbb{E}\left[\prod_{j=1}^{k}W^{q_j}_{i_1+\sum_{l=1}^{j-1}e_l,h_n}\right]\right| \\
\le\ &\sum_{i_1=1}^{n}\sum_{e_s=1}^{u_n}\sum_{1\le e_j\le e_s, j\neq s}\left|\mathbb{E}\left[\prod_{j=1}^{k}W^{q_j}_{i_1+\sum_{l=1}^{j-1}e_l,h_n}\right]\right| \\
&+\sum_{i_1=1}^{n}\sum_{e_s=u_n+1}^{n}\sum_{1\le e_j\le e_s, j\neq s}\left|\mathbb{E}\left[\prod_{j=1}^{k}W^{q_j}_{i_1+\sum_{l=1}^{j-1}e_l,h_n}\right]\right| \\
=\ &A+B. \qquad (7.26)
\end{aligned}$$

On va étudier et borner séparément A et B. En ce qui concerne A, comme les indices des variables sont pas très éloignés on utilise les hypothèses (7.9) et (7.10).

$$\begin{aligned}
A &= \sum_{i_1=1}^{n}\sum_{e_s=1}^{u_n}\sum_{1\le e_j\le e_s, j\neq s}\left|\mathbb{E}\left[\mathbb{E}\left[\prod_{j=1}^{k}W^{q_j}_{i_1+\sum_{l=1}^{j-1}e_l,h_n}\mid X_{i_1}, X_{i_1+e_1}, \ldots, X_{i_1+\sum_{l=1}^{k-1}e_l}\right]\right]\right| \\
&\le C\sum_{i_1=1}^{n}\sum_{e_s=1}^{u_n}\sum_{1\le e_j\le e_s, j\neq s}\Theta_k(h_n) \\
&\le Cnu_n^{k-1}F^{1+\nu_k}(h_n). \qquad (7.27)
\end{aligned}$$

Passons maintenant à une majoration de B qui utilise l'inégalité (7.15).

$$\begin{aligned}
&B \\
&\le \sum_{i_1=1}^{n}\sum_{e_s=u_n+1}^{n}\sum_{1\le e_j\le e_s, j\neq s}\Bigg(\left|\mathbb{E}\left[\prod_{j=1}^{s}W^{q_j}_{i_1+\sum_{v=1}^{j-1}e_v,h_n}\right]\right|\left|\mathbb{E}\left[\prod_{j=s+1}^{k}W^{q_j}_{i_1+\sum_{v=1}^{j-1}e_v,h_n}\right]\right| \\
&+C\alpha(i_s)^{1-\frac{\sum_{v=1}^{s}q_v}{p}-\frac{\sum_{v=s+1}^{k}q_v}{p}}\left\|\prod_{j=1}^{s}W^{q_j}_{i_1+\sum_{v=1}^{j-1}e_v,h_n}\right\|_{\frac{p}{\sum_{v=1}^{s}q_v}}\left\|\prod_{j=s+1}^{k}W^{q_j}_{i_1+\sum_{v=1}^{j-1}e_v,h_n}\right\|_{\frac{p}{\sum_{l=s+1}^{k}q_l}}\Bigg) \\
&\le \sum_{i_1=1}^{n}\sum_{1\le e_j\le n, j<s}\left|\mathbb{E}\left[\prod_{j=1}^{s}W^{q_j}_{i_1+\sum_{v=1}^{j-1}e_v,h_n}\right]\right|\sum_{i_0=1}^{n}\sum_{1\le e_j\le n, j>s}\left|\mathbb{E}\left[\prod_{j=s+1}^{k}W^{q_j}_{i_0+\sum_{v=s+1}^{j-1}e_v,h_n}\right]\right| \\
&+C\sum_{i_1=1}^{n}\sum_{e_s=u_n+1}^{n}\sum_{1\le e_j\le e_s, j\neq s}\alpha(e_s)^{\frac{p-q}{p}}F^{\frac{q}{p}}(h_n) \\
&\le B_1+B_2
\end{aligned}$$

Pour continuer et majorer B_1 on fait le constat suivant : de deux choses l'une, soit $s=1$ et $q_1=1$ ou $s=k-1$ et $q_k=1$ et alors l'une des deux espérances est nulle car les $W_{i,h}$ sont centrées, soit l'hypothèse de récurrence (H_m) permet d'obtenir les

majorations suivantes où les K_j sont des constantes indépendantes de n :

$$\sum_{i_1=1}^{n}\sum_{1\leq e_j\leq n, j<s}\left|\mathbb{E}\left[\prod_{j=1}^{s}W^{q_j}_{i_1+\sum_{v=1}^{j-1}e_l,h_n}\right]\right| \leq K_1\sqrt{nF(h_n)}^{\sum_{v=1}^{s}q_v}, \tag{7.28}$$

$$\sum_{i_0=1}^{n}\sum_{1\leq e_j\leq n, j>s}\left|\mathbb{E}\left[\prod_{j=s+1}^{k}W^{q_j}_{i_0+\sum_{v=s+1}^{j-1}e_v,h_n}\right]\right| \leq K_2\sqrt{nF(h_n)}^{\sum_{v=s+1}^{k}q_v}. \tag{7.29}$$

En effet, on peut montrer que dans le cas où aucune des espérences n'est nulle, on a $s\leq 2m$. On a directement $s\leq k-1\leq 2m+2-1=2m+1$. Si $k=2m+2$ et $s=k-1$ alors $q_j=1,\forall 1\leq j\leq 2m+2$ et alors la seconde espérance est nulle. On a donc $k<2m+2$ ou $s<k-1$ et donc $s\leq 2m$. On peut montrer de la même manière que $k-s\leq 2m$. On montre également que lorsque $B_1\neq 0$, on a $\sum_{i=1}^{s}q_j\leq 2m$ (on sait que $\sum_{i=1}^{k}q_j\leq 2m+2$ et que $\sum_{j=s+1}^{k}q_j\geq 2$ puisque soit $s<k-1$, soit $q_k\geq 2$) et $\sum_{i=s+1}^{k}q_j\leq 2m$. A partir de ce qui précède on remarque que soit $B_1=0$, soit $q\geq 4$ (puisque $\sum_{j=s+1}^{k}q_j\geq 2$ et $\sum_{j=1}^{s}q_j\geq 2$). Tout cela nous permet d'appliquer (H_m) pour obtenir les majorations précédentes. A partir des inégalités (7.28) et (7.29) on aboutit à :

$$B_1 \leq K_3\sqrt{nF(h_n)}^{q} \tag{7.30}$$

La dernière ligne provient du fait que $q=\sum_{i=1}^{k}q_i$. L'hypothèse (7.10) permet d'obtenir que si $\eta>0$ est assez petit,alors $k-1-a\frac{p-q}{p}+\eta<0$ et d'aboutir à la majoration suivante (où C ne dépend pas de n) :

$$\begin{aligned}
B_2 &\leq Cn\sum_{i_s=u_n}^{n} i_s^{k-2}\alpha(i_s)^{\frac{p-q}{p}}F^{\frac{q}{p}}(h_n)\\
&\leq Cn\sum_{i=u_n}^{n} i^{k-2-a\frac{p-q}{p}+1+\eta}i^{-1-\eta}F^{\frac{q}{p}}(h_n)\\
&\leq Cn(u_n)^{k-2-a\frac{p-q}{p}+1+\eta}F^{\frac{q}{p}}(h_n)\sum_{i=u_n}^{n} i^{-1-\eta}\\
&\leq Cn(u_n)^{k-1-a\frac{p-q}{p}+\eta}F^{\frac{q}{p}}(h_n). \qquad (7.31)
\end{aligned}$$

On peut facilement voir que les majorations (7.27) et (7.31) sont respectivement croissantes et décroissantes par rapport à u_n. L'idée est maintenant de choisir u_n de façon à équilibrer les ordres de grandeur de ces majorations. On choisit donc

$$u_n=(F(h_n))^{\frac{q-p(1+\nu_k)}{a(p-q)-\eta p}}.$$

Avec ce choix de u_n on aboutit à la majoration suivante :

$$A+B_2\leq Cn(F(h_n))^{\frac{(k-1)[q-p(1+\nu_k)]}{a(p-q)-\eta p}+1+\nu_k}.$$

En utilisant (7.10) on peut voir que $(k-1)\left[q-p\left(1+\nu_k\right)\right]+\nu_k\left(a\left(p-q\right)-\eta p\right)>0$ si η est assez petit et on a donc

$$A+B_2\leq CnF\left(h_n\right). \tag{7.32}$$

On déduit aisément de (7.25), (7.26), (7.30) et (7.32) la majoration suivante :

$$\begin{aligned}\sum_{i_0=1}^{n}\sum_{1\leq i_1,i_2,\ldots,i_{k-1}\leq n}\left|\mathbb{E}\left[\prod_{j=1}^{k}W_{i_0+\sum_{v=1}^{j-1}i_v,h_n}^{q_j}\right]\right| &\leq K_4\sqrt{nF\left(h_n\right)}^{q}+K_5nF\left(h_n\right)\\ &\leq C_1\sqrt{nF\left(h_n\right)}^{q}. \qquad (7.33)\end{aligned}$$

La dernière ligne provient du fait que $q\geq 2$ et $nF\left(h_n\right)\to+\infty$. □

On en conclut donc que la propriété (H_m) est héréditaire et donc qu'elle est vraie pour tout $m\in\mathbb{N}^*$ tel que $m\leq\left[\frac{min(t,u)}{2}\right]$ sous les hypothèses (7.9) et (7.10). On conclut directement à partir de (7.33) et (7.21) que

$$\mathbb{E}\left[\left|\sum_{i=1}^{n}W_{i,h_n}\right|^{2m}\right]\leq C_1\left(nF\left(h_n\right)\right)^m.$$

□

□

Ainsi (7.17) est établie et la preuve de ce lemme peut être achevée comme celle du Lemme 7.2.3. □

Preuve du Théorème 7.3.1 :. — On remarque pour commencer que l'on a $\mathbb{E}\left[\hat{r}\left(x\right)\right]=r\left(x\right)+B_n+o\left(h_n\right)+O\left(\frac{1}{nF(h_n)}\right)$ (voir [167] pour plus de détails dans le cas indépendant). De plus ceci reste vrai sous nos hypothèses lorsque $u\geq 1$ (par conséquent $\mathbb{E}\left[\hat{r}\left(x\right)|X_1,\ldots,X_n\right]\leq M$ p.s.) avec la même décomposition et les résultats de [134] concernant la variance de $\hat{f}\left(x\right)$ et la covariance de $\hat{f}\left(x\right)$ et de $\hat{g}\left(x\right)$. Donc

$$Z_n=\sqrt{\frac{nF\left(h_n\right)M_1^2}{M_2\sigma_\epsilon^2}}\left(\hat{r}\left(x\right)-\mathbb{E}\left[\hat{r}\left(x\right)\right]\right)+o\left(h_n\sqrt{nF\left(h_n\right)}\right)+O\left(\frac{1}{\sqrt{nF\left(h_n\right)}}\right).$$

On a donc que $U_n:=\sqrt{\frac{nF(h_n)M_1^2}{M_2\sigma_\epsilon^2}}\left(\hat{r}\left(x\right)-\mathbb{E}\left[\hat{r}\left(x\right)\right]\right)=Z_n+o\left(1\right)$ puisque $\sqrt{nF\left(h_n\right)}h_n\leq C$ et par conséquent, lorsque $l\in\mathbb{N}$, U_n^l est uniformément intégrable et converge en loi vers W^l où W est une variable aléatoire gaussienne centrée réduite puisque $\left(Z_n\right)^l$ possède ces propriétés. En effet, $\left|U_n\right|^l\leq 2^l\left|Z_n\right|^l+o\left(1\right)$ et ce majorant est uniformément intégrable puisque $\left|Z_n\right|^l$ l'est . De plus il est clair que U_n converge vers W en loi à cause de la convergence de Z_n et comme $x\mapsto x^l$ est continue sur $\mathbb{R}$

on en déduit que U_n^l converge en loi vers W^l. On obtient donc de la même manière que dans la preuve des Théorèmes 7.2.1 et 7.2.4 :

$$\mathbb{E}\left[\left(\sqrt{\frac{nF\left(h_n\right)M_1^2}{M_2\sigma_\epsilon^2}}\left(\hat{r}\left(x\right)-\mathbb{E}\left[\hat{r}\left(x\right)\right]\right)\right)^l\right] \overset{n\to+\infty}{\to} \mathbb{E}\left[W^l\right].$$

On peut obtenir des résultats similaires avec les valeurs absolues et sans supposer que l est entier. En effet, pour tout $l \in [0;p[$, $|U_n|^l$ est uniformément intégrable et converge en loi vers $|W|^l$ où W est une variable aléatoire gaussienne centrée réduite puisque $|Z_n|^l$ a ces propriétés. On obtient donc de la même manière que dans la preuve des Thérèmes 7.2.1 et 7.2.4 :

$$\mathbb{E}\left[\left|\sqrt{\frac{nF\left(h_n\right)M_1^2}{M_2\sigma_\epsilon^2}}\left(\hat{r}\left(x\right)-\mathbb{E}\left[\hat{r}\left(x\right)\right]\right)\right|^l\right] \overset{n\to+\infty}{\to} \mathbb{E}\left[\left|W\right|^l\right].$$

□

Preuve du Théorème 7.3.2 :. — **Preuve de (i) :**. — On part du fait que

$$T_n := \left|\sqrt{\frac{nF\left(h_n\right)M_1^2}{M_2\sigma_\epsilon^2}}\left(\hat{r}\left(x\right)-r\left(x\right)\right)\right|^q = \left|Z_n+\sqrt{\frac{nF\left(h_n\right)M_1^2}{M_2\sigma_\epsilon^2}}B_n\right|^q.$$

Les Lemmes 7.2.2 et 7.2.3 ou 7.2.5 et 7.2.6 montrent que $|Z_n|^q$ est uniformément intégrable et que Z_n converge en loi vers W. L'hypothèse (7.7) permet d'obtenir que $\zeta_n = \sqrt{\frac{nF(h_n)M_1^2}{M_2\sigma_\epsilon^2}}B_n$ est borné par une constante C. Une des conséquences de ceci est que T_n est uniformément intégrable puisque l'on peut montrer que

$$T_n \leq 2^q\left|\sqrt{\frac{nF\left(h_n\right)M_1^2}{M_2\sigma_\epsilon^2}}B_n\right|^q + 2^q\left|Z_n\right|^q,$$

et remarquer que le majorant est uniformément intégrable. D'autre part on se fixe $\epsilon > 0$ et on fait la décomposition suivante :

$$\begin{aligned} & \mathbb{E}\left[\left|Z_n+\zeta_n\right|^q\right]-\mathbb{E}\left[\left|W+\zeta_n\right|^q\right] \\ = \;& \mathbb{E}\left[\left|Z_n+\zeta_n\right|^q 1_{|Z_n|\leq M}\right]-\mathbb{E}\left[\left|W+\zeta_n\right|^q 1_{|W|\leq M}\right] \\ + \;& \mathbb{E}\left[\left|Z_n+\zeta_n\right|^q 1_{|Z_n|> M}\right]-\mathbb{E}\left[\left|W+\zeta_n\right|^q 1_{|W|> M}\right] \end{aligned} \tag{7.34}$$

Dans un premier temps on remarque que $|Z_n| > M$ (respectivement $|W| > M$) implique que $|Z_n+\zeta_n| > M-C$ (respectivement $|W+\zeta_n| > M-C$). On peut donc majorer la dernière ligne de (7.34) :

$$\begin{aligned} & \left|\mathbb{E}\left[\left|Z_n+\zeta_n\right|^q 1_{|Z_n|>M}\right]-\mathbb{E}\left[\left|W+\zeta_n\right|^q 1_{|W|>M}\right]\right| \\ \leq \;& \mathbb{E}\left[\left|Z_n+\zeta_n\right|^q 1_{|Z_n+\zeta_n|>M-C}\right]+\mathbb{E}\left[\left|W+\zeta_n\right|^q 1_{|W+\zeta_n|>M-C}\right]. \end{aligned}$$

Enfin, l'uniforme intégrabilité de T_n et de $|W+\zeta_n|^q$ permet d'obtenir pour M assez grand ($M \geq M_0$), les majorations suivantes :

(7.35)

$$\limsup_{n\to+\infty}\mathbb{E}\left[\left|Z_n+\zeta_n\right|^q 1_{|Z_n+\zeta_n|>M-C}\right] < \frac{\epsilon}{3},\ \limsup_{n\to+\infty}\mathbb{E}\left[\left|W+\zeta_n\right|^q 1_{|W+\zeta_n|>M-C}\right] < \frac{\epsilon}{3}.$$

On introduit la famille de fonctions $\mathcal{F}_{C,M} = \left\{ f_{\zeta,M}(x) = 1_{|x|\leq M} |x+\zeta|^q, |\zeta| \leq C \right\}$. On utilisera le lemme suivant :

Lemme 7.5.3. — *Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction continue et soit*

$$\mathcal{F}_{C,M,f} = \left\{ f_{\zeta,M}(x) = 1_{|x|\leq M} f(x+\zeta), |\zeta| \leq C \right\}.$$

Si Z_n converge en loi vers Z (dont la loi n'a pas d'atome) alors

$$\sup_{|\zeta|\leq C} |\mathbb{E}[f_{\zeta,M}(Z_n)] - \mathbb{E}[f_{\zeta,M}(Z)]| \to 0.$$

Preuve du Lemme 7.5.3 :. — On va s'inspirer très fortement de la preuve du Théorème de Helly-Bray ou Portemanteau. Par soucis de clarté on note $f_\zeta = f_{\zeta,M}$. Remarquons tout d'abord que puisque f est continue, elle est bornée et uniformément continue sur tout compact donc notamment sur $\mathcal{K} = [-M-C; M+C]$:

$$\forall \epsilon > 0, \exists \delta_{f,\epsilon} > 0, \forall z_1, z_2 \in \mathcal{K}, |z_1 - z_2| < \delta_{f,\epsilon} \Rightarrow |f(z_1) - f(z_2)| < \epsilon.$$

En notant que $f_\zeta(z) = f(z+\zeta)$ si $|z| \leq M$, on a pour tous $z_1, z_2 \in [-M; +M]$:
(7.36)
$$|z_1 + \zeta - \zeta - z_2| = |z_1 - z_2| < \delta_{f,\epsilon} \Rightarrow |f_\zeta(z_1) - f_\zeta(z_2)| = |f(z_1+\zeta) - f(z_2+\zeta)| < \epsilon$$

Soit $\epsilon > 0$ fixé. Comme $[-M; +M]$ est un compact, on peut le recouvrir par un nombre fini d'intevales A_j de longeur strictement plus petite que $\delta_{f,\epsilon}$:

$$[-M; +M] \subset_{j=1}^{J} A_j, \forall j, z_1, z_2 \in A_j \Rightarrow |z_1 - z_2| < \delta_{f,\epsilon}.$$

Pour toute fonction f_ζ l'idée est de construire à partir de la remarque (7.36) une fonction en escaliers $f_{\zeta,\epsilon}$ telle que $\sup_z |f_\zeta(z) - f_{\zeta,\epsilon}(z)| < \epsilon$. On la définit de la manière suivante :

$$f_{\zeta,\epsilon}(z) = \sum_{j=1}^{J} f_\zeta(z_j) 1_{[-M;+M]}(z) 1_{A_j}(z),$$

où z_j est un point de $A_j \cap [-M; M]$. Pour obtenir le résultat recherché on fait la décomposition suivante :

$$\begin{aligned}
&\sup_{|\zeta|<C} |\mathbb{E}[f_\zeta(Z_n)] - \mathbb{E}[f_\zeta(Z)]| \\
&= \sup_{|\zeta|<C} |\mathbb{E}[f_\zeta(Z_n)] - \mathbb{E}[f_{\zeta,\epsilon}(Z_n)] + \mathbb{E}[f_{\zeta,\epsilon}(Z_n)] - \mathbb{E}[f_{\zeta,\epsilon}(Z)] + \mathbb{E}[f_{\zeta,\epsilon}(Z)] - \mathbb{E}[f_\zeta(Z)]| \\
&\leq \sup_{|\zeta|<C} |\mathbb{E}[f_\zeta(Z_n)] - \mathbb{E}[f_{\zeta,\epsilon}(Z_n)]| + \sup_{|\zeta|<C} |\mathbb{E}[f_{\zeta,\epsilon}(Z_n)] - \mathbb{E}[f_{\zeta,\epsilon}(Z)]| \\
&+ \sup_{|\zeta|<C} |\mathbb{E}[f_{\zeta,\epsilon}(Z)] - \mathbb{E}[f_\zeta(Z)]| \\
&\leq 2\epsilon + \sup_{|\zeta|<C} |\mathbb{E}[f_{\zeta,\epsilon}(Z_n)] - \mathbb{E}[f_{\zeta,\epsilon}(Z)]|.
\end{aligned}$$

(7.37)

On obtient la dernière ligne en utilisant que $f_{\zeta,\epsilon}$ est proche de f_ζ. Pour conclure il faut montrer que le dernier supremum tend vers 0.

$$\begin{aligned}
&\sup_{|\zeta|<C} |\mathbb{E}\left[f_{\zeta,\epsilon}\left(Z_n\right)\right] - \mathbb{E}\left[f_{\zeta,\epsilon}\left(Z\right)\right]| \\
= &\sup_{|\zeta|<C} \left|\mathbb{E}\left[\sum_{j=1}^{J} f_\zeta\left(z_j\right)\left(1_{A_j\cap[-M;+M]}\left(Z_n\right) - 1_{A_j\cap[-M;+M]}\left(Z\right)\right)\right]\right| \\
= &\sup_{|\zeta|<C} \sum_{j=1}^{J} \left|f_\zeta\left(z_j\right)\right| \left|\mathbb{P}\left(Z_n \in A_j \cap [-M;+M]\right) - \mathbb{P}\left(Z \in A_j \cap [-M;+M]\right)\right| \\
= &\, O\left(\sum_{j=1}^{J} \left|\mathbb{P}\left(Z_n \in A_j \cap [-M;+M]\right) - \mathbb{P}\left(Z \in A_j \cap [-M;+M]\right)\right|\right) \\
= &\, o\left(1\right).
\end{aligned}$$

Les dernières lignes proviennent du fait que f est bornée sur $\mathcal{K}$ et que puisque Z_n converge vers Z en loi chacune des quantités

$$\mathbb{P}\left(Z_n \in A_j \cap [-M;+M]\right) - \mathbb{P}\left(Z \in A_j \cap [-M;+M]\right)$$

converge vers 0 lorsque $n \to +\infty$ car la loi de Z n'a pas d'atomes. On en conclut donc que :

$$\forall \epsilon > 0, \exists N_0, \forall n \geq N_0, \ \sup_{|\zeta|<C} |\mathbb{E}\left[f_{\zeta,\epsilon}\left(Z_n\right)\right] - \mathbb{E}\left[f_{\zeta,\epsilon}\left(Z\right)\right]| < \epsilon. \tag{7.38}$$

On obtient directement en utilisant (7.37) et (7.38) que

$$\forall \epsilon > 0, \exists N_0, \forall n \geq N_0, \ \sup_{|\zeta|<C} |\mathbb{E}\left[f_{\zeta}\left(Z_n\right)\right] - \mathbb{E}\left[f_{\zeta}\left(Z\right)\right]| < 3\epsilon,$$

ce qui achève la preuve.

□

En appliquant le lemme précédent à la famille $\mathcal{F}_{C,M}$ on en conclut que, pour tout M :

$$\begin{aligned}
|\mathbb{E}\left[f_{\zeta_n,M}\left(Z_n\right)\right] - \mathbb{E}\left[f_{\zeta_n,M}\left(W\right)\right]| &\leq \sup_{|\zeta|\leq C} |\mathbb{E}\left[f_{\zeta,M}\left(Z_n\right)\right] - \mathbb{E}\left[f_{\zeta,M}\left(W\right)\right]| \\
&= o\left(1\right).
\end{aligned}$$

Par conséquent il existe N_0 tel que pour tout $n \geq N_0$ on a

$$|\mathbb{E}\left[f_{\zeta_n,M}\left(Z_n\right)\right] - \mathbb{E}\left[f_{\zeta_n,M}\left(W\right)\right]| \leq \frac{\epsilon}{3}. \tag{7.39}$$

On aboutit ansi à partir de (7.34), (7.35) et (7.39) à la conclusion suivante :

$$\forall \epsilon > 0, \ \exists N_0, \forall n \geq N_0, \ |\mathbb{E}\left[|Z_n + \zeta_n|^q\right] - \mathbb{E}\left[|W + \zeta_n|^q\right]| < \epsilon.$$

Ceci achève la preuve du (i). Avec un raisonnement identique on peut obtenir des résultats semblables sans les valeurs absolues lorsque $q \in \mathbb{N}$. Il suffit d'utiliser le lemme avec la famille $\left\{f_{\zeta,M}\left(x\right) = 1_{|x|\leq M}\left(x+\zeta\right)^q, |\zeta| \leq C\right\}$ et de remarquer que

$$\left|\mathbb{E}\left[\left(Z_n + \zeta_n\right)^q 1_{|Z_n|>M}\right]\right| \leq \mathbb{E}\left[\left|Z_n + \zeta_n\right|^q 1_{|Z_n|>M}\right]$$

□

Preuve de (ii) ∴ — Dans le cas où q est un entier pair l'erreur $\mathbb{L}^q$ est le moment d'ordre q de $\hat{r}(x) - r(x)$; on peut donc se passer des valeurs absolues. On va montrer un résultat plus général que (ii) : sans les valeurs absolues et pour $q = l$ entier quelconque. On a comme conséquence de la deuxième partie de la preuve de (i) le résultat :

$$\begin{aligned}\mathbb{E}\left[(\hat{r}(x) - r(x))^l\right] &= \mathbb{E}\left[\left(W\sqrt{\frac{M_2\sigma_\epsilon^2}{nF(h_n)M_1^2}} + B_n\right)^l\right] + o\left(\frac{1}{(nF(h_n))^{\frac{l}{2}}}\right) \\ &= \sum_{k=0}^{l} C_l^k \mathbb{E}\left[W^k\right]\left(\sqrt{\frac{M_2\sigma_\epsilon^2}{nF(h_n)M_1^2}}\right)^k (B_n)^{l-k} + o\left(\frac{1}{(nF(h_n))^{\frac{l}{2}}}\right)\end{aligned}$$

On peut à présent remarquer que les moments impairs de W sont nuls tandis que ceux d'odre $2p$ valent $\frac{(2p)!}{2^p p!}$. On en déduit aisément que l'on a :

$$\begin{aligned}&\mathbb{E}\left[(\hat{r}(x) - r(x))^l\right] \\ =& \sum_{p=0}^{\left[\frac{l}{2}\right]} C_l^{2p}\left(\frac{M_2\sigma_\epsilon^2}{M_1^2 nF(h_n)}\right)^p \frac{(2p)!}{2^p p!}\left(\phi'(0)\frac{M_0}{M_1}h_n\right)^{l-2p} + o\left(\frac{1}{(nF(h_n))^{\frac{l}{2}}}\right) \\ =& \sum_{p=0}^{\left[\frac{l}{2}\right]}\left(\frac{M_2\sigma_\epsilon^2}{M_1^2}\right)^p\left(\frac{M_0}{M_1}\phi'(0)\right)^{l-2p}\frac{l!}{(l-2p)!p!2^p}\frac{h_n^{l-2p}}{(nF(h_n))^p} + o\left(\frac{1}{(nF(h_n))^{\frac{l}{2}}}\right) \\ =& \sum_{p=0}^{\left[\frac{l}{2}\right]}\frac{M_2^p M_0^{l-2p}(\phi'(0))^{l-2p}(\sigma_\epsilon^2)^p}{M_1^l}\frac{l!}{(l-2p)!p!2^p}\frac{h_n^{l-2p}}{(nF(h_n))^p} + o\left(\frac{1}{(nF(h_n))^{\frac{l}{2}}}\right).\end{aligned}$$

Ceci achève la preuve. En effet il suffit de remplacer l par $2m$ pour obtenir (ii). □

Preuve de (iii) ∴ — On déduit du (i) que

$$\begin{aligned}\mathbb{E}\left[|\hat{r}(x) - r(x)|^{2m+1}\right] &= \left(\sqrt{\frac{M_2\sigma_\epsilon^2}{M_1^2 nF(h_n)}}\right)^{2m+1}\mathbb{E}\left[\left|\sqrt{\frac{M_1^2 nF(h_n)}{M_2\sigma_\epsilon^2}}B_n + W\right|^{2m+1}\right] \\ &\quad + o\left(\frac{1}{(nF(h_n))^{m+\frac{1}{2}}}\right).\end{aligned}$$

Pour simplifier les calculs on introduit la notation

$$U_n = \sqrt{\frac{M_1^2 nF(h_n)}{M_2\sigma_\epsilon^2}}B_n.$$

Il nous faut donc exprimer plus précisément l'espérance de droite.

$$\begin{aligned}\mathbb{E}\left[|U_n+W|^{2m+1}\right] &= -\int_{-\infty}^{-U_n}(U_n+w)^{2m+1}\frac{e^{-\frac{w^2}{2}}}{\sqrt{2\pi}}dw+\int_{-U_n}^{+\infty}(U_n+w)^{2m+1}\frac{e^{-\frac{w^2}{2}}}{\sqrt{2\pi}}dw\\ &= \sum_{k=0}^{2m+1}C_{2m+1}^{k}(U_n)^{2m+1-k}I_k,\end{aligned}$$

où

$$I_k=-\int_{-\infty}^{-U_n}\frac{w^k e^{-\frac{w^2}{2}}}{\sqrt{2\pi}}dw+\int_{-U_n}^{+\infty}\frac{w^k e^{-\frac{w^2}{2}}}{\sqrt{2\pi}}dw.$$

On utilisera le lemme suivant :

Lemme 7.5.4. — *Pour tout k entier naturel on a :*

$$\begin{aligned}I_{2k+1} &= 2g(U_n)\sum_{j=0}^{k}(U_n)^{2(k-j)}\frac{2^k k!}{2^{k-j}(k-j)!},\\ I_{2k} &= -2g(U_n)\sum_{j=0}^{k-1}(U_n)^{2(k-j)-1}\frac{(2k)!2^{k-j}(k-j)!}{2^k k!(2(k-j))!}+[2G(U_n)-1]\frac{(2k)!}{2^k k!},\end{aligned}$$

avec la convention qu'une somme sur un ensemble vide d'indices est nulle.

Preuve du Lemme 7.5.4 :. — Par des arguments de parité, on voit aisément que :

$$I_{2k}=2\int_0^{U_n}w^{2k}\frac{e^{-\frac{w^2}{2}}}{\sqrt{2\pi}}dw \text{ et } I_{2k+1}=2\int_{U_n}^{+\infty}w^{2k+1}\frac{e^{-\frac{w^2}{2}}}{\sqrt{2\pi}}dw.$$

On va raisonner par récurrence sur k. Etudions tout d'abord les I_{2k}.
Pour $k=0$, il est évident que $I_0=2G(U_n)-1$.
On suppose que la propriété est vraie pour $k=r$, et on veut montrer qu'elle est vraie pour $k=r+1$.

$$\begin{aligned}I_{2(r+1)} &= 2\int_0^{U_n}w^{2(r+1)}\frac{e^{-\frac{w^2}{2}}}{\sqrt{2\pi}}dw\\ &= 2\left[-w^{2r+1}\frac{e^{-\frac{w^2}{2}}}{\sqrt{2\pi}}\right]_0^{U_n}+(2r+1)I_{2r}\\ &= (2r+1)\left(-2g(U_n)\sum_{j=0}^{r-1}(U_n)^{2(r-j)-1}\frac{(2r)!2^{r-j}(r-j)!}{2^r r!(2(r-j))!}+[2G(U_n)-1]\frac{(2r)!}{2^r r!}\right)\\ &\quad -2g(U_n)U_n^{2r+1}\\ &= -2g(U_n)\sum_{j=0}^{r}(U_n)^{2(r+1-j)-1}\frac{(2(r+1))!2^{r+1-j}(r+1-j)!}{2^{r+1}(r+1)!(2(r+1-j))!}\\ &\quad +[2G(U_n)-1]\frac{(2(r+1))!}{2^{r+1}(r+1)!}.\end{aligned}$$

La propriété est donc héréditaire, vraie pour $k = 0$ donc vraie pour tout $k \in \mathbb{N}$. On établit le résultat pour I_{2k+1} de la même manière. Il est clair que $I_1 = 2g(U_n)$ et d'autre part

$$I_{2(r+1)+1} = 2U_n^{2(r+1)} g(U_n) + 2(r+1) I_{2r+1}.$$

□

En remplaçant I_k par son expression, on obtient en gardant comme convention qu'une somme sur un ensemble d'indices vide est nulle :

$$\begin{aligned}
\mathbb{E}\left[|U_n + W|^{2m+1}\right] &= \sum_{k=0}^{m} \left[C_{2m+1}^{2k} (U_n)^{2(m-k)+1} I_{2k} + C_{2m+1}^{2k+1} (U_n)^{2(m-k)} I_{2k+1} \right] \\
&= [2G(U_n) - 1] \sum_{k=0}^{m} C_{2m+1}^{2k} \frac{(2k)!}{2^k k!} (U_n)^{2(m-k)+1} \\
&+ 2g(U_n) \Big\{ \sum_{j=0}^{m} (U_n)^{2(m-j)} \sum_{k=j}^{m} C_{2m+1}^{2k+1} \frac{2^k k!}{2^{k-j}(k-j)!} \\
&\quad - \sum_{j=0}^{m-1} (U_n)^{2(m-j)} \sum_{k=j+1}^{m} C_{2m+1}^{2k} \frac{(2k)! 2^{k-j}(k-j)!}{2^k k! (2(k-j))!} \Big\} \\
&= [2G(U_n) - 1] P_{2m+1}(U_n) + 2g(U_n) Q_{2m+1}(U_n).
\end{aligned}$$

On en déduit facilement le résultat recherché. □

Ainsi s'achève la preuve du Théorème 7.3.2. □

CHAPITRE 8

ENGLISH NOTE ON CLT AND $\mathbb{L}^q$ ERRORS IN NONPARAMETRIC FUNCTIONAL REGRESSION

The content of the present chapter corresponds to the one of the short English note Delsol (2007a) published in the "Comptes rendus de l'Académie des Sciences" (Proceedings of the French Academy of Sciences). The aims of this chapter are both providing an English version of the main result of Chapter 7 and summarizing the main results of Chapters 6 and 7.

Abstract. We study a nonparametric functional regression model and we provide an asymptotic law with explicit constants under α-mixing assumptions. Then we establish both pointwise confidence bands for the regression operator and asymptotic $\mathbb{L}^q$ errors for its kernel estimator.

Résumé. Loi et erreurs $\mathbb{L}^q$ asymptotiques dans un modèle de régression non-paramétrique On étudiera dans cet article la normalité asymptotique de l'estimateur à noyau pour des données α-mélangeantes fonctionnelles . L'explicitation des constantes apparaissant dans la loi asymptotique permet d'établir des intervalles de confiance ponctuels pour l'opérateur de régression ainsi que l'expression des erreurs $\mathbb{L}^q$.

8.1. Introduction

We focus on the usual regression model : $Y = r(X) + \epsilon$, where Y is a real random variable and X a random variable which takes values on a semi-metric space (E, d). Only regularity assumptions are made on r that is why the model is called nonparametric. Then, we consider a dataset of n pairs (X_i, Y_i) identically distributed as (X, Y) which may be dependent.

One considers an element x of E and estimates $r(x)$ by the following kernel estimator :

$$\hat{r}(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{d(X_i,x)}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{d(X_i,x)}{h_n}\right)},$$

where K is a kernel function and h_n a smoothing parameter. This estimator has been introduced by Ferraty and Vieu [176] to generalise to the functional case the classical Nadaraya-Watson estimator. The results on $\mathbb{L}^q$ errors given below are new even in the independent and multivariate case.

8.2. Notation and main assumptions

Firstly, we introduce some key-functions which take an important place in our approach :

$$\phi(s) = \mathbb{E}\left[(r(X) - r(x))/d(X,x) = s\right], \; F(h) = \mathbb{P}(d(X,x) \leq h), \; \tau_h(s) = \frac{F(hs)}{F(h)}.$$

We start with some assumptions on the model (already introduced in [167]) :

(8.1) r is bounded on a neighbourhood of x,

(8.2) $F(0) = 0$, $\phi(0) = 0$ and $\phi'(0)$ exists,
$\sigma_\epsilon^2(x) := \mathbb{E}\left[\epsilon^2 \mid X = x\right]$ is continuous on a neighbourhood of x

(8.3) and $\sigma_\epsilon^2 := \sigma_\epsilon^2(x) > 0$,

(8.4) $\forall s \in [0;1], \lim_{n\to+\infty} \tau_{h_n}(s) = \tau_0(s)$ with $\tau_0(s) \neq 1_{[0;1]}(s)$.

Assumption (8.2) enables to get the exact expression of the constants where a standard Lipschitz condition would only give upper bounds. Then, one may note that assumptions dealing with the law of X only concern small ball probabilities. Besides, many standard processes fulfill assumption (8.4). In order to control the sum of covariances we propose the following assumptions :

(8.5) $\exists p > 2, \exists M > 0, \mathbb{E}\left[|\epsilon|^p \mid X\right] \leq M$ $a.s.$,

(8.6) $\exists C, \forall i,j \in \mathbb{Z}$ $\max\left(\mathbb{E}\left[|\epsilon_i\epsilon_j| \mid X_i, X_j\right], \mathbb{E}\left[|\epsilon_i| \mid X_i, X_j\right]\right) \leq C$ $a.s.$.

Finally we make some assumptions on the functional kernel estimator :

(8.7) $h_n = O\left(\frac{1}{\sqrt{nF(h_n)}}\right)$, $\lim_{n\to+\infty} nF(h_n) = +\infty$,

(8.8) K has a compact support $[0;1]$, is C^1 and non-increasing on $]0;1[$, $K(1) > 0$.

Our results will be expressed using the following constants :

$$M_0 = K(1) - \int_0^1 (sK(s))' \tau_0(s)\, ds \quad , \quad M_j = K^j(1) - \int_0^1 \left(K^j\right)'(s)\, \tau_0(s)\, ds, \; j = 1,2,$$

and the random variable Z_n defined with $B_n = h_n\phi'(0)\frac{M_0}{M_1}$ by :

$$Z_n := \frac{M_1}{\sqrt{M_2\sigma_\epsilon^2}}\sqrt{nF(h_n)}\left(\hat{r}(x) - r(x) - B_n\right).$$

If (X_i, Y_i) are dependent, we assume them to be α-mixing (see [376], p37, Definition 2, Notation 1.2). The mixing coefficients are denoted by $\{\alpha(n), n \in \mathbb{N}\}$ and we

introduce the notations :

$$\forall k \geq 2,\, \Theta_k(s) := \max\left(\max_{1\leq i_1<\cdots<i_k\leq n} P\left(d\left(X_{i_j},x\right)\leq s,\, 1\leq j\leq k\right),\, F^k(s)\right),$$

$$\Gamma_i := Y_i K\left(\frac{d(X_i,x)}{h_n}\right),\, \Delta_i := K\left(\frac{d(X_i,x)}{h_n}\right), U_{i,n} = \frac{\Gamma_i \mathbb{E}[\Delta_i] - \Delta_i \mathbb{E}[\Gamma_i]}{F(h_n)\sqrt{nF(h_n)}}.$$

We need the following assumptions :

$$\exists (u_n)_{n\in\mathbb{N}} \in \mathbb{N}^{\mathbb{N}},\, O\left(\frac{n\left[\alpha(u_n)\right]^{\frac{p-2}{p}}}{F(h_n)^{\frac{p-2}{p}}}\right) + O\left(u_n \frac{\Theta_2(h_n)}{F(h_n)}\right) \overset{n\to+\infty}{\to} 0, \qquad (H1)$$

$$I_n := n\int_0^1 \alpha^{-1}\left(\frac{x}{2}\right) Q^2_{U_{1,n}}(x)\, inf\left(\frac{3M_1\sqrt{\sigma_\epsilon^2 M_2}}{2}, \alpha^{-1}\left(\frac{x}{2}\right) Q_{U_{1,n}}(x)\right) dx \to 0, \quad (H2)$$

where $Q_{U_{i,n}}(x) = inf\left\{t,\, \mathbb{P}\left(|U_{i,n}|>t\right)\leq x\right\}$, and $\alpha^{-1}\left(\frac{x}{2}\right) = inf\left\{t,\, \alpha\left([t]\right)\leq \frac{x}{2}\right\}$ (see [376]). These assumptions will be simplified in the particular case of arithmetic α-mixing coefficients (see (8.9) and (8.10) below).

8.3. Asymptotic normality

Theorem 8.3.1. — *Under assumptions (8.1)-(8.8), $(H1)$ and $(H2)$ we get*

$$Z_n \to N(0,1).$$

Remark : If the α-mixing coefficients are arithmetic of order a : $\alpha(i) \leq Ci^{-a}$, assumptions (H1) and (H2) may be replaced, thanks to [287], by the following ones :

$$\exists \nu > 0,\, \Theta(h_n) = O\left(F(h_n)^{1+\nu}\right), \text{ with } a > \frac{(1+\nu)p-2}{\nu(p-2)}, \qquad (8.9)$$

$$\exists \gamma > 0,\, nF^{1+\gamma}(h_n) \to +\infty \text{ and } a > \frac{2}{\gamma}+1. \qquad (8.10)$$

8.4. Pointwise asymptotic confidence bands

To get asymptotic confidence bands one needs to estimate the constants appearing in Theorem 8.3.1. Whereas M_1 and M_2 seem to be easily estimated, the bias term is more difficult to study. To avoid this problem one makes an additional assumption on h_n which will make the bias negligible with respect to $\sqrt{nF(h_n)}^{-1}$. Taking

$$\hat{M}_2(x) := \frac{1}{n\hat{F}(h_n)}\sum_{i=1}^n K^2\left(\frac{d(X_i,x)}{h_n}\right), \hat{M}_1(x) := \frac{1}{n\hat{F}(h_n)}\sum_{i=1}^n K\left(\frac{d(X_i,x)}{h_n}\right),$$

where $\hat{F}(t) = \frac{1}{n}\sum_{i=1}^n 1_{[d(X_i,x),+\infty[}(t)$, we get the following corollary :

Corollary 8.4.1. — *Under assumptions of Theorem 8.3.1, if* $h_n\sqrt{nF(h_n)} \to 0$, *then :*

$$\frac{\hat{M}_1}{\sqrt{\hat{M}_2\hat{\sigma}_\epsilon^2}}\sqrt{n\hat{F}(h_n)}\left(\hat{r}(x) - r(x)\right) \to N(0,1),$$

where $\hat{\sigma}_\epsilon^2$ *is an estimator converging in probability to* σ_ϵ^2.

8.5. Asymptotic expressions of $\mathbb{L}^q$ errors

We get directly from the asymptotic law of Z_n and the uniform integrability of $|Z_n|^q$ the asymptotic expressions of the moments of order q and the $\mathbb{L}^q$ errors of the kernel estimator. To get the uniform integrability of $|Z_n|^q$ we need additional assumptions :

$$\exists t,\ 2 \le t < p,\ \forall k \le t,\ \exists \nu_k > 0, \Theta_k(s) = O\left(F(s)^{1+\nu_k}\right)$$

$$\text{with } \nu_k \le \nu_{k-1} + 1 \text{ and } a > \max_{2\le k\le t}(k-1)\frac{(1+\nu_k)p - t}{\nu_k(p-t)}, \tag{8.11}$$

$$\exists u,\ 2 \le u \le p,\ \exists M,\ \max_{i,n}\mathbb{E}\left[|\epsilon_i|^u \mid X_1,\ldots,X_n\right] \le M \ a.s.. \tag{8.12}$$

From a statistical point of view, to choose the smoothing parameter, one often focuses on the $\mathbb{L}^q$ errors of the kernel estimator. That is the aim of the following theorem in which we explicit their asymptotic dominant terms whether q is even or not. In particular the case q odd, leads us to introduce the two following sequences of polynomials : $P_{2m+1}(u) = \sum_{l=0}^m a_{m,l}u^{2l+1}$ and $Q_{2m+1}(u) = \sum_{l=0}^m b_{m,l}u^{2l}$, where

$$\begin{aligned}
a_{m,l} &= \frac{(2m+1)!}{(2l+1)!2^{(m-l)}(m-l)!},\\
b_{m,l} &= \sum_{j=m-l+1}^{m}\left[C_{2m+1}^{2j+1}\frac{2^j j!}{2^{j+l-m}(j+l-m)!} - C_{2m+1}^{2j}\frac{(2j)!2^{j+l-m}(j+l-m)!}{2^j j!\left(2(j+l-m)\right)!}\right]\\
&\quad + C_{2m+1}^{2(m-l)+1}2^{m-l}(m-l)!,
\end{aligned}$$

and the function $\psi_m(u) = (2G(u)-1)P_{2m+1}(u) + 2g(u)Q_{2m+1}(u)$ (with the convention that a sum on an empty set of indices equals zero).

Theorem 8.5.1. — *Under assumptions (8.1)-(8.8) and taking* $\ell = p$ *in the independent case or (8.1)-(8.12) and taking* $\ell = 2\left[\frac{\min(t,u)}{2}\right]$ *in the arithmetic* α*-mixing case, one gets :*

(i) $\forall\ 0 \le q < \ell,\ \mathbb{E}\left[|\hat{r}(x) - r(x)|^q\right] = \mathbb{E}\left[\left|h_n B + W\frac{V}{\sqrt{nF(h_n)}}\right|^q\right] + o\left(\frac{1}{(nF(h_n))^{\frac{q}{2}}}\right),$

(ii) $\forall m \in \mathbb{N},\ 0 \le 2m < \ell,$

$$\mathbb{E}\left[|\hat{r}(x) - r(x)|^{2m}\right] = \sum_{k=0}^{m}\frac{V^{2k}B^{2(m-k)}(2m)!}{(2(m-k))!k!2^k}\frac{h_n^{2(m-k)}}{(nF(h_n))^k} + o\left(\frac{1}{(nF(h_n))^m}\right),$$

(iii) $\forall m \in \mathbb{N},\ 0 \leq 2m+1 < \ell$,

$$\mathbb{E}\left[\left|\hat{r}\left(x\right) - r\left(x\right)\right|^{2m+1}\right] = \frac{1}{\left(nF\left(h_n\right)\right)^{m+\frac{1}{2}}}\left(V^{2m+1}\psi_m\left(\frac{Bh_n\sqrt{nF\left(h_n\right)}}{V}\right) + o\left(1\right)\right),$$

where $B = \phi'\left(0\right)\frac{M_0}{M_1}$ *and* $V = \sqrt{\frac{M_2\sigma_\epsilon^2}{M_1^2}}$.

8.6. Conclusions and perspectives

The above results complete [167] and [314] and give explicit constants in the asymptotic law and $\mathbb{L}^q$ errors instead of giving only upper bounds. Consequently, they may be used to provide confidence bands or choose the smoothing parameter that is asymptotically optimal with regard to the $\mathbb{L}^q$ error. Besides they generalise former works on the $\mathbb{L}^1$ (see [421]) and $\mathbb{L}^2$ errors with multivariate variables to the functional case and for other orders.

For the future, it would be of interest to get an "integrated version" of asymptotic normality of $\hat{r}\left(x\right)$ notably to construct some specification testing procedures. It would also be interesting to get similar results for the integrated errors. Finally, we might attempt to extend our results to a more general case of α-mixing variables introduced in [129].

CHAPITRE 9

ADDITIONAL COMMENTS AND APPLICATIONS

The aim of this chapter is to gather remarks and an application example that we have not written in Chapter 6.

9.1. Some comments on Theorem 6.2.1 assumptions

Firstly, assumption (6.2) replaces the usual Lipschitz assumption on r. If r is continuous, and under assumptions on the distribution of X then we obtain $\phi(0) = 0$. Existence of $\phi'(0)$ enables us to give exact values of dominant terms, what would not permit a standard Lipschitz assumption. Hypothesis (6.1) and (6.2) mix assumptions on r, that is to say the link between X and Y, semimetric d and the distribution of X. Finally, focusing on assumption (6.4), we can note that τ_0 is measurable and undecreasing as limit of such functions. We can also note that (6.4) holds when X corresponds to standard processes :

- **Unsmooth processes.** It has been shown that for these processes we have $\mathbb{P}(d(X,x) < s) \overset{s\to 0}{\sim} C_{1,x} s^{-\alpha} \exp\left(-\frac{C_2}{s^\beta}\right)$ with $\alpha \geq 0$, β, $C_{1,x}$, $C_2 > 0$ and using $\mathbb{L}^p$ or $\mathbb{L}^\infty$ metric. Consequently, (6.4) holds with $\tau_0 \equiv 1_{\{1\}}$.
- **Fractal processes.** For such processes we have $\mathbb{P}(d(X,x) < s) \overset{s\to 0}{\sim} C_{3,x} s^\delta$ with δ and $C_{3,x} > 0$ thus (6.4) holds with $\tau_0(s) = s^\delta$.
- **Finite dimensional processes.** If X takes values on $\mathbb{R}^d$ and if X has a continuous density f with regard to Lebesgues measure such that $f(x) > 0$, then we are in the previous case with $\gamma = d$, for any metric on $\mathbb{R}^d$. In conclusion hypothesis (6.8) is weaker than the standard finite dimensional one.

As we say in the proof, assumption (6.9) does not depend on the value of $F(h_n)$ but on the link between Θ and F. However assumptions (6.10) and (8.10) may be specified when we know the expression of F. Here are some examples to show what assumptions (6.10) and (8.10) mean in some standard cases.

– **Fractal processes.** For such processes, condition $nF^{1+\gamma}(h_n) \to +\infty$ together with $h_n\sqrt{nF(h_n)} = O(1)$ imply that $\gamma < \frac{2}{\delta}$. Consequently assumptions (6.10) and (8.10) become respectively $a > \max\left(2\delta, \frac{p+\delta(p-1)}{p-2}\right)$ and $a > \delta + 1$.

– **Unsmooth processes.** For such processes, conditions $nF^{1+\gamma}(h_n) \to +\infty$ and $h_n\sqrt{nF(h_n)} = O(1)$ are contradictory. Indeed, if we assume that the condition $nF^{1+\gamma}(h_n) \to +\infty$ holds, then the variance term is negligible with regard to the bias one. Consequently, we cannot obtain a result like Theorem 6.2.1. However, one gets the asymptotic normality result

$$\sqrt{\frac{nF(h_n)M_1^2}{M_2\sigma_\epsilon^2}}\left(\hat{r}(x) - \frac{\mathbb{E}[\hat{g}(x)]}{\mathbb{E}\left[\hat{f}(x)\right]}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Moreover, under some additional assumption (for instance if the residuals are assumed to be bounded or independent of $(X_1, \ldots, X_n)$ for all n), one gets

$$\frac{\mathbb{E}[\hat{g}(x)]}{\mathbb{E}\left[\hat{f}(x)\right]} - \mathbb{E}[\hat{r}(x)] = o\left(\frac{1}{\sqrt{nF(h_n)}}\right),$$

and hence one states

$$\sqrt{\frac{nF(h_n)M_1^2}{M_2\sigma_\epsilon^2}}\left(\hat{r}(x) - \mathbb{E}[\hat{r}(x)]\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

9.2. An other application example : Electricity Consumption

We now consider a data corresponding to electricity consumption in the United States of America during 28 years. Since the original process is not stationary, we make a log-differentiation. As before we cut our new process into 28 annual data and we try to study the 28^{th} year knowing the others.

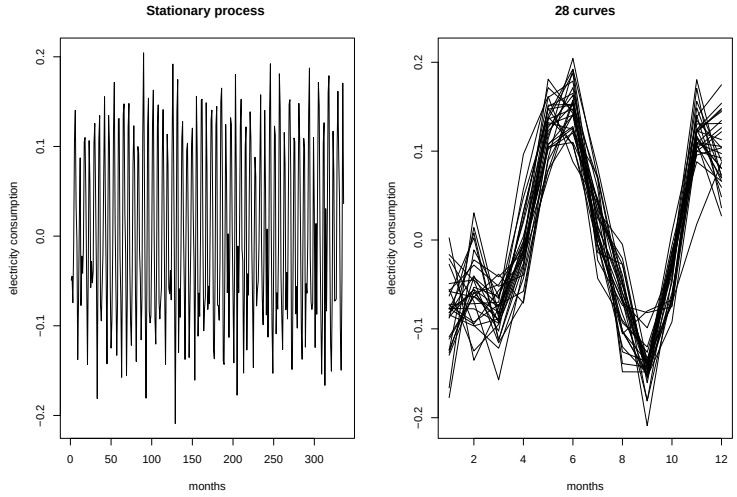

FIGURE 9.1. Complete data and 28 curves obtained

We take the semimetric and the kernel used by F. Ferraty and P. Vieu in their study. A first study consists in plotting the predicted and the real 28^{th} year. The prediction appears through the dotted line.

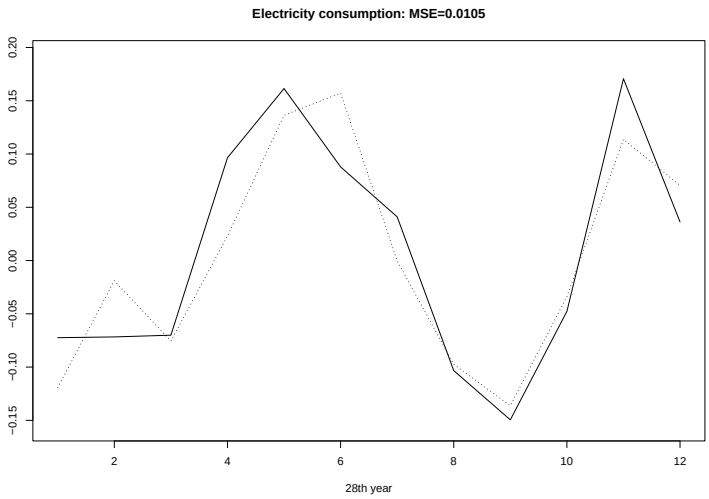

FIGURE 9.2. Prediction of the 28^{th} year

We obtain the following figure which gathers asymptotic confidence bands study.

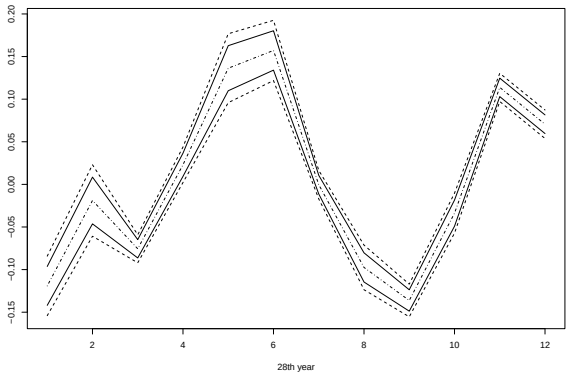

FIGURE 9.3. Confidence bands for electricity consumption during the 28^{th} year

PARTIE III

TESTS DE STRUCTURE EN RÉGRESSION SUR VARIABLE FONCTIONNELLE

In this part we are interested in structural testing procedures in regression on functional variable. We firstly provide in Chapter 10 a theoretical framework to construct innovative structural testing procedures. In practice, we have to adapt our method and propose a way to estimate the critical value of the test. We discuss in Chapter 11 the use of bootstrap procedures to estimate the critical value. We also study the empirical level and power of various bootstrap procedures in the case of a no-effect test.

CHAPITRE 10

STRUCTURAL TEST IN REGRESSION ON FUNCTIONAL VARIABLES

The content of the present chapter is the one of the paper Delsol *et al.* (2008) submitted for publication.

Abstract. Many papers deal with structural testing procedures in multivariate regression. More recently, various estimators have been proposed for regression models involving functional explanatory variables. Thanks to these new estimators, we propose a theoretical framework for structural testing procedures adapted to functional regression. The procedures introduced in this paper are innovative and make the link between former works on functional regression and others on structural testing procedures in multivariate regression. We prove asymptotic properties of the level and the power of our procedures under general assumptions that cover a large scope of possible applications : tests for no effect, linearity, dimension reduction, ...

10.1. Introduction

For many years statisticians have worked on models designed for multivariate random variables. However, the improvement of measuring apparatus provides data discretized on a thinner and thinner grid. Consequently, these data become intrinsically functional. Spectrometric curves, satellite images, annual electricity consumptions or sounds records are few examples, among others, of such intrinsically functional variables. This has led to a new statistical way of thinking in which we are interested in models where variables may belong to a functional space. Hence some new fields like operatorial statistics (see [378] and [122] among others) have emerged. Moreover, various topics and methods have been revisited in a functional context for instance through works dealing with principal component analysis for functional variables (see [125], [363] and [214] for more references), curves discrimination (see [145], [215] and [174]), estimation of some characteristics of the distribution (see [121] (density) and [116] (conditional mode) and the references therein), conditional quantiles estimation (see [70] and [166]) and regression models with functional explanatory variables (see for instance the various models studied in [39], [104], [75] and [164]). To get more references on the state of art in functional statistics the reader may refer

to the synthetic books [366], [367] and [368] that gather a large scope of statistical methods adapted to functional data study, while [46] focuses on dependent functional random variables and [176] resumes some recent advances with nonparametric methods.
In this paper one focuses more precisely on a functional regression model where the response variable Y is real-valued while the explanatory variable X belongs to a functional space. In other words one considers the following model

$$Y = r(X) + \epsilon \tag{10.1}$$

where the regression operator r is unknown. Such models have already been widely studied in the linear case (i.e. when r is linear) and still are topical issues (see for instance [363], [75], [77] and [102]). In the nonparametric case (i.e. only under regularity assumptions on r) the first results come from [172] in which a generalization of the well-known Nadaraya-Watson estimator is introduced. Many results have been obtained on this estimator concerning almost sure convergence (see [176] for more references), asymptotic normality (see [314], [167] and [134]) and $\mathbb{L}^q$ errors (see [132]).
In the multivariate case, many structural testing procedures have been proposed based on nonparametric methods, empirical likelihood ratio, orthogonal projection ... Various procedures have been introduced to test for no effect that means to test if the explanatory variable X has any effect on the response variable Y (see for instance [372], [85], or [148]). Other papers were devoted to test if the data come from a parametric model (see for instance [220], [227], [401], [404], [219], [249], [157] and [86]) or more precisely from a linear model (see for instance [150], [19] and [244]). Furthermore other papers were devoted to test if the data come from a semiparametric model (see for instance [427] and [405]) and more recently, [87] proposed a test based on both empirical likelihood and nonparametric methods to check simultaneously various semiparametric models. In the context of variable selection some testing procedures have been proposed to test if the effect of a multivariate explanatory variable is significative or not (see for instance [198], [195], and [276]). According to the intensive literature on this topic, this is not an overview but it attests that testing procedures have been introduced in a wide scope of problems and it gives a good idea on the usefulness of developing testing procedures.
In this paper we are interested in the potential use of nonparametric tools created for functional regression to extend structural testing procedures from the multivariate case to the functional case. Despite the abundant literature devoted to functional regression models and structural testing procedures in multivariate regression, there are very few papers on structural testing procedures in functional regression. As far as we know the existing literature is reduced to papers dealing with tests for no-effect (see for instance [186]) or tests of $\mathcal{H}_0 : \{r = r_0\}$ (where r_0 is a known operator) in the functional linear model (see for instance [74]). There is no test to check if the regression model is linear or not, and more generally to test if the true regression operator belongs to a given family of operators. In this paper we present two general structural testing procedures adapted to functional regression that allow to check if r is a known operator r_0 or if r belongs to a given family $\mathcal{F}$. The proposed approach consists in a comparison between a general nonparametric estimator and a particular one that converges quicker under the null hypothesis. This idea is similar to the one used in [220] in the multivariate case where a nonparametric and a parametric

estimator are compared to check for a parametric model. This work extends the previous abundant existing literature on structural testing procedures to the case of a functional explanatory variable. Indeed, the general assumptions used through this paper allow to cover a large scope of testing procedures such as, for instance, tests for no effect, tests for linearity, tests for dimension reduction, tests for semiparametric or parametric models...
Section 10.2 presents the model we are interested in, a discussion on the hypotheses tested and empirical motivation for constructing the test statistic. The main results of this paper are given in Section 10.3. After the study, of a first test to check if r is equal to a given operator r_0, one focuses on a structural test to check if r belongs to a given family $\mathcal{F}$. The results of Theorems 10.3.1 and 10.3.2 state the asymptotic normality (respectively divergence) of the considered test statistics under the null (respectively alternative) hypothesis. Section 10.4 is devoted to particular and complementary situations. To illustrate various interesting and practical settings when this test statistic can be used one focuses in Section 10.5 on a few examples : tests for no effect, tests for linearity or tests for dimension reduction. All the proofs of this paper are given in Section 10.7.

10.2. The model and the test statistic

In this paper one focuses on the functional regression model (10.1) presented above. One considers a set of independent pairs of random variables $(X_i, Y_i)_{1\leq i\leq n}$ identically distributed as (X, Y). One also assumes that these variables follow the model (10.1) i.e. that there exists a sequence of independent variables $(\epsilon_i)_{1\leq i\leq n}$ identically distributed as ϵ such that

$$\forall 1 \leq i \leq n,\, Y_i = r\left(X_i\right) + \epsilon_i.$$

Moreover one assumes that $\mathbb{E}\left[\epsilon_i | X_i\right] = 0$ in such a way that the regression operator is defined almost surely by $r\left(x\right) = \mathbb{E}\left[Y_i | X_i = x\right]$. Hence the regression operator r reflects the effect of the explanatory variable X on the response variable Y. The study of this link is a main issue in statistics. In this paper one is interested in testing if the true regression operator r belongs to a given family $\mathcal{F}$ of operators. If the family $\mathcal{F}$ is a nonempty, closed and convex subset of $\mathbb{L}^2$, the projection r_0 of r on $\mathcal{F}$ may be defined. Then, a standard idea to check if r belongs to $\mathcal{F}$ is to compute the distance between r and $\mathcal{F}$ (i.e. the $\mathbb{L}^2$-distance between r and r_0). Indeed, r belongs to $\mathcal{F}$ if and only if the distance between r and r_0 is null. When both r and r_0 are known, we can compute the $\mathbb{L}^2$-distance between r and r_0 and conclude that r belongs to $\mathcal{F}$ if and only if

$$\int \left(r - r_0\right)^2 \left(x\right) dP_X\left(x\right) = 0. \tag{10.2}$$

However, in our case the regression operator r and its projection r_0 are unknown hence one has to estimate them. Because one wants to make no structural assumptions on r in order to make possible the use of the test in various cases, one considers

a nonparametric kernel pseudo-estimator that converges almost surely towards r under regularity assumptions :

$$\tilde{r}(x) = \frac{1}{n\mathbb{E}\left[K\left(\frac{d(x,X)}{h_n}\right)\right]} \sum_{i=1}^{n} Y_i K\left(\frac{d(x,X_i)}{h_n}\right).$$

In the same way we introduce the following pseudo-estimator of r_0

$$\tilde{r_0}(x) = \frac{1}{n\mathbb{E}\left[K\left(\frac{d(x,X)}{h_n}\right)\right]} \sum_{i=1}^{n} r_0^*(X_i) K\left(\frac{d(x,X_i)}{h_n}\right),$$

in which r_0^* is a particular estimator of r_0 computed on a dataset independent of $(X_i, Y_i)_{1\leq i\leq n}$. Then, equality (10.2) might be accepted as soon as the following quantity is small enough :

$$\int \frac{1}{\left(n\mathbb{E}\left[K\left(\frac{d(x,X)}{h_n}\right)\right]\right)^2} \left(\sum_{i=1}^{n} (Y_i - r_0^*(X_i)) K\left(\frac{d(x,X_i)}{h_n}\right)\right)^2 dP_X(x).$$

For simplicity arguments we introduce a weight function w with bounded support. Consequently, the test detects differences between r and r_0 only on the bounded support of w. The use of a weight function w is a standard tool in structural testing procedures (see for instance [195] and [87]) as an alternative to assuming that the law of X has a bounded support (see for instance [220]). Furthermore, because under assumptions (10.5) and (10.9) introduced below the denominator may be bounded, independently of x, by two positive sequences of same order, one removes the denominator. Finally, one actually tests if r belongs to $\mathcal{F}$ through the asymptotic law of the following statistic (already used in the multivariate case in [195]) :

$$T_n^* = \int \left(\sum_{i=1}^{n} (Y_i - r_0^*(X_i)) K\left(\frac{d(x,X_i)}{h_n}\right)\right)^2 w(x) dP_X(x). \tag{10.3}$$

In the next section two structural testing procedures are presented. The first one is devoted to the case where $\mathcal{F}$ is reduced to the singleton $\{r_0\}$. In this special case r_0 is known and one takes $r_0^* = r_0$. The second test is devoted to the general case when $\mathcal{F}$ is a more complex family. In this case, the crucial issue of the choice of the particular estimator r_0^* is discussed.

10.3. Main results

In this section, we focus on two testing procedures. The first one enables to test if the true regression function r is a given function r_0. The second one consists in a structural testing procedure, using an other estimator of the regression function that converges quicker on a subset $\mathcal{F}$ of E.

10.3.1. Testing $r = r_0$. — This paragraph is devoted to a testing procedure which aims to test the null hypothesis

$$\mathcal{H}_0 : \{\mathbb{P}(r(X) = r_0(X)) = 1\}$$

against the alternative

$$\mathcal{H}_1 : \left\{ \|r - r_0\|_{\mathbb{L}^2(wdP_X)} \geq \eta_n \right\}.$$

It is worth noting that $\mathcal{H}_1$ depends on n. The sequence η_n reflects the capacity to detect smaller and smaller differences between the true regression function r and r_0 when n grows. We propose to construct a testing procedure from the asymptotic behaviour of the following statistic :

$$T_n = \int \left(\sum_{i=1}^{n} (Y_i - r_0(X_i)) K\left(\frac{d(x, X_i)}{h_n} \right) \right)^2 w(x) \, dP_X(x).$$

We will show that under the null hypothesis $\frac{1}{\sqrt{V_n}}(T_n - B_n)$ is asymptotically normal (V_n and B_n will be specified later) whereas the same statistic tends to infinity under the alternative.

10.3.1.1. Deccomposition of the statistic. — The theoretical analysis of T_n is not so easy, and we introduce the following decomposition :

$$T_n = T_{1,n} + T_{2,n} + T_{3,n} + T_{4,n} + 2T_{5,n} + 2T_{6,n},$$

where $T_{i,n}$ are specified below :

$$\begin{aligned}
T_{1,n} &= \int \sum_{i=1}^{n} K^2\left(\frac{d(X_i, x)}{h_n} \right) \epsilon_i^2 w(x) \, dP_X(x), \\
T_{2,n} &= \int \sum_{1 \leq i \neq j \leq n} K\left(\frac{d(X_i, x)}{h_n} \right) K\left(\frac{d(X_j, x)}{h_n} \right) \epsilon_i \epsilon_j w(x) \, dP_X(x), \\
T_{3,n} &= \int \sum_{i=1}^{n} K^2\left(\frac{d(X_i, x)}{h_n} \right) (r(X_i) - r_0(X_i))^2 w(x) \, dP_X(x), \\
T_{4,n} &= \int \sum_{1 \leq i \neq j \leq n} \prod_{\ell = i,j} K\left(\frac{d(X_\ell, x)}{h_n} \right) (r(X_\ell) - r_0(X_\ell)) w(x) \, dP_X(x), \\
T_{5,n} &= \int \sum_{i=1}^{n} K^2\left(\frac{d(X_i, x)}{h_n} \right) (r(X_i) - r_0(X_i)) \epsilon_i w(x) \, dP_X(x), \\
T_{6,n} &= \int \sum_{1 \leq i \neq j \leq n} K\left(\frac{d(X_i, x)}{h_n} \right) K\left(\frac{d(X_j, x)}{h_n} \right) \epsilon_i (r(X_j) - r_0(X_j)) w(x) \, dP_X(x).
\end{aligned}$$

Among these variables, we will see that $T_{1,n}$ gives the bias dominant term (see Lemma 10.3.1), $T_{2,n}$ has the dominant variance (see Lemma 10.3.2) and the other $T_{i,n}$ are null under the null hypothesis $\mathcal{H}_0$. On the other hand, under the alternative $\mathcal{H}_1$, the dominant term $T_{4,n}$ tends to infinity quicker than the variance of $T_{2,n}$. So, the main steps of our proofs consist in a detailed study of the random variables $T_{i,n}$.

More precisely bias and variance of each of these terms are studied under both $\mathcal{H}_0$ and $\mathcal{H}_1$.

10.3.1.2. Assumptions. — In order to obtain the asymptotic normality of T_n, we introduce some assumptions. We start with assumptions on the statistic T_n :

w is nonnegative, not null P_X a.s., has a bounded support W,
(10.4) and is bounded.

K has a compact support on $[0,\, 1]$, is nonincreasing and $\mathcal{C}^1$ on $]0, 1[$
(10.5) and $K(1) > 0$.

Let us make some comments on these assumptions. Firstly, concerning (10.4), the support of a weight function is often assumed to be compact. In $\mathbb{R}^p$ a set is compact if and only if it is closed and bounded. However, in general semi-metric spaces it is well-known that the unit ball is not always compact. There exists some useful results that give conditions for a functional set to be compact. For example, on a Hilbert space the unit ball is compact for the weak topology. In order to be more general, we only assume that W is a bounded set. Secondly, (10.5) is a standard assumption in kernel functional statistics (see [167] or [176]). The regularity of the kernel function enables to explicit asymptotic expressions of the moments of our statistic and assumption $K(1) > 0$ is very useful to obtain lower bounds.

Let us now introduce the notation

$$W_\gamma = \{x \in E,\, \exists y \in W,\, d(x, y) < \gamma\}$$

and focus on assumptions concerning the model :

(10.6) $\exists \gamma_0 > 0,\ \ r$ and r_0 are Hölderian of order β on W_{γ_0}, with respect to d.

(10.7) $\exists M > 0,\, \mathbb{E}\left[\epsilon^4 | X\right] \leq M$ a.s. and $\mathbb{E}\left[\epsilon^2 \mid X\right] = \sigma_\epsilon^2 > 0$.

Assumption (10.6) is very standard in nonparametric statistics to control the regularity of the true regression function r. Since r_0 is equal to r under $\mathcal{H}_0$, r_0 must have at least the same regularity as r.

Finally, we introduce some notations for key elements and sets that appear in our study :

$$F_x(s) = \mathbb{P}(d(x, X) \leq s),\ F_{x,y}(s,t) = \mathbb{P}(d(x, X) \leq s,\, d(y, X) \leq t),$$
$$\Omega_1(s) = \int_E F_x(s)\, w(x)\, dP_X(x),$$
$$\Omega_4(s) = \int_{E\times E} F_{x,y}^2(s,s)\, w(x)\, w(y)\, dP_X(x)\, dP_X(y).$$

The following assumptions are linking the sequence η_n, the smoothing parameter h_n and small ball probabilities :

$$(10.8)\ \exists C>0,\ \theta_n := Cv_n\left(\frac{1}{n^{\frac{1}{2}}\Phi^{\frac{1}{4}}(h_n)}+h_n^{\beta}\right)\leq \eta_n,\ \text{ where } v_n\to+\infty \text{ and } \theta_n\to 0,$$

$$(10.9)\ \exists C_1,C_2,\gamma>0,\ \exists\Phi,\ \forall x\in W_\gamma,\ \forall n\in\mathbb{N},\ C_1\Phi(h_n)\leq F_x(h_n)\leq C_2\Phi(h_n),$$

$$(10.10)\ \exists C_3>0,\ \Omega_4(h_n)\geq C_3\Phi^{3+l}(h_n)\ \text{ with } l<\frac{1}{2} \text{ and } n\Phi^{1+2l}(h_n)\to+\infty.$$

Firstly, it is worth noting that the assumptions made on the law of X only concern small ball probabilities. No assumption is made on X concerning the existence of a density with respect to a given dominant measure. Then, in assumption (10.8), v_n may be any sequence that tends to infinity slower than $\left(n^{-\frac{1}{2}}\Phi^{-\frac{1}{4}}(h_n)+h_n^{\beta}\right)^{-1}$ as for instance $v_n=\left(n^{-\frac{1}{2}}\Phi^{-\frac{1}{4}}(h_n)+h_n^{\beta}\right)^{-\gamma}$ with $0<\gamma<1$. Assumption (10.8) implies that under $\mathcal{H}_1$ the bias given by $T_{4,n}$ leads to the divergence of T_n. Keeping in mind the definition of $\mathcal{H}_1$, the rate of convergence of $v_n\left(n^{-\frac{1}{2}}\Phi^{-\frac{1}{4}}(h_n)+h_n^{\beta}\right)$ to 0 reflects the capacity for the test to detect smaller and smaller differences between the two regression functions tested when n grows. Assumption (10.9) has been chosen to make easier the reading of the proofs given in Section 10.7 . Assumption (10.10) is very important to compare the various rates of convergence of the mean and variance terms. We will see later in Section 10.4.1 that this assumption holds for fractal processes with $l=0$. In this case the second part of assumption (10.10) is very standard in nonparametric statistics (see for instance [167]).

10.3.1.3. Asymptotic law of T_n. — This paragraph is devoted to the study of the asymptotic law of the test statistic T_n whether the null hypothesis or the alternative one holds. Previous assumptions enable to prove that under $\mathcal{H}_0$, T_n (well normalized) is asymptotically gaussian whereas it tends to infinity under the alternative $\mathcal{H}_1$. That is what the following theorem highlights.

Theorem 10.3.1. — *Under assumptions (10.4)-(10.10) one gets :*

- *Under ($\mathcal{H}_0$),* $\frac{1}{\sqrt{Var(T_{2,n})}}\left(T_n-\mathbb{E}[T_{1,n}]\right)\xrightarrow{\mathcal{L}}\mathcal{N}(0,1)$,
- *Under ($\mathcal{H}_1$),* $\frac{1}{\sqrt{Var(T_{2,n})}}\left(T_n-\mathbb{E}[T_{1,n}]\right)\xrightarrow{P}+\infty$.

This result enables to construct an asymptotic test based on the variable

$$I_n=\frac{1}{\sqrt{Var\,(T_{2,n})}}\left(T_n-\mathbb{E}[T_{1,n}]\right)$$

with the quantiles of a standard gaussian law. General expressions of $\mathbb{E}[T_{1,n}]$ and $Var\,(T_{2,n})$ are given along the proof in Lemmas 10.3.1 and 10.3.2. However, to fix ideas, here are their asymptotic expressions in the simplest case when $K\equiv 1_{[0,1]}$

$$\mathbb{E}[T_{1,n}]=n\Omega_1(h_n)\,\sigma_\epsilon^2 \text{ and } Var\,(T_{2,n})\sim 2n^2\left(\sigma_\epsilon^2\right)^2\Omega_4(h_n).$$

Proof of Theorem 10.3.1 The proof will come directly from the following lemmas. Note that the variables $T_{1,n}$ and $T_{2,n}$ have the same law whether $\mathcal{H}_0$ holds or not.

Lemma 10.3.1. — *If assumptions (10.4),(10.5),(10.7) and (10.9) hold then one gets :*

$$\mathbb{E}\left[T_{1,n}\right] = n\Omega_1\left(h_n\right)\sigma_\epsilon^2\left[K^2\left(1\right) - \int_0^1 \left(K^2\right)'\left(s\right)\frac{\Omega_1\left(sh_n\right)}{\Omega_1\left(h_n\right)}ds\right],$$
$$Var\left(T_{1,n}\right) = O\left(n\Phi^2\left(h_n\right)\right).$$

Lemma 10.3.2. — *If assumptions (10.4),(10.5),(10.7),(10.9) and (10.10) hold, then* $\frac{1}{\sqrt{Var(T_{2,n})}}T_{2,n}$ *is asymptotically gaussian with zero mean and variance* 1. *Moreover, there exist two positive constants* C *and* C' *such that*

$$Cn^2\Phi^{3+l}\left(h_n\right) \leq Var\left(T_{2,n}\right) \sim 2\left(\sigma_\epsilon^2\right)^2\Gamma_n n^2\Omega_4\left(h_n\right) \leq C'n^2\Phi^3\left(h_n\right),$$

for some Γ_n *such that* $K^2\left(1\right) \leq \Gamma_n \leq K^2\left(0\right)$. *See formula (10.32) in the proof for the explicit expression of* Γ_n.

Lemma 10.3.3. — *Under* $\mathcal{H}_1$, *if (10.4)-(10.6) and (10.8)-(10.10) hold then one gets :*

$$T_{3,n} \geq 0 \text{ a.s. and } \mathbb{E}\left[T_{4,n}\right] \geq Cn^2\Phi^2\left(h_n\right)\left\|r - r_0\right\|^2_{\mathbb{L}^2(wdP_X)} \geq C\eta_n^2 n^2\Phi^2\left(h_n\right).$$

Lemma 10.3.4. — *Under* $\mathcal{H}_1$, *if assumptions (10.4)-(10.7) and (10.9) hold then one gets :*

$$Var\left(T_{5,n}\right) \leq Cn\Phi^2\left(h_n\right)\left\|r - r_0\right\|^2_{\mathbb{L}^2(wdP_X)}, Var\left(T_{4,n}\right) = O\left(n^3\Phi^4\left(h_n\right)\left\|r - r_0\right\|^2_{\mathbb{L}^2(wdP_X)}\right),$$
$$\text{and } Var\left(T_{6,n}\right) = O\left(n^3\Phi^4\left(h_n\right)\left\|r - r_0\right\|^2_{\mathbb{L}^2(wdP_X)}\right).$$

Lemma 10.3.5. — *Under assumptions (10.4) and (10.9) one gets :*

$$\Omega_4\left(h_n\right) = O\left(\Phi^3\left(h_n\right)\right).$$

– **Under** $\mathcal{H}_0$ **:** Under $\mathcal{H}_0$ and assumptions of Theorem 10.3.1, the bias dominant term comes from $T_{1,n}$, $T_{2,n}$ is asymptotically gaussian and variances of $T_{j,n}, j \neq 2$, are null. From Lemmas 10.3.1 and 10.3.2, we conclude that

$$\begin{aligned} T_n &= \left[\mathbb{E}\left[T_{1,n}\right] + O_p\left(\sqrt{Var\left(T_{1,n}\right)}\right)\right] + T_{2,n} \\ &= \mathbb{E}\left[T_{1,n}\right] + O_p\left(\sqrt{n}\Phi\left(h_n\right)\right) + T_{2,n}. \end{aligned}$$

Assumptions (10.10) gives

$$\frac{\sqrt{n}\Phi\left(h_n\right)}{n\Phi^{\frac{3+l}{2}}\left(h_n\right)} = \frac{1}{\sqrt{n\Phi^{1+l}\left(h_n\right)}} \to 0 \text{ since } n\Phi^{1+2l}\left(h_n\right) \to +\infty.$$

and by using the second part of Lemma 10.3.2 we arrive at :

$$T_n = \mathbb{E}\left[T_{1,n}\right] + T_{2,n} + o_p\left(\sqrt{Var\left(T_{2,n}\right)}\right).$$

Finally, we have shown the following statement

$$\frac{T_n - \mathbb{E}\left[T_{1,n}\right]}{\sqrt{Var\left(T_{2,n}\right)}} = \frac{T_{2,n}}{\sqrt{Var\left(T_{2,n}\right)}} + o_p\left(1\right).$$

The first part of Lemma 10.3.2 ensures the asymptotic normality of $\frac{T_{2,n}}{\sqrt{Var(T_{2,n})}}$ and we conclude with Slutsky's theorem.

– **Under $\mathcal{H}_1$:** Lemmas 10.3.3 and 10.3.4 reflect that under $\mathcal{H}_1$ many changes occur. In fact, under the alternative hypothesis, variables $T_{j,n}$, $j = 3, \ldots, 6$, are not null as under $\mathcal{H}_0$ and consequently their asymptotic behaviour is different. More precisely, the bias dominant term no longer comes from $T_{1,n}$ but from $T_{4,n}$ and variances of $T_{4,n}$ and $T_{6,n}$ are no longer negligible with regard to the variance of $T_{2,n}$. However we will see that these variances are negligible with regard to the bias dominant term that tends to infinity. From Lemmas 10.3.1-10.3.4, since $n\Phi(h_n) \to +\infty$, one gets :

$$\begin{aligned} T_n &= \mathbb{E}[T_{1,n}] + T_{2,n} + T_{3,n} + \mathbb{E}[T_{4,n}] + O_p\left(n^{\frac{3}{2}}\Phi^2(h_n)\|r - r_0\|_{\mathbb{L}^2(wdP_X)}\right) \\ &\quad + o_p\left(\sqrt{Var(T_{2,n})}\right) \\ &\geq \mathbb{E}[T_{1,n}] + T_{2,n} + C\|r - r_0\|^2_{\mathbb{L}^2(wdP_X)} n^2\Phi^2(h_n) \\ &\quad + O_p\left(n^{\frac{3}{2}}\Phi^2(h_n)\|r - r_0\|_{\mathbb{L}^2(wdP_X)}\right) + o_p\left(\sqrt{Var(T_{2,n})}\right), \end{aligned}$$

where C is a positive constant. The rate of the O_p is negligible with regard to the bias dominant term. Indeed, it comes from assumption (10.8) that :

$$\begin{aligned} \frac{n^{\frac{3}{2}}\Phi^2(h_n)\|r - r_0\|_{\mathbb{L}^2(wdP_X)}}{n^2\Phi^2(h_n)\|r - r_0\|^2_{\mathbb{L}^2(wdP_X)}} &\leq \frac{1}{\|r - r_0\|_{\mathbb{L}^2(wdP_X)}\sqrt{n}} \\ &\leq \frac{1}{\eta_n\sqrt{n}} \\ &\leq C\frac{1}{\left(v_n n^{-\frac{1}{2}}\phi^{-\frac{1}{4}}(h_n)\right)\sqrt{n}} \to 0. \end{aligned}$$

Hence one obtains with Lemma 10.3.5 :

$$\begin{aligned} \frac{T_n - \mathbb{E}[T_{1,n}]}{\sqrt{Var(T_{2,n})}} &\geq \frac{T_{2,n}}{\sqrt{Var(T_{2,n})}} + o_p(1) + C\frac{\eta_n^2 n^2\Phi^2(h_n)}{n\Phi^{\frac{3}{2}}(h_n)} + o_p\left(\frac{\eta_n^2 n^2\Phi^2(h_n)}{n\Phi^{\frac{3}{2}}(h_n)}\right) \\ &\geq \frac{T_{2,n}}{\sqrt{Var(T_{2,n})}} + o_p(1) + o_p\left(\eta_n^2 n\Phi^{\frac{1}{2}}(h_n)\right) + C\eta_n^2 n\Phi^{\frac{1}{2}}(h_n). \end{aligned}$$

Assumption (10.8) on η_n implies that the last term in the previous minoration tends to infinity :

$$\eta_n^2 n\Phi^{\frac{1}{2}}(h_n) \geq Cv_n^2 n^{-1}\Phi^{-\frac{1}{2}}(h_n)\, n\Phi^{\frac{1}{2}}(h_n) \geq Cv_n^2.$$

The first part of Lemma 10.3.2 ensures that $\frac{T_{2,n}}{\sqrt{Var(T_{2,n})}} = O_p(1)$ since it converges in law and we conclude with Slutsky's theorem.

□

10.3.2. Testing $r \in \mathcal{F}$. — In this paragraph we focus on a modification of the previous statistic in order to construct a structural testing procedure. In a few words, given a set $\mathcal{F}$ of operators, the aim is to test if the true regression function r is in this set or not. Here the idea is to replace, in the previous test statistic, the function r_0 by an estimator r_0^* of the projection of the regression function on $\overline{\mathcal{F}}^w$ (this notation will be specified later). If this estimator converges with a better rate of convergence than the general kernel estimator does on E, it will be possible to prove that if $r \in \mathcal{F}$ then the statistic is asymptotically gaussian, but converges to infinity if r does not belong to $\overline{\mathcal{F}}^w$. Before going on, let us introduce the following notation for any subset A of $\mathbb{L}^2_{dP_X}(W_\alpha), \alpha > 0$:

$$\overline{A}^w = \overline{\mathcal{A}}^{\mathbb{L}^2_{(wdP_X)}(E)} \text{ and } \overline{A} = \overline{\mathcal{A}}^{\mathbb{L}^2_{(dP_X)}(W_\alpha)}.$$

10.3.2.1. Decomposition of the statistic. — This paragraph is devoted to a testing procedure which aims, for a given family $\mathcal{F}$ of operators, to test the null hypothesis

$$\mathcal{H}_0 : \{r \in \mathcal{F}\}$$

against the alternative

$$\mathcal{H}_1 : \left\{ \inf_{r_0 \in \overline{\mathcal{F}}^w} \|r - r_0\|_{\mathbb{L}^2(wdP_X)} \geq \eta_n \right\}.$$

As before, the sequence η_n reflects the capacity to detect smaller and smaller differences between the true regression function r and its projection r_0 on $\overline{\mathcal{F}}^w$ when n grows.

In the previous part, we have introduced a statistic T_n to test if the true regression operator is a given known operator r_0. In this kind of test, we fix an operator r_0 and try to construct a test to know if our choice seems to be correct or not. In such situations we can use a statistic that depends on the choice of r_0. However, when we are interested in testing if r belongs to some family of operators we have to change our approach. To test if r belongs to some given family $\mathcal{F}$ of operators we would like to use the projection r_0 of r into the family $\overline{\mathcal{F}}^w$. Indeed, the difference $r - r_0$ would be a good way to quantify the distance from r to the family $\mathcal{F}$ since it would be null if r belongs to $\overline{\mathcal{F}}^w$ and positive if it does not. Unfortunately, we cannot use directly this idea because the projection r_0 is still unknown. That is why we propose to introduce a new statistic T_n^* that differs from the previous one only through the fact that r_0 is replaced by an estimator r_0^* :

$$T_n^* = \int \left(\sum_{i=1}^n \left(Y_i - r_0^*(X_i) \right) K\left(\frac{d(x, X_i)}{h_n} \right) \right)^2 w(x)\, dP_X(x).$$

Then we propose the same kind of decomposition as for T_n :

$$T_n^* = T_{1,n} + T_{2,n} + T_{3,n}^* + T_{4,n}^* + 2T_{5,n}^* + 2T_{6,n}^*,$$

where the $T_{i,n}^*$'s are specified below :

$$\begin{aligned}
T_{3,n}^* &= \int \sum_{i=1}^{n} K^2\left(\frac{d(X_i,x)}{h_n}\right)\left(r(X_i)-r_0^*(X_i)\right)^2 w(x)\,dP_X(x),\\
T_{4,n}^* &= \int \sum_{1\le i\ne j\le n}\ \prod_{\ell=i,j} K\left(\frac{d(X_\ell,x)}{h_n}\right)\left(r(X_\ell)-r_0^*(X_\ell)\right) w(x)\,dP_X(x),\\
T_{5,n}^* &= \int \sum_{i=1}^{n} K^2\left(\frac{d(X_i,x)}{h_n}\right)\left(r(X_i)-r_0^*(X_i)\right)\epsilon_i w(x)\,dP_X(x),\\
T_{6,n}^* &= \int \sum_{1\le i\ne j\le n} K\left(\frac{d(X_i,x)}{h_n}\right) K\left(\frac{d(X_j,x)}{h_n}\right)\epsilon_i\left(r(X_j)-r_0^*(X_j)\right) w(x)\,dP_X(x).
\end{aligned}$$

10.3.2.2. Complementary assumptions. — We introduce the following assumption on the family $\mathcal{F}$:

(10.11) $\mathcal{F}$ is nonempty and convex, $\exists \alpha>0, \mathcal{F}\subset \mathbb{L}^2_{(dP_X)}(W_\alpha)$,

in order to define r_0 the projection of r into $\overline{\mathcal{F}}^w$ with respect to the inner-product

$$(u,v)=\int u(x)\,v(x)\,w(x)\,dP_X(x).$$

Furthermore, for any operator u, we can define the distance between u and $\overline{\mathcal{F}}$ by :

$$d_{\overline{\mathcal{F}}}(u)=\inf_{u_0\in\overline{\mathcal{F}}}\|u-u_0\|_{\mathbb{L}^2_{(dP_X)}(W_\alpha)}.$$

We introduce the following notation for Hölderian functions sets :

$$Hol(C,\beta)=\left\{u: E\to\mathbb{R},\ \forall x,y\in W_\alpha,\ |u(x)-u(y)|\le C d^\beta(x,y)\right\}.$$

To get asymptotic results we assume that the estimator r_0^* has some good asymptotic properties resumed in the following assumptions in which l comes from assumption (10.10) :

(10.12) r_0^* is constructed from $(X_i^*,Y_i^*)_{1\le i\le u_n}$ independent of $(X_i,Y_i)_{1\le i\le n}$,

(10.13) under $\mathcal{H}_0$, $n\Phi^{\frac{1-l}{2}}(h_n)\,\mathbb{E}\left[\left(r_0^*(X_1)-r_0(X_1)\right)^2 1_{W_\alpha}(X_1)\right]\to 0$,

(10.14) under $\mathcal{H}_1$, $\eta_n^{-2}\mathbb{E}\left[d^2_{\overline{\mathcal{F}}}(r_0^*)\right]\to 0$ and $\exists C>0$, $\eta_n^{-2}\mathbb{E}\left[d^2_{Hol(C,\beta)}(r_0^*)\right]\to 0$,

(10.15) under $\mathcal{H}_1$, $\mathbb{E}\left[(r_0^*(X))^4 1_{W_\alpha}(X)\right]<+\infty$.

Let us make some remarks about these assumptions. Firstly, an easy way to fulfill the first assumption is to split the original sample of size $2n$ into two samples of size n, compute r_0^* from the second sample and keep the first sample to compute the statistic T_n^*. However, in some cases it may be interesting to split differently the dataset and take $u_n\ne n$. Moreover, if r_0^* belongs to the family $\mathcal{F}$ (or even $\overline{\mathcal{F}}^w$) almost surely, then the first part of assumption (10.14) is fulfilled since $d_{\overline{\mathcal{F}}}(r_0^*)=0$ *a.s.*. Now we have to study the asymptotic behaviour of each variable $T_{j,n}^*$, $j=3,\dots,6$ (we have already studied $T_{1,n}$ and $T_{2,n}$ in the previous paragraph). To do that we will use the same arguments as in the first part and prove that under $\mathcal{H}_0$ variables $T_{j,n}^*$, $j=3,\dots,6$ are negligible with regard to the dominant terms $T_{1,n}$ and $T_{2,n}$. On

the other hand, under the alternative $\mathcal{H}_1$, the dominant term $T_{3,n}^* + T_{4,n}^*$ tends to infinity and leads to the divergence of the test statistic. The main issue still is to study precisely the conditional mean and variance of each variable $T_{j,n}^*$.

10.3.2.3. Asymptotic law of T_n^*. — Previous assumptions enable to get the same kind of result as Theorem 10.3.1 for T_n^*. That is to say that under the null hypothesis $\mathcal{H}_0$, T_n^* (correctly normalized) is asymptotically gaussian but tends to infinity under the alternative $\mathcal{H}_1$. That is what the next theorem highlights.

Theorem 10.3.2. — *Under assumptions (10.4)-(10.10) and (10.11)-(10.15) one gets :*

- *Under ($\mathcal{H}_0$),* $\frac{1}{\sqrt{Var(T_{2,n})}}\left(T_n^* - \mathbb{E}\left[T_{1,n}\right]\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,1\right),$
- *Under ($\mathcal{H}_1$),* $\frac{1}{\sqrt{Var(T_{2,n})}}\left(T_n^* - \mathbb{E}\left[T_{1,n}\right]\right) \xrightarrow{P} +\infty.$

Proof of Theorem 10.3.2 :
The proof comes from the following lemmas :

Lemma 10.3.6. — *Under $\mathcal{H}_0$, if assumptions (10.4)-(10.10) and (10.11)-(10.14) hold then one gets :*

$$T_{3,n}^* + T_{4,n}^* + 2T_{5,n}^* + 2T_{6,n}^* = o_p\left(\sqrt{Var\left(T_{2,n}\right)}\right).$$

Lemma 10.3.7. — *Under $\mathcal{H}_1$, if assumptions (10.4)-(10.10) and (10.11)-(10.14) hold then one gets :*

$$T_{3,n}^* + T_{4,n}^* + 2T_{5,n}^* + 2T_{6,n}^* \geq Cn^2\Phi^2\left(h_n\right)\eta_n^2\left(1 + o_p\left(1\right)\right).$$

– Under $\mathcal{H}_0$, Lemmas 10.3.1, 10.3.2 and 10.3.6 imply that

$$\begin{aligned}
& \frac{1}{\sqrt{Var\left(T_{2,n}\right)}}\left(T_n^* - \mathbb{E}\left[T_{1,n}\right]\right) \\
= & \frac{1}{\sqrt{Var\left(T_{2,n}\right)}}\left(T_{1,n} - \mathbb{E}\left[T_{1,n}\right] + T_{2,n} + T_{3,n}^* + T_{4,n}^* + 2T_{5,n}^* + 2T_{6,n}^*\right) \\
= & \frac{1}{\sqrt{Var\left(T_{2,n}\right)}}\left(O_p\left(\sqrt{n}\Phi\left(h_n\right)\right) + T_{2,n} + o_p\left(\sqrt{Var\left(T_{2,n}\right)}\right)\right) \\
= & \frac{T_{2,n}}{\sqrt{Var\left(T_{2,n}\right)}} + O_p\left(\frac{1}{\sqrt{n\Phi^{1+l}\left(h_n\right)}}\right) + o_p\left(1\right) \\
= & \frac{T_{2,n}}{\sqrt{Var\left(T_{2,n}\right)}} + o_p\left(1\right).
\end{aligned}$$

We apply Slutsky's theorem together with Lemma 10.3.2 to conclude.

– Under $\mathcal{H}_1$, Lemmas 10.3.1, 10.3.2 and 10.3.7 imply that

$$\begin{aligned}
&\frac{1}{\sqrt{Var\left(T_{2,n}\right)}}\left(T_n^* - \mathbb{E}\left[T_{1,n}\right]\right)\\
= &\frac{1}{\sqrt{Var\left(T_{2,n}\right)}}\left(T_{1,n} - \mathbb{E}\left[T_{1,n}\right] + T_{2,n} + T_{3,n}^* + T_{4,n}^* + 2T_{5,n}^* + 2T_{6,n}^*\right)\\
\geq &\frac{1}{\sqrt{Var\left(T_{2,n}\right)}}\left(O_p\left(\sqrt{n}\Phi\left(h_n\right)\right) + T_{2,n} + Cn^2\Phi^2\left(h_n\right)\eta_n^2\left(1 + o_p\left(1\right)\right)\right)\\
\geq &\frac{T_{2,n}}{\sqrt{Var\left(T_{2,n}\right)}} + Cn\Phi^{\frac{1}{2}}\left(h_n\right)v_n^2\left(\frac{1}{n^{\frac{1}{2}}\Phi^{\frac{1}{4}}\left(h_n\right)}\right)^2\left(1 + o_p\left(1\right)\right)\\
&+O_p\left(\frac{1}{\sqrt{n\Phi^{1+l}\left(h_n\right)}}\right)\\
\geq &O_p\left(1\right) + o_p\left(1\right) + Cv_n^2\left(1 + o_p\left(1\right)\right)\\
\geq &Cv_n^2\left(1 + o_p\left(1\right)\right) \xrightarrow{P} +\infty.
\end{aligned}$$

Prohorov's theorem together with Lemma 10.3.2 ensure that

$$\left(\sqrt{Var\left(T_{2,n}\right)}\right)^{-1} T_{2,n} = O_p\left(1\right).$$

Finally we apply Slutsky's theorem together with assumption (10.8) to conclude.

□

10.4. Some comments on main assumptions

10.4.1. The particular case of fractal variables. — In order to understand better previous assumptions and results, we choose to consider the particular case in which we assume the variable X to be fractal with respect to the semimetric d. We remind the definition of uniformly fractal processes introduced in [164] as a uniform version of the definition of fractal processes given in [172] :

$$\exists\delta > 0,\ \exists c,\ \inf_{x\in S} c\left(x\right) > 0,\ \lim_{h\to 0^+}\sup_{x\in S}\left|\frac{F_x\left(h\right)}{h^\delta} - c\left(x\right)\right| = 0. \tag{10.16}$$

In this paragraph we make the following assumptions :

(10.17) $\left(X_i\right)_i$ is uniformly fractal on the compact S and $\sup_S c\left(x\right) < +\infty$,

(10.18) w is continuous on E and has a compact support W with $d\left(W, S^c\right) > 0$,

(10.19) $nh_n^\delta \to +\infty$,

(10.20) under $\mathcal{H}_0$, $nh_n^{\frac{\delta}{2}}\mathbb{E}\left[\left(r - r_0^*\right)^2\left(X_1\right)1_{W_\alpha}\left(X_1\right)\right] \to 0$.

It is obvious that since S is compact the second part of assumption (10.17) is fulfilled when c is continuous on S. Assumption (10.18) is not restrictive with regard to the

model. To be more general, one may take S bounded but not necessary compact and w uniformly continuous on E. Under assumptions (10.17)-(10.19), many simplifications occur. For instance, assumption (10.9) is now always fulfilled, whereas (10.10) is equivalent to (10.19) in this setting.

All these simplifications give the following corollary of Theorem 10.3.2.

Corollary 10.4.1. — *If assumptions (10.5)-(10.8), (10.11)-(10.12), (10.14)-(10.15) and (10.17)-(10.20) hold, then one gets :*

- *Under* $(\mathcal{H}_0)$, $\frac{1}{\sqrt{Var(T_{2,n})}}\left(T_n^* - \mathbb{E}\left[T_{1,n}\right]\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,1\right)$,
- *Under* $(\mathcal{H}_1)$, $\frac{1}{\sqrt{Var(T_{2,n})}}\left(T_n^* - \mathbb{E}\left[T_{1,n}\right]\right) \xrightarrow{\mathcal{L}} +\infty$.

Proof of Corollary 10.4.1 The proof comes directly from the following lemmas.

Lemma 10.4.1. — *If (10.17) and (10.18) hold then (10.9) holds with* $\Phi\left(s\right) = s^{\delta}$.

Lemma 10.4.2. — *If (10.17)-(10.19) hold then (10.10) holds with* $l = 0$ *and assumption (10.20) is a particular case of (10.13) with* $\Phi\left(s\right) = s^{\delta}$ *and* $l = 0$.

□

Of course one also gets a corollary of Theorem 10.3.1 under assumptions (10.5)-(10.8) and (10.17)-(10.19). Assumptions (10.17)-(10.19) are more easy to understand than assumptions (10.9)-(10.10), but assumption (10.17) may seem very restrictive. However we will see now that any process on a separable Hilbert space fulfills assumption (10.17) for a good semimetric choice. Firstly, when the explanatory variable is in $\mathbb{R}^p$ and has a continuous and positive density f, the statement (10.16) always holds, for any norm, with $\delta = p$. Furthermore, as explained in [176] (Part V, Lemma 13.6), for any functional variable X on a separable Hilbert space, a good semi-metric choice, as for instance a projection semimetric, leads to the fractal case. The following lemma illustrates these facts and hence shows that assumption (10.17) is not as restrictive as it seems at a glance. For any norm $\|\ \|$ on $\mathbb{R}^p$, we introduce the following notation :

$$V_{\|\ \|}\left(p\right) = \int_{\mathbb{R}^p} 1_{\|u\| \leq 1} d\lambda\left(u\right),$$

in which λ is the Lebesgue measure on $\mathbb{R}^p$.

Lemma 10.4.3. — *Let E a separable Hilbert space with inner product $< .,. >$ and an orthonormal basis $\{e_j, j = 1, \ldots, \infty\}$. Let $F = Vect\left\{e_j, j = 1, \ldots, p\right\}$ and $\|\ .\ \|$ a norm on $\mathbb{R}^p$. We define $x_F = \left(< x, e_1 >, \ldots, < x, e_p >\right)^T$ and a projection semimetric d_F from E to F by $d_F\left(x,y\right) = \|x_F - y_F\|$. If the variable X_F has a continuous and positive density f on a compact set C, then, for any $S \subset C$ such that $d_F\left(S, C^c\right) > 0$, X is uniformly fractal on S with $\delta = p$ and $c\left(x\right) = V_{\|\ \|}\left(p\right) f\left(x_F\right)$ for the semimetric d_F.*
Besides, if f is continuous and positive on F, then X is uniformly fractal on any compact set of (E, d_F).

Moreover, under assumptions of Lemma 10.4.3, we have $\sup_C c(x) < +\infty$. Indeed, $c(x) = V_{\|\,\|}(p) f(x_F)$ is continuous (because f is continuous) on the compact C hence c is bounded on C. So, under assumptions of Lemma 10.4.3, assumption (10.17) holds.

10.4.2. Alternative sets of assumptions. — Assumptions introduced in Part 10.3 have been chosen to make a compromise between simplicity and generality. In this paragraph we present a few variations on some previous assumptions. Our aim is to enable the use of both previous tests in cases where Theorem 10.3.1 or Theorem 10.3.2 cannot be applied directly. We discuss more general assumptions for which the result of Theorem 10.3.2 is still true. Then, we will show how other alternatives than $\mathcal{H}_{1,n}$ give less restrictive lower bounds for η_n. However, making some assumptions more general has a cost, in the sense that some other assumptions become more restrictive. One has to weight the pros and cons of each set of assumptions. To be more concise we note $\mathbb{E}_*$ when we take the mean conditionally to $(X_i^*, Y_i^*)_{1\leq i\leq u_n}$.

Firstly, assumption (10.9) may seem fairly restrictive. Previous comments on the fractal case have shown that this assumption is not as restrictive as it seems, however the result of Theorem 10.3.1 may be obtained under the more general assumption :

$$\exists \gamma, \mu, \forall x \in W_\gamma, \ \frac{F_x(h_n)}{\Phi(h_n)} \overset{n\to+\infty}{\to} p(x), \ 0 < \mu \leq p(x) \leq M,$$
$$\text{and } \frac{F_x(h_n)}{p(x)\,\Phi(h_n)} \leq v(x), \ \text{ with } v \in \mathbb{L}^7_{(wdP_X)}.$$

There exists an important literature devoted to the study of asymptotic equivalents for small ball probabilities (see [166] for a review or [44] and [285] for gaussian processes) hence this alternatives may be theoretically more interesting.

Secondly we want to focus on some alternatives for $\mathcal{H}_{1,n}$. In the next lines, are presented two alternative sets of assumptions for which Theorem 10.3.2 still holds. Because Theorem 10.3.1 is a direct application of Theorem 10.3.2 with $\mathcal{F} = \{r_0\}$ and $r_0^* = r_0$ these alternative assumptions are also relevant for Theorem 10.3.1. Through the present section we use for any operator u the notation $u_\eta(x) = u(x) 1_{\{u(x)>\eta\}}$. First of all, here are alternatives for $\mathcal{H}_{1,n}$ and assumptions (10.4), (10.6), (10.8) and (10.14) that avoid Hölder asumptions on r and r_0^* :

$$\mathcal{H}_{1,n} : \left\{ r \in \cup_{\mu>0} \left\{ m, \ \inf_{r_0\in\mathcal{F}} \|m - r_0\|^2_{\mathbb{L}^2(w_\mu dP_X)} \geq \eta_n^2 \right\} \right\}, \tag{10.21}$$

$$w \text{ is uniformly continuous on } E, \tag{10.22}$$

$$\mathbb{E}\left[r^4(X) 1_{W_\alpha}(X)\right] < +\infty, \tag{10.23}$$

$$\eta_n \geq v_n n^{\frac{1}{2}} \Phi^{\frac{3}{4}}(h_n) \text{ with } v_n \to +\infty \text{ and } v_n^2 n \Phi^{\frac{3}{2}}(h_n) \to 0, \tag{10.24}$$

$$\text{under } \mathcal{H}_1, \ \eta_n^{-2}\mathbb{E}\left[d^2_{\mathcal{F}}(r_0^*)\right] \to 0. \tag{10.25}$$

Assumptions (10.23) and (10.25) are clearly less restrictive than assumptions (10.6) and (10.14) because they do not need Hölderian regularity assumptions on r and r_0^*. However, the alternative hypothesis is a little more restrictive and the weight function w is assumed to have more regularity. It is worth noting that because w is a parameter of our statistic, assumption (10.22) is not restrictive with respect to the model. Finally, the sequence η_n have to fulfill a different lower bound than the one given by (10.8). This new bound is not really bad because when small ball probabilities are concentrated, the condition made on η_n in (10.24) is less restrictive than the one made in (10.8). For example, it is the case for fractal small ball probabilities as soon as $\delta > 4\beta$. A direct consequence of this remark is to say that if we assume (10.6) and (10.14) in place of (10.23) and (10.25), assumption (10.24) might be replaced by :

$$\eta_n = v_n \min\left(n^{\frac{1}{2}}\Phi^{\frac{3}{4}}(h_n), h_n^{\beta} + \frac{1}{n^{\frac{1}{2}}\Phi^{\frac{1}{4}}(h_n)} \right) \text{ with } v_n \to +\infty \text{ and } \eta_n \to 0. \quad (*)$$

Nevertheless, we may prove Theorem 10.3.2 under a less restrictive condition on η_n than $(*)$ if assumptions (10.6) and (10.14) hold. Theorem 10.3.2 may be proved if the following alternatives hold instead of $\mathcal{H}_{1,n}$ and assumption (10.8) :

$$(10.26)\quad \mathcal{H}_{1,n} : \left\{ r \in \cup_{\nu>0} \left\{ m,\ \inf_{r_0 \in \mathcal{F}} \int_E \left((m - r_0)^2 \right)_{\nu}(x)\, w(x)\, dP_X(x) \geq \eta_n^2 \right\} \right\},$$

$$(10.27)\quad \eta_n = v_n \frac{1}{n^{\frac{1}{2}}\Phi^{\frac{1}{4}}(h_n)} \text{ with } v_n \to +\infty \text{ and } v_n \frac{1}{n^{\frac{1}{2}}\Phi^{\frac{1}{4}}(h_n)} \to 0.$$

The alternative $\mathcal{H}_{1,n}$ is more restrictive and detect divergences that are locally greater than a threshold ν. However the less restrictive assumption made on η_n is classical in nonparametric structural testing procedures (see for instance [87] and [220]).

Important remark : The results of this paper may be obtained when $r = r_n$ and r_n satisfies the previous complementary alternatives for each n with μ and ν independent of n.

Finally, assumptions made on the estimator r_0^* are always used to get upper bounds, in probability, for conditional means as a direct consequence of the following inequality (valid for any real random variable U such that $\mathbb{E}_*[U] \geq 0$) :

$$\mathbb{E}_*[U] = O_p(\mathbb{E}[U]).$$

In fact, we only need upper bounds in probability for conditional means thus Theorem 10.3.2 still holds if assumptions (10.13)-(10.15) are replaced by the following conditional assumptions :

$$(10.28)\qquad \text{under } \mathcal{H}_0,\ n\Phi^{\frac{1-l}{2}}(h_n)\, \mathbb{E}_*\left[(r_0^*(X_1) - r_0(X_1))^2 1_{W_\alpha}(X_1) \right] \xrightarrow{P} 0,$$

$$(10.29)\qquad \text{under } \mathcal{H}_1,\ \eta_n^{-2} d^2_{\mathcal{F}}(r_0^*) \xrightarrow{P} 0 \text{ and } \exists C > 0, \eta_n^{-2} d^2_{Hol(C,\beta)}(r_0^*) \xrightarrow{P} 0,$$

$$(10.30)\qquad \mathbb{E}_*\left[(r_0^*(X))^4 1_{W_\alpha}(X) \right] = O_p(1).$$

10.5. Examples of use

In this paragraph we present some examples of structural testing procedures covered by the results given above. One focuses on three examples selected to cover various different kinds of situations. However, this enumeration is actually not exhaustive since our test may be employed in many other situations. We will see among these applications that the alternative sets of assumptions discussed in Section 10.4.2 may be helpful to allow the use of our testing procedure. As we will see the key issue consists in finding a estimator r_0^* that fulfills assumptions (10.13)-(10.15).

– At first one focuses on the issue of testing if there is a link between Y and X. There exists a relationship between these random variables if and only if $r \neq \mathbb{E}[Y]$. This problem has been widely studied in the multivariate case (see for instance [372], [153], [152], [85] and [148]). However, in the case of a functional explanatory variable, there are few results. The problem has been studied in a linear regression model in [74] and more recently in a nonparametric model assuming that the mean of Y is known (see [186]). The results given above complete these tests in a nonparametric and more general way. Indeed, if the mean of Y is a known constant c, one can apply Theorem 10.3.1 to test if $r \equiv c$. Moreover, if the mean of the response variable is unknown one can use Theorem 10.3.2 to test if r is constant (by taking $\mathcal{F} = \{r, \exists c, r \equiv c\}$) with $r_0^* \equiv {u_n}^{-1} \sum_{i=1}^{u_n} Y_i$. This test is useful to check the existence of a relationship between the response and the explanatory variables before any estimation of the regression function.

– Secondly, one may be interested in testing if r is linear or not. This may be usefull ever to justify some linear modelizations already used or to test if one can use more powerful results that only hold in the linear case. In the multivariate case, many papers focus on testing a parametric model with nonparametric methods (see for instance [150], [151], [220], [19], [402], [227] and [249] in the independent case or [242], [243], [244] and [157] in time series context) or via empirical likelihood (see for instance [156], [86] and [159]). However, in the case of a functional explanatory variable it seems that no result exists in the literature. The lack of such structural testing procedures seems to be a more important problem in the case of functional explanatory variable since it is not possible, as in the real case, to plot the data. For instance, before any application of a linear regression estimator, one should want to check if the data may correspond to a linear model. In the case of a real explanatory variable, in addition to other methods, one often plots the data and check if it seems linear or not. When the explanatory variable is functional this is not possible and one needs a structural testing procedure to check linearity. An application of Theorem 10.3.2 answers to the issue of testing linearity in this context if one gets an estimator r_0^* that fulfills assumptions of Theorem 10.3.2. In [102], it has been proposed an estimator that may, under appropriate assumptions, fulfill the conditional assumptions (10.28)-(10.30) discussed in the previous

paragraph (see Theorem 3 in [102] for more details).

– In the third example, one deals with the issue of testing if the effect of the functional explanatory variable X may be resumed in the effect of a vector $V(X)$ deduced from this functional variable with a known operator V. In other words one aims to test if the true regression operator may be rewritten as a smooth function defined on $\mathbb{R}^p$. In a few words, the goal is to check if a dimension reduction is possible or not. In the case of multivariate explanatory variable (i.e. $E = \mathbb{R}^m$), this problem has been widely studied notably in [198], [275], [195] and more recently [276] or [273] in a different context. However, in the more general case where E is an infinite dimensional space it seems that no result exists. Here, Theorem 10.3.2 may be applied in order to answer to this problem with a good choice of nonparametric estimator r_0^*. For instance, if the regression function between Y and $V(X)$ is smooth enough, some kernel estimator may fulfill, under apropriate assumptions, the conditions of Theorem 10.3.2. Moreover, one may imagine to adapt the refinement introduced in [276] to Theorem 10.3.2 in order to have a better detection of small local alternatives.

10.6. Conclusion

The results of this paper provide a theoretical framework for structural testing procedures in functional nonparametric regression models. The interest of such procedures in the case of a functional explanatory variable is very important all the more so in that scatter plot is not possible. The tests proposed above are innovative in this context and complete the former results established in multivariate regression models. They enable to test if the true regression operator r is a given operator r_0 or if r belongs to a family of operators $\mathcal{F}$. Theorems 10.3.1 and 10.3.2 have been established under general assumptions allowing a large scope of potential applications. The main issue to apply the results above consists in finding an "appropriate" regression estimator under the structural submodel $\mathcal{F}$.
Finally, in order to avoid the estimation of the constants involved in the asymptotic law of our test statistic one should in practice use bootstrap arguments to get empirical quantiles of the law of our statistic. From a theoretical point of view, the justification of bootstrap methods is often very similar to the initial proof. Besides, from a practical point of view bootstrap methods can be easily implemented. Hence, the natural prospects of this paper are to implement bootstrap methods and study their practical and theoretical behaviour on a particular structural test (for instance test of linearity, no effect, dimension reduction,...).

10.7. Appendix : proofs

Proof of Lemma 10.3.1 The asymptotic expression of the mean of $T_{1,n}$ comes directly from a straightforward calculus using Fubini-Tonelli's theorem and the fact

that variables (X_i, Y_i) are independent and identically distributed.

$$\begin{aligned}\mathbb{E}[T_{1,n}] &= \sum_{i=1}^{n}\mathbb{E}\left[\int \epsilon_i^2 K^2\left(\frac{d(X_i,x)}{h_n}\right) w(x)\, dP_X(x)\right] \\ &= n\sigma_\epsilon^2 \int \mathbb{E}\left[K^2\left(\frac{d(X_i,x)}{h_n}\right)\right] w(x)\, dP_X(x).\end{aligned}$$

Now we use the method introduced in [167] to get a more explicit expression :

$$\begin{aligned}&\mathbb{E}[T_{1,n}] \\ =\ & n\sigma_\epsilon^2 \int \mathbb{E}\left[1_{\{d(X,x)\le h_n\}}\left(K^2(1) - \int_{\frac{d(X,x)}{h_n}}^{1}\left(K^2\right)'(s)\, ds\right)\right] w(x)\, dP_X(x) \\ =\ & n\sigma_\epsilon^2 \int \Big(K^2(1)\,\mathbb{P}(d(x,X)\le h_n) \\ & \qquad - \int_0^1 \left(K^2\right)'(s)\,\mathbb{P}(d(x,X)\le s h_n)\, ds\Big) w(x)\, dP_X(x) \\ =\ & n\sigma_\epsilon^2\Big(K^2(1)\int \mathbb{P}(d(x,X)\le h_n)\, w(x)\, dP_X(x) \\ & \qquad - \int_0^1 \left(K^2\right)'(s)\int \mathbb{P}(d(x,X)\le s h_n)\, w(x)\, dP_X(x)\, ds\Big) \\ =\ & n\sigma_\epsilon^2\Omega_1(h_n)\left(K^2(1) - \int_0^1 \left(K^2\right)'(s)\,\frac{\Omega_1(s h_n)}{\Omega_1(h_n)} ds\right).\end{aligned}$$

Moreover, if $K = 1_{[0,1]}$, it comes directly that $\mathbb{E}[T_{1,n}] = n\Omega_1(h_n)\sigma_\epsilon^2$.

To end the proof we now have to bound the variance of $T_{1,n}$. One starts from the fact that variables (X_i, Y_i) are independent and identically distributed and consequently (X_i, ϵ_i) have these properties. Then, variance may be bounded by the second order moment and the third line comes from assumption (10.7) and Fubini-Tonelli's theorem.

$$\begin{aligned}&Var(T_{1,n}) \\ =\ & nVar\left(\int \epsilon^2 K^2\left(\frac{d(X,x)}{h_n}\right) w(x)\, dP_X(x)\right) \\ \le\ & n\mathbb{E}\left[\epsilon^4\left(\int K^2\left(\frac{d(X,x)}{h_n}\right) w(x)\, dP_X(x)\right)^2\right] \\ \le\ & Cn\int\int \mathbb{E}\left[K^2\left(\frac{d(X,x)}{h_n}\right) K^2\left(\frac{d(X,y)}{h_n}\right)\right] w(x)\, w(y)\, dP_X(x)\, dP_X(y) \\ \le\ & Cn\int\int \mathbb{P}(d(x,X)\le h_n, d(X,y)\le h_n)\, w(x)\, w(y)\, dP_X(x)\, dP_X(y).\end{aligned}$$

By simplicity we introduce the following notation

$$\Omega_3(h_n) = \int\int \mathbb{P}(d(x,X)\le h_n, d(X,y)\le h_n)\, w(x)\, w(y)\, dP_X(x)\, dP_X(y)$$

The following Lemma ends the proof.

Lemma 10.7.1. — *If assumptions (10.4) and (10.9) hold then one gets :*

$$\exists\, C > 0,\, \Omega_3\left(h_n\right) \leq C\Phi^2\left(h_n\right).$$

□

Proof of Lemma 10.3.2 This result is an application of Theorem 1 of [205] that consists in a central limit theorem for degenerated U-Statistics. Indeed, $T_{2,n}$ may be written as a degenerated U-statistic :

$$\begin{aligned} T_{2,n} &= \sum_{1\leq i\neq j\leq n} \int \epsilon_i\epsilon_j K\left(\frac{d\left(X_i,x\right)}{h_n}\right) K\left(\frac{d\left(X_j,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right) \\ &= \sum_{1\leq i\neq j\leq n} \mathcal{H}_n\left(\left(\epsilon_i,X_i\right),\left(\epsilon_j,X_j\right)\right), \end{aligned} \tag{10.31}$$

where

$$\mathcal{H}_n\left(\left(a_1,f_1\right),\left(a_2,f_2\right)\right) = \int a_1a_2K\left(\frac{d\left(f_1,x\right)}{h_n}\right) K\left(\frac{d\left(f_2,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right).$$

The operator $\mathcal{H}_n$ is by definition symmetric. Then, it has to be shown that the general term $\mathcal{H}_n\left(\left(\epsilon_i,X_i\right),\left(\epsilon_j,X_j\right)\right)$ have zero mean conditionally to $\left(\epsilon_i,X_i\right)$. To do that we take the mean conditionally to X_j and conclude with the independence of the variables $\left(X_i,\epsilon_i\right)$ and that ϵ is not correlated with X.

$$\begin{aligned} &\mathbb{E}\left[\mathcal{H}_n\left(\left(\epsilon_i,X_i\right),\left(\epsilon_j,X_j\right)\right) \mid \left(\epsilon_i,X_i\right)\right] \\ &= \mathbb{E}\left[\epsilon_i\epsilon_j \int K\left(\frac{d\left(X_i,x\right)}{h_n}\right) K\left(\frac{d\left(X_j,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right) \mid \left(\epsilon_i,X_i\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\epsilon_i\epsilon_j \int K\left(\frac{d\left(X_i,x\right)}{h_n}\right) K\left(\frac{d\left(X_j,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right) \mid \left(\epsilon_i,X_i\right),X_j\right] \mid \left(\epsilon_i,X_i\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\epsilon_j \mid \left(\epsilon_i,X_i\right),X_j\right]\epsilon_i \int K\left(\frac{d\left(X_i,x\right)}{h_n}\right) K\left(\frac{d\left(X_j,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right) \mid \left(\epsilon_i,X_i\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\epsilon_j \mid X_j\right]\epsilon_i \int K\left(\frac{d\left(X_i,x\right)}{h_n}\right) K\left(\frac{d\left(X_j,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right) \mid \left(\epsilon_i,X_i\right)\right] \\ &= 0 \end{aligned}$$

The next step of the proof consists in a detailed study of the second order moment of $\mathcal{H}_n$. We start with Fubini-Tonelli's theorem to reverse mean and integral, then we use the fact that $\left(X_i,\epsilon_i\right)$ are independent and identically distributed and assumption

(10.7).

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{H}_n^2\left(\left(\epsilon_1,X_1\right),\left(\epsilon_2,X_2\right)\right)\right]\\
=\;&\mathbb{E}\left[\int\int\left(\prod_{\ell=1}^{2}\epsilon_\ell^2 K\left(\frac{d\left(X_\ell,x\right)}{h_n}\right)K\left(\frac{d\left(X_\ell,y\right)}{h_n}\right)\right)w\left(x\right)w\left(y\right)dP_X\left(x\right)dP_X\left(y\right)\right]\\
=\;&\int\int\left(\mathbb{E}\left[\epsilon_1^2 K\left(\frac{d\left(X_1,x\right)}{h_n}\right)K\left(\frac{d\left(X_1,y\right)}{h_n}\right)\right]\right)^2 w\left(x\right)w\left(y\right)dP_X\left(x\right)dP_X\left(y\right)\\
=\;&\left(\sigma_\epsilon^2\right)^2\int\int\left(\mathbb{E}\left[K\left(\frac{d\left(X_1,x\right)}{h_n}\right)K\left(\frac{d\left(X_1,y\right)}{h_n}\right)\right]\right)^2 w\left(x\right)w\left(y\right)dP_X\left(x\right)dP_X\left(y\right)
\end{aligned}$$

Because the kernel function K has a compact support on $[0,1]$ one gets :

$$K^2\left(1\right)\left(\sigma_\epsilon^2\right)^2\Omega_4\left(h_n\right)\leq\mathbb{E}\left[\mathcal{H}_n^2\left(\left(\epsilon_1,X_1\right),\left(\epsilon_2,X_2\right)\right)\right]\leq K^2\left(0\right)\left(\sigma_\epsilon^2\right)^2\Omega_4\left(h_n\right)$$

However, it might be more interesting to have an explicit expression of the second order moment of $\mathcal{H}_n$. In this expression we will use the following notation :

$$\Omega_4\left(s_1,s_2,s_3,s_4\right)=\int_{E\times E}F_{x,y}\left(s_1,s_2\right)F_{x,y}\left(s_3,s_4\right)w\left(x\right)w\left(y\right)dP_X\left(x\right)dP_X\left(y\right).$$

$$\begin{aligned}
&\mathbb{E}\left[K\left(\frac{d\left(X,x\right)}{h_n}\right)K\left(\frac{d\left(X,y\right)}{h_n}\right)\right]\\
=\;&\mathbb{E}\left[1_{\{d(X,x)\leq h_n,\,d(X,y)\leq h_n\}}\left(K\left(1\right)-\int_{\frac{d(X,x)}{h_n}}^{1}K'\left(s\right)ds\right)\left(K\left(1\right)-\int_{\frac{d(X,y)}{h_n}}^{1}K'\left(t\right)dt\right)\right]\\
=\;&K^2\left(1\right)F_{x,y}\left(h_n,h_n\right)+\int_{[0;1]^2}K'\left(s\right)K'\left(t\right)F_{x,y}\left(sh_n,th_n\right)dsdt\\
&-K\left(1\right)\int_0^1 K'\left(s\right)F_{x,y}\left(sh_n,h_n\right)ds-K\left(1\right)\int_0^1 K'\left(t\right)F_{x,y}\left(h_n,th_n\right)ds.
\end{aligned}$$

Consequently one gets

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{H}_n^2\left(\left(\epsilon_1,X_1\right),\left(\epsilon_2,X_2\right)\right)\right]\\
=\;&\left(\sigma_\epsilon^2\right)^2\left(K^4\left(1\right)\Omega_4\left(h_n\right)+2K^2\left(1\right)\int_{[0;1]^2}K'\left(s\right)K'\left(t\right)\Omega_4\left(sh_n,h_n,th_n,h_n\right)dsdt\right.\\
&+\int_{[0;1]^4}K'\left(s_1\right)K'\left(t_1\right)K'\left(s_2\right)K'\left(t_2\right)\Omega_4\left(s_1h_n,t_1h_n,s_2h_n,t_2h_n\right)ds_1ds_2dt_1dt_2\\
&-4K^3\left(1\right)\int_0^1 K'\left(s\right)\Omega_4\left(sh_n,h_n,h_n,h_n\right)ds\\
&+2K^2\left(1\right)\int_{[0;1]^2}K'\left(s\right)K'\left(t\right)\Omega_4\left(sh_n,th_n,h_n,h_n\right)dsdt\\
&+2K^2\left(1\right)\int_{[0;1]^2}K'\left(s\right)K'\left(t\right)\Omega_4\left(sh_n,h_n,h_n,th_n\right)dsdt\\
(10.32)\;&\left.-4K\left(1\right)\int_{[0;1]^3}K'\left(r\right)K'\left(s\right)K'\left(t\right)\Omega_4\left(rh_n,sh_n,th_n,h_n\right)drdsdt\right)\\
=\;&\left(\sigma_\epsilon^2\right)^2\Omega_4\left(h_n\right)\Gamma_n,
\end{aligned}$$

where $\Gamma_n \to \Gamma$ if $\alpha_{h_n}(s_1,s_2,s_3,s_4) = \frac{\Omega_4(s_1h_n,s_2h_n,s_3h_n,s_4h_n)}{\Omega_4(h_n)} \to \alpha_0(s_1,s_2,s_3,s_4)$ (with dominated convergence theorem).

Now we have to bound the fourth order moment of $\mathcal{H}_n$ using assumption (10.7) and the fact that variables (X_i,ϵ_i) are independent.

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{H}_n^4\left((\epsilon_1,X_1),(\epsilon_2,X_2)\right)\right] \\
=\;& \mathbb{E}\left[\epsilon_1^4\epsilon_2^4\int_{E^4}\Pi_{i=1}^4\left[K\left(\frac{d(X_1,x_i)}{h_n}\right)K\left(\frac{d(X_2,x_i)}{h_n}\right)w(x_i)\,dP_X(x_i)\right]\right] \\
\leq\;& C\int_{E^4}\mathbb{E}\left[\Pi_{i=1}^4\left[K\left(\frac{d(X_1,x_i)}{h_n}\right)K\left(\frac{d(X_2,x_i)}{h_n}\right)\right]\right]\Pi_{i=1}^4\left[w(x_i)\,dP_X(x_i)\right] \\
\leq\;& C\int_{E^4}F^2_{x_1,x_2,x_3,x_4}(h_n,h_n,h_n,h_n)\,\Pi_{i=1}^4\left[w(x_i)\,dP_X(x_i)\right].
\end{aligned}$$

By simplicity we introduce the following notation :

$$\Omega_6(h_n) = \int_{E^4}F^2_{x_1,x_2,x_3,x_4}(h_n,h_n,h_n,h_n)\,\Pi_{i=1}^4\left[w(x_i)\,dP_X(x_i)\right].$$

Finally we have to study the function

$$G_n\left((a_1,f_1),(a_2,f_2)\right) = \mathbb{E}\left[\mathcal{H}_n\left((\epsilon_1,X_1),(a_1,f_1)\right)\mathcal{H}_n\left((\epsilon_1,X_1),(a_2,f_2)\right)\right].$$

Firstly we explicit this function using Fubini-Tonelli's theorem and assumption (10.7).

$$\begin{aligned}
&G_n\left((a_1,f_1),(a_2,f_2)\right) \\
=\;& \mathbb{E}\Big[\int_{E\times E}\epsilon_1^2a_1a_2K\left(\frac{d(f_1,x)}{h_n}\right)K\left(\frac{d(f_2,y)}{h_n}\right)K\left(\frac{d(X_1,x)}{h_n}\right)K\left(\frac{d(X_1,y)}{h_n}\right) \\
&\quad w(x)\,w(y)\,dP_X(x)\,dP_X(y)\Big] \\
=\;& \sigma_\epsilon^2a_1a_2\int_{E\times E}K\left(\frac{d(f_1,x)}{h_n}\right)K\left(\frac{d(f_2,y)}{h_n}\right)\mathbb{E}\left[K\left(\frac{d(X_1,x)}{h_n}\right)K\left(\frac{d(X_1,y)}{h_n}\right)\right] \\
&\qquad w(x)\,w(y)\,dP_X(x)\,dP_X(y).
\end{aligned}$$

Consequently one obtains the following bounds :

$$\begin{aligned}
&\left|G_n\left((a_1,f_1),(a_2,f_2)\right)\right| \\
\leq\;& K^2(0)\,\sigma_\epsilon^2|a_1a_2|\int_{E\times E}K\left(\frac{d(f_1,x)}{h_n}\right)K\left(\frac{d(f_2,y)}{h_n}\right)F_{x,y}(h_n,h_n)\left(\prod_{z=x,y}w(z)\,dP_X(z)\right).
\end{aligned}$$

Hence one gets an upper bound for the second order moment of G_n conditioning by (X_i,X_j), using Fubini-Tonelli's theorem, the variables independence and finally

assumption (10.5) :

$$\begin{aligned}
&\mathbb{E}\left[G_n^2\left(\left(\epsilon_1,X_1\right),\left(\epsilon_2,X_2\right)\right)\right]\\
&\leq C\mathbb{E}\big[\epsilon_1^2\epsilon_2^2\int_{E^4}K\left(\frac{d\left(X_1,x_1\right)}{h_n}\right)K\left(\frac{d\left(X_2,x_2\right)}{h_n}\right)K\left(\frac{d\left(X_1,x_3\right)}{h_n}\right)K\left(\frac{d\left(X_2,x_4\right)}{h_n}\right)\\
&\qquad F_{x_1,x_2}\left(h_n,h_n\right)F_{x_3,x_4}\left(h_n,h_n\right)\Pi_{i=1}^4\left[x\left(x_i\right)dP_X\left(x_i\right)\right]\big]\\
&\leq C\int_{E^4}\mathbb{E}\left[K\left(\frac{d\left(X_1,x_1\right)}{h_n}\right)K\left(\frac{d\left(X_1,x_3\right)}{h_n}\right)\right]\mathbb{E}\left[K\left(\frac{d\left(X_2,x_2\right)}{h_n}\right)K\left(\frac{d\left(X_2,x_4\right)}{h_n}\right)\right]\\
&\qquad F_{x_1,x_2}\left(h_n,h_n\right)F_{x_3,x_4}\left(h_n,h_n\right)\Pi_{i=1}^4\left[x\left(x_i\right)dP_X\left(x_i\right)\right]\\
&\leq C\int_{E^4}F_{x_1,x_2}\left(h_n,h_n\right)F_{x_3,x_4}\left(h_n,h_n\right)F_{x_1,x_3}\left(h_n,h_n\right)F_{x_2,x_4}\left(h_n,h_n\right)\prod_{i=1}^4\left[w\left(x_i\right)dP_X\left(x_i\right)\right].
\end{aligned}$$

By simplicity we introduce the following notation :

$$\begin{aligned}
&\Omega_5\left(h_n\right)\\
&=\int_{E^4}F_{x_1,x_2}\left(h_n,h_n\right)F_{x_3,x_4}\left(h_n,h_n\right)F_{x_1,x_3}\left(h_n,h_n\right)F_{x_2,x_4}\left(h_n,h_n\right)\Pi_{i=1}^4\left[x\left(x_i\right)dP_X\left(x_i\right)\right].
\end{aligned}$$

In order to apply the central limit theorem to the degenerated U-statistic defined by (10.31) we have to show that :

$$\frac{\Omega_5\left(h_n\right)+\frac{1}{n}\Omega_6\left(h_n\right)}{\left(\Omega_4\left(h_n\right)\right)^2}\to 0.$$

To do that, we use the following Lemma :

Lemma 10.7.2. — *Under assumptions (10.4) and (10.9) one gets :*

$$\Omega_5\left(h_n\right)=O\left(\Phi^7\left(h_n\right)\right)\ \text{ and }\Omega_6\left(h_n\right)=O\left(\Phi^5\left(h_n\right)\right).$$

Hence, we have :

$$\begin{aligned}
\frac{\Omega_5\left(h_n\right)+\frac{1}{n}\Omega_6\left(h_n\right)}{\left(\Omega_4\left(h_n\right)\right)^2} &= O\left(\frac{\Phi^7\left(h_n\right)+\frac{1}{n}\Phi^5\left(h_n\right)}{\Phi^{6+2l}\left(h_n\right)}\right)\\
&= O\left(1\right)\left(\Phi^{1-2l}\left(h_n\right)+\frac{1}{n\Phi^{1+2l}\left(h_n\right)}\right).
\end{aligned}$$

The last line tends to zero because of assumption (10.10) and hence $T_{2,n}$ is asymptotically gaussian with variance given by

$$2n^2\mathbb{E}\left[\mathcal{H}_n^2\left(\left(\epsilon_1,X_1\right),\left(\epsilon_2,X_2\right)\right)\right]=2n^2\left(\sigma_\epsilon^2\right)^2\Omega_4\left(h_n\right)\Gamma_n,$$

where Γ_n is given by (10.32).

□

Proof of Lemma 10.3.3 By definition $T_{3,n}\geq 0$ hence the first statement of this lemma is trivial.

To deal with $T_{4,n}$ we introduce the following decomposition :

$$\begin{aligned}
& \mathbb{E}\left[T_{4,n}\right] \\
= \; & \mathbb{E}\Big[\int \sum_{1\leq i\neq j\leq n} K\left(\frac{d\left(X_i,x\right)}{h_n}\right)\left(r\left(X_i\right)-r_0\left(X_i\right)\right) K\left(\frac{d\left(X_j,x\right)}{h_n}\right)\left(r\left(X_j\right)-r_0\left(X_j\right)\right) \\
& \qquad w\left(x\right) dP_X\left(x\right)\Big] \\
= \; & n\left(n-1\right)\int\left(\mathbb{E}\left[K\left(\frac{d\left(X,x\right)}{h_n}\right)\left(r\left(X\right)-r_0\left(X\right)\right)\right]\right)^2 w\left(x\right) dP_X\left(x\right) \\
= \; & n\left(n-1\right)\int\left(\mathbb{E}\left[K\left(\frac{d\left(X,x\right)}{h_n}\right)\left(\left(r-r_0\right)\left(X\right)-\left(r-r_0\right)\left(x\right)\right)\right]\right)^2 w\left(x\right) dP_X\left(x\right) \\
+ \; & n\left(n-1\right)\int\left(\mathbb{E}\left[K\left(\frac{d\left(X,x\right)}{h_n}\right)\right]\right)^2 \left(r\left(x\right)-r_0\left(x\right)\right)^2 w\left(x\right) dP_X\left(x\right) \\
+ \; & 2n\left(n-1\right)\int \mathbb{E}\left[K\left(\frac{d\left(X,x\right)}{h_n}\right)\left(\left(r-r_0\right)\left(X\right)-\left(r-r_0\right)\left(x\right)\right)\right]\mathbb{E}\left[K\left(\frac{d\left(X,x\right)}{h_n}\right)\right] \\
& \qquad \left(r\left(x\right)-r_0\left(x\right)\right) w\left(x\right) dP_X\left(x\right) \\
= \; & U_{1,n}+U_{2,n}+U_{3,n}.
\end{aligned}$$

Since $r-r_0$ is of Hölder-type and assumptions (10.5) and (10.9) hold, the first integral may be bounded in the following way :

$$\begin{aligned}
U_{1,n} &\leq Ch_n^{2\beta}n^2\int\left(\mathbb{E}\left[K\left(\frac{d\left(X,x\right)}{h_n}\right)\right]\right)^2 w\left(x\right) dP_X\left(x\right) \\
&\leq Ch_n^{2\beta}n^2K^2\left(0\right)\int F_x^2\left(h_n\right) w\left(x\right) dP_X\left(x\right) \\
&= O\left(h_n^{2\beta}n^2\Phi^2\left(h_n\right)\right).
\end{aligned}$$

Afterwards, assumption (10.8) enables to get the following inequality.

$$\begin{aligned}
U_{2,n} &\geq Cn^2\int\left(F_x\left(h_n\right)\right)^2\left(r\left(x\right)-r_0\left(x\right)\right)^2 w\left(x\right) dP_X\left(x\right) \\
&\geq Cn^2\Phi^2\left(h_n\right)\eta_n^2.
\end{aligned}$$

Consequently, the upper bound of $U_{1,n}$ is negligible with respect to the lower bound of $U_{2,n}$:

$$\frac{h_n^{2\beta}n^2\Phi^2\left(h_n\right)}{n^2\Phi^2\left(h_n\right)\eta_n^2} \leq C\frac{h_n^{2\beta}}{v_n^2\left(h_n^{\beta}\right)^2} \leq C\frac{1}{v_n^2} \to 0.$$

Finally, a Cauchy-Schwartz inequality implies that $|U_{3,n}| \leq \sqrt{U_{1,n}}\sqrt{U_{2,n}}$, hence $U_{3,n}$ is also negligible with respect to $U_{2,n}$. Consequently the lower bound shown for $U_{2,n}$ still holds for $\mathbb{E}_{\mathcal{H}_1}\left[T_{4,n}\right]$ with a constant C small enough.

□

Proof of Lemma 10.3.4 To obtain bounds for variances, we use the same arguments as in the second part of the proof of Lemma 10.3.1. The only difference is that we use the fact that r_0 and r are of Hölder-type (assumption (10.6)), instead of assumption (10.7), to get that $|r(X) - r_0(X)| 1_{d(x,X)\leq h_n} 1_W(x)$ is bounded. Indeed, if x_0 is a fixed point of W, because the diameter of W is finite since W is bounded, assumption (10.6) gives :

$$\begin{aligned} & |r(X) - r_0(X)| 1_{d(x,X)\leq h_n} 1_W(x) \\ \leq\ & |r(x_0) - r_0(x_0)| + Cd^\beta(X, x_0) 1_{d(x,X)\leq h_n} 1_W(x) \\ \leq\ & |r(x_0) - r_0(x_0)| + Ch_n^\beta + Cdiam^\beta(W) \\ =\ & O(1). \end{aligned}$$

Let us focus now on the way to bound the variance of $T_{4,n}$. Similar arguments may be employed to bound the variance of $T_{6,n}$. To reverse mean and integral we can't use Fubini-Tonelli's theorem since the integrated function is not nonnegative. However it is easy to see that since r_0 is of Hölder-type on W, r_0 is bounded on W, hence Fubini's theorem may be employed. Consequently it comes with Fubini's theorem and symmetry arguments :

$$\begin{aligned} & Var(T_{4,n}) \\ & = \mathbb{E}\Big[\Big(\int \sum_{1\leq i\neq j\leq n} (r(X_i) - r_0(X_i))(r(X_j) - r_0(X_j)) K\left(\frac{d(X_i,x)}{h_n}\right) K\left(\frac{d(X_j,x)}{h_n}\right) \\ & \quad w(x)\, dP_X(x)\Big)^2\Big] \\ & -\Big(\mathbb{E}\Big[\int \sum_{1\leq i\neq j\leq n} (r(X_i) - r_0(X_i))(r(X_j) - r_0(X_j)) K\left(\frac{d(X_i,x)}{h_n}\right) K\left(\frac{d(X_j,x)}{h_n}\right) \\ & \quad w(x)\, dP_X(x)\Big]\Big)^2 \\ & = n(n-1)(n-2)(n-4)\left(\int \left(\mathbb{E}\left[(r(X) - r_0(X)) K\left(\frac{d(X,x)}{h_n}\right)\right]\right)^2 w(x)\, dP_X(x)\right)^2 \\ & +4n(n-1)(n-2) A_n \\ & +2n(n-1) B_n \\ & -\left(n(n-1)\int \left(\mathbb{E}\left[(r(X) - r_0(X)) K\left(\frac{d(X,x)}{h_n}\right)\right]\right)^2 w(x)\, dP_X(x)\right)^2 \\ & \leq 4n(n-1)(n-2) A_n + 2n(n-1) B_n, \end{aligned}$$

with

$$\begin{aligned} A_n \ =\ & \int_{E\times E} \mathbb{E}\left[(r(X) - r_0(X)) K\left(\frac{d(X,x)}{h_n}\right)\right] \mathbb{E}\left[(r(X) - r_0(X)) K\left(\frac{d(X,y)}{h_n}\right)\right] \\ & \mathbb{E}\left[(r(X) - r_0(X))^2 K\left(\frac{d(X,x)}{h_n}\right) K\left(\frac{d(X,y)}{h_n}\right)\right] w(x)\, dP_X(x)\, w(y)\, dP_X(y), \end{aligned}$$

and

$$B_n = \int_{E\times E}\left(\mathbb{E}\left[\left(r\left(X\right)-r_0\left(X\right)\right)^2 K\left(\frac{d\left(X,x\right)}{h_n}\right)K\left(\frac{d\left(X,y\right)}{h_n}\right)\right]\right)^2 w\left(x\right)dP_X\left(x\right)w\left(y\right)dP_X\left(y\right).$$

Under assumptions (10.5) and (10.6) one gets :

$$\begin{aligned}
&A_n\\
&\leq C\int_{E\times E}\mathbb{E}\left[\left(r\left(X\right)-r_0\left(X\right)\right)^2 K\left(\frac{d\left(X,x\right)}{h_n}\right)K\left(\frac{d\left(X,y\right)}{h_n}\right)\right]\\
&\quad\mathbb{E}\left[K\left(\frac{d\left(X,x\right)}{h_n}\right)\right]\mathbb{E}\left[K\left(\frac{d\left(X,y\right)}{h_n}\right)\right]w\left(x\right)dP_X\left(x\right)w\left(y\right)dP_X\left(y\right)\\
&\leq C\int_{E\times E}\mathbb{E}\left[\left(r\left(x\right)-r_0\left(x\right)\right)^2 K\left(\frac{d\left(X,x\right)}{h_n}\right)K\left(\frac{d\left(X,y\right)}{h_n}\right)\right]\\
&\qquad F_x\left(h_n\right)F_y\left(h_n\right)w\left(x\right)dP_X\left(x\right)w\left(y\right)dP_X\left(y\right)\\
&+C\int_{E\times E}\mathbb{E}\left[\left(\left(r-r_0\right)\left(X\right)-\left(r-r_0\right)\left(x\right)\right)^2 K\left(\frac{d\left(X,x\right)}{h_n}\right)K\left(\frac{d\left(X,y\right)}{h_n}\right)\right]\\
&\qquad F_x\left(h_n\right)F_y\left(h_n\right)w\left(x\right)dP_X\left(x\right)w\left(y\right)dP_X\left(y\right)\\
&\leq C\int_{E\times E}\left(h_n^{2\beta}+\left(r-r_0\right)^2\left(x\right)\right)F_x\left(h_n\right)F_y\left(h_n\right)F_{x,y}\left(h_n,h_n\right)w\left(x\right)dP_X\left(x\right)w\left(y\right)dP_X\left(y\right)\\
&\leq C\Phi^2\left(h_n\right)\int_E\left(h_n^{2\beta}+\left(r-r_0\right)^2\left(x\right)\right)\int_E F_{x,y}\left(h_n,h_n\right)w\left(y\right)dP_X\left(y\right)w\left(x\right)dP_X\left(x\right)\\
&\leq C\Phi^4\left(h_n\right)\left(h_n^{2\beta}+\int_E\left(r-r_0\right)^2\left(x\right)w\left(x\right)dP_X\left(x\right)\right)\\
&\leq C\Phi^4\left(h_n\right)\int_E\left(r-r_0\right)^2\left(x\right)w\left(x\right)dP_X\left(x\right).
\end{aligned}$$

The last line is a direct consequence of assumption (10.8). With similar arguments B_n may be bounded as follows :

$$\begin{aligned}
B_n &\leq C\int_{E\times E}\left(h_n^{2\beta}+\left(r-r_0\right)^2\left(x\right)\right)F_{x,y}^2\left(h_n,h_n\right)w\left(x\right)dP_X\left(x\right)w\left(y\right)dP_X\left(y\right)\\
&\leq C\Phi^3\left(h_n\right)\int_E\left(r-r_0\right)^2\left(x\right)w\left(x\right)dP_X\left(x\right).
\end{aligned}$$

Then, since $n\Phi\left(h_n\right)\to+\infty$, $n^2\Phi^3\left(h_n\right)$ is negligible with respect to $n^3\Phi^4\left(h_n\right)$. Hence the bound obtained for $2n\left(n-1\right)B_n$ is negligible with regard to the upper bound of $4n\left(n-1\right)\left(n-2\right)A_n$. Consequently

$$Var\left(T_{4,n}\right)\leq Cn^3\Phi^4\left(h_n\right)\left\|r-r_0\right\|^2_{\mathbb{L}^2(wdP_X)}.$$

Finally, variance of $T_{5,n}$ may be bounded in the same way with the Hölder assumption to bound $(r(X) - r_0(X))^2$ and assumption (10.7) to bound $\mathbb{E}[\epsilon^2 \mid X]$.

$$
\begin{aligned}
&Var(T_{5,n}) \\
&= nVar\left(\int \epsilon (r(X) - r_0(X)) K^2\left(\frac{d(X,x)}{h_n}\right) w(x)\, dP_X(x)\right) \\
&\leq Cn\mathbb{E}\left[\left(\int |(r-r_0)(X) - (r-r_0)(x)| K^2\left(\frac{d(X,x)}{h_n}\right) w(x)\, dP_X(x)\right)^2\right] \\
&+Cn\mathbb{E}\left[\left(\int |(r-r_0)(x)| K^2\left(\frac{d(X,x)}{h_n}\right) w(x)\, dP_X(x)\right)^2\right] \\
&\leq Cnh_n^{2\beta}\mathbb{E}\left[\left(\int K^2\left(\frac{d(X,x)}{h_n}\right) w(x)\, dP_X(x)\right)^2\right] \\
&+Cn\mathbb{E}\left[\int_{E\times E} (r-r_0)^2(x) K^2\left(\frac{d(X,x)}{h_n}\right) K^2\left(\frac{d(X,y)}{h_n}\right) w(x)\, dP_X(x)\, w(y)\, dP_X(y)\right] \\
&\leq Cnh_n^{2\beta}\mathbb{E}\left[F_X^2(h_n) 1_{W_\gamma}(X)\right] \\
&+Cn\int (r-r_0)^2(x)\, \mathbb{E}\left[K^2\left(\frac{d(X,x)}{h_n}\right)\int K^2\left(\frac{d(X,y)}{h_n}\right) w(y)\, dP_X(y)\right] w(x)\, dP_X(x) \\
&\leq Cn\Phi^2(h_n)\left(h_n^{2\beta} + \|r-r_0\|^2_{\mathbb{L}^2(wdP_X)}\right).
\end{aligned}
$$

Assumption (10.8) is enough to conclude.

□

The next lemmas provide upper bounds for the sequences $(\Omega_i(h_n))_n$. However if one makes an assumption on the asymptotic behaviour of $F_{x,y}(h_n, h_n)$, then one will obtain other bounds and so assumptions (10.8)-(10.10) will be different.

Proof of Lemma 10.3.5 This proof lies on assumptions (10.4) and (10.9). Assumption (10.9) notably implies that for n large enough ($n > N_0$) one gets :

$$
\forall u \in W_{h_n},\ C_1\Phi(h_n) \leq F_u(h_n) \leq C_2\Phi(h_n). \tag{10.33}
$$

$$
\begin{aligned}
&\Omega_4(h_n) \\
&= \int_{E^4} 1_{\{d(x,u)\leq h_n,\, d(y,u)\leq h_n,\, d(x,v)\leq h_n,\, d(y,v)\leq h_n\}} w(x)\, w(y)\, dP_X(x)\, dP_X(y)\, dP_X(u)\, dP_X(v) \\
&\leq \int_{E^4} 1_{\{d(x,u)\leq h_n,\, d(y,u)\leq h_n,\, d(x,v)\leq h_n\}} w(x)\, w(y)\, dP_X(x)\, dP_X(y)\, dP_X(u)\, dP_X(v) \\
&\leq C\int_{E^2} F_x(h_n) F_u(h_n) 1_{\{d(x,u)\leq h_n\}} w(x) 1_{W_{h_n}}(u)\, dP_X(x)\, dP_X(u).
\end{aligned}
$$

Hence one gets :

$$\begin{aligned}\Omega_4(h_n) &\leq C\Phi^2(h_n)\int_{E^2} 1_{\{d(x,u)\leq h_n\}} w(x)\, dP_X(x)\, dP_X(u)\\ &\leq C\Phi^2(h_n)\int_E F_x(h_n)\, w(x)\, dP_X(x)\, dP_X(u)\\ &= O\left(\Phi^3(h_n)\right).\end{aligned}$$

□

Proof of Lemma 10.7.1 This proof, very similar to the proof of Lemma 10.3.5, lies on the statement (10.33) and assumption (10.4).

$$\begin{aligned}\Omega_3(h_n) &= \int_{E^3} 1_{\{d(x,u)\leq h_n,\, d(y,u)\leq h_n\}} w(x)\, w(y)\, dP_X(x)\, dP_X(y)\, dP_X(u)\\ &\leq C\int_{E^3} 1_{\{d(x,u)\leq h_n,\, d(y,u)\leq h_n\}} 1_{W_{h_n}}(u)\, dP_X(x)\, dP_X(y)\, dP_X(u)\\ &\leq C\int_E F_u^2(h_n)\, 1_{W_{h_n}}(u)\, dP_X(u)\\ &= O\left(\Phi^2(h_n)\right).\end{aligned}$$

□

Proof of Lemma 10.7.2 This proof is similar to the proofs of Lemmas 10.3.5 and 10.7.1. We use many times assumption (10.4) and the statement (10.33).

$$\begin{aligned}&\Omega_5(h_n)\\ &= \int_{E^4} F_{x_1,x_2}(h_n,h_n)\, F_{x_3,x_4}(h_n,h_n)\, F_{x_1,x_3}(h_n,h_n)\, F_{x_2,x_4}(h_n,h_n) \prod_{i=1}^4 \left[x(x_i)\, dP_X(x_i)\right]\\ &= \int_{E^8} 1\begin{cases} d(x_1,u_1)\leq h_n,\, d(x_2,u_1)\leq h_n,\, d(x_1,u_2)\leq h_n,\, d(x_3,u_2)\leq h_n\\ d(x_2,u_3)\leq h_n,\, d(x_4,u_3)\leq h_n,\, d(x_3,u_4)\leq h_n,\, d(x_4,u_4)\leq h_n\end{cases}\\ &\qquad \prod_{i=1}^4 \left[w(x_i)\, dP_X(x_i)\, dP_X(u_i)\right]\\ &\leq \int_{E^8} 1\begin{cases} d(x_1,u_1)\leq h_n,\, d(x_2,u_1)\leq h_n,\, d(x_1,u_2)\leq h_n,\, d(x_3,u_2)\leq h_n\\ d(x_2,u_3)\leq h_n,\, d(x_4,u_3)\leq h_n,\, d(x_3,u_4)\leq h_n\end{cases}\\ &\qquad \prod_{i=1}^4 \left[w(x_i)\, dP_X(x_i)\, dP_X(u_i)\right]\end{aligned}$$

In the last line x_4 and u_4 appear only once hence the idea is to integrate by x_4 and u_4.

$$
\begin{aligned}
&\Omega_5(h_n)\\
&\leq C\int_{E^6} F_{u_3}(h_n)\,F_{x_3}(h_n)\,1_{\begin{cases} d(x_1,u_1)\leq h_n,\, d(x_2,u_1)\leq h_n,\, d(x_1,u_2)\leq h_n \\ d(x_3,u_2)\leq h_n,\, d(x_2,u_3)\leq h_n \end{cases}} 1_{W_{h_n}}(u_3)\\
&\qquad \prod_{i=1}^{3}\left[w(x_i)\,dP_X(x_i)\,dP_X(u_i)\right]\\
&\leq C\Phi^2(h_n)\int_{E^6} 1_{\begin{cases} d(x_1,u_1)\leq h_n,\, d(x_2,u_1)\leq h_n,\, d(x_1,u_2)\leq h_n \\ d(x_3,u_2)\leq h_n,\, d(x_2,u_3)\leq h_n \end{cases}}\\
&\qquad\qquad \prod_{i=1}^{3}\left[w(x_i)\,dP_X(x_i)\,dP_X(u_i)\right]\\
&\leq C\Phi^2(h_n)\int_{E^4} F_{x_2}(h_n)\,F_{u_2}(h_n)\,1_{\begin{cases} d(x_1,u_1)\leq h_n,\, d(x_2,u_1)\leq h_n \\ d(x_1,u_2)\leq h_n \end{cases}} 1_{W_{h_n}}(u_2)\\
&\qquad\qquad \prod_{i=1}^{2}\left[w(x_i)\,dP_X(x_i)\,dP_X(u_i)\right]\\
&\leq C\Phi^4(h_n)\int_{E^2} F_{x_1}(h_n)\,F_{u_1}(h_n)\,1_{\{d(x_1,u_1)\leq h_n\}}1_{W_{h_n}}(u_1)\left[w(x_1)\,dP_X(x_1)\,dP_X(u_1)\right]\\
&\leq C\Phi^6(h_n)\int_{E} F_{x_1}(h_n)\,w(x_1)\,dP_X(x_1)\\
&= O\left(\Phi^7(h_n)\right).
\end{aligned}
$$

Let us focus now on Ω_6 in the same way :

$$
\begin{aligned}
\Omega_6(h_n) &= \int_{E^6}\prod_{i=1}^{4}\left(1_{d(x_i,u)\leq h_n,\, d(x_i,v)\leq h_n} w(x_i)\,dP_X(x_i)\right)dP_X(u)\,dP_X(v)\\
&\leq \int_{E^6}\prod_{i=1}^{4}\left(1_{d(x_i,u)\leq h_n} w(x_i)\,dP_X(x_i)\right)1_{d(x_1,v)\leq h_n} w^4(x_1)\,dP_X(u)\,dP_X(v)\\
&\leq C\int_{E^2} F_u^3(h_n)\,F_{x_1}(h_n)\,1_{d(x_1,u)\leq h_n}1_{W_{h_n}}(u)\,w(x_1)\,dP_X(u)\,dP_X(x_1)\\
&\leq C\Phi^4(h_n)\int_{E^2} 1_{d(x_1,u)\leq h_n} w(x_1)\,dP_X(u)\,dP_X(x_1)\\
&\leq C\Phi^4(h_n)\int_{E} F_{x_1}(h_n)\,w(x_1)\,dP_X(x_1)\\
&= O\left(\Phi^5(h_n)\right).
\end{aligned}
$$

□

Proof of Lemma 10.3.6

Firstly, the aim is to bound the mean of $T_{3,n}^* + T_{4,n}^*$ under $\mathcal{H}_0$.

$$\begin{aligned}
\mathbb{E}\left[T_{3,n}^*\right] &= \mathbb{E}\left[\int_E \sum_{i=1}^n \left(r\left(X_i\right) - r_0^*\left(X_i\right)\right)^2 K^2\left(\frac{d\left(X_i,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^n \left(r\left(X_i\right) - r_0^*\left(X_i\right)\right)^2 \int_E K^2\left(\frac{d\left(X_i,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right)\right] \\
&\leq Cn\mathbb{E}\left[\left(r\left(X_1\right) - r_0^*\left(X_1\right)\right)^2 \int_E 1_{d(X_1,x)\leq h_n} dP_X\left(x\right)\right] \\
&\leq Cn\mathbb{E}\left[\left(r\left(X_1\right) - r_0^*\left(X_1\right)\right)^2 F_{X_i}\left(h_n\right) 1_{d(X_1,W)\leq h_n}\right] \\
&\leq Cn\Phi\left(h_n\right)\mathbb{E}\left[\left(r\left(X_i\right) - r_0^*\left(X_i\right)\right)^2 1_{W_\alpha}\left(X_1\right)\right]. \qquad (10.34)
\end{aligned}$$

To obtain the last line we used successively the fact that the sample $\left(X_i, Y_i\right)_i$ is identically distributed, assumptions (10.4), (10.5) and (10.9). Now we have to bound the mean of $T_{4,n}^*$ with similar arguments :

$$\begin{aligned}
&\left|\mathbb{E}\left[T_{4,n}^*\right]\right| \\
&= \left|\mathbb{E}\Big[\int_E \sum_{1\leq i\neq j\leq n} \left(r\left(X_i\right) - r_0^*\left(X_i\right)\right)\left(r\left(X_j\right) - r_0^*\left(X_j\right)\right) K\left(\frac{d\left(X_i,x\right)}{h_n}\right)\right. \\
&\qquad \left. K\left(\frac{d\left(X_j,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right)\Big]\right| \\
&\leq Cn^2\left|\mathbb{E}\Big[\int_E K\left(\frac{d\left(X_1,x\right)}{h_n}\right) K\left(\frac{d\left(X_2,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right)\left(r\left(X_1\right) - r_0^*\left(X_1\right)\right)\right. \\
&\qquad \left.\left(r\left(X_2\right) - r_0^*\left(X_2\right)\right)\Big]\right| \\
&\leq Cn^2\mathbb{E}\left[\left(r\left(X_1\right) - r_0^*\left(X_1\right)\right)^2 \int_E K\left(\frac{d\left(X_1,x\right)}{h_n}\right) K\left(\frac{d\left(X_2,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right)\right] \\
&\leq Cn^2 \int_E \mathbb{E}\left[\left(r\left(X_1\right) - r_0^*\left(X_1\right)\right)^2 K\left(\frac{d\left(X_1,x\right)}{h_n}\right)\right]\mathbb{E}\left[K\left(\frac{d\left(X_2,x\right)}{h_n}\right)\right] w\left(x\right) dP_X\left(x\right) \\
&\leq Cn^2\Phi\left(h_n\right)\mathbb{E}\left[\left(r\left(X_1\right) - r_0^*\left(X_1\right)\right)^2 \int_E K\left(\frac{d\left(X_1,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right)\right] \\
&\leq Cn^2\Phi^2\left(h_n\right)\mathbb{E}\left[\left(r\left(X_1\right) - r_0^*\left(X_1\right)\right)^2 1_{W_\alpha}\left(X_1\right)\right]. \qquad (10.35)
\end{aligned}$$

The third line is obtained from the second one with a Cauchy-Swartz inequality applied to

$$f\left(X_1, X_2\right) = \left|r\left(X_1\right) - r_0^*\left(X_1\right)\right| \sqrt{\int_E K\left(\frac{d\left(X_1,x\right)}{h_n}\right) K\left(\frac{d\left(X_2,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right)}$$

and

$$g(X_1,X_2)=|r(X_2)-r_0^*(X_2)|\sqrt{\int_E K\left(\frac{d(X_1,x)}{h_n}\right)K\left(\frac{d(X_2,x)}{h_n}\right)w(x)\,dP_X(x)}.$$

Because $n\Phi(h_n)\to+\infty$, we obtain from assumption (10.13) and the previous statements (10.34) and (10.35) the following bounds :

$$0\leq\mathbb{E}\left[T_{3,n}^*+T_{4,n}^*\right]=O\left(o\left(n\Phi^{\frac{3+l}{2}}(h_n)\right)\right). \tag{10.36}$$

Furthermore, $T_{3,n}^*+T_{4,n}^*$ is by definition nonnegative hence we can use a Bienaimé-Tchebichev inequality, (10.36) and assumption (10.10) to obtain :

$$\begin{aligned}T_{3,n}^*+T_{4,n}^* &= O_p\left(\mathbb{E}\left[T_{3,n}^*+T_{4,n}^*\right]\right)\\ &= O_p\left(o\left(n\Phi^{\frac{3+l}{2}}(h_n)\right)\right)\\ &= O_p\left(o\left(\sqrt{Var\left(T_{2,n}\right)}\right)\right)\\ &= o_p\left(\sqrt{Var\left(T_{2,n}\right)}\right). \qquad (10.37)\end{aligned}$$

Remark : With similar arguments it can be shown that

$$\begin{aligned}0 &\leq \mathbb{E}\left[T_{3,n}^*+T_{4,n}^*\mid (X_i^*,Y_i^*),\,1\leq i\leq u_n\right]\\ &\leq Cn^2\Phi^2(h_n)\,\mathbb{E}\left[\left(r(X_1)-r_0^*(X_1)\right)^2\mid (X_i^*,Y_i^*),\,1\leq i\leq u_n\right].\end{aligned}$$

Now we want to prove that :

$$Var\left(T_{5,n}^*\right)=o\left(\Phi^{\frac{3+l}{2}}(h_n)\right)\text{ and }Var\left(T_{6,n}^*\right)=o\left(n^2\Phi^{\frac{7+l}{2}}(h_n)\right).$$

Because variables $T_{5,n}^*$ and $T_{6,n}^*$ are centered, we only have to consider their second order moment :

$$\begin{aligned}&\mathbb{E}\left[\left(T_{5,n}^*\right)^2\right]\\ =\;&\sum_{i=1}^n\mathbb{E}\Big[\int_{E\times E}K^2\left(\frac{d(X_i,x)}{h_n}\right)K^2\left(\frac{d(X_i,y)}{h_n}\right)\left(r(X_i)-r_0^*(X_i)\right)^2\epsilon_i^2w(x)\,w(y)\\ &\qquad dP_X(x)\,dP_X(y)\Big]\\ =\;&n\mathbb{E}\Big[\mathbb{E}\left[\epsilon_1^2\mid X_1,\,(X_k^*,Y_k^*)_{1\leq k\leq u_n}\right]\left(\int_E K^2\left(\frac{d(X_1,x)}{h_n}\right)w(x)\,dP_X(x)\right)^2\\ &\qquad\left(r(X_1)-r_0^*(X_1)\right)^2\Big]\\ \leq\;&Cn\Phi^2(h_n)\,\mathbb{E}\left[\left(r(X_1)-r_0^*(X_1)\right)^2 1_{W_\alpha}(X_1)\right].\end{aligned}$$

The second and the last lines come directly from the fact that ϵ_1 is independent of $(X_k^*,Y_k^*)_{1\leq k\leq u_n}$, centered and not correlated with X_1 and hence one gets :

$$\mathbb{E}\left[\epsilon_1^k\mid X_1,\,(X_k^*,Y_k^*)_{1\leq k\leq u_n}\right]=\mathbb{E}\left[\epsilon_1^k\mid X_1\right],\,k=1,2.$$

To get the previous lines we have successively used the fact that variables $(X_i, Y_i)_{1\leq i\leq n}$ are independent and assumptions (10.9), (10.4),(10.5),(10.7) and (10.11).
Now, assumption (10.12) implies that the second order moment of $T^*_{5,n}$ is bounded as follows :

$$\mathbb{E}\left[\left(T^*_{5,n}\right)^2\right] = o\left(\Phi^{\frac{3+l}{2}}\left(h_n\right)\right). \tag{10.38}$$

Afterwards we have to bound the second order moment of $T^*_{6,n}$. In order to make the next expressions more easy to read we will note $\mathbb{E}_*$ when we take the mean conditionally to $(X^*_i, Y^*_i)_{1\leq i\leq u_n}$.

$$\begin{aligned}
&\mathbb{E}_*\left[\left(T^*_{6,n}\right)^2\right] \\
&= \mathbb{E}_*\Big[\int_{E\times E} \sum_{\substack{1\leq i\neq j\leq n \\ 1\leq k\neq l\leq n}} K\left(\frac{d\left(X_i,x\right)}{h_n}\right) K\left(\frac{d\left(X_j,x\right)}{h_n}\right) K\left(\frac{d\left(X_k,y\right)}{h_n}\right) \\
&\qquad K\left(\frac{d\left(X_l,y\right)}{h_n}\right) \epsilon_i\epsilon_k \left(r-r^*_0\right)\left(X_j\right)\left(r-r^*_0\right)\left(X_l\right) w\left(x\right) w\left(y\right) dP_X\left(x\right) dP_X\left(y\right)\Big] \\
&= \sum_{\substack{1\leq i_1,i_2,i_3,i_4\leq n \\ i_j\neq i_k, j\neq k}} \Big(\int_E \mathbb{E}_*\left[K\left(\frac{d\left(X_1,x\right)}{h_n}\right)\epsilon_1\right]\mathbb{E}_*\left[K\left(\frac{d\left(X_1,x\right)}{h_n}\right)\left(r-r^*_0\right)\left(X_1\right)\right] \\
&\qquad w\left(x\right) dP_X\left(x\right)\Big)^2 \\
&+ \sum_{\substack{1\leq i_1,i_2,i_3\leq n \\ i_j\neq i_k, j\neq k}} \Big(\int_{E\times E} \prod_{z=x,y}\left(\mathbb{E}_*\left[K\left(\frac{d\left(X_1,z\right)}{h_n}\right)\left(r-r^*_0\right)\left(X_1\right)\right]\right) \\
&\qquad \mathbb{E}_*\left[K\left(\frac{d\left(X_1,x\right)}{h_n}\right) K\left(\frac{d\left(X_1,y\right)}{h_n}\right)\epsilon_1^2\right] w\left(x\right) dP_X\left(x\right) w\left(y\right) dP_X\left(y\right)\Big) \\
&+ \sum_{1\leq i\neq j\leq n} \Big(\int_{E\times E} \mathbb{E}_*\left[K\left(\frac{d\left(X_1,x\right)}{h_n}\right) K\left(\frac{d\left(X_1,y\right)}{h_n}\right)\left(r-r^*_0\right)^2\left(X_1\right)\right] \\
&\qquad \mathbb{E}_*\left[K\left(\frac{d\left(X_1,x\right)}{h_n}\right) K\left(\frac{d\left(X_1,y\right)}{h_n}\right)\epsilon_1^2\right] w\left(x\right) dP_X\left(x\right) w\left(y\right) dP_X\left(y\right)\Big) \\
&= A + B + D
\end{aligned} \tag{10.39}$$

The first term A is null because $\mathbb{E}\left[\epsilon|X\right] = 0$. To get the second equality we have to use that for the same reason, terms with three distinct indices, have a zero mean unless $i = k$. Moreover, terms with two distinct indices also have zero mean unless $i = k$ and $j = l$. Consequently B and D are the only remaining terms. The next step of the proof consists in providing an upper bound for B. To do that we will

use symmetry arguments together with assumptions (10.7), (10.9), (10.4),(10.5) and (10.11).

$$\begin{aligned}
|B| &\leq Cn^3 \int_{E\times E} \mathbb{E}_* \big[K\left(\frac{d(X_1,x)}{h_n}\right) K\left(\frac{d(X_1,y)}{h_n}\right) K\left(\frac{d(X_2,x)}{h_n}\right) K\left(\frac{d(X_3,y)}{h_n}\right) \\
&\qquad |\left(r(X_2)-r_0^*(X_2)\right)\left(r(X_3)-r_0^*(X_3)\right)| \big] w(x)\, dP_X(x)\, w(y)\, dP_X(y) \\
&\leq Cn^3 \Big(\int_{E\times E} \mathbb{E}_* \big[K\left(\frac{d(X_1,x)}{h_n}\right) K\left(\frac{d(X_1,y)}{h_n}\right) K\left(\frac{d(X_2,x)}{h_n}\right) K\left(\frac{d(X_3,y)}{h_n}\right) \\
&\qquad \left(r(X_2)-r_0^*(X_2)\right)^2 \big] w(x)\, dP_X(x)\, w(y)\, dP_X(y) \Big)^{\frac{1}{2}} \\
&\qquad \times \Big(\int_{E\times E} \mathbb{E}_* \big[K\left(\frac{d(X_1,x)}{h_n}\right) K\left(\frac{d(X_1,y)}{h_n}\right) K\left(\frac{d(X_2,x)}{h_n}\right) K\left(\frac{d(X_3,y)}{h_n}\right) \\
&\qquad \left(r(X_3)-r_0^*(X_3)\right)^2 \big] w(x)\, dP_X(x)\, w(y)\, dP_X(y) \Big)^{\frac{1}{2}} \\
&\leq Cn^3 \int_{E\times E} \mathbb{E}_* \left[K\left(\frac{d(X_1,y)}{h_n}\right)\right] \mathbb{E}_* \left[K\left(\frac{d(X_1,x)}{h_n}\right) \left(r(X_1)-r_0^*(X_1)\right)^2 \right] \\
&\qquad \mathbb{E}_* \left[K\left(\frac{d(X_1,x)}{h_n}\right) K\left(\frac{d(X_1,y)}{h_n}\right)\right] w(x)\, dP_X(x)\, w(y)\, dP_X(y).
\end{aligned}$$

Now the idea is to use the fact that assumptions (10.5) and (10.9) and the independence between X_i and $(X_i^*, Y_i^*)_{1\leq i\leq n}$ give $\mathbb{E}_* \left[K\left(\frac{d(X_2,y)}{h_n}\right)\right] \leq C\Phi(h_n)$ and then integrate with respect to y and use many time assumptions (10.4), (10.5) and (10.9).

$$\begin{aligned}
|B| &\leq Cn^3 \Phi(h_n) \int_E \mathbb{E}_* \left[K\left(\frac{d(X_1,x)}{h_n}\right) \int_E K\left(\frac{d(X_1,y)}{h_n}\right) w(y)\, dP_X(y)\right] \\
&\qquad \times \mathbb{E}_* \left[K\left(\frac{d(X_1,x)}{h_n}\right) \left(r(X_1)-r_0^*(X_1)\right)^2 \right] w(x)\, dP_X(x) \\
&\leq Cn^3 \Phi^2(h_n) \Big(\int_E \mathbb{E}_* \left[K\left(\frac{d(X_1,x)}{h_n}\right) \left(r(X_1)-r_0^*(X_1)\right)^2 \right] \\
&\qquad \mathbb{E}_* \left[K\left(\frac{d(X_1,x)}{h_n}\right)\right] w(x)\, dP_X(x) \Big) \\
&\leq Cn^3 \Phi^3(h_n)\, \mathbb{E}_* \left[\int_E K\left(\frac{d(X_1,x)}{h_n}\right) w(x)\, dP_X(x) \left(r(X_1)-r_0^*(X_1)\right)^2 \right] \\
&\leq Cn^3 \Phi^4(h_n)\, \mathbb{E}_* \left[\left(r(X_1)-r_0^*(X_1)\right)^2 1_{W_\alpha}(X_1)\right]. \qquad (10.40)
\end{aligned}$$

Now we have to provide an upper bound for D with similar arguments.

$$\begin{aligned} D &\leq Cn^2 \int_{E\times E} \mathbb{E}_* \left[K\left(\frac{d(X_1,x)}{h_n}\right) K\left(\frac{d(X_1,y)}{h_n}\right) (r(X_1) - r_0^*(X_1))^2 \right] \\ &\quad \times \mathbb{E}_* \left[K\left(\frac{d(X_1,x)}{h_n}\right) K\left(\frac{d(X_1,y)}{h_n}\right) \right] w(x)\, dP_X(x)\, w(y)\, dP_X(y) \\ &\leq Cn^2 \Phi(h_n)\, \mathbb{E}_* \left[(r(X_1) - r_0^*(X_1))^2 \left(\int_E K\left(\frac{d(X_1,x)}{h_n}\right) w(x)\, dP_X(x) \right)^2 \right] \\ &\leq Cn^2 \Phi^3(h_n)\, \mathbb{E}_* \left[(r(X_1) - r_0^*(X_1))^2 1_{W_\alpha}(X_1) \right]. \end{aligned} \tag{10.41}$$

We get directly from (10.39)-(10.41) and assumption (10.12) the following inequality :

$$\begin{aligned} \mathbb{E}\left[\left(T_{6,n}^*\right)^2 \right] &= O\left(n^3 \Phi^4(h_n)\, \mathbb{E}\left[(r(X_1) - r_0^*(X_1))^2 1_{W_\alpha}(X_1) \right] \right) \\ &= o\left(n^2 \Phi^{\frac{7+l}{2}}(h_n) \right). \end{aligned} \tag{10.42}$$

Moreover, variables $T_{5,n}^*$ and $T_{6,n}^*$ are centered. Hence by a Bienaimé-Tchebychev inequality, assumption (10.10) and the statement (10.38) one gets :

$$\begin{aligned} T_{5,n}^* &= O_p\left(\sqrt{Var\left(T_{5,n}^*\right)} \right) \\ &= O_p\left(o\left(\Phi^{\frac{3+l}{4}}(h_n) \right) \right) \\ &= O_p\left(o\left(\frac{\sqrt{Var(T_{2,n})}}{n\Phi^{\frac{3+l}{4}}} \right) \right) \\ &= O_p\left(o\left(\sqrt{Var(T_{2,n})} \right) \right) \\ &= o_p\left(\sqrt{Var(T_{2,n})} \right), \end{aligned} \tag{10.43}$$

because $l < 1$ and $n\Phi(h_n) \to +\infty$. We can bound $T_{6,n}$ with the same arguments and (10.42) :

$$\begin{aligned} T_{6,n}^* &= O_p\left(o\left(n\Phi^{\frac{7+l}{4}}(h_n) \right) \right) \\ &= O_p\left(o\left(\Phi^{\frac{1-l}{4}}(h_n) \sqrt{Var(T_{2,n})} \right) \right) \\ &= o_p\left(\sqrt{Var(T_{2,n})} \right). \end{aligned} \tag{10.44}$$

Statements (10.37),(10.43) and (10.44) enable to end the proof :

$$T_{3,n}^* + T_{4,n}^* + 2T_{5,n}^* + 2T_{6,n}^* = o_p\left(\sqrt{Var(T_{2,n})} \right).$$

□

Proof of Lemma 10.3.7 In order to get asymptotic inequalities for variables $T_{j,n}^*$ we have chosen to study asymptotic properties of their first and second order conditional moments. In fact, the aim is to prove that the conditional mean of $T_{3,n}^* + T_{4,n}^*$ converges to infinity quicker than the square root of the conditional variance of any $T_{j,n}^*$, $j = 3, \ldots, 6$.

The first step consists in getting a lower bound for the mean of $T_{3,n}^* + T_{4,n}^*$. Firstly, the second part of assumption (10.14) implies that there exists a sequence of random operators $r_{\overline{L},n}^*$ in $\overline{Hol\,(C,\beta)}$ (the projection of r_0^*) such that

$$\eta_n^{-2}\mathbb{E}\left[\left\|r_0^* - r_{\overline{L},n}^*\right\|^2_{\mathbb{L}^2_{dP_X}(W_\alpha)}\right] \to 0.$$

By density arguments, there exists a sequence of random operators $\left(r_{L,n}^*\right)_n$ in $Hol\,(C,\beta)$ such that

$$\eta_n^{-2}\mathbb{E}\left[\left\|r_{L,n}^* - r_{\overline{L},n}^*\right\|^2_{\mathbb{L}^2_{dP_X}(W_\alpha)}\right] \to 0.$$

Consequently there exists a sequence of random operators $\left(r_{L,n}^*\right)_n$ in $Hol\,(C,\beta)$ such that

$$\eta_n^{-2}\mathbb{E}\left[\left\|r_0^* - r_{L,n}^*\right\|^2_{\mathbb{L}^2_{dP_X}(W_\alpha)}\right] \to 0. \tag{10.45}$$

Secondly, we use the previous statement to obtain a lower bound for the conditional mean :

$$\begin{aligned}
&\mathbb{E}_*\left[T_{3,n}^* + T_{4,n}^*\right] \\
&= \mathbb{E}_*\left[\int_E \left(\sum_{i=1}^n \left(r\left(X_i\right) - r_0^*\left(X_i\right)\right) K\left(\frac{d\left(X_i,x\right)}{h_n}\right)\right)^2 w\left(x\right) dP_X\left(x\right)\right] \\
&= \mathbb{E}_*\Big[\int_E \Big(\sum_{i=1}^n \left(r\left(X_i\right) - r_{L,n}^*\left(X_i\right)\right) K\left(\frac{d\left(X_i,x\right)}{h_n}\right) \\
&\qquad + \sum_{i=1}^n \left(r_{L,n}^*\left(X_i\right) - r_0^*\left(X_i\right)\right) K\left(\frac{d\left(X_i,x\right)}{h_n}\right)\Big)^2 w\left(x\right) dP_X\left(x\right)\Big] \\
&\geq \Bigg(\sqrt{\mathbb{E}_*\left[\int_E \left(\sum_{i=1}^n \left(r\left(X_i\right) - r_{L,n}^*\left(X_i\right)\right) K\left(\frac{d\left(X_i,x\right)}{h_n}\right)\right)^2 w\left(x\right) dP_X\left(x\right)\right]} \\
&\quad - \sqrt{\mathbb{E}_*\left[\int_E \left(\sum_{i=1}^n \left(r_{L,n}^*\left(X_i\right) - r_0^*\left(X_i\right)\right) K\left(\frac{d\left(X_i,x\right)}{h_n}\right)\right)^2 w\left(x\right) dP_X\left(x\right)\right]}\Bigg)^2 \\
&= \left(\sqrt{A_n} - \sqrt{B_n}\right)^2
\end{aligned} \tag{10.46}$$

We will show later that $B_n = o_p(A_n)$. Now we use the Hölder property of operators r and $r^*_{L,n}$ (see (10.6) and (10.45)) to get the following lower bound for A_n :

$$\begin{aligned}
A_n &= \mathbb{E}_*\Big[\int_E \Big(\sum_{i=1}^n \left(r - r^*_{L,n}\right)(x)\, K\left(\frac{d(X_i,x)}{h_n}\right) \\
&\quad + \sum_{i=1}^n \Big(\left(r - r^*_{L,n}\right)(X_i) - \left(r - r^*_{L,n}\right)(x)\Big) K\left(\frac{d(X_i,x)}{h_n}\right)\Big)^2 w(x)\, dP_X(x)\Big] \\
&= \mathbb{E}_*\left[\int_E \left(\sum_{i=1}^n \left(r - r^*_{L,n}\right)(x)\, K\left(\frac{d(X_i,x)}{h_n}\right)\right)^2 w(x)\, dP_X(x)\right] \\
&+ \mathbb{E}_*\left[\int_E \left(\sum_{i=1}^n \Big(\left(r - r^*_{L,n}\right)(X_i) - \left(r - r^*_{L,n}\right)(x)\Big) K\left(\frac{d(X_i,x)}{h_n}\right)\right)^2 w(x)\, dP_X(x)\right] \\
&+ 2\mathbb{E}_*\Big[\int_E \left(\sum_{i=1}^n \Big(\left(r - r^*_{L,n}\right)(X_i) - \left(r - r^*_{L,n}\right)(x)\Big) K\left(\frac{d(X_i,x)}{h_n}\right)\right) \\
&\quad \times \left(\sum_{i=1}^n \left(r - r^*_{L,n}\right)(x)\, K\left(\frac{d(X_i,x)}{h_n}\right)\right) w(x)\, dP_X(x)\Big] \\
&= A_{1,n} + A_{2,n} + 2A_{3,n}.
\end{aligned} \tag{10.47}$$

Now it is possible to give an almost sure lower bound for $A_{1,n}$:

$$\begin{aligned}
&A_{1,n} \\
&= \mathbb{E}_*\left[\int_E \sum_{i=1}^n K^2\left(\frac{d(X_i,x)}{h_n}\right)\left(r - r^*_{L,n}\right)^2(x)\, w(x)\, dP_X(x)\right] \\
&+ \mathbb{E}_*\left[\int_E \sum_{1\le i\ne j\le n} K\left(\frac{d(X_i,x)}{h_n}\right) K\left(\frac{d(X_j,x)}{h_n}\right)\left(r - r^*_{L,n}\right)^2(x)\, w(x)\, dP_X(x)\right] \\
&\ge \int_E \sum_{1\le i\ne j\le n} \mathbb{E}_*\left[K\left(\frac{d(X_i,x)}{h_n}\right)\right]\mathbb{E}_*\left[K\left(\frac{d(X_j,x)}{h_n}\right)\right]\left(r - r^*_{L,n}\right)^2(x)\, w(x)\, dP_X(x) \\
&\ge Cn^2\Phi^2(h_n)\int_E \left(r - r^*_{L,n}\right)^2(x)\, w(x)\, dP_X(x)
\end{aligned} \tag{10.48}$$

On one hand, with similar arguments as those used to get (10.45), assumption (10.14) implies that there exist a sequence of random operators $r^*_{F,n}$ in $\mathcal{F}$ such that

$$\eta_n^{-2}\mathbb{E}\left[\left\|r^*_0 - r^*_{F,n}\right\|^2_{\mathbb{L}^2_{dP_X}(W_\alpha)}\right] \to 0. \tag{10.49}$$

Consequently, (10.45) together with (10.49) give by triangular inequality :

$$\eta_n^{-2}\mathbb{E}\left[\left\|r^*_{L,n} - r^*_{F,n}\right\|^2_{\mathbb{L}^2_{dP_X}(W_\alpha)}\right] \to 0. \tag{10.50}$$

A direct consequence of (10.50) is that by Markov's inequality together with assumption (10.4) one gets :

$$\begin{aligned}\int_E \left(r^*_{F,n} - r^*_{L,n}\right)^2 (x)\, w(x)\, dP_X(x) &= O_p\left(\mathbb{E}\left[\int_E \left(r^*_{F,n} - r^*_{L,n}\right)^2 (x)\, w(x)\, dP_X(x)\right]\right) \\ &= O_p\left(\mathbb{E}\left[\int_E \left(r^*_{F,n} - r^*_{L,n}\right)^2 (x)\, 1_{W_\alpha}(x)\, dP_X(x)\right]\right) \\ &= O_p\left(o\left(\eta_n^2\right)\right) \\ &= o_p\left(\eta_n^2\right) \qquad (10.51)\end{aligned}$$

On the other hand, if r_0 denotes the projection of r on $\overline{\mathcal{F}}^w$, since $r^*_{F,n}$ belongs to $\mathcal{F}$ one gets :

$$\begin{aligned}\left\|r - r^*_{F,n}\right\|^2_{\mathbb{L}^2_{(wdP_X)}(E)} &\geq \|r - r_0\|^2_{\mathbb{L}^2_{(wdP_X)}(E)} + \left\|r^*_{F,n} - r_0\right\|^2_{\mathbb{L}^2_{(wdP_X)}(E)} \\ &\geq \|r - r_0\|^2_{\mathbb{L}^2_{(wdP_X)}(E)} \\ &\geq \eta_n^2 \qquad (10.52)\end{aligned}$$

Finally, with (10.51) and (10.52), the statement (10.48) becomes :

$$\begin{aligned}A_{1,n} &\geq Cn^2\Phi^2(h_n)\Big(\sqrt{\int_E \left(r - r^*_{F,n}\right)^2 (x)\, w(x)\, dP_X(x)} \\ &\qquad - \sqrt{\int_E \left(r^*_{F,n} - r^*_{L,n}\right)^2 (x)\, w(x)\, dP_X(x)}\Big)^2 \\ &\geq Cn^2\Phi^2(h_n)\left(\sqrt{\eta_n^2} - \sqrt{o_p\left(\eta_n^2\right)}\right)^2 \\ &\geq Cn^2\Phi^2(h_n)\,\eta_n^2\left(1 + o_p(1)\right)^2. \qquad (10.53)\end{aligned}$$

Now we want to show that the term $A_{2,n}$ is negligible with respect to $A_{1,n}$. Because $r - r^*_{L,n}$ is Hölderian of order β and $n\Phi(h_n) \to +\infty$, it comes :

$$\begin{aligned}A_{2,n} &\leq Ch_n^{2\beta}\mathbb{E}_*\left[\int_E \left(\sum_{i=1}^n K\left(\frac{d(X_i,x)}{h_n}\right)\right)^2 w(x)\, dP_X(x)\right] \\ &\leq Ch_n^{2\beta}\sum_{i=1}^n \int_E \mathbb{E}_*\left[K^2\left(\frac{d(X_i,x)}{h_n}\right)\right] w(x)\, dP_X(x) \\ &+ Ch_n^{2\beta}\sum_{1\leq i\neq j\leq n} \int_E \mathbb{E}_*\left[K\left(\frac{d(X_i,x)}{h_n}\right)\right]\mathbb{E}_*\left[K\left(\frac{d(X_j,x)}{h_n}\right)\right] w(x)\, dP_X(x) \\ (10.54) &\leq Ch_n^{2\beta}n^2\Phi^2(h_n)\end{aligned}$$

Now (10.53) together with (10.54) and assumption (10.8) give :

$$\frac{A_{2,n}}{A_{1,n}} \leq C\frac{h_n^{2\beta}}{\eta_n^2}\left(1 + o_p(1)\right) \leq C\frac{1}{v_n^2}\left(1 + o_p(1)\right) \xrightarrow{P} 0. \qquad (10.55)$$

Furthermore, a standard Cauchy-Schwartz inequality together with (10.55) imply that $A_{3,n}$ is also negligible with respect to $A_{1,n}$:

$$|A_{3,n}| \leq \sqrt{A_{1,n}}\sqrt{A_{2,n}} = o_p\left(A_{1,n}\right). \tag{10.56}$$

Consequently, with (10.53), (10.55), (10.56) and (10.47) one gets the following lower bound for A_n :

$$A_n \geq Cn^2\Phi^2\left(h_n\right)\eta_n^2 \max\left(\left(1+o_p\left(1\right)\right),0\right). \tag{10.57}$$

At this step of the proof we want to give an upper bound of B_n in order to show that it is negligible with respect to A_n. We will use the statement (10.45) to bound B_n.

$$\begin{aligned}
&B_n \\
&= O_p\left(\mathbb{E}\left[\mathbb{E}_*\left[\int_E\left(\sum_{i=1}^n\left(r_{L,n}^*\left(X_i\right)-r_0^*\left(X_i\right)\right)K\left(\frac{d\left(X_i,x\right)}{h_n}\right)\right)^2 w\left(x\right)dP_X\left(x\right)\right]\right]\right) \\
&= O_p\big(\mathbb{E}\big[\mathbb{E}_*\left[\int_E\sum_{i=1}^n\left(r_{L,n}^*\left(X_i\right)-r_0^*\left(X_i\right)\right)^2K^2\left(\frac{d\left(X_i,x\right)}{h_n}\right)w\left(x\right)dP_X\left(x\right)\right] \\
&+\mathbb{E}_*\big[\int_E\sum_{1\leq i\neq j\leq n}\left(r_{L,n}^*\left(X_i\right)-r_0^*\left(X_i\right)\right)\left(r_{L,n}^*\left(X_j\right)-r_0^*\left(X_j\right)\right)K\left(\frac{d\left(X_i,x\right)}{h_n}\right) \\
&\qquad K\left(\frac{d\left(X_j,x\right)}{h_n}\right)w\left(x\right)dP_X\left(x\right)\big]\big]\big) \\
&= O_p\big(\mathbb{E}\big[\sum_{i=1}^n\mathbb{E}_*\left[\left(r_{L,n}^*\left(X_i\right)-r_0^*\left(X_i\right)\right)^2\int_E K^2\left(\frac{d\left(X_i,x\right)}{h_n}\right)w\left(x\right)dP_X\left(x\right)\right] \\
&+\sum_{1\leq i\neq j\leq n}\mathbb{E}_*\left[\int_E\left(r_{L,n}^*\left(X_i\right)-r_0^*\left(X_i\right)\right)^2K\left(\frac{d\left(X_i,x\right)}{h_n}\right)K\left(\frac{d\left(X_j,x\right)}{h_n}\right)w\left(x\right)dP_X\left(x\right)\right]\big]\big) \\
&= O_p\left(O\left(\left(n\Phi\left(h_n\right)+n^2\Phi^2\left(h_n\right)\right)\mathbb{E}\left[\left\|r_0^*-r_{L,n}^*\right\|^2_{\mathbb{L}^2_{dP_X}(W_\alpha)}\right]\right)\right) \\
&= O_p\left(n^2\Phi^2\left(h_n\right)o\left(\eta_n^2\right)\right) \\
&= o_p\left(n^2\Phi^2\left(h_n\right)\eta_n^2\right) \\
&= o_p\left(A_n\right)
\end{aligned} \tag{10.58}$$

The third equality is obtained by Cauchy-Schwartz inequality whereas the last lines come directly from (10.45) and (10.57). Finally, (10.46), (10.57) and (10.58) are enough to conclude.

$$\mathbb{E}_*\left[T_{3,n}^*+T_{4,n}^*\right] \geq Cn^2\Phi^2\left(h_n\right)\eta_n^2\max\left(\left(1+o_p\left(1\right)\right),0\right). \tag{10.59}$$

The second step of the proof consists in finding an upper bound for the conditional variances of $T_{3,n}^*$ and $T_{4,n}^*$. We will show that the square roots of these variances are negligible with respect to the conditional mean $\mathbb{E}_*\left[T_{3,n}^*+T_{4,n}^*\right]$. We note Var_* when the variance is taken conditionally to $\left(\left(X_i^*,Y_i^*\right)_{1\leq i\leq u_n}\right.$. We start with the conditional

variance of $T_{3,n}^*$.

$$\begin{aligned} Var_*\left(T_{3,n}^*\right) &= nVar_*\left(\left(r-r_0^*\right)^2(X)\int_E K^2\left(\frac{d(X,x)}{h_n}\right)w(x)\,dP_X(x)\right) \\ &\leq n\mathbb{E}_*\left[\left(r-r_0^*\right)^4(X)\left(\int_E K^2\left(\frac{d(X,x)}{h_n}\right)w(x)\,dP_X(x)\right)^2\right] \\ &\leq n\Phi^2(h_n)\,\mathbb{E}_*\left[\left(r-r_0^*\right)^4(X)\,1_{W_{h_n}}(X)\right] \\ &\leq Cn\Phi^2(h_n)\,O_p\left(\mathbb{E}\left[\mathbb{E}_*\left[\left(r-r_0^*\right)^4(X)\,1_{W_\alpha}(X)\right]\right]\right) \\ &\leq Cn\Phi^2(h_n)\,O_p(1). \qquad (10.60) \end{aligned}$$

The last line comes from assumptions (10.6) and (10.15) :

$$\begin{aligned} \mathbb{E}\left[\left(r-r_0^*\right)^4(X)\,1_{W_\alpha}(X)\right] &= O\left(\mathbb{E}\left[(r)^4(X)\,1_{W_\alpha}(X)\right]\right)+O\left(\mathbb{E}\left[\left(r_0^*\right)^4(X)\,1_{W_\alpha}(X)\right]\right) \\ &< +\infty. \end{aligned}$$

It comes directly from (10.59) and (10.60) that the square root of the conditional variance of $T_{3,n}^*$ is negligible with regard to the conditional mean of $T_{3,n}^*+T_{4,n}^*$:

$$\frac{\sqrt{Var_*\left(T_{3,n}^*\right)}}{\mathbb{E}_*\left[T_{3,n}^*+T_{4,n}^*\right]} \leq O_p\left(\frac{\sqrt{n}\Phi(h_n)}{n^2\Phi^2(h_n)\,\eta_n^2}\right) \leq O_p\left(\frac{1}{\sqrt{n\Phi(h_n)}v_n^2}\right)=o_p(1) \qquad (10.61)$$

Now the aim is to bound the conditional variance of $T_{4,n}^*$. We will use a method similar to the one used to bound the variance of $T_{4,n}$ in the proof of Lemma 10.3.4.

$$\begin{aligned} &Var_*\left(T_{4,n}^*\right) \\ &= \mathbb{E}_*\Big[\Big(\int \sum_{1\leq i\neq j\leq n}\left(r(X_i)-r_0^*(X_i)\right)\left(r(X_j)-r_0^*(X_j)\right)K\left(\frac{d(X_i,x)}{h_n}\right) \\ &\qquad K\left(\frac{d(X_j,x)}{h_n}\right)w(x)\,dP_X(x)\Big)^2\Big] \\ &\quad -\Big(\mathbb{E}_*\Big[\int \sum_{1\leq i\neq j\leq n}\left(r(X_i)-r_0^*(X_i)\right)\left(r(X_j)-r_0^*(X_j)\right)K\left(\frac{d(X_i,x)}{h_n}\right) \\ &\qquad K\left(\frac{d(X_j,x)}{h_n}\right)w(x)\,dP_X(x)\Big]\Big)^2 \\ &= n(n-1)(n-2)(n-4)\left(\int\left(\mathbb{E}_*\left[\left(r-r_0^*\right)(X)\,K\left(\frac{d(X,x)}{h_n}\right)\right]\right)^2 w(x)\,dP_X(x)\right)^2 \\ &\quad +4n(n-1)(n-2)\,C_n \\ &\quad +2n(n-1)\,D_n \\ &\quad -\left(n(n-1)\int\left(\mathbb{E}_*\left[(r(X)-r_0^*(X))\,K\left(\frac{d(X,x)}{h_n}\right)\right]\right)^2 w(x)\,dP_X(x)\right)^2 \\ &\leq 4n(n-1)(n-2)\,C_n+2n(n-1)\,D_n, \end{aligned}$$

with

$$C_n = \int_{E\times E} \mathbb{E}_* \left[\left(r\left(X\right) - r_0^*\left(X\right)\right) K\left(\frac{d\left(X,x\right)}{h_n}\right)\right] \mathbb{E}_* \left[\left(r\left(X\right) - r_0^*\left(X\right)\right) K\left(\frac{d\left(X,y\right)}{h_n}\right)\right]$$
$$\mathbb{E}_* \left[\left(r\left(X\right) - r_0^*\left(X\right)\right)^2 K\left(\frac{d\left(X,x\right)}{h_n}\right) K\left(\frac{d\left(X,y\right)}{h_n}\right)\right] w\left(x\right) dP_X\left(x\right) w\left(y\right) dP_X\left(y\right),$$

and

$$D_n = \int_{E\times E} \left(\mathbb{E}_* \left[\left(r\left(X\right) - r_0^*\left(X\right)\right)^2 K\left(\frac{d\left(X,x\right)}{h_n}\right) K\left(\frac{d\left(X,y\right)}{h_n}\right)\right]\right)^2$$
$$w\left(x\right) dP_X\left(x\right) w\left(y\right) dP_X\left(y\right).$$

To bound C_n we use Cauchy-Schwartz inequality twice, Fubini's theorem, then assumptions (10.14) and (10.15) to get :

$$\begin{aligned}
& |C_n| \\
\leq\ & \int_{E\times E} \mathbb{E}_* \left[K\left(\frac{d\left(X,x\right)}{h_n}\right)\right] \mathbb{E}_* \left[\left(r\left(X\right) - r_0^*\left(X\right)\right)^2 K\left(\frac{d\left(X,y\right)}{h_n}\right)\right] \\
& \mathbb{E}_* \left[\left(r\left(X\right) - r_0^*\left(X\right)\right)^2 K\left(\frac{d\left(X,x\right)}{h_n}\right) K\left(\frac{d\left(X,y\right)}{h_n}\right)\right] w\left(x\right) dP_X\left(x\right) w\left(y\right) dP_X\left(y\right) \\
\leq\ & \int_{E\times E} \mathbb{E}_* \left[K\left(\frac{d\left(X,x\right)}{h_n}\right)\right] \sqrt{\mathbb{E}_* \left[\left(r\left(X\right) - r_0^*\left(X\right)\right)^4 1_{W_\alpha}\left(X\right)\right]} \sqrt{\mathbb{E}\left[K^2\left(\frac{d\left(X,y\right)}{h_n}\right)\right]} \\
& \mathbb{E}_* \left[\left(r\left(X\right) - r_0^*\left(X\right)\right)^2 K\left(\frac{d\left(X,x\right)}{h_n}\right) K\left(\frac{d\left(X,y\right)}{h_n}\right)\right] w\left(x\right) dP_X\left(x\right) w\left(y\right) dP_X\left(y\right) \\
\leq\ & CO_p\left(1\right) \Phi^{\frac{3}{2}}\left(h_n\right) \mathbb{E}_* \left[\left(r\left(X\right) - r_0^*\left(X\right)\right)^2 \left(\int_E K\left(\frac{d\left(X,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right)\right)^2\right] \\
\leq\ & O_p\left(1\right) \Phi^{\frac{5}{2}}\left(h_n\right) \mathbb{E}_* \left[\left(r\left(X\right) - r_0^*\left(X\right)\right)^2 \int_E K\left(\frac{d\left(X,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right)\right] \\
\leq\ & O_p\left(1\right) \Phi^{\frac{5}{2}}\left(h_n\right) \left(\mathbb{E}_* \left[\left(r - r_{L,n}^*\right)^2\left(X\right) \int_E K\left(\frac{d\left(X,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right)\right]\right. \\
& \left. + \mathbb{E}_* \left[\left(r_{L,n}^* - r_0^*\right)^2\left(X\right) \int_E K\left(\frac{d\left(X,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right)\right]\right) \\
\leq\ & O_p\left(1\right) \Phi^{\frac{5}{2}}\left(h_n\right) \left(\mathbb{E}_* \left[\int_E \left(r - r_{L,n}^*\right)^2\left(x\right) K\left(\frac{d\left(X,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right)\right]\right. \\
& + \mathbb{E}_* \left[\left(\left(r_{L,n}^* - r\right)\left(X\right) - \left(r_{L,n}^* - r\right)\left(x\right)\right)^2 \int_E K\left(\frac{d\left(X,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right)\right] \\
& \left. + \mathbb{E}_* \left[\left(r_{L,n}^* - r_0^*\right)^2\left(X\right) \int_E K\left(\frac{d\left(X,x\right)}{h_n}\right) w\left(x\right) dP_X\left(x\right)\right]\right).
\end{aligned}$$

Hölder assumptions on r and $r^*_{L,n}$ together with the previous statement (10.45) give :

$$\begin{aligned}|C_n| &\leq |O_p(1)|\,\Phi^{\frac{7}{2}}(h_n)\Big(\int_E \left(r-r^*_{L,n}\right)^2(x)\,w(x)\,dP_X(x)+h_n^{2\beta}\\ &\qquad +\mathbb{E}_*\left[\left(r^*_{L,n}-r^*_0\right)^2(X)\,1_{W_\alpha}(X)\right]\Big)\\ &\leq |O_p(1)|\,\Phi^{\frac{7}{2}}(h_n)\left(\frac{A_{1,n}}{n^2\Phi^2(h_n)}+h_n^{2\beta}+O_p\left(\mathbb{E}\left[\mathbb{E}_*\left[\left(r^*_{L,n}-r^*_0\right)^2(X)\,1_{W_\alpha}(X)\right]\right]\right)\right)\\ &\leq |O_p(1)|\,\Phi^{\frac{7}{2}}(h_n)\left(\frac{\mathbb{E}_*\left[T^*_{3,n}+T^*_{4,n}\right](1+o_p(1))}{n^2\Phi^2(h_n)}+o_p\left(\eta_n^2\right)\right)\end{aligned} \tag{10.62}$$

Now, D_n may be bounded in the same way :

$$\begin{aligned}|D_n| &\leq \int_{E\times E}\mathbb{E}_*\left[\left(r(X)-r^*_0(X)\right)^4 K\left(\frac{d(X,x)}{h_n}\right)K\left(\frac{d(X,y)}{h_n}\right)\right]\\ &\quad\times\mathbb{E}_*\left[K\left(\frac{d(X,x)}{h_n}\right)K\left(\frac{d(X,y)}{h_n}\right)\right]w(x)\,dP_X(x)\,w(y)\,dP_X(y)\\ &\leq C\Phi(h_n)\,\mathbb{E}_*\left[\left(r(X)-r^*_0(X)\right)^4\left(\int_E K\left(\frac{d(X,x)}{h_n}\right)w(x)\,dP_X(x)\right)^2\right]\\ &\leq C\Phi^3(h_n)\,\mathbb{E}_*\left[\left(r(X)-r^*_0(X)\right)^4 1_{W_\alpha}(X)\right]\\ &\leq C\Phi^3(h_n)\,O_p\left(\mathbb{E}\left[\mathbb{E}_*\left[\left(r(X)-r^*_0(X)\right)^4 1_{W_\alpha}(X)\right]\right]\right)\\ &= O_p\left(\Phi^3(h_n)\right)\end{aligned} \tag{10.63}$$

Finally, it comes from (10.62) and (10.63) the following bound for the conditional variance of $T^*_{4,n}$:

$$\begin{aligned}&Var_*\left(T^*_{4,n}\right)\\ &= O_p\left(n\Phi^{\frac{3}{2}}(h_n)\,\mathbb{E}_*\left[T^*_{3,n}+T^*_{4,n}\right]\right)+o_p\left(n^3\eta_n^2\Phi^{\frac{7}{2}}(h_n)\right)+O_p\left(n^2\Phi^3(h_n)\right).\end{aligned} \tag{10.64}$$

From (10.59) and (10.64) one gets that the square root of the conditional variance of $T^*_{4,n}$ is negligible with respect to the conditional mean of $T^*_{3,n}+T^*_{4,n}$:

$$\begin{aligned}\frac{\sqrt{Var_*\left(T^*_{4,n}\right)}}{\mathbb{E}_*\left[T^*_{3,n}+T^*_{4,n}\right]} &= O_p\left(\frac{\sqrt{n}\Phi^{\frac{3}{4}}(h_n)}{n\Phi(h_n)\,\eta_n}\right)+o_p\left(\frac{n^{\frac{3}{2}}\eta_n\Phi^{\frac{7}{4}}(h_n)}{n^2\Phi^2(h_n)\,\eta_n^2}\right)+O_p\left(\frac{n\Phi^{\frac{3}{2}}(h_n)}{n^2\Phi^2(h_n)\,\eta_n^2}\right)\\ &= O_p\left(\frac{1}{n^{\frac{1}{2}}\Phi^{\frac{1}{4}}(h_n)\,\eta_n}\right)+o_p\left(\frac{1}{n^{\frac{1}{2}}\Phi^{\frac{1}{4}}(h_n)\,\eta_n}\right)+O_p\left(\frac{1}{n\Phi^{\frac{1}{2}}(h_n)\,\eta_n^2}\right)\\ &= O_p\left(\frac{1}{v_n}\right)+o_p\left(\frac{1}{v_n}\right)+O_p\left(\frac{1}{v_n^2}\right)\\ &= o_p(1).\end{aligned} \tag{10.65}$$

The following lower bound for $T_{3,n}^* + T_{4,n}^*$ comes directly from (10.59), (10.61) and (10.65).

$$\begin{aligned} T_{3,n}^* + T_{4,n}^* &= \mathbb{E}_* \left[T_{3,n}^* + T_{4,n}^* \right] + O_p \left(\sqrt{Var_* \left(T_{3,n}^* \right)} \right) + O_p \left(\sqrt{Var \left(T_{4,n}^* \right)} \right) \\ &\geq Cn^2 \Phi^2 \left(h_n \right) \eta_n^2 \left(1 + o_p \left(1 \right) \right). \end{aligned} \tag{10.66}$$

At this step of the proof we have not yet studied the asymptotic behaviour of $T_{5,n}^*$ and $T_{6,n}^*$. This study is easier because these variables are centered. Firstly, the variance of $T_{5,n}^*$ may be bounded in the same way as the variance of $T_{3,n}^*$:

$$\begin{aligned} Var \left(T_{5,n}^* \right) &= nVar \left(\left(r - r_0^* \right) (X) \epsilon \int_E K^2 \left(\frac{d\left(X, x \right)}{h_n} \right) w\left(x \right) dP_X \left(x \right) \right) \\ &= n\mathbb{E} \left[\left(r - r_0^* \right)^2 (X) \epsilon^2 \left(\int_E K^2 \left(\frac{d\left(X, x \right)}{h_n} \right) w\left(x \right) dP_X \left(x \right) \right)^2 \right] \\ &\leq Cn\Phi^2 \left(h_n \right) \sqrt{\mathbb{E} \left[\left(r - r_0^* \right)^4 (X) 1_{W_{h_n}} (X) \right]} \\ &\leq Cn\Phi^2 \left(h_n \right) \end{aligned} \tag{10.67}$$

Consequently, one gets, with the same arguments as to get (10.60), that $T_{5,n}^*$ is negligible with respect to the conditional mean of $T_{3,n}^* + T_{4,n}^*$:

$$T_{5,n}^* = O_p \left(\sqrt{Var \left(T_{5,n}^* \right)} \right) = o_p \left(\mathbb{E}_* \left[T_{3,n}^* + T_{4,n}^* \right] \right). \tag{10.68}$$

To end the proof we now have to bound the variance of $T_{6,n}^*$. Because ϵ is centered, $T_{6,n}^*$ is centered and its conditional variance is its conditional second order moment. Moreover, this conditional second order moment is fairly simple since many terms are centered. In fact one gets :

$$\begin{aligned} & Var_* \left(T_{6,n}^* \right) \\ = \; & \mathbb{E}_* \left[\left(T_{6,n}^* \right)^2 \right] \\ \leq \; & Cn^3 \int_{E \times E} \mathbb{E}_* \left[\left| r - r_0^* \right| (X) K \left(\frac{d\left(X, x \right)}{h_n} \right) \right] \mathbb{E}_* \left[\left| r - r_0^* \right| (X) K \left(\frac{d\left(X, y \right)}{h_n} \right) \right] \\ & \qquad \mathbb{E}_* \left[\epsilon^2 K \left(\frac{d\left(X, x \right)}{h_n} \right) K \left(\frac{d\left(X, y \right)}{h_n} \right) \right] w\left(x \right) w\left(y \right) dP_X \left(x \right) dP_X \left(y \right) \\ & + Cn^2 \int_{E \times E} \mathbb{E}_* \left[\left(r - r_0^* \right)^2 (X) K \left(\frac{d\left(X, x \right)}{h_n} \right) K \left(\frac{d\left(X, y \right)}{h_n} \right) \right] \\ & \qquad \mathbb{E}_* \left[\epsilon^2 K \left(\frac{d\left(X, x \right)}{h_n} \right) K \left(\frac{d\left(X, y \right)}{h_n} \right) \right] w\left(x \right) w\left(y \right) dP_X \left(x \right) dP_X \left(y \right). \end{aligned} \tag{10.69}$$

Then we obtain with Cauchy-Schwartz inequality :

$$
\begin{aligned}
&Var_*\left(T_{6,n}^*\right)\\
&\leq Cn^3\int_{E\times E}\mathbb{E}_*\left[K\left(\frac{d\left(X,x\right)}{h_n}\right)\right]\mathbb{E}_*\left[\left(r-r_0^*\right)^2\left(X\right)K\left(\frac{d\left(X,y\right)}{h_n}\right)\right]\\
&\quad\times\mathbb{E}_*\left[\mathbb{E}_*\left[\epsilon^2|\ X\right]K\left(\frac{d\left(X,x\right)}{h_n}\right)K\left(\frac{d\left(X,y\right)}{h_n}\right)\right]w\left(x\right)w\left(y\right)dP_X\left(x\right)dP_X\left(y\right)\\
&+Cn^2\int_{E\times E}\mathbb{E}_*\left[\left(r-r_0^*\right)^2\left(X\right)K\left(\frac{d\left(X,x\right)}{h_n}\right)K\left(\frac{d\left(X,y\right)}{h_n}\right)\right]\\
&\quad\times\mathbb{E}_*\left[\mathbb{E}_*\left[\epsilon^2|\ X\right]K\left(\frac{d\left(X,x\right)}{h_n}\right)K\left(\frac{d\left(X,y\right)}{h_n}\right)\right]w\left(x\right)w\left(y\right)dP_X\left(x\right)dP_X\left(y\right)\\
&\leq Cn^3\Phi\left(h_n\right)\int_E\mathbb{E}_*\left[\left(r-r_0^*\right)^2\left(X\right)K\left(\frac{d\left(X,y\right)}{h_n}\right)\right]\\
&\quad\times\mathbb{E}_*\left[\int_E K\left(\frac{d\left(X,x\right)}{h_n}\right)w\left(x\right)dP_X\left(x\right)K\left(\frac{d\left(X,y\right)}{h_n}\right)\right]w\left(y\right)dP_X\left(y\right)\\
&+Cn^2\Phi\left(h_n\right)\mathbb{E}_*\left[\left(r-r_0^*\right)^2\left(X\right)1_{W_{h_n}}\left(\int_E K\left(\frac{d\left(X,x\right)}{h_n}\right)w\left(x\right)dP_X\left(x\right)\right)^2\right]\\
&\leq Cn^3\Phi^3\left(h_n\right)\int_E\mathbb{E}_*\left[\left(r-r_0^*\right)^2\left(X\right)K\left(\frac{d\left(X,y\right)}{h_n}\right)\right]w\left(y\right)dP_X\left(y\right)\\
&+Cn^2\Phi^3\left(h_n\right)\mathbb{E}_*\left[\left(r-r_0^*\right)^2\left(X\right)1_{W_\alpha}\left(X\right)\right]\\
&\leq Cn^3\Phi^4\left(h_n\right)\left(\frac{\mathbb{E}_*\left[T_{3,n}^*+T_{4,n}^*\right]\left(1+o_p\left(1\right)\right)}{n^2\Phi^2\left(h_n\right)}+o_p\left(\eta_n^2\right)\right)+n^2\Phi^3\left(h_n\right)O_p\left(1\right).
\end{aligned}
\tag{10.70}
$$

The last line comes from bounds already used to bound the conditional variance of $T_{4,n}^*$ (notably to bound C_n). As a consequence $T_{6,n}^*$ is negligible with respect to the conditional mean of $T_{3,n}^*+T_{4,n}^*$:

$$
\begin{aligned}
&T_{6,n}^*\\
&=O_p\left(\sqrt{Var_*\left(T_{6,n}^*\right)}\right)\\
&=O_p\left(n^{\frac{1}{2}}\Phi\left(h_n\right)\sqrt{\mathbb{E}_*\left[T_{3,n}^*+T_{4,n}^*\right]}\right)+o_p\left(n^{\frac{3}{2}}\Phi^2\left(h_n\right)\eta_n\right)+O_p\left(n\Phi^{\frac{3}{2}}\left(h_n\right)\right)\\
&=\mathbb{E}_*\left[T_{3,n}^*+T_{4,n}^*\right]\left(O_p\left(\frac{n^{\frac{1}{2}}\Phi\left(h_n\right)}{n\Phi\left(h_n\right)\eta_n}\right)+o_p\left(\frac{n^{\frac{3}{2}}\Phi^2\left(h_n\right)\eta_n}{n^2\Phi^2\left(h_n\right)\eta_n^2}\right)+O_p\left(\frac{n\Phi^{\frac{3}{2}}\left(h_n\right)}{n^2\Phi^2\left(h_n\right)\eta_n^2}\right)\right)\\
&=\mathbb{E}_*\left[T_{3,n}^*+T_{4,n}^*\right]\left(O_p\left(\frac{1}{n^{\frac{1}{2}}\eta_n}\right)+o_p\left(\frac{1}{n^{\frac{1}{2}}\eta_n}\right)+O_p\left(\frac{1}{n\Phi^{\frac{1}{2}}\left(h_n\right)\eta_n^2}\right)\right)\\
&=o_p\left(\mathbb{E}_*\left[T_{3,n}^*+T_{4,n}^*\right]\right).
\end{aligned}
\tag{10.71}
$$

The end of the proof is a direct application of (10.59), (10.61), (10.65), (10.66), (10.68) and (10.71).

□

Proof of Lemma 10.4.1 In this proof we note $\mu = \inf_{x\in S} c(x)$. The uniform convergence of $F_x(h_n)$ on S implies that :

$$\exists N_0,\ \forall n \geq N_0,\ \sup_{x\in S}\left|\frac{F_x(h_n)}{h_n^\delta} - c(x)\right| \leq \frac{\mu}{2}.$$

Hence, for $n \geq N_0$ and for all $x \in S$ one gets :

$$h_n^\delta\left(c(x) - \frac{\mu}{2}\right) \leq F_x(h_n) \leq h_n^\delta\left(c(x) + \frac{\mu}{2}\right).$$

Moreover, since $d(W, S^c) > 0$, if one takes $\gamma = \frac{d(W,S^c)}{2}$, then $W_\gamma \subset S$. Finally, assumption (10.17) provides upper and lower bounds for c on S. Consequently, since $W_\gamma \subset S$, there exists a positive constant M such that :

$$\forall x \in W_\gamma,\ \mu \leq c(x) \leq M.$$

Therefore, for n large enough ($n \geq N_0$), one gets :

$$\forall x \in W_\gamma,\ \frac{\mu}{2}h_n^\delta \leq F_x(h_n) \leq h_n^\delta\left(M + \frac{\mu}{2}\right).$$

Because the previous inequalities are true for n large enough, there exist two positive constants C_1 and C_2 such that :

$$\forall x \in W_\gamma,\ C_1 h_n^\delta \leq F_x(h_n) \leq C_2 h_n^\delta.$$

□

Proof of Lemma 10.4.2 We have to show that assumption (10.10) holds with $l = 0$. Firstly, our aim is to obtain the following inequality :

$$\Omega_4(h_n) \geq C\Phi^3\left(\frac{h_n}{2}\right) \tag{10.72}$$

With triangular inequalities one gets :

$$1_{\left\{d(x,y)\leq\frac{h_n}{2},\, d(x,u)\leq\frac{h_n}{2},\, d(x,v)\leq\frac{h_n}{2}\right\}} \leq 1_{\{d(x,u)\leq h_n,\, d(y,u)\leq h_n,\, d(x,v)\leq h_n,\, d(y,v)\leq h_n\}} \tag{10.73}$$

Then, we introduce the notations :

$$\mu_n = \sup_{\left\{d(x,y)\leq\frac{h_n}{2},\, x,y\in E\right\}} |w(x) - w(y)|.$$

Since w is uniformly continuous on its compact support, w is uniformly continuous on E and μ_n tends to 0. Moreover, one obtains the following inequality :

$$\begin{aligned} w(y)\, 1_{d(x,y)\leq\frac{h_n}{2}} &= w(x)\, 1_{\left\{d(x,y)\leq\frac{h_n}{2}\right\}} + (w(y) - w(x))\, 1_{\left\{d(x,y)\leq\frac{h_n}{2}\right\}} \\ &\geq w(x)\, 1_{\left\{d(x,y)\leq\frac{h_n}{2}\right\}} - \mu_n 1_{\left\{d(x,y)\leq\frac{h_n}{2}\right\}}. \end{aligned} \tag{10.74}$$

We now start from the definition of $\Omega_4(h_n)$ and then use (10.73), (10.74) and Lemma 10.4.1.

$$
\begin{aligned}
&\Omega_4(h_n)\\
&= \int_{E^4} 1_{\{d(x,u)\leq h_n,\, d(y,u)\leq h_n,\, d(x,v)\leq h_n,\, d(y,v)\leq h_n\}} w(x)\, w(y)\, dP_X(x)\, dP_X(y)\, dP_X(u)\, dP_X(v)\\
&\geq \int_{E^4} 1_{\left\{d(x,u)\leq \frac{h_n}{2},\, d(y,x)\leq \frac{h_n}{2},\, d(x,v)\leq \frac{h_n}{2}\right\}} w(x)\, w(y)\, dP_X(x)\, dP_X(y)\, dP_X(u)\, dP_X(v)\\
&\geq \int_{E^2} F_x^2\left(\frac{h_n}{2}\right) 1_{\left\{d(x,y)\leq \frac{h_n}{2}\right\}} w(x)\, w(y)\, dP_X(x)\, dP_X(y)\\
&\geq C_1^2\Phi^2\left(\frac{h_n}{2}\right)\int_{E^2} 1_{\left\{d(x,y)\leq \frac{h_n}{2}\right\}} w^2(x)\, dP_X(x)\, dP_X(y)\\
&\quad -C_2^2\mu_n\Phi^2\left(\frac{h_n}{2}\right)\int_{E^2} 1_{\left\{d(x,y)\leq \frac{h_n}{2}\right\}} w(x)\, dP_X(x)\, dP_X(y)\\
&\geq C_1^2\Phi^2\left(\frac{h_n}{2}\right)\int_{E} F_x\left(\frac{h_n}{2}\right) w^2(x)\, dP_X(x)\\
&\quad -C_2^2\mu_n\Phi^2\left(\frac{h_n}{2}\right)\int_{E} F_x\left(\frac{h_n}{2}\right) w(x)\, dP_X(x)\\
&\geq C_1^3\Phi^3\left(\frac{h_n}{2}\right)\left(\int_E w^2(x)\, dP_X(x) - \left(\frac{C_2}{C_1}\right)^3 \mu_n \int_E w(x)\, dP_X(x)\right)\\
&\geq C\Phi^3\left(\frac{h_n}{2}\right).
\end{aligned}
$$

The last line comes from the fact that both integral are finite, $\mu_n \to 0$ and the first integral is stricly positive because $w \neq 0$ a.s.. Finally, Lemma 10.4.1 implies that $\Phi(s) = s^\delta$ hence one gets :

$$\Phi\left(\frac{h_n}{2}\right) = \left(\frac{h_n}{2}\right)^\delta = \Phi(h_n)\, 2^{-\delta}.$$

Consequently, (10.72) implies that

$$\Omega_4(h_n) \geq C\Phi^3(h_n).$$

Assumption (10.19) implies $n\Phi(h_n) \to +\infty$ hence assumption (10.10) holds with $l = 0$.

□

Proof of Lemma 10.4.3 We will note $\nu = d_F(S, C^c)$. Since X_F has a continuous and positive density on C, if λ is the Lebesgue measure on $\mathbb{R}^p$, for all $x \in S \subset C$

one gets :

$$\begin{aligned} F_x(h_n) &= \int_{\mathbb{R}^p} 1_{\|x_F - y_F\| \le h_n} f(y_F)\, d\lambda(y_F) \\ &= f(x_F) \int_{\mathbb{R}^p} 1_{\|x_F - y_F\| \le h_n} d\lambda(y_F) + \int_{\mathbb{R}^p} 1_{\|x_F - y_F\| \le h_n} (f(y_F) - f(x_F))\, d\lambda(y_F) \\ (10.75) &= \lambda(B(0,h_n)) \left(f(x_F) + O\left(\sup_{x_F \in S,\, \|x_F - y_F\| \le h_n} |f(x_F) - f(y_F)| \right) \right) \end{aligned}$$

The last line comes from the fact that the Lebesgue measure is invariant by translation. On one hand, f is continuous hence uniformly continuous on C and consequently one gets :

$$\sup_{x_F, y_F \in C,\, \|x_F - y_F\| \le h_n} |f(x_F) - f(y_F)| \to 0. \tag{10.76}$$

Moreover, since $h_n \to 0$, for n large enough ($n \ge N_0$), one has $h_n \le \frac{\nu}{2}$ hence $S_{h_n} \subset C$. Indeed for all $y \in S_{h_n}$, there exists $x \in S$ such that $d_F(x,y) \le h_n$ and a simple triangular inequality gives :

$$d_F(y, C^c) \ge d_F(x, C^c) - d_F(x,y) \ge \frac{\nu}{2} > 0.$$

Consequently, for $n \ge N_0$, it comes :

$$\sup_{x_F \in S,\, \|x_F - y_F\| \le h_n} |f(x_F) - f(y_F)| \le \sup_{x_F, y_F \in C,\, \|x_F - y_F\| \le h_n} |f(x_F) - f(y_F)|. \tag{10.77}$$

Then, with (10.76) and (10.77), one gets :

$$\sup_{x_F \in S,\, \|x_F - y_F\| \le h_n} |f(x_F) - f(y_F)| \to 0. \tag{10.78}$$

On the other hand, we make a variable change to show that $\lambda(B(0,h_n)) = V_{\|\,\|}(p)\, h_n^p$.

$$\begin{aligned} \lambda(B(0,h_n)) &= \int_{\mathbb{R}^p} 1_{d_F(0,s) \le h_n} d\lambda(s) \\ &= \int_{\mathbb{R}^p} 1_{d_F\left(0, \frac{s}{h_n}\right) \le 1} d\lambda(s) \\ &= \int_{\mathbb{R}^p} 1_{d_F(0,u) \le 1} h_n^p d\lambda(u) \\ &= V_{\|\,\|}(p)\, h_n^p. \end{aligned} \tag{10.79}$$

From (10.75), (10.78) and (10.79) we get the following asymptotic equivalence :

$$\sup_S \left| \frac{F_x(h_n)}{h_n^p} - V_{\|\,\|}(p)\, f(x_F) \right| \to 0$$

Moreover, since f is continuous and positive on the compact C, f is bounded on C and achieves its extreme values. Consequently, one obtains

$$\inf_C c(x) = V_{\|\,\|}(p) \inf_C f(x_F) = V_{\|\,\|}(p)\, f(x_0) > 0,$$

hence X is uniformly fractal on S for the semimetric d_F with $\delta = p$.

□

CHAPITRE 11

BOOTSTRAP PROCEDURES AND APPLICATIONS FOR NO EFFECT TESTS

In this chapter, we want to use the results of Chapter 10 to construct structural testing procedures. However, we are faced with some practical issues dealing with the estimation of the critical value of the test and the approximation of the integral used in (10.3) to define the test statistic. We introduce in this chapter bootstrap procedures to estimate the critical value of our general structural testing procedures. Then, we focus on no effect tests and observe on simulated datasets the good level and power properties of our tests. We also discuss the issue of the choice of both smoothing parameter and bootstrap iterations number. Moreover we show that wild bootstrap methods stay relevent in the case of heteroscedastic errors. Finally, we consider two real datasets coming from the food industry. The results we obtain on the Tecator dataset are strengthen by the previous studies of this reference datasets by other authors. The study of the Corn dataset illustrates the interest and the efficiency of our procedures even on a small length dataset. This chapter is the base of an applied article that is still in progress. Its final version will be submitted soon.

To answer to the problem of integral approximation, we propose the following method. We assume that we have a dataset $D^{**} := (X_k^{**})_{1\leq k\leq l_n}$ independent from the sample D and D^*. In practice, the datasets D, D^* and D^{**} may for instance come from a decomposition of the initial independent dataset. Because the integral (with respect to dP_X) used in (10.3) to define the test statistic corresponds to an expectation, it may be approximated, in straightforward way, by the empirical mean of

$$\left(\left(\sum_{i=1}^{n}\left(Y_i - r_0^*\left(X_i\right)\right)K\left(\frac{d\left(X_i, X_k^{**}\right)}{h_n}\right)\right)^2 w\left(X_k^{**}\right)\right)_{1\leq k\leq l_n}.$$

11.1. Bootstrap procedures

In practice one has to estimate the critical value of the test. The estimation of bias and variance terms seems difficult and it is often irrelevant to use directly the quantiles of the asymptotic law to estimate the critical value. Instead of doing so,

we propose various residual-based bootstrap procedures. From the pioneer work of Efron (1979), bootstrap methods have encountered a strong interest. Bootstrap methods have been extensively sutied in the context of nonparametric regression with scalar or multivariate covariate. They allow to improve the performences of confidence bands (see for instance Härdle and Marron, 1990, Cao, 1991, Hall, 1992) ans testing procedures (see for instance Hall and Hart,1990, Härdle and Mammen, 1993, Stute *et al.*, 1998a, Härdle *et al.*, 2005) and can also be useful in the issue of smoothing parameter choice (see for instance Hall, 1990, Gonzalez Manteiga *et al.*, 2004). In the functional context, bootstrap methods have been less developped (see Cuevas et Fraiman, 2004, Fernández de Castro *et al.*, 2005 and Cuevas *et al.*, 2006). In the functional nonparametric regression model, the recent work of Ferraty *et al.* (2008b) focuses on the use of residual-based bootstrap methods to estimate confidence bands. They provide both theoretical and practical interesting results. We will use a similar bootstrap approach.

We denote $\hat{r}^*$ the kernel estimator constructed from the sample D^*. We propose to repeat the following procedure for b in $\{1, \ldots, N_{boot}\}$ to construct N_{boot} bootstrap values of T_n^*.

Bootstrap Procedure :

We compute successively the values of :

1. Estimated residuals : $\hat{\epsilon}_i = Y_i - \hat{r}\left(X_i\right), 1 \leq i \leq n$ and $\hat{\epsilon}_i = Y_i - \hat{r}^*\left(X_i\right), n+1 \leq i \leq N$.
2. Estimated centered residuals : $\hat{\tilde{\epsilon}}_i = \hat{\epsilon}_i - \overline{\hat{\epsilon}}_n, 1 \leq i \leq n$, $\hat{\tilde{\epsilon}}_i = \hat{\epsilon}_i - \overline{\hat{\epsilon}}_{N-n}, n+1 \leq i \leq N$.
3. Bootstrap residuals (three alternative methods) :
   a) Resampling or Naive bootstrap : $\left(\tilde{\epsilon}_i^b\right)_{1\leq i\leq n}$, respectively $\left(\tilde{\epsilon}_i^b\right)_{n+1\leq i\leq N}$, are drawn with replacement from $\left\{\hat{\tilde{\epsilon}}_i, 1 \leq i \leq n\right\}$, respectively from $\left\{\hat{\tilde{\epsilon}}_i, n+1 \leq i \leq N\right\}$.
   or
   b) Smooth Naive bootstrap : $\left(\tilde{\epsilon}_i^b\right)_{1\leq i\leq n}$, respectively $\left(\tilde{\epsilon}_i^b\right)_{n+1\leq i\leq N}$, are drawn from a smoothed version of the empirical cumulative distribution function of $\left(\hat{\tilde{\epsilon}}_i\right)_{1\leq i\leq n}$, respectively $\left(\hat{\tilde{\epsilon}}_i\right)_{n+1\leq i\leq N}$.
   or
   c) Wild bootstrap : $\tilde{\epsilon}_i^b = \hat{\tilde{\epsilon}}_i U_i, 1 \leq i \leq N$, where $(U_i)_{1\leq i\leq n}$ are i.i.d.$\sim P_W$, independent of $(X_i, Y_i)_{1\leq i\leq N}$ and fulfill $\mathbb{E}\left[U_1\right] = 0$, $\mathbb{E}\left[U_1^j\right] = 1, j = 2, 3$.
4. Bootstrap responses : $\tilde{Y}_i^b = r_0^*\left(X_i\right) + \tilde{\epsilon}_i^b, 1 \leq i \leq N$.
5. Bootstrap test statistic : $\tilde{T}_n^{b*}$ computed from the sample $\left(\tilde{Y}_i^b, X_i\right)$.

Finally, if α is the level of the test, we reject assumption $\mathcal{H}_0$ if our test statistic T_n^* is greater than the value of the empirical $(1-\alpha)$-quantile of the family $\left(\tilde{T}_n^{b*}\right)_{1\leq b\leq N_{boot}}$.

Remark : We study three wild bootstrap procedures constructed from three distributions P_W^1, P_W^2, P_W^3 initially introduced in Mammen (1993) :

- $P_W^1 = \frac{\sqrt{5}+1}{2\sqrt{5}}\delta_{\frac{1-\sqrt{5}}{2}} + \frac{\sqrt{5}-1}{2\sqrt{5}}\delta_{\frac{1+\sqrt{5}}{2}}$ where δ is the Dirac function.
- P_W^2 is the law of the random variable U defined by $U = \frac{V}{sqrt2} + \frac{\left(V^2-1\right)}{2}$, where $V \sim \mathcal{N}(0,1)$.
- P_W^3 is the law of the variable U defined by $U = \left(\zeta_1 + \frac{V_1}{\sqrt{2}}\right)\left(\zeta_2 + \frac{V_2}{\sqrt{2}}\right) - \zeta_1\zeta_2$, where V_1 and V_2 are independent $\mathcal{N}(0,1)$ random variables and $\zeta_1 = \sqrt{\frac{3}{4} + \frac{\sqrt{17}}{12}}$ and $\zeta_2 = \sqrt{\frac{3}{4} - \frac{\sqrt{17}}{12}}$.

However, it is possible to apply the previous algorithm with any law P_W. In the following we compare various procedures and use the notations :

Res.	Resampling procedure
S.N.B.	Smooth Naive Bootstrap procedure
*W.B.*1	Wild Bootstrap procedure with P_W^1
*W.B.*2	Wild Bootstrap procedure with P_W^2
*W.B.*3	Wild Bootstrap procedure with P_W^3

11.2. Simulation studies : No effect tests

We compare level and power of these bootstrap procedures on simulation studies. We also study how many bootstrap iterations we need to get relevant results and focus on the problem of choosing the smoothing parameter. In the specific context of no effect tests, we propose to take $r_0^* \equiv \frac{1}{m_n}\sum_{i=n+1}^{n+m_n} Y_i$ and our test statistic have the following expression :

$$\int \left(\sum_{i=1}^{n}\left(Y_i - \frac{1}{m_n}\sum_{i=n+1}^{n+m_n} Y_i\right) K\left(\frac{d(x,X_i)}{h_n}\right)\right)^2 w(x)\, dP_X(x).$$

11.2.1. A First simulated dataset. — We simulate 300 curves

$$X_i(t) = a_i cos(2\pi t) + b_i sin(3\pi t) + c_i (t-0.45)(t-0.75) e^{d_i t},$$

with $a_i \sim \mathcal{U}([-1;1]), b_i \sim \mathcal{N}(1;1), c_i \sim \mathcal{U}([1;5])$ and $d_i \sim \mathcal{U}([-1.5;1.5])$ and consider the following model where $\epsilon_i \sim \mathcal{N}(0;1)$

$$Y_i = k(a_i + b_i + c_i + d_i) + 2 + \epsilon_i.$$

We split this dataset in three independent datasets of size 100. The first one corresponds to D, the second one to D^* while the third one is used to approximate the integral.

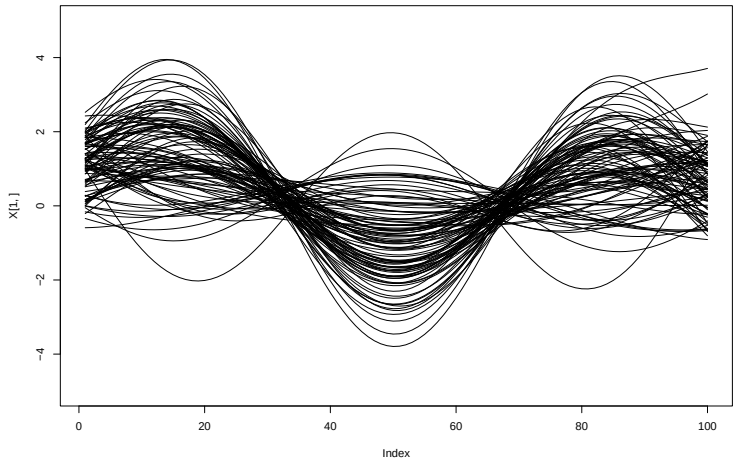

FIGURE 11.1. Some simulated curves

In this simulation we use the $\mathbb{L}^2([0;1])$ metric, a quadratic kernel and choose the smoothing parameter h_0 (its definition is given in the next paragraph). In the following table we give the empirical probabilities of rejecting the no-effect assumption for various values of k computed on 10000 tests with $N_{boot} = 100$. R represents the empirical signal-to-noise ratio. The parameter k quantifies the effect of the explanatory variable X on the response variable Y. When $k = 0$, the variable X has no effect on Y, we are under the null hypothesis hence the empirical rejection probabilities correspond to the empirical level of the test. On the contrary, when $k > 0$ we are under the alternative hypothesis and the empirical rejection probabilities represent the empirical power of the test. Furthermore, the greater k is, the more the effect of X on Y is important, and the more the empirical power of the test grows (see Table 1).

Table 1. —

k	*Res.*	*S.N.B.*	*W.B.*1	*W.B.*2	*W.B.*3	R
0	0.0618	0.0463	0.0614	0.0437	0.0534	1
0.2	0.1609	0.1275	0.1601	0.1133	0.1373	1.14
0.4	0.5180	0.4603	0.5437	0.4447	0.4999	1.56
0.59	0.8293	0.7806	0.8517	0.7836	0.8231	2.21
0.8	0.9557	0.9420	0.9659	0.9445	0.9591	3.23
1	0.9862	0.9775	0.9921	0.9827	0.9886	4.47

In Figure 11.2 we represent the power of our bootstrap testing procedures in function of the value of the signal to noise ratio R. The solid curve corresponds to the naive bootstrap procedure, the dotted curve stands for the resampling method and the dashed curves correspond to the wild bootstrap methods. The horizontal line indicates the value of the level 0.05 of the test.

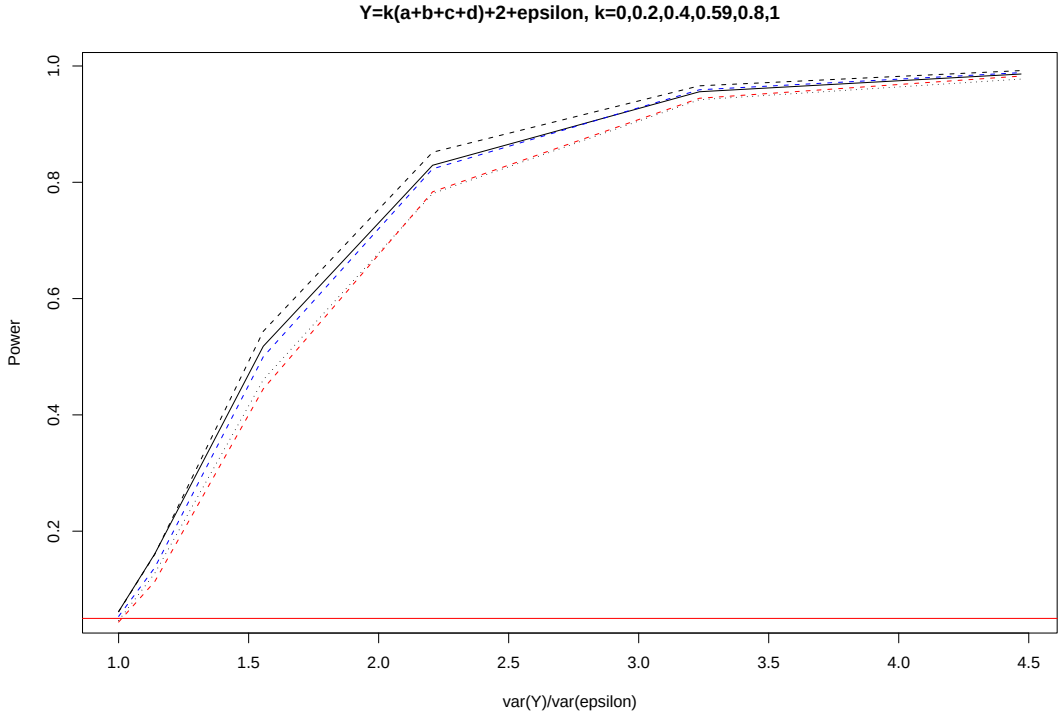

FIGURE 11.2. Power of bootstrap procedures relatively to the value of the signal-to-noise ration R

We can see through Table 1 and Figure 11.2 that all the proposed methods give fairly good results in terms of level and power. However, it seems that the naive bootstrap and the third wild bootstrap procedures provide better results than the other methods. Consequently we only focus on these methods for the remaining simulation studies.

We now consider the problem of the choice of the smoothing parameter in pratice. We have simulated 1000 datasets of 300 pairs (X_i, Y_i) in the same way as before. We note x_i and x_k^{**} the observed curves corresponding to X_i and X_k^{**}. For each dataset we compute an approximation h_0 of the value h_{min} defined by :

$$h_{min} := \max_{1\leq k\leq l_n} \left(\min_{1\leq i\leq n} \left(d\left(x_i, x_k^{**}\right)\right)\right).$$

In other words, h_{min} corresponds to the smallest value of the smoothing parameter h for which

$$\min_{1\leq k\leq l_n} \sum_{i=1}^{n} K\left(\frac{d\left(x_i, x_k^{**}\right)}{h}\right) > 0.$$

For each dataset, we have made a no effect test (with the third wild bootstrap procedure and 100 iterations) for different values of the smoothing parameter. The results (empirical rejection probabilities) are presented in the following Table 2 where h_{CV} is the smoothing parameter given by cross-validation and $h_j := h_0 + \frac{j}{5}\left(6h_{CV} - h_0\right)$, $0 \leq j \leq 5$.

Table 2. —

k	h_0	h_1	h_2	h_3	h_4	h_5	h_{CV}
0	0.052	0.055	0.059	0.057	0.060	0.059	0.051
0.59	0.843	0.638	0.338	0.193	0.138	0.119	0.82
1	0.989	0.901	0.649	0.407	0.258	0.191	0.984

It seems that the level of our test is not influenced by the choice of the smoothing parameter. However, when we take great values of the smoothing parameter the power of our test is low. Finally, it seems that the choice of the smoothing parameter by cross-validation provides fairly good results. The value h_0 also seems to give relevant results.

We now analyse the influence of the number of bootstrap iterations on the empirical level and power of our test. We simulate 1000 datasets in a similar way as before and take the smoothing parameter given by h_0. For each dataset we have made various no-effect tests based on different numbers of bootstrap iterations. We get the following results :

Table 3. —

Method	k	5	10	15	20	50	100	200	500	700	1000
$W.B.3$	0	0.170	0.076	0.056	0.077	0.041	0.048	0.040	0.038	0.037	0.034
$W.B.3$	0.59	0.886	0.782	0.718	0.860	0.810	0.819	0.829	0.836	0.828	0.825
$W.B.3$	1	0.991	0.957	0.948	0.994	0.988	0.993	0.992	0.994	0.992	0.993
$S.N.B.$	0	0.147	0.078	0.055	0.090	0.059	0.063	0.048	0.047	0.047	0.046
$S.N.B.$	0.59	0.851	0.764	0.693	0.835	0.762	0.779	0.77	0.765	0.775	0.775

It appears that the power of both testing procedures and the level of the naive bootstrap procedure become stable when the iteration number is greater than 50. However, it seems that when the number of iterations grows the level of the wild bootstrap procedure decreases. Maybe, an other choice of the smoothing parameter could improve these results.

11.2.2. Nonparametrically generated growth curves. — We are now interested in models in which the simulated curves have a nonparametric nature. To simulate each functional random variable X_i, we propose the following procedure :

1. Simulate 1000 standard gaussian random variables $(\epsilon_j)_{1\leq j\leq 1000}$,
2. Compute $U_\ell = \sum_{j=1}^{\ell} \epsilon_j$ for $1 \leq \ell \leq 1000$,
3. Compute $U_\ell^+ = U_\ell + |\min(\min_{1\leq \ell\leq 1000} U_\ell, 0)|$, for $1 \leq \ell \leq 1000$
4. Compute $X_{i,\ell} = \frac{\sum_{j=1}^{\ell} U_j^+}{1000}$ for $1 \leq \ell \leq 1000$.

The values $(U_\ell)_{1\leq\ell\leq 1000}$ may be viewed as the discretization of a Brownian motion $B_i(t)$ defined on $[0;1]$ with $B_i(0) = 0$. We then introduce $W_i(t) = B_i(t) + |\min\left(\inf_{s\in[0;1]} B(s), 0\right)$ that corresponds to a vertical translation of $B_i(t)$ which only takes nonnegative values. Finally we define $X_i(t)$ as the integral of W_i between 0 and t. Each value $X_{i,l}$ corresponds to approximation of the value of X_i at the discretization point $j/1000$. The simulated curves $(X_i)_{1\leq i\leq n}$ may be viewed as growth curves (see Figure 11.3).

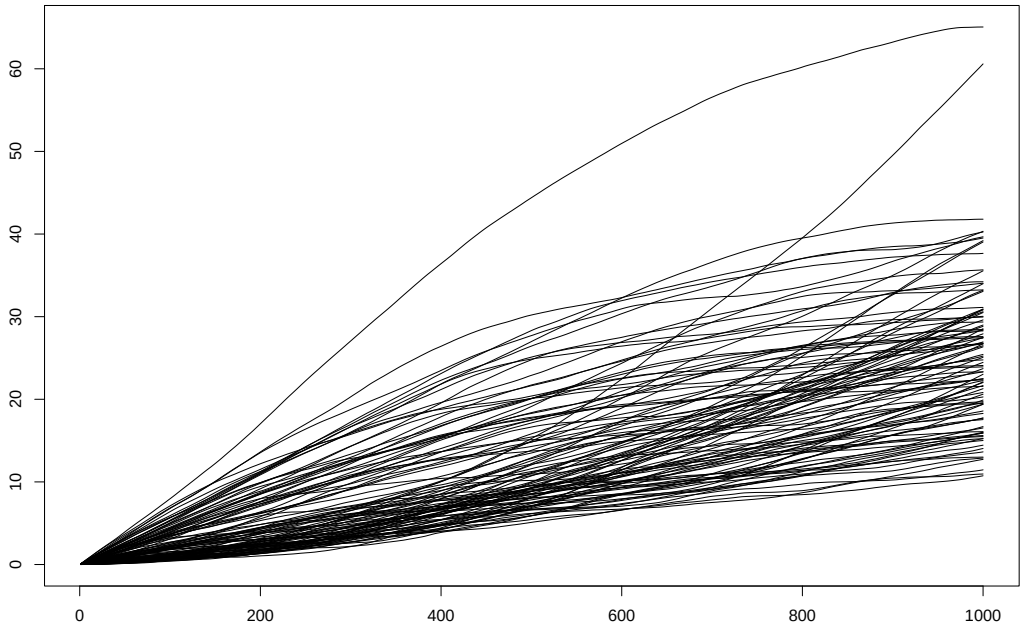

FIGURE 11.3. Sample of simulated curves

We use in this section the smmothing parameter h_0 (see Table 6 to get a justification) and the $\mathbb{L}^2$ metric. We simulate 10000 datasets containing 300 independent pairs (X_i, ϵ_i), where $\epsilon_i \sim \mathcal{N}(0,1)$. For each dataset, we consider various models of the form $Y = k\int_0^1 X(t)\, cos(7.5t)\, dt + \epsilon$ and construct for each model three subdatasets containing 100 pairs (X_i, Y_i). We use each subdataset in a similar way as in the previous section and use the same notation R for the signal to noise ratio. Finally, we have compared (see Table 4) the behaviour of our bootstrap procedures with $N_{boot} = 100$ on these 10000 simulated samples.

Table 4. —

	Res	SNB	WB1	WB2	WB3	R
$Y = \epsilon$	0.0616	0.0498	0.0629	0.0565	0.0596	1
$Y = \frac{1}{3}\int_0^1 X(t)\, cos(7.5t)\, dt + \epsilon$	0.1962	0.1715	0.2073	0.1845	0.1961	1.152766
$Y = \frac{2}{3}\int_0^1 X(t)\, cos(7.5t)\, dt + \epsilon$	0.5554	0.5301	0.5901	0.5644	0.5805	1.610272
$Y = \int_0^1 X(t)\, cos(7.5t)\, dt + \epsilon$	0.7996	0.7928	0.8340	0.8268	0.8336	2.372517
$Y = \frac{4}{3}\int_0^1 X(t)\, cos(7.5t)\, dt + \epsilon$	0.8889	0.8900	0.9175	0.9149	0.9140	3.439501
$Y = \frac{5}{3}\int_0^1 X(t)\, cos(7.5t)\, dt + \epsilon$	0.9221	0.9283	0.9475	0.9450	0.9473	4.811225

One can read the previous table has Table 1 : the first line corresponds to the level of our no effect testing procedures while the others lines contain the power of our testing procedures for various alternatives. It appears that the results of the proposed methods are fairely good and have a similar nature. However the smooth naive bootstrap procedure seems a little better in terms of level while bootstrap methods seem more powerfull. We have also compared our methods under other alternatives on 1000 datasets. Table 5 presents the results obtained with $N_{boot} = 100$. They can be read as before and the comparison of the methods leads to the same comments.

Table 5. —

	Res	SNB	WB1	WB2	WB3	R
$Y = \epsilon$	0.066	0.056	0.068	0.058	0.060	1
$Y = 5\exp\left(-\int_0^1 X(t)\, cos(7.5t)\, dt\right) + \epsilon$	0.195	0.172	0.203	0.186	0.197	1.147120
$Y = 10\exp\left(-\int_0^1 X(t)\, cos(7.5t)\, dt\right) + \epsilon$	0.574	0.528	0.582	0.540	0.564	1.589047
$Y = 15\exp\left(-\int_0^1 X(t)\, cos(7.5t)\, dt\right) + \epsilon$	0.766	0.755	0.769	0.769	0.775	2.325781
$Y = 20\exp\left(-\int_0^1 X(t)\, cos(7.5t)\, dt\right) + \epsilon$	0.841	0.845	0.860	0.858	0.854	3.357322
$Y = 25\exp\left(-\int_0^1 X(t)\, cos(7.5t)\, dt\right) + \epsilon$	0.882	0.883	0.901	0.887	0.902	4.68367

We choose to use in the remaining of this section the third wild bootstrap procedure because it seems to make the balance between good level and power properties. We are now interested in exploring how the choice of the smoothing parameter has an influence on the level and the power of our testing procedures. We simulate 1000 datasets in the same way as before. For each dataset, we make a no-effect test for various values of the smoothing parameter under both null and alternative hypothesis. We present here in Table (6) the empirical rejection probabilities obtained with $N_{boot} = 100$ from these 1000 datasets.

Table 6. —

h	h0	$\frac{2h0+hCV}{3}$	$\frac{h0+2hCV}{3}$	hCV	$\frac{-h0+4hCV}{3}$	$\frac{3}{2}hCV$
$Y = e$	0.056	0.056	0.055	0.056	0.054	0.057
$Y = \int_0^1 X(t)\, cos(7.5t)\, dt + e$	0.833	0.787	0.738	0.693	0.645	0.475

We observe, as on the previous simulated datasets, that the choice of the smoothing parameter does not seem to have an important influence on the level of our test. However, it is clear that the more the smoothing parameter grows, the less our test is powerfull. In particular, one notes that if we use the smoothing parameter given by cross-validation we get a lack of power of 14 percents. Consequently, we choose to use the smoothing parameter h_0 in the remaining of this chapter.

When one wants to use our testing procedures in practice, one is also faced with the issue of choosing the value of the bootstrap iterations number and may wonder how many bootstrap iterations are necessary. To explore how the choice of the bootstrap iterations number has an effect on the power and the level of our testing procedure, we have compared the results obtained with our procedure for various bootstrap iteration numbers (N_{boot}) on 1000 datasets. Table 7 gives the results we have obtained under both null and alternative hypothesis.

Table 7. —

Nboot	5	10	15	20	50	100	200	500	1000	10000
Level	0.181	0.097	0.064	0.101	0.063	0.061	0.060	0.058	0.057	0.056
Power	0.894	0.865	0.850	0.892	0.863	0.881	0.874	0.875	0.875	0.874

We see in this table that a number of bootstrap iterations (N_{boot}) around 100 or 200 is enough to have a good approximation of the quantiles. Taking a higher number of bootstrap iterations leads to similar results and takes more time.

Finally, we have only considered homoscedastic errors until now. In the multivariate case, it is well-known that wild bootstrap methods are adapted to the context of heteroscedastic errors. It would be interesting to study the behaviour of our methods in the heteroscedastic context. Let $a : t \mapsto \cos(7.5t)$, $b : t \mapsto sin(7.5t)$ and define heteroscedastic errors : $\epsilon_i^k \sim \mathcal{N}\left(0, \left(1 + k|\int_0^1 X(t)\, b(t)\, dt|\right)\right)$. We compare the results obtained by the smooth naive bootstrap and the third wild bootstrap methods on 1000 datasets, with $N_{boot} = 100$ when the heteroscedasticity effect (i.e. k) grows. The way we construct our alternative model implies that when k grows, the signal to noise ratio R decreases. Hence the power of the test should decrease when k grows.

Table 8. —

k	0	0.25	0.5	0.75	1	1.25
$Y = \epsilon^k$ SNB	0.052	0.049	0.034	0.036	0.031	0.024
$Y = \epsilon^k$ WB3	0.062	0.058	0.052	0.055	0.057	0.053
$Y = \int_0^1 X(t)\, a(t)\, dt + \epsilon^k$ SNB	0.774	0.665	0.528	0.425	0.322	0.246
$Y = \int_0^1 X(t)\, a(t)\, dt + \epsilon^k$ WB3	0.836	0.759	0.659	0.566	0.487	0.399
R	2.3631	1.9492	1.6869	1.5152	1.3985	1.3164

We observe in Table 8 a significative difference between the results obtained by the third wild bootstrap method and thoses obtained by the smooth naive bootstrap procedure. It seems that the wild bootstrap method is adapted to the case of heteroscedastic errors since its power seems to be comparable to the one obtained in the homoscedastic model for the same signal to noise ratio (see Table 4). Moreover, the level of the wild bootstrap stays near the theoretical level 0.05 while the smooth naive bootstrap procedure leads to much worse results. As a conclusion, if one has no information on the heteroscedasticity or homoscedasticity of the residuals, one would rather use a wild bootsrap procedure. If one knows that the model under study is homoscedastic, the use of both wild or naive bootstrap will lead to comparable results.

11.3. Application on real datasets

We now consider two concrete datasets coming from spectrometric studies. Previous studies of such datasets (see for instance Ferraty and Vieu, 2006, section 7.2) have highlighted the efficiency of semimetrics based on derivatives. In spectrometric sudies, one often wants to predict the quantity of a specific chemical element present in an object from the spectrometric curve corresponding to the whole object. An usual question is then to find which derivative has an effect on the quantity we want to predict. We propose to use our no effect tests taking as explanatory variables the successives derivatives of spectrometric curves. We will use in the remaining of the manuscript the $\mathbb{L}^2$ metric. Note that the choice of the semimetric has a direct influence on the nature of the null and alternative hypothesis of our tests.

11.3.1. Tecator dataset. — One now focuses on the "Tecator" dataset already presented at the beginning of this dissertation (cf Chapter 3). This dataset has been studied by many authors. It can be downloaded from the Statlib website :

$$http : //lib.stat.cmu.edu/datasets/tecator,$$

where one may also find a deeper description of the data. The results we obtain are satisfacting and are corroborated by former studies made on this "reference" dataset. For each one of the 215 meat pieces under study, one gets a spectrometric curve (discretized in 100 frequencies) together with a chemical analysis of moisture, fat and protein contents. Chemical analysis need time and money while spectrometric curves may be obtained more easily. Consequently, it is worth considering the issue of predicting moisture, fat, or protein contents from spectrometric curves. A first step is to test if the spectrometric curve (or one of its first deriatives) has an effect on the moisture, fat or protein content. We propose to make a no effect test with five different explanatory variables : the successive derivatives of the spectrometric curves (see Figures 3.4 and 11.4).

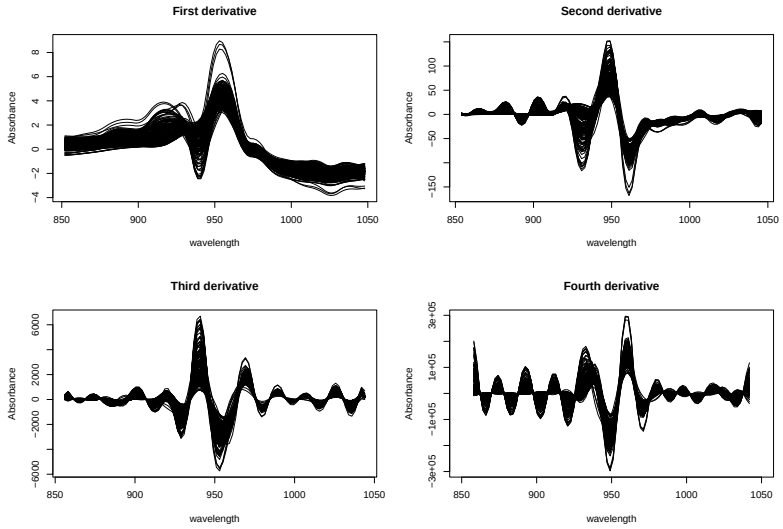

FIGURE 11.4. Sample of derivatives of spectrometric curves

We randomly split the original dataset into three datasets of respective length $n = 90$, $m_n = 90$, and $l_n = 35$. Then we compute the empirical signification degree of five no effect tests where the covariate is respectively the derivative of order $k \in \{0, 1, 2, 3, 4\}$ of the spectrometric curves. From the results obtained in the simulation study we choose the third wild bootstrap method, with smoothing parameter h_0 and we take $N_{boot} = 1000$. We repeat one hundred times the previous study and present in the next figure the boxplot of the computed empirical signification degrees. The horizontal solid line stands for the level 0.05 of the test.

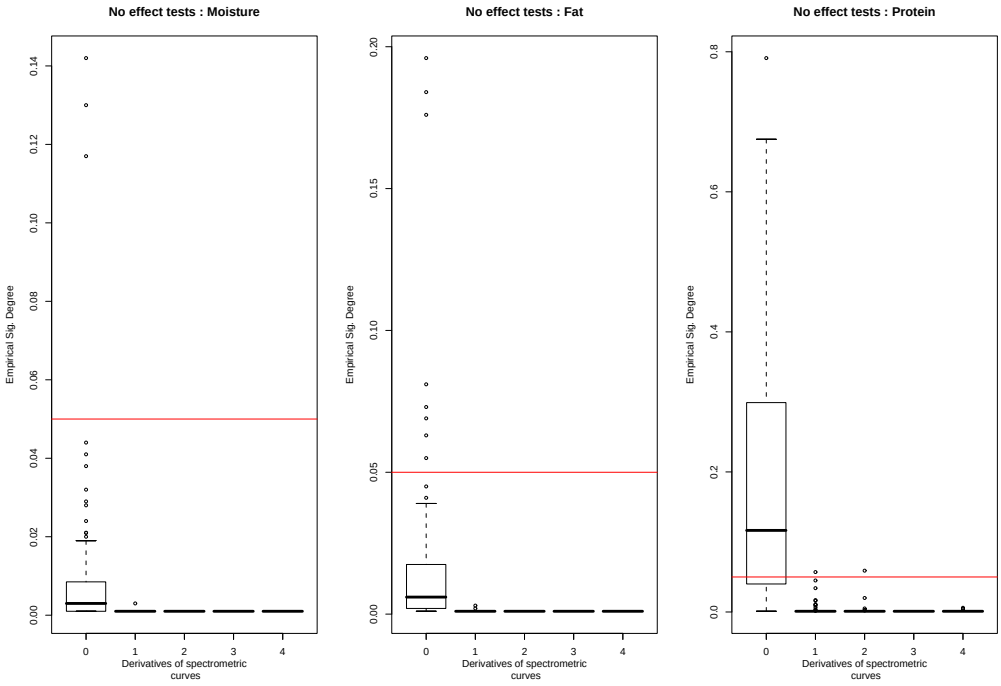

FIGURE 11.5. No effect tests

It appears in Figure 11.5 that the successive derivatives of spectrometric curves seem to have an effect on the moisture, fat and protein content of meat pieces. The original spectrometric curves seems to have an effect only in the context of moisture and fat content prediction. Moreover, the effect of original spectrometric curves seems to be less important than the effect their derivatives. Now, we want to know which derivative is the best to predict moisture, fat or protein content. To do that, we propose to split the dataset into a learning dataset of length 125 and a testing dataset of length 90 and compute the mean squared errors linked with pedicting moisture, fat, or protein content from each derivative. We repeat this procedure 100 times and compute the mean over the 100 mean squared errors we obtained. The results are gathered in Table 9.

Table 9. —

Derivation order	0	1	2	3	4
Moisture	72.88396	23.53545	**5.252226**	8.844904	20.08812
Fat	125.8555	42.50953	**6.283264**	11.05759	30.45783
Protein	8.245794	3.824866	2.754601	**2.277398**	3.006949

We conclude from the previous table that the second derivative of spectrometric curves seems to be the best explanatory variable in the context of moisture and fat content prediction while the third one appears as the best covariate in the case of protein content prediction. The successive derivatives of spectrometric curves are linked by nature. Consequently, one can wonder if the effect of all the derivatives is resumed by the effect of the best predictor (the one that leads to the best mean squared error). We consider the regression models

$$Y_{Moisture} = m_M\left(X''\right) + \epsilon_M,\ Y_{Fat} = m_F\left(X''\right) + \epsilon_F,\ \text{ and } Y_{Protein} = m_P\left(X'''\right) + \epsilon_P.$$

We propose to estimate the residuals

$$R_{Moisture} := Y_{Moisture} - \hat{m_M}\left(X''\right),\ \text{ and } R_{Fat} := Y_{Fat} - \hat{m_F}\left(X''\right)$$

and we are interested in testing if the derivatives of order $\{0, 1, 3, 4\}$ have an effect on the residuals $R_{Moisture}$ and R_{Fat}. We also want to test if the derivatives of order $\{0, 1, 2, 4\}$ have an effect on $R_{Protein} := Y_{Protein} - \hat{m_P}\left(X'''\right)$. For each regression model, we estimate the operator m nonparametrically with kernel methods and use the smoothing parameter given by cross-validation. Then we make no effect tests on the residuals in the same way as before. Our results are presented in Figure 11.6.

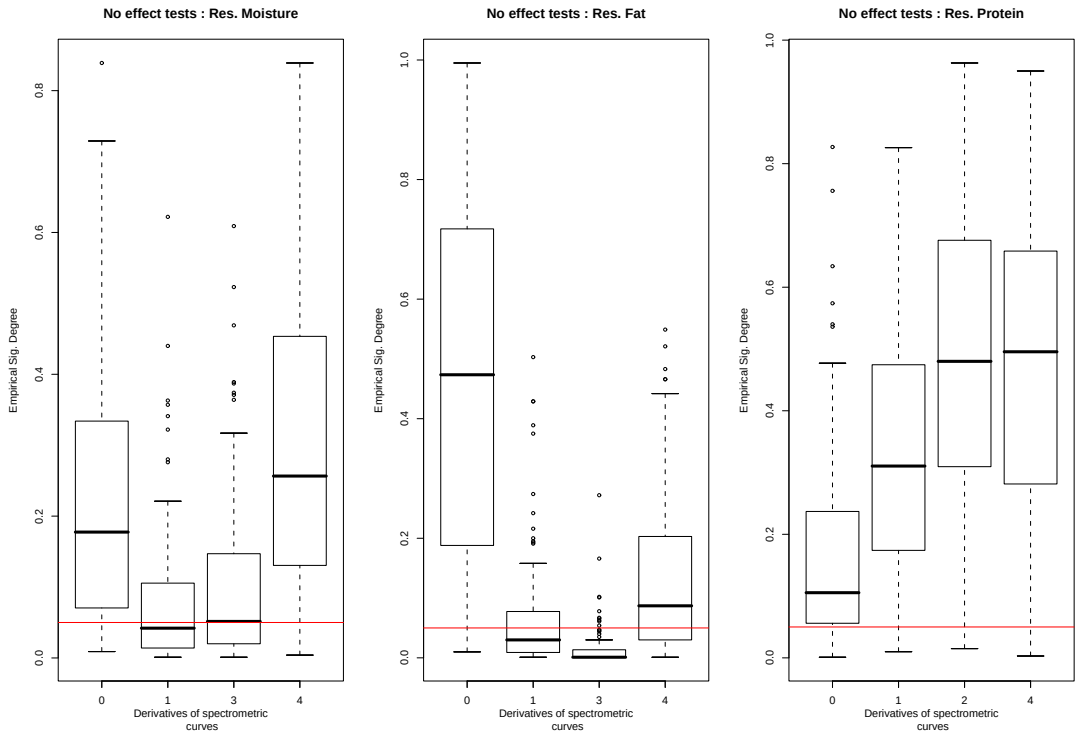

FIGURE 11.6. No effect tests on first residuals

In the context of moisture content prediction, the first and third derivatives seem to have a "little" effect on the residual. In the context of fat content prediction, the first derivative seems to have a "little" effect on the residual but the effect of the third derivative seems stronger and significant. Finally, in the context of protein content prediction, no significant effect is detected on the residuals. We follow the same ideas as before and compute the mean squared error for the prediction of the residuals $R_{Moisture}$, R_{Fat}, and $R_{Protein}$. We get the results given in Table 10.

Table 10. —

Derivation order	0	1	3	4
$R_{Moisture}$	4.660818	**3.764095**	4.155877	4.261119
R_{Fat}	5.415786	4.509586	**3.995835**	4.976787
Derivation order	0	1	2	4
$R_{Protein}$	36.35755	14.01741	**4.868826**	11.51740

Remark : In the third case, the second derivative is the most adapted in terms of prediction of the residual $R_{Protein}$ (because it is linked to the smallest mean squared error). However, the results presented in Fig. 11.6 lead to the conclusion that its effect on the residual is not significant.

To end this study, we use similar notations as before and estimate the new residuals

$$R_{Moisture,2} = R_{Moisture} - m\hat{}_{M,2}\left(X'\right), \text{ and } R_{Fat,2} = R_{Fat} - m\hat{}_{F,2}\left(X'''\right),$$

and make no effect tests to see if other derivatives have an effect on these residuals. We proceed in the same way as before and obtain the following results presented in Fig 11.7.

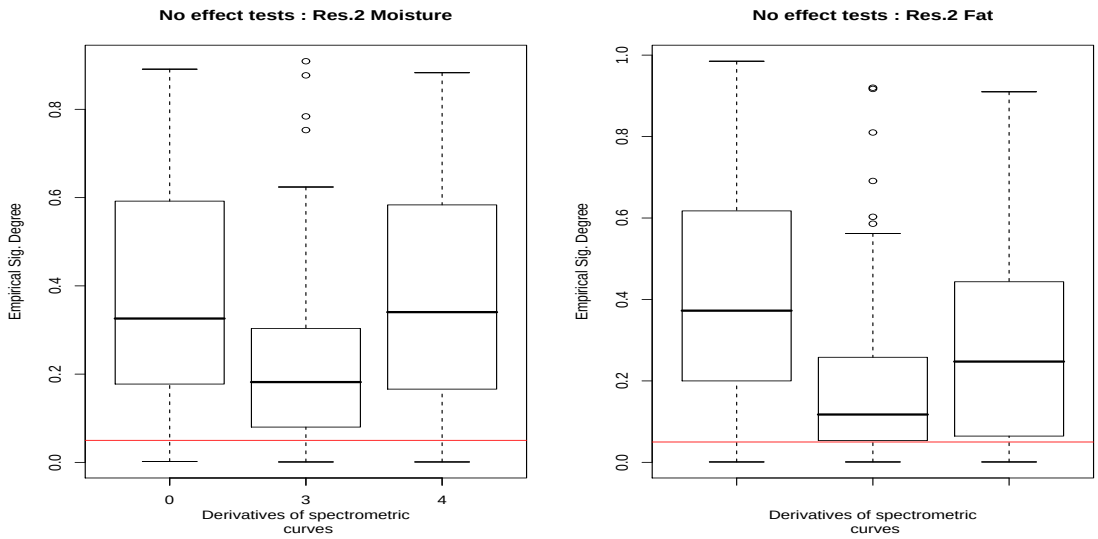

FIGURE 11.7. No effect tests on second estimated residuals

It seems that the remaining derivatives do not have a significant effect on the residuals $R_{Moisture,2}$, and $R_{Fat,2}$.

11.3.2. Corn dataset. — We now consider another dataset coming from the spectrometric analysis of 80 corn samples. For each corn sample, the dataset contains a spectrometric curve (the wavelength range is 1100-2498 nm at 2 nm intervals) and the moisture, oil, protein and starch values.

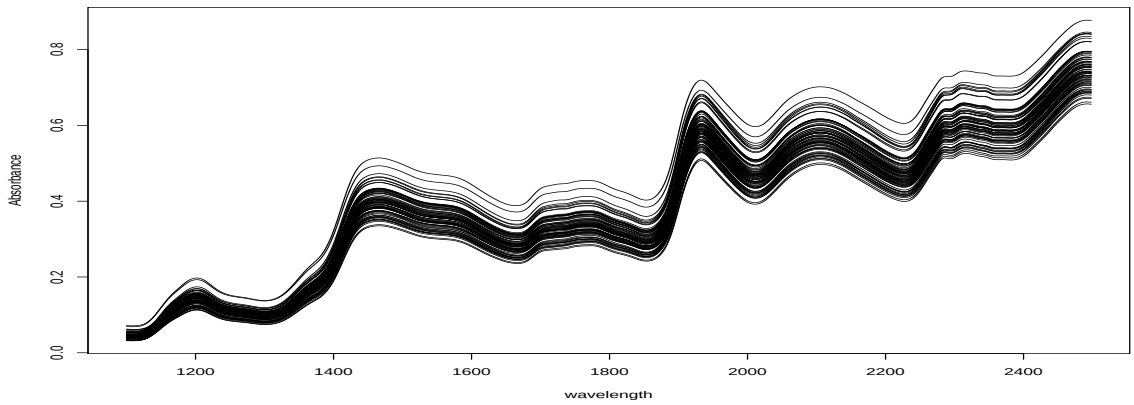

FIGURE 11.8. Sample of spectrometric curves

This dataset is available online on the webpage

$$http://software.eigenvector.com/Data/Corn/index.html$$

where more information about the data is given. This dataset has a fairly small size. However we have observed in simulation studies (not presented in this manuscript) that our testing procedures still have fairly good level and power properties even on small datasets. The study of the Corn dataset allow to present interesting ideas and methods that illustrate the fact that our tests are relevent and efficient even on small datasets. Of course, our results could be improved if we had to our disposal a more important dataset. We start with the same kind of analysis as the one we made on the Tecator dataset. We use the same method as the one presented before and randomly split our original dataset into three subdatasets of respective length $n = 40$, $m_n = 20$ and $l_n = 20$. We consider the issue of predicting moisture, oil, protein and starch content. Figure 11.5 presents the results we have obtained.

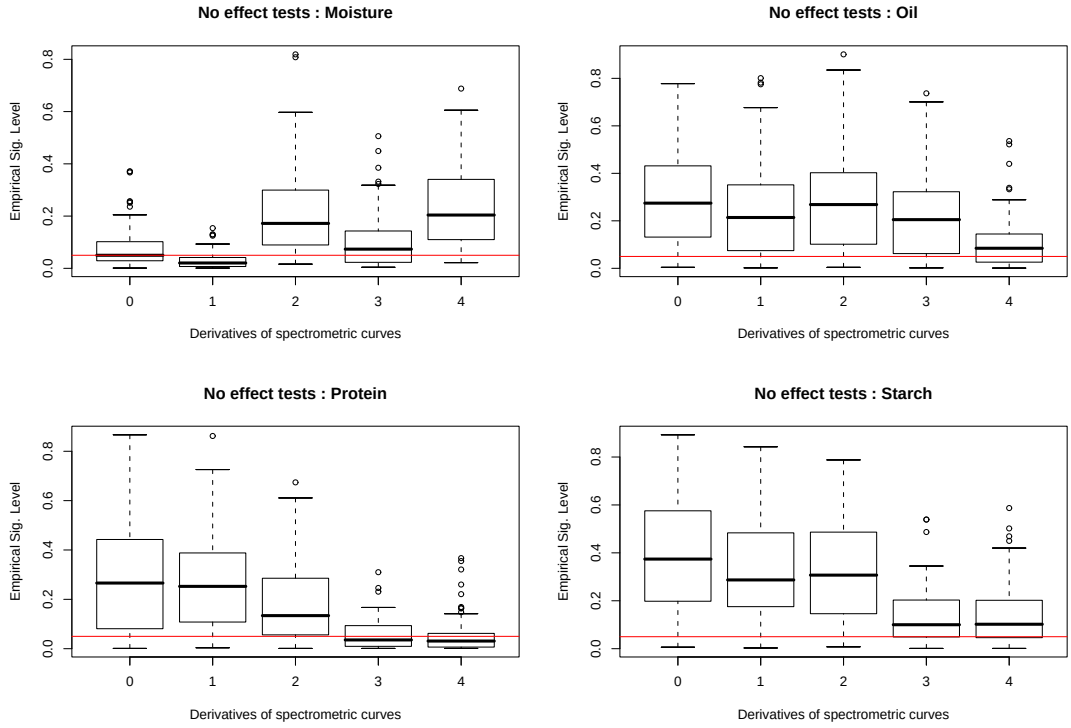

FIGURE 11.9. No effect tests in moisture, oil, protein and stach content prediction

No significant effect appears in the context of oil and starch content prediction. In the context of the moisture content prediction, the spectrometric curves seems to have a "little" effect while its first derivative has a more significant effect. In the context of protein content prediction it seems that the third and fourth derivative have a "little" effect. We now focus on the issue of predicting moisture and protein content. As in the previous section, we want to know which derivative is the best for the prediction. Hence, as in the previous section, we randomly split the dataset into a learning dataset of length 60 and a testing dataset of length 20 and we compute

the mean squared error. Table 11 presents the results obtained (average over 100 mean squared errors).

Table 11. —

Derivation order	0	1	2	3	4
Moisture	0.08916295	**0.06759695**	0.1182824	0.07248743	0.1114330
Protein	0.2214153	0.1989325	0.205501	**0.1281586**	0.1480457

As in the previous section, we are now interested in testing if the derivatives of spectrometric curves have an effect on the estimated residuals

$$R_{Moisture} = Y_{Moisture} - \hat{m_M}\left(X'\right), \text{ and } R_{Protein} = Y_{Protein} - \hat{m_P}\left(X'''\right).$$

The same testing procedures lead to the results presented on Figure 11.10. It seems that the derivatives have no significant additional effect on the residuals.

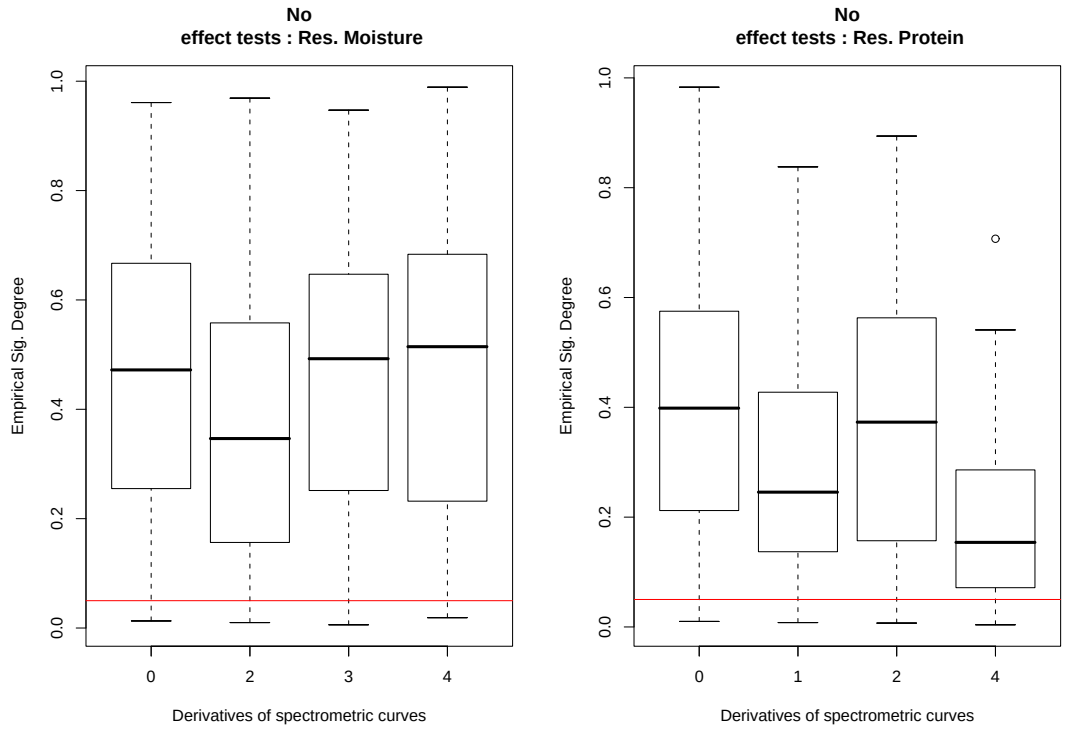

FIGURE 11.10. No effect tests on the residuals in the context of moisture and protein content prediction

To end this study, we propose to focus on the moisture content prediction. We split the whole spectrometric curve into seven consecutive curves (see Fig. 11.11). We have chosen a specific way to split the curve but other decompositions could be considered.

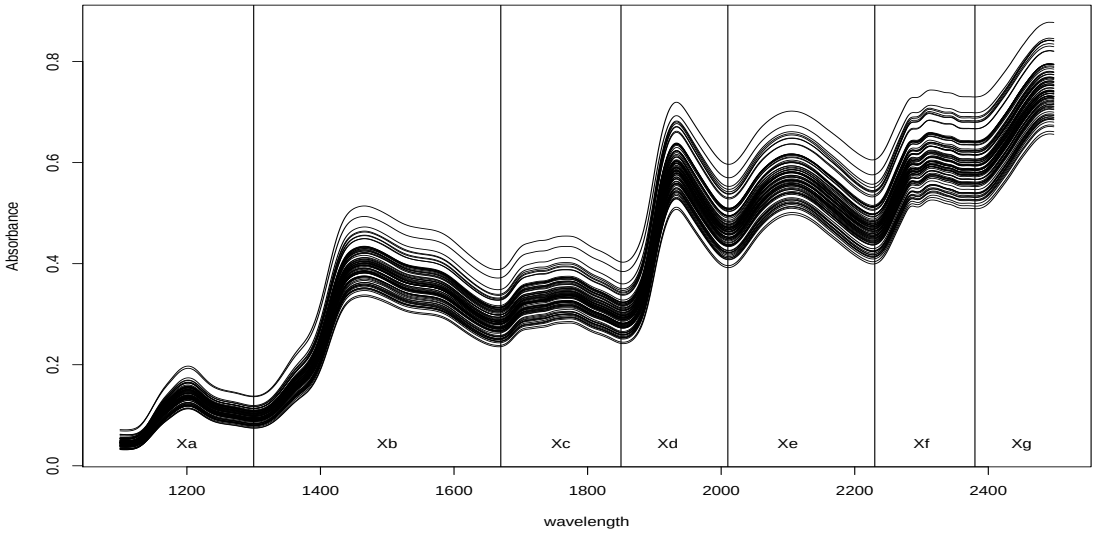

FIGURE 11.11. Spectrometric curves decomposition

In the previous study, we have noted that the first derivative of spectrometric curve has an effect on moisture prediction. We are now interested in finding which parts of this curve have an effect. We hence make no effect tests of each part $(X'_a, X'_b, X'_c, X'_d, X'_e, X'_f,$ and $X'_g)$ of the first derivative curve on the moisture content. We get the resulsts presented in Figure 11.12

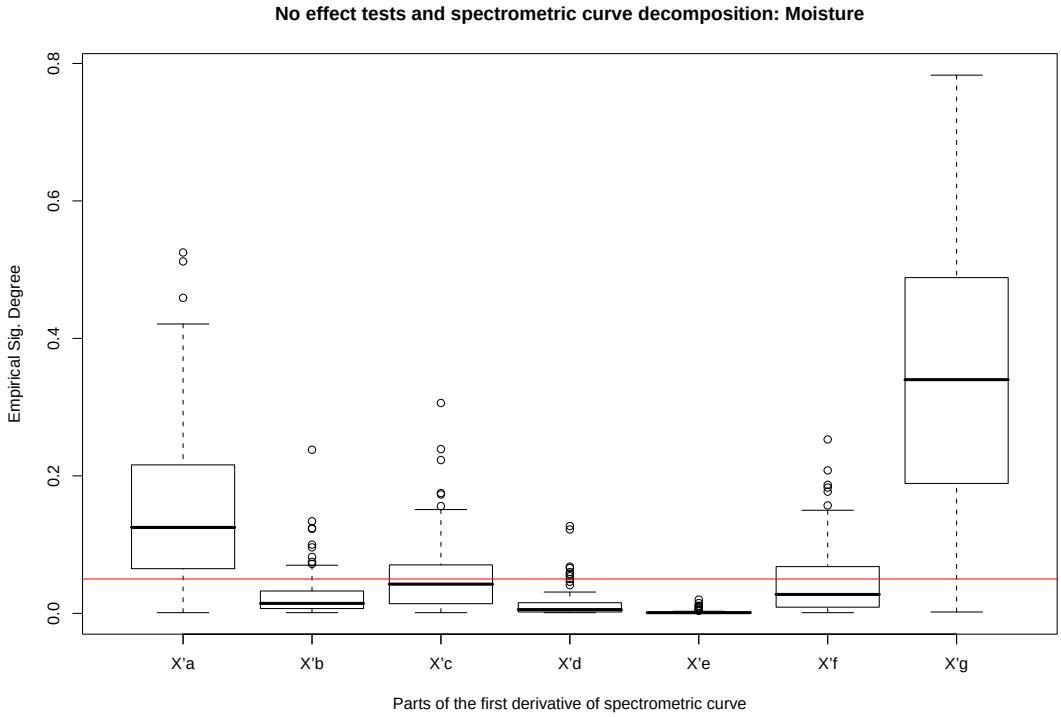

FIGURE 11.12. No effect tests and Spectrometric curves decomposition

We can read on Fig. 11.12 that the curves X'_b, X'_d and X'_e have a significant effect while the curves X'_c and X'_f seem to have a "little" effect. Morever the effect of the curve X'_e seems to be the strongest. This is confirmed by the mean squared errors presented in Table 12.

Table 12. —

Curves	X'_a	X'_b	X'_c	X'_d	X'_e	X'_f	X'_g
$Moisture$	0.10244	0.068818	0.082374	0.07395	**0.043999**	0.066727	0.11578

We have also studied the estimated residuals $R_{Moisture} = Y_{Moisture} - \hat{m_M}(X'_e)$. It seems that the curves $X'_a, X'_b, X'_c, X'_d, X'_f,$ and X'_g do not have a significant additional effect on moisture content (see Fig. 11.13).

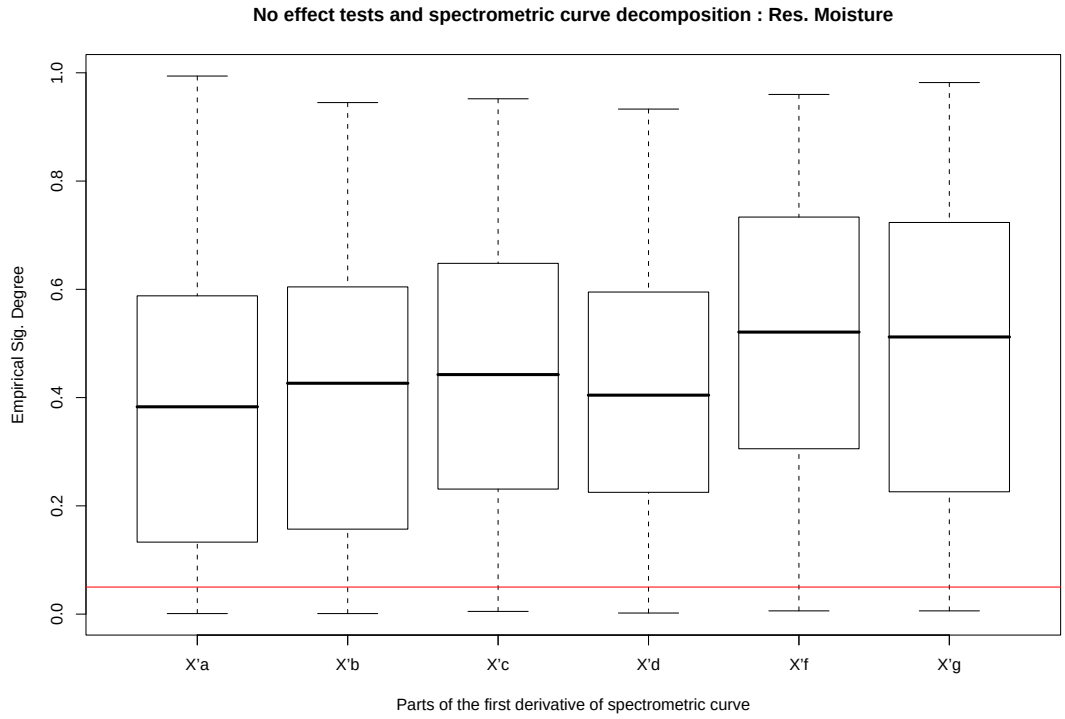

FIGURE 11.13. No effect tests and Spectrometric curves decomposition

In a more general context, one can imagine to develop an iterative testing procedure to detect which part of a spectrometric curve has an effect to predict the proportion of a specific substance. This would allow decreasing the necessary wavelength range we have to study to predict this proportion.

11.4. Conclusion

In this chapter we have considered the practical issues linked to the application of structural testing procedures introduced in Chapter 10. We have introduced various residual-based bootstrap methods to estimate the threshold of our testing procedures. We have considered in the remaining of the chapter the special case of no effect tests. Various bootstrap no effect tests were compared on two simulated datasets. We have obtained interesting level and power properties. We have also considered the issue of the choice of the smoothing parameter and the bootstrap iterations number. Finally, we have seen that wild bootstrap methods are adapted to study models with heteroscedastic errors. We have chosen two datasets to illustrate the efficiency and the interest of our testing procedures. The results obtained on the tecator dataset are relevent and corroborated by former studies of this "reference" dataset. Moreover, the last application illustrates that our testing procedure leads to interesting and efficient results even in the case of small datasets. This chapter constitutes the body of an applied article that will be submitted soon. We have made a deep study of our testing procedures in the context of no effect tests. It would be interesting to do similar studies for other strutural tests and more precisely for linearity tests. Simulations studies for linearity tests are in progress. The first comparative results (obtained on 1000 datasets with the smoothing parameter h_0 and $N_{boot} = 100$) are encouraging. They are presented in Table 13. We have used as specific estimator r_0^* the one proposed by Crambes *et al.* (2008) and considered the first simulated dataset presented in the case of no effect tests.

Table 13. —

	Res	SNB	WB1	WB2	WB3
$Y = \int_0^1 X(t)\, e^t dt + \epsilon$	0.056	0.052	0.069	0.049	0.058
$Y = 2\cos\left(3\int_0^1 X(t)\, e^t dt\right) + \epsilon$	0.667	0.594	0.706	0.645	0.700
$Y = 3\cos\left(3\int_0^1 X(t)\, e^t dt\right) + \epsilon$	0.884	0.824	0.879	0.875	0.7846

Both theoretical justification of bootstrap methods and practical study of more general structural testing procedures are important and interesting challenges for the future.

BIBLIOGRAPHIE

[1] Abraham, C., Biau, G. et Cadre, B. (2006) On the kernel rule for function classification. *Ann. Inst. Statist. Math.* **58** (3) 619-633.

[2] Abraham, C., Cornillon, P. A., Matzner-Løber, E. et Molinari, N. (2003) Unsupervised curve clustering using B-splines. *Scand. J. Statist.* **30** (3) 581-595.

[3] Abramovich, F. et Angelini, C. (2006) Bayesian maximum a posteriori multiple testing procedure. *Sankhyā* **68** (3) 436-460.

[4] Aguilera, A.M., Ocaña, F.A. et Valderrama, M.J. (1997) An approximated principal component prediction model for continuous-time stochastic processes. *Appl. Stochastic Models Data Anal.* **13** (2) 61-72.

[5] Aguilera, A.M., Ocaña, F.A. et Valderrama, M.J. (1999) Forecasting time series by functional PCA. Discussion of several weighted approaches. *Comput. Statist.* **14** (3) 443-467.

[6] Aguilera, A.M., Bouzas, P.R. et Ruiz-Fuentes, N. (2002) Functional Principal Component Modelling of the Intensity of a Doubly Stochastic Poisson Process *COMPSTAT 2002 (Berlin)* 373-376 Physica, Heidelberg.

[7] Ait Saïdi, A., Ferraty, F. et Kassa, R. (2005) Single functional index model for a time series. *Rev. Roumaine Math. Pures Appl.* **50** (4) 321-330.

[8] Ait-Saïdi, A., Ferraty, F., Kassa, R. et Vieu, P. (2008) Cross-validated estimations in the single functional index model. *soumis*

[9] Allam, A. et Mourid, T. (2002) Geometric absolute regularity of Banach space-valued autoregressive processes. *Statist. Probab. Lett.* **60** (3) 241-252.

[10] Aneiros-Perez, G., Cardot, H., Estevez-Perez, G. et Vieu, P. (2004) Maximum Ozone Concentration Forecasting by Functional Nonparametric Approaches. *Environmetrics* **15** 675-685.

[11] Aneiros-Pérez, G. et Vieu, P. (2006) Semi-functional partial linear regression. *Statist. Probab. Lett.* 76 (11) 1102-1110.

[12] Antoniadis, A. et Sapatinas, T. (2003) Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *J. Multivariate Anal.* **87** (1) 133-158.

[13] Antoniadis, A. et Sapatinas, T. (2007) Estimation and inference in functional mixed-effect models *Comp. Stat. & Data Anal.* **51** (10) 4793-4813.

[14] Antoniadis, A., Paparoditis, E. et Sapatinas, T. (2006) A functional wavelet-kernel approach for time series prediction. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** (5) 837-857.

[15] Araki, Y., Konishi, S. et Imoto, S. Functional discriminant analysis for microarray gene expression data via radial basis function networks. (English summary) *COMPSTAT 2004—Proceedings in Computational Statistics, Physica, Heidelberg* 613-620.

[16] Aspirot, L., Bertin, K., and Perera, G. (2008) Asymptotic normality of the Nadaraya-Watson estimator for non-stationary functional data and applications to telecommunications. (Submitted)

[17] Attouch, M., Laksaci, A. et Ould-Saïd, E. (2007) Strong uniform convergence rate of robust estimator of the regression function for functional and dependent processes. *Technical Report L.M.P.A.*.

[18] Aurzada, F. et Simon, T. (2007) Small ball probabilities for stable convolutions. *ESAIM Probab. Stat.* **11** 327-343 (electronic).

[19] Azzalini, A., Bowman, A. (1993) On the use of Nonparametric Regression for Checking Linear Relationships *J.R. Statist. Soc. B* **55,2** 549-557.

[20] Azzedine, N., Laksaci, A. and Ould-Saïd, E. (2006) On the robust nonparametric regression estimation for functional regressor. *Technical Report L.M.P.A.*

[21] Benhenni, K., Ferraty, F., Rachdi, M. et P. Vieu (2007). Local smoothing regression with functional data. *Comp. Statistics* **22** 353-369.

[22] Benko, M. (2006) Functional Data Analysis with Applications in Finance *Mémoire de Thèse.*

[23] Benko, M., Härdle, W. et Kneip, A. (2006) Common Functional Principal Components *SFB 649 Discussion Papers SFB649DP2006-010, Humboldt University, Berlin, Germany.*

[24] Beran, J. (1994) *Statistics for long-memory processes.* Monographs on Statistics and Applied Probability **61** Chapman and Hall, New York, x+315 pp.

[25] Besse, P.C. et Cardot, H. (1996) Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. (French. English, French summary) [Spline approximation of the prediction of a first-order autoregressive functional process] *Canad. J. Statist.* **24** (4) 467-487.

[26] Besse, P., Cardot, H. et Stephenson D. (2000) Autoregressive Forecasting of Some Functional Climatic Variations. *Scandinavian Journal of Statistics* **27** 673-687.

[27] Besse, P.C., Cardot, H., Faivre, R. et Goulard, M. (2005) Statistical modelling of functional data. *Appl. Stoch. Models Bus. Ind.* **21** (2) 165-173.

[28] Besse, P. et Ramsay, J.O. (1986) Principal components analysis of sampled functions. *Psychometrika* **51** (2) 285-311.

[29] Biau, G., Bunea, F. et Wegkamp, M.H. (2005) Functional classification in Hilbert spaces. *IEEE Trans. Inform. Theory* **51** (6) 2163-2172.

[30] Bigot, J., Gamboa, F. et Vimond, M. (2007) Estimation of translation, rotation and scaling between noisy images using the Fourier Mellin transform, soumis.

[31] Blanke, D. and Pumo, B. (2003) Optimal sampling for density estimation in continuous time. (English summary) *J. Time Ser. Anal.* **24** (1) 1-23.

[32] Boente, G., Fraiman, R. (2000) Kernel-based functional principal components. *Statist. Probab. Lett.* **48** (4) 335-345.

[33] Bogachev, V.I. (1998) *Gaussian measures. Mathematical Surveys and Monographs* **62** American Mathematical Society, Providence, RI. xii+433 pp.

[34] Borovkova, S., Burton, R. et Dehling, H. (2001) Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Trans. Amer. Math. Soc.* **353** (11) 4261-4318 (electronic).

[35] Borggaard, C. et Thodberg, H.H. Optimal minimal neural interpretation of spectra *Analytical chemistry* **64** (5) pp 545 - 551.

[36] Bosq, D. (1975) Inégalité de Bernstein pour les processus stationnaires et mélangeants. Applications. *C. R. Acad. Sci. Paris* **281** (24) 1095-1098.

[37] Bosq, D. (1989) Propriétés des opérateurs de covariance empiriques d'un processus stationnaire hilbertien. (French) [Properties of empirical covariance operators for a Hilbertian stationary process] *C. R. Acad. Sci. Paris Sér. I Math.* **309** (14) 873-875.

[38] Bosq, D. (1990) Modèle autorégressif hilbertien. Application à la prévision du comportement d'un processus à temps continu sur un intervalle de temps donné. (French) [Hilbertian autoregressive model : application to the prediction of the future behavior of a continuous-time process on a given time interval] *C. R. Acad. Sci. Paris Sér. I Math.* **310** (11) 787-790.

[39] Bosq, D. (1991) Modelization, nonparametric estimation and prediction for continuous time processes. In Nonparametric functional estimation and related topics (Spetses, 1990), 509-529, *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.* **335**, Kluwer Acad. Publ., Dordrecht.

[40] Bosq, D. (1993a) Propriétés asymptotiques des processus autorégressifs banachiques. Applications. (French. English, French summary) [Asymptotic properties of autoregressive processes in Banach spaces. Applications] *C. R. Acad. Sci. Paris Sér. I Math.* **316** (6) 607-610.

[41] Bosq, D. (1993b) Bernstein-type large deviations inequalities for partial sums of strong mixing processes. *Statistics* **24** (1) 59-70.

[42] Bosq, D. (1996a) Limit theorems for Banach-valued autoregressive processes. Applications to real continuous time processes. *Bull. Belg. Math. Soc. Simon Stevin* **3** (5) 537-555.

[43] Bosq, D. (1996b) *Nonparametric statistics for stochastic processes. Estimation and prediction.* Lecture Notes in Statistics **110** Springer-Verlag, New York, xii+169 pp.

[44] Bosq, D. (1998) *Nonparametric statistics for stochastic processes. Estimation and prediction.* Second edition. Lecture Notes in Statistics **110** Springer-Verlag, New York. xvi+210 pp.

[45] Bosq, D. (1999) Représentation autorégressive de l'opérateur de covariance empirique d'un ARH(1). Applications. (French) [Autoregressive representations for the empirical covariance operator of an ARH(1). Applications] *C. R. Acad. Sci. Paris Sér. I Math.* **329** (6) 531-534 .

[46] Bosq, D. (2000) *Linear Processes in Function Spaces : Theory and Applications* Lecture Notes in Statistics **149** Springer-Verlag, New York.

[47] Bosq, D. (2002) Estimation of mean and covariance operator of autoregressive processes in Banach spaces. *Stat. Inference Stoch. Process.* **5** (3) 287-306.

[48] Bosq, D. (2003a) Berry-Esseen inequality for linear processes in Hilbert spaces. *Statist. Probab. Lett.* **63** (3) 243-247.

[49] Bosq, D. (2003b) Processus linéaires vectoriels et prédiction. (French) [Linear functional processes and prediction] *C. R. Math. Acad. Sci. Paris* **337** (2) 115-118.

[50] Bosq, D. (2004) Moyennes mobiles hilbertiennes standard. (French) [Standard moving averages in Hilbert spaces] *Ann. I.S.U.P.* **48** (3) 17-28.

[51] Bosq, D. (2007) General linear processes in Hilbert spaces and prediction. *J. Statist. Plann. Inference* **137** (3) 879-894.

[52] Bosq, D. and Blanke, D. (2007) *Inference and prediction in large dimensions.* Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester ; Dunod, Paris. x+316 pp.

[53] Bosq, D. et Delecroix, M. (1985) Nonparametric prediction of a Hilbert-space valued random variable. *Stochastic Process. Appl.* **19** (2) 271-280.

[54] Bouzas, P.R., Aguilera, A.M., Valderrama, M.J. et Ruiz-Fuentes, N. (2006a) On the structure of the stochastic process of mortgages in Spain. *Comput. Statist.* **21** (1) 73-89.

[55] Bouzas, P.R., Valderrama, M.J., Aguilera, A.M. et Ruiz-Fuentes, N. (2006b) Modelling the mean of a doubly stochastic Poisson process by functional data analysis. *Comput. Statist. Data Anal.* **50** (10) 2655-2667.

[56] Bradley, R.C. (2005) Basic properties of strong mixing conditions. A survey and some open questions *Probability Surveys* **2** 107-144 (electronic).

[57] Breiman, L. et Friedman, J. (1985) Estimating optimal transformations for multiple regression and correlation. With discussion and with a reply by the authors. *J. Amer. Statist. Assoc.* **80** (391) 580-619.

[58] Brumback, B. et Rice, J.A (1998) Smoothing spline models for the analysis of nested and crossed samples of curves. (English summary) With comments and a rejoinder by the authors. *J. Amer. Statist. Assoc.* **93** (443) 961-994.

[59] Burba, F., Ferraty, F. et Vieu, P. (2008) Propriétés asymptotiques de l'estimateur kNN en régression fonctionnelle non-paramétrique. *C. R. Math. Acad. Sci.* Ser. I **346** 339-342.

[60] Cadre, B. (2001) Convergent estimators for the L_1-median of a Banach valued random variable. *Statistics* **35** (4) 509-521.

[61] Cai, T.T. et Hall, P. (2006) Prediction in functional linear regression. (English summary) *Ann. Statist.* **34** (5) 2159-2179.

[62] Cao, J. et Ramsay, J.O. (2007) Parameter cascades ans profiling in functional data analysis *Comp. Stat.* **22** (3) 335-351.

[63] Cao, R. (1991) Rate of convergencefor the wild bootstrap in nonparametric regression *Annals Statist.* **19** 2226-2231.

[64] Carbon, M. (1983) Inégalité de Bernstein pour les processus fortement mélangeants, non nécessairement stationnaires. Applications. (French. English summary) [Bernstein inequality for strong mixing, nonstationary processes. Applications] *C. R. Math. Acad. Sci. Paris* **297** (5) 303-306.

[65] Cardot, H. (1998) Convergence du lissage spline de la prévision des processus autorégressifs fonctionnels. (French) [Consistency of the spline approximation of the prediction of functional space valued autoregressive processes] *C. R. Acad. Sci. Paris Sér. I Math.* **326** (6) 755-758.

[66] Cardot, H. (2000) Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *J. Nonparametr. Statist.* **12** (4) 503-538.

[67] Cardot, H. (2007) Conditional functional principal components. analysis. *Scandinavian J. of Statistics* **34** (2) 317-335.

[68] Cardot, H., Crambes, C, et Sarda, P. (2004a) Estimation spline de quantiles conditionnels pour variables explicatives fonctionnelles. (French) [Spline estimation of conditional quantiles for functional covariates] *C. R. Math. Acad. Sci. Paris* **339** (2) 141-144.

[69] Cardot, H., Crambes, C. and Sarda, P. (2004b) Conditional Quantiles with Functional Covariates : an Application to Ozone Pollution Forecasting. Contributed paper in *Compstat Prague 2004 Proceedings* 769-776.

[70] Cardot, H., Crambes, C. and Sarda, P. (2005) Quantile regression when the covariates are functions. *J. Nonparametr. Stat.* **17** (7) 841-856.

[71] Cardot, H., Crambes, C. and Sarda, P. (2006) Ozone pollution forecasting using conditional mean and conditional quantiles with functional covariates. *Statistical methods for biostatistics and related fields, Härdle, Mori and Vieu (Eds.)*, Springer.

[72] Cardot, H., Crambes, C., Kneip, A. and Sarda, P (2007) Smoothing splines estimators in functional linear regression with errors-in-variables. *Computational Statistics and Data Analysis, special issue on functional data analysis* **51** (10) 4832-4848.

[73] Cardot, H., Faivre, R. et Goulard, M. (2003) Functional approach for predicting land use with the temporal evolution of coarse resolution remote sensing data *Journal of Applied Statistics* **30** (10) 1185 - 1199.

[74] Cardot, H., Ferraty, F., Mas, A. and Sarda, P. (2003) Testing Hypothesys in the Functional Linear Model *Scandinavian Journal of Statistics* **30** 241-255.

[75] Cardot, H., Ferraty, F. and Sarda, P. (1999) Functional Linear Model *Statist. and Prob. Letters* **45** 11-22.

[76] Cardot, H., Ferraty, F. et Sarda, P. (2000) Étude asymptotique d'un estimateur spline hybride pour le modèle linéaire fonctionnel. (French) [Asymptotic study of a hybrid spline estimator for the functional linear model] *C. R. Acad. Sci. Paris Sér. I Math.* **330** (6) 501-504.

[77] Cardot, H., Ferraty, F. and Sarda, P. (2003) Spline Estimators for the Functional Linear Model *Statistica Sinica* **13** (3) 571-591.

[78] Cardot, H., Goia, A. et Sarda, P. (2004) Testing for no effect in functional linear regression models, some computational approaches. *Comm. Statist. Simulation Comput.* **33** (1) 179-199.

[79] Cardot, H., Mas, A. et Sarda, P. (2007) CLT in functional linear regression models. (English summary) *Probab. Theory Related Fields* **138** (3-4) 325-361.

[80] Cardot, H., Prchal, L. et Sarda, P. (2007) Noeffect and lack-of-fit permutation tests for functional regression *Comp. Statist.* **22** (3) 371-390.

[81] Cardot, H. et Sarda, P. (2005) Estimation in generalized linear models for functional data via penalized likelihood. *J. Multivariate Anal.* **92** (1) 24-41.

[82] Cardot, H. et Sarda, P. (2006) Linear regression models for functional data. *The art of semiparametrics* 49-66, Contrib. Statist., Physica-Verlag/Springer, Heidelberg.

[83] Cérou, F. et Guyader, A. (2006) Nearest neighbor classification in infinite dimension. *ESAIM Probab. Stat. 10* 340-355 (electronic).

[84] Chanda, K.C. (1974) Strong mixing properties of linear stochastic processes. *J. Appl. Probability* **11** 401-408.

[85] Chao, C.J. (1994) Testing for no effect in nonparametric regression via spline smoothing technics *Ann. Inst. Statist. Math.* bf 46 (2) 251-265

[86] Chen, S.X., Hardle, W. and Li, M. (2003) An empirical likelihood goodness-of-fit test for time series *J. R. Statist. Soc.- Series B* **65** 663-678.

[87] Chen, S.X. and Van Keilegom I. (2006) A goodness-of-fit test for parametric and semiparametric models in muliresponse regression. *Inst de Statistique, U.C.L., Discussion paper* **0616**

[88] Chiou, J.M. and Müller H.-G. (2007) Diagnostics for functional regression via residual processes *Computational Statistics & Data Analysis* **51**, (10) 4849-4863.

[89] Chiou, J-M., Müller, H.-G. et Wang, J.-L. (2003) Functional quasi-likelihood regression models with smooth random effects. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65** (2) 405-423.

[90] Chiou, J.-M., Müller, H.-G. et Wang, J.-L. (2004) Functional response models. *Statist. Sinica* **14** (3) 675-693.

[91] Chiou, J-M., Müller, H.-G., Wang, J.-L. et Carey, J.R. (2003) A functional multiplicative effects model for longitudinal data, with application to reproductive histories of female medflies. *Statist. Sinica* **13** (4) 1119-1133.

[92] Clot, D. (2002) Using functional PCA for cardiac motion exploration *Proceedings of the. IEEE International Conference on Data Mining* 91-98.

[93] Collomb, G. (1976) *Estimation nonparamétrique de la régression* (in french). PhD Thesis, Université Paul Sabatier, Toulouse.

[94] Collomb, G. (1981) Estimation non-paramétrique de la régression : revue bibliographique. (French. English summary) *Internat. Statist. Rev.* **49** (1) 75-93.

[95] Collomb, G. (1984) Propriétés de convergence presque complète du prédicteur à noyau. (French) [Almost complete convergence properties of kernel predictors] *Z. Wahrsch. Verw. Gebiete* **66** (3) 441-460.

[96] Collomb, G. (1985) Nonparametric regression : an up-to-date bibliography. *Statistics* **16** (2) 309-324.

[97] Costanzo, D., Preda, C. et Saporta, G. (2006) Anticipated prediction in discriminant analysis on functional data for binary response . *COMPSTAT2006, 17th Symposium on Computational Statistics, Rome* 821-828, Physica-Verlag.

[98] Coulon-Prieur, C. et Doukhan, P. (2000) A triangular central limit theorem under a new weak dependence condition. *Statist. Probab. Lett.* **47** (1) 61-68.

[99] Crambes, C. (2005). Total Least Squares for Functional Data. *Invited paper in ASMDA Brest 2005 Proceedings* 619-626.

[100] Crambes, C. (2006) *Modèle de régression linéaire pour variables explicatives fonctionnelles*. PhD Thesis.

[101] Crambes, C. (2007) Régression fonctionnelle sur composantes principales pour variable explicative bruitée. *Comptes Rendus de l'Académie des Sciences* **345** (9) 519-522.

[102] Crambes, C., Kneip, A. and Sarda, P. (2007) Smoothing splines estimators for functional linear regression *submitted*.

[103] Cuesta-Albertos, J.A. et Fraiman, R. (2007) Impartial trimmed k-means for functional data. *Comp. Statist. & Data Anal.* **51** (10) 4864-4877.

[104] Cuevas, A., Febrero, M. and Fraiman, R. (2002) Linear function regression : the case of fixed design and functional response *Canad. J. of Statistics* **30** 285-300.

[105] Cuevas, A., Febrero, M. et Fraiman, R. (2004) An anova test for functional data. *Comput. Statist. Data Anal.* **47** (1) 111-122.

[106] Cuevas, A., Febrero, M. et Fraiman, R. (2006) On the use of the bootstrap for estimating functions with functional data. *Comput. Statist. & Data Anal.* **51** (2) 1063-1074.

[107] Cuevas, A., Febrero, M. et Fraiman, R. (2007) Robust estimation and classification for functional data via projection-based depth notions *Comput. Statist.* **22** (3) 481-496.

[108] Cuevas, A. et Fraiman, R. (2004) On the bootstrap methodology for functional data. (English summary) *COMPSTAT 2004—Proceedings in Computational Statistics* 127-135 Physica, Heidelberg.

[109] Dabo-Niang, S. (2002) Sur l'estimation de la densité dans un espace de dimension infinie : application aux diffusions. (in french) [On Density estimation in an infinite-dimensional space : application to diffusion processes] *PhD Paris VI*.

[110] Dabo-Niang, S. (2004a) Density estimation by orthogonal series in an infinite dimensional space : application to process diffusion type I. The International Conference on Recent Trends and Directions in Nonparametric Statistics. *J. Nonparametr. Stat.* **16** 171-186.

[111] Dabo-Niang, S. (2004b) Kernel density estimator in an infinite-dimensional space with a rate of convergence in the case of diffusion process. *Appl. Math. Lett.* **17** (4) 381-386.

[112] Dabo-Niang, S., Ferraty, F. et Vieu, P. (2004a) Estimation du mode dans un espace vectoriel semi-normé. (French) [Mode estimation in a semi-normed vector space] *C. R. Math. Acad. Sci. Paris* **339** (2004) **9** 659-662.

[113] Dabo-Niang, S., Ferraty, F. et Vieu, P. (2004b) Nonparametric unsupervised classification of satellite wave altimeter forms. In *Proceedings in Computational Statistics*, Ed. J. Antoch, Physica-Verlag, Heidelberg New-York, 879-886.

[114] Dabo-Niang, S., Ferraty, F. et Vieu, P. (2006) Mode estimation for functional random variable and its application for curves classification. *Far East J. Theor. Stat.* **18** (1) 93-119.

[115] Dabo-Niang, S., Ferraty, F. et Vieu, P. (2007) On the using of modal curves for radar wave curves classification. *Comp. Statist. & Data Anal.* **51** (10) 4878-4890.

[116] Dabo-Niang, S. and Laksaci, A. (2007) Estimation non paramétrique de mode conditionnel pour variable explicative fonctionnelle. (French) [Nonparametric estimation of the conditional mode when the regressor is functional] *C. R. Math. Acad. Sci. Paris* **344** (1) 49-52.

[117] Dabo-Niang, S. and Rhomari, N. (2003) Estimation non paramétrique de la régression avec variable explicative dans un espace métrique. (French. English, French summary) [Kernel regression estimation when the regressor takes values in metric space] *C. R. Math. Acad. Sci. Paris* **336** (1) 75–80.

[118] Damon, J. et Guillas, S. (2002) The inclusion of exogenous variables in functional autoregressive ozone forecasting *Environmetrics* **13** 759-774.

[119] Damon, J. et Guillas, S. (2005) Estimation and simulation of autoregressive Hilbertian processes with exogenous variables. *Stat. Inference Stoch. Process.* **8** (2) 185-204.

[120] Dauxois, J., Ferré, L. et Yao, A.-F. (2001) Un modèle semi-paramétrique pour variables aléatoires hilbertiennes. (French) [A semiparametric model for Hilbertian random variables] *C. R. Acad. Sci. Paris Sér. I Math.* **333** (10) 947-952.

[121] Dauxois, J. et Nkiet, G.M. (2002) Measures of association for Hilbertian subspaces and some applications. *J. Multivariate Anal.* **82** (2) 263-298.

[122] Dauxois, J., Nkiet, G.M. and Romain, Y. (2001) Projecteurs orthogonaux, opérateurs associés et Statistique multidimensionnelle (in french) *Annales de l'ISUP* **45** (1) 31-54.

[123] Dauxois, J., Nkiet, G.M. et Romain, Y. (2004) Canonical analysis relative to a closed subspace. *Linear Algebra Appl.* **388** 119-145.

[124] Dauxois, J. et Pousse, A. (1975) Une extension de l'analyse canonique. Quelques applications. *Ann. Inst. H. Poincaré Sect. B (N.S.)* **11** (4) 355-379.

[125] Dauxois, J., Pousse, , A. and Romain, Y. (1982) Asymptotic theory for the principal component analysis of a vector random function : some applications to statistical inference *J. Multivariate Anal.* **12** (1) 136-154.

[126] Davidian, M., Lin, X. et Wang, J.-L. (2004) Introduction [Emerging issues in longitudinal and functional data analysis] *Statistica Sinica* **14** (3) 613-614.

[127] Davydov, Y.A. (1973) Mixing conditions for Markov chains. *Theory Probab. Appl.* **18** 312-328.

[128] Dedecker, J., Doukhan, P., Lang, G., León, R., José R., Louhichi, S. et Prieur, C. (2007) *Weak dependence : with examples and applications.* Lecture Notes in Statistics **190** Springer, New York. xiv+318 pp.

[129] Dedecker, J. et Prieur, C. (2005) New dependence coefficients. Examples and applications to statistics. *Probab. Theory Related Fields* **132** (2) 203-236.

[130] Delicado, P. Functional k-sample problem when data are density functions *Comp. Stat.* **22** (3) 391-410.

[131] Delsol, L. (2007a) CLT and $\mathbb{L}^q$ errors in nonparametric functional regression *C. R. Math. Acad. Sci.* **345** (7) 411-414.

[132] Delsol, L. (2007b) Régression non-paramétrique fonctionnelle : Expressions asymptotiques des moments *Annales de l'I.S.U.P.* **LI** (3) 43-67.

[133] Delsol, L. (2008a) Tests de structure en régression sur variable fonctionnelle. *C. R. Math. Acad. Sci. Paris* **346** (5-6) 343-346.

[134] Delsol, L. (2008b) Advances on asymptotic normality in nonparametric functional Time Series Analysis *Statistics* (à paraître)

[135] Delsol, L., Ferraty, F., Vieu, P. (2008) Structural test in regression on functional variables (submitted)

[136] Dereich, S. (2003) *High resolution coding of stochastic processes and small ball probabilities.* PhD Thesis.

[137] Deville, J. (1974) Méthodes statistiques et numériques de l'analyse harmonique. *Annales de l'INSEE* **15** 3-97.

[138] Diaconis, P. et Shashahani, M. (1984) On nonlinear functions of linear combinations. *SIAM, J. Sci. Statist. Comput.* **5** 175-191.

[139] Doob, J.L. (1953) *Stochastic Processes* John Wiley and Sons, Inc New York

[140] Doukhan, P. et Louhichi, S. (1999) A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* **84** (2) 313-342.

[141] Doukhan, P., Massart, P. et Rio, E. (1994) The functional central limit theorem for strongly mixing processes. *Ann. Inst. H. Poincaré Probab. Statist.* **30** (1) 63-82.

[142] Doukhan, P. et Neumann, M.H. (2007) Probability and moment inequalities for sums of weakly dependent random variables, with applications. *Stochastic Process. Appl.* **117** (7) 878-903.

[143] Dryden, I.L. et Mardia, K.V. (1998) *Statistical shape analysis.* (English summary) Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons, Ltd., Chichester.

[144] (1979) Bootstrap Methods : Another Look at the Jackknife. *Annals Statist.* **7** (1) 1-26.

[145] Eilers, P.H.C. et Marx, B.D. (1996) Flexible smoothing with B-splines and penalties. With comments and a rejoinder by the authors. *Statist. Sci.* **11** (2) 89–121.

[146] Escabias, M., Aguilera, A. M. et Valderrama, M. J. (2004) Principal component estimation of functional logistic regression : discussion of two different approaches. *J. Nonparametr. Stat.* **16** (3-4) 365-384.

[147] Escabias, M., Aguilera, A.M. et Valderrama, M.J. (2005) Modeling environmental data by functional principal component logistic regression. *Environmetrics* **16** (1) 95-107.

[148] Eubank, R.L.(2000) Testing for No Effect by Cosine Series Methods *Scandinavian Journal of Statistics* **27**, (4) 747-763.

[149] Eubank, R. L., Huang, C., Muñoz Maldonado, Y., Wang, N., Wang, S. et Buchanan, R. J. (2004) Smoothing spline estimation in varying-coefficient models. (English summary) *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66** (3) 653-667.

[150] Eubank, R.L. and Spiegelmann, C.H. (1990) Testing the goodness of fit of a linear model via nonparametric techniques *J. Amer. Statist. Assoc.***85** 387-392

[151] Eubank, R.L. and Hart, J.D. (1992) Testing goodness of fit in regression via order selection criteria *Ann. Statist* **20** 1412-1425

[152] Eubank, R.L. and Hart, J.D. (1993) Commonality of cosum, von Neuman and smoothing-based goodness-of-fit tests *Biometrika* **80** (1) 89-98

[153] Eubank, R.L. and LaRiccia V.N. (1993) Testing for no effect in non-parametric regression *Journal of Statistical planning and inference* **36**, (1) 1-14.

[154] Ezzahrioui, M. (2007) Prévision dans les modèles conditionnels en dimension infinie. (French) *PhD Thesis*.

[155] Fan, Y. et Li, Q. (1999) Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. *J. Nonparametr. Statist.* **10** (3) 245-271.

[156] Fan, J., Zhang, C. and Zhang, J. (2001) Generalized likelihood ratio statistics and Wilks phenomenon *Ann. Statist.* **29** 153-193.

[157] Fan, J. and Yao, Q. (2003) *Nonlinear Time Series : Nonparametric and Parametric methods* Springer, New York.

[158] Fan, J. et Zhang, J.-T. (2000) Two-step estimation of functional linear models with applications to longitudinal data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **62** (2) 303-322.

[159] Fan, J. and Zhang, J. (2004) Sieve empirical likelihood ratio tests for nonparametric functions *Ann. Statist.* **32** 1858-1907.

[160] Faraway, J. (1997) Regression analysis for a functional response. (English summary) *Technometrics* **39** (3) 254-261.

[161] Febrero, M., Galeano, P. et González-Manteiga, W. (2007) A functional analysis of NOx levels : location and scale estimation and outlier detection *Comp. Statist.* **22** (3) 411-428.

[162] Fernández de Castro, B., Guillas, S., and González Manteiga, W. (2005) Functional Samples and Bootstrap for Predicting Sulfur Dioxide Levels *Technometrics* **47** (2) 212-222.

[163] Ferraty, F., Goia, A. et Vieu, P. (2002a) Régression non-paramétrique pour des variables aléatoires fonctionnelles mélangeantes. (French) [Nonparametric regression for mixing functional random variables] *C. R. Math. Acad. Sci. Paris* **334** (3) 217-220.

[164] Ferraty F., Goia A. and Vieu P. (2002b) Functional nonparametric model for time series : a fractal approach for dimension reduction. *Test*, **11**, (2) 317-344

[165] Ferraty, F., Laksaci, A. et Vieu, P. (2005) Functional time series prediction via conditional mode estimation. *C. R. Math. Acad. Sci. Paris* **340** (5) 389-392.

[166] Ferraty, F., Laksaci, A. and Vieu, P. (2006) Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat. Inference Stoch. Process* **9** (1) 47-76.

[167] Ferraty, F., Mas, A. and Vieu, P. (2007) Advances on nonparametric regression for fonctionnal data. *ANZ Journal of Statistics* **49** 267-286.

[168] Ferraty, F., Peuch, A. et Vieu, P. (2003) Modèle à indice fonctionnel simple. (French) [Single functional index model] *C. R. Math. Acad. Sci. Paris* **336** (12) 1025-1028.

[169] Ferraty, F., Rabhi, A. et Vieu, P. (2005) Conditional quantiles for dependent functional data with application to the climatic El Niño phenomenon. *Sankhyā* **67** (2) 378-398.

[170] Ferraty, F., Rabhi, A. et Vieu, P. (2008) Estimation non-paramétrique de la fonction de hasard avec variable explicative fonctionnelle. *Rom. J. Pure & Applied Math.* (to appear).

[171] Ferraty, F., Van Keilegom, I., and Vieu, P. (2008b) On the validity of the bootstrap in nonparametric functionl regression (submited).

[172] Ferraty, F. and Vieu, P. (2000) Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés *Compte Rendus de l'Académie des Sciences* Paris, **330**, 403-406.

[173] Ferraty, F. et Vieu, P. (2002) The functional nonparametric model and application to spectrometric data. *Comput. Statist.* **17** (4) 545-564.

[174] Ferraty, F. and Vieu, P. (2003) Curves discrimination : a nonparametric functional approach *Computational Statistics and Data Analysis* **44** (1-2) 161-173.

[175] Ferraty, F. et Vieu, P. (2004) Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination. The International Conference on Recent Trends and Directions in Nonparametric Statistics. *J. Nonparametr. Stat.* **16** (1-2) 111-125.

[176] Ferraty F. and Vieu P. (2006a) *Nonparametric modelling for functional data.* Springer-Verlag, New York.

[177] Ferraty, F. et Vieu, P. (2006b) Functional nonparametric statistics in action. *The art of semiparametrics* 112-129, Contrib. Statist., Physica-Verlag/Springer, Heidelberg.

[178] Ferraty, F, Vieu, P. et Viguier-Pla, S. (2007) Factor-based comparison of groups of curves *Comp. Stat. & Data Anal.* **51** (10) 4903-4910.

[179] Ferré, L. et Villa, N. (2005) Discrimination de courbes par régression inverse fonctionnelle *RSA* **LIII** (1)

[180] Ferré, L. et Villa, N. (2006) Multi-Layer perceptron with functional inputs : an inverse regression approach. *Scandinavian Journal of Statistics* **33** (4).

[181] Ferré, L. et Yao, A. F. Functional sliced inverse regression analysis. *Statistics* **37** (6) 475-488.

[182] Ferré, L. ; Yao, A.-F. (2005) Smoothed functional inverse regression. *Statist. Sinica* **15** (3) 665-683.

[183] Fraiman, R. et Muniz, G. (2001) Trimmed means for functional data. (English summary) *Test* **10** (2) 419-440.

[184] Frank, I.E. et Friedman, J.H. (1993). A statistical view of some chemometrics regression tools. *Technometrics* **35** 109-135.

[185] Gamboa, F., Loubes, J.-M. et Maza, E. (2007) Semi-parametric estimation of shifts *Electronic Journal of Statistics* **1** 616-640 (electronic).

[186] Gadiaga, D. and Ignaccolo, R.(2005) Test of no-effect hypothesis by nonparametric regression. *Afr. Stat.* **1** (1) 67-76.

[187] Gao, F. et Li, W.V. (2007) Small ball probabilities for the Slepian Gaussian fields. *Trans. Amer. Math. Soc.* **359** (3) 1339-1350 (electronic).

[188] Gasser, T., Hall, P. et Presnell, B. (1998) Nonparametric estimation of the mode of a distribution of random curves. *J. R. Statomptes. Soc. Ser. B Stat. Methodol.* **60** (4) 681-691.

[189] Gasser, T. et Kneip, A. (1995) Searching for structure in curve samples. *Journal of the American Statistical Association* **90** 1179-1188.

[190] Gasser, T. et Müller, H.-G. (1979) Kernel estimation of regression functions. Smoothing techniques for curve estimation (Proc. Workshop, Heidelberg, 1979) 23-68 *Lecture Notes in Math.* **757** Springer, Berlin.

[191] Gasser, T., Müller, H.-G. et Mammitzsch, V. (1985) Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B* **47** (2) 238-252.

[192] Gervini, D. (2006) Free-knot spline smoothing for functional data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** (4) 671-687.

[193] Glasbey, C.A. et Mardia, K.V. (1998) A review of image warping methods. *J. Appl. Statistics* **25** (2) 155-171.

[194] González-Manteiga, W., Martínez Miranda, M.D., and Pérez González, A. (2004) The choice of smoothing parameter in nonparametric regression through Wild Bootstrap *Comp. Stat. and data Analysis* **47** 487-515.

[195] González-Manteiga, W., Quintela-del-Río, A. and Vieu, P. (2002) A note on variable selection in nonparametric regression with dependent data *Statistics and Probability Letters* **57** 259-268.

[196] González-Manteiga W. and Vieu P., 2007, [Editorial] Statistics for Functional Data. *Comp. Stat. Data Analysis* **51**, (10), 4788-4792.

[197] Gou, Z. et Fyfe, C. (2004) A canonical correlation neural network for multicollinearity and functional data. *Neural Networks* **17** (2) 285-293.

[198] Gozalo, P.L. (1993) A consistent model specification test for nonparametric estimation of regression function models *Econometric Theory* **9** (3) 451-477.

[199] Grenander, U. (1981) *Abstract inference.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York. ix+526 pp.

[200] Guillas, S. (2000) Non-causalité et discrétisation fonctionnelle, théorèmes limites pour un processus ARHX(1). (French) [Noncausality and functional discretization ; limit theorems for an ARHX(1) process] *C. R. Acad. Sci. Paris Sér. I Math.* **331** (1) 91-94.

[201] Guillas, S. (2001) Rates of convergence of autocorrelation estimates for autoregressive Hilbertian processes. (English summary) *Statist. Probab. Lett.* **55** (3) 281-291.

[202] Guillas, S. (2002) Doubly stochastic Hilbertian processes. *J. Appl. Probab.* **39** (3) 566-580.

[203] Guo, W. (2002) Functional mixed effects models. (English, French summary) *Biometrics* **58** (1) 121-128.

[204] Györfi, L., Härdle, W., Sarda, P. et Vieu, P. (1989) Nonparametric curve estimation from time series. *Lecture Notes in Statistics* **60** Springer-Verlag, Berlin. viii+153 pp.

[205] Hall, P.(1984) Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators *Academic Press Inc* **0047-259X/84**.

[206] Hall, P. (1990) Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *J. Multivariate Anal.* **32** 177-203.

[207] Hall, P. (1992) On bootstrap confidence intervals in nonparametric regression. *Ann. Statist.* **20** 695-711.

[208] Hall, P. and Hart, J. (1990) Bootstrap test for differene between means in nonparametric regression *J. Amer. Statist. Assoc.* **85** 1039-1049.

[209] Hall, P., Lahiri, S.N., Polzehl, J. (2002) On bandwidth choice in nonparametric regression with short- and long-range dependent errors. *Annals of Statistics*, **23**, (6) 1921-1936

[210] Hall, P. et Cai, T.T. (2006) Prediction in functional linear regression. (English summary) *Ann. Statist.* **34** (5) 2159-2179.

[211] Hall, P. et Heckman, N.E. (2002) Estimating and depicting the structure of a distribution of random functions. *Biometrika* **89** (1) 145-158.

[212] Hall, P. et Horowitz, J.L. (2007) Methodology and convergence rates for functional linear regression. (English summary) *Ann. Statist.* **35** (1) 70-91.

[213] Hall, P., Müller, H.G. et Wang, J.L. (2006) Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34** (3) 1493-1517.

[214] Hall, P., Poskitt, P. and Presnell, D. (2001) A functional data-analytic approach to signal discrimination *Technometrics* **43** 1-9

[215] Hall, P. and Hosseini-Nasab, M. (2006) On properties of functional principal components analysis *J.R. Stat. Soc. B* **68** (1) 109-126.

[216] Hall, P. et Vial, C. (2006a) Assessing extrema of empirical principal component functions. *Ann. Statist.* **34** (3) 1518-1544.

[217] Hall, P. et Vial, C. (2006b) Assessing the finite dimensionality of functional data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** (4) 689-705.

[218] Härdle, W. (1990) *Applied nonparametric regression. Econometric Society Monographs* **19**. Cambridge University Press, Cambridge. xvi+333 pp.

[219] Härdle, W. and Kneip, A. (1999) Testing a Regression Model When We Have Smooth Alternatives in Mind *Scandinavian Journal of Statistics* **26** 221-238

[220] Härdle, W. and Mammen, E. (1993) Comparing Nonparametric Versus Parametric Regression Fits *The Annals of Statistics* **21**, (4) 1926-1947.

[221] Härdle, W., Mammen, E. and Proenca, I. (2005) A Bootstrap Test for Single Index Models *Econometrics* **0508007** EconWPA.

[222] Härdle, W. et Marron, J.S. (1990) Semiparametric comparison of regression curves. *Ann. Statist.* **18** (1) 63-89.

[223] Härdle, W. and Marron, J.S. (1991) Bootstrap Simultaneous Error Bars for Nonparametric Regression *The Annals of Statistics* **19** (2) 778-796.

[224] Harel, M. et Elharfaoui, E. (2003) La convergence faible des U-statistiques multivariées pour des processus non stationnaires. (French) [Weak convergence of multivariate U-statistics for nonstationary processes] *C. R. Math. Acad. Sci. Paris* **337** (12) 801-804.

[225] Harel, M. et Puri, M.L. (1990) Weak invariance of generalized U-statistics for nonstationary absolutely regular processes. *Stochastic Process. Appl.* **34** (2) 341-360.

[226] Harezlak, J., Coull, B.A., Laird, N.M., Magari, S.R. et Christiani, D.C. Penalized solutions to functional regression problems *Comp. Stat. & Data. Anal.* **51** (10) 4911-4925.

[227] Hart, J. (1997) *Nonparametric Smoothing and Lack-of-fit Tests* Springer, New York

[228] Hastie, T., Buja, A. et Friedman, J. (2001) *The elements of statistical Learning. Data mining, inference, and prediction.* Springer-Verlag, New-York.

[229] Hastie, T., Buja, A. et Tibshirani, R. (1995) Penalized discriminant analysis. *Ann. Statist.* **13** 435-475.

[230] Hastie, T. et Mallows, C. (1993) Discussion of "A statistical view of some chemometrics regression tools." by Frank, I.E. and Friedman, J.H. *Technometrics* **35** 140-143.

[231] Hastie, T. et Tibshirani, R. (1986) Generalized additive models. With discussion. *Statist. Sci.* **1** (3) 297-318.

[232] Hastie, T. et Tibshirani, R. (1990) *Generalized additive models.* Monographs on Statistics and Applied Probability **43** Chapman and Hall, Ltd., London.

[233] Hastie, T. et Tibshirani, R. (1993) Varying-coefficient models. (English summary) With discussion and a reply by the authors. *J. Roy. Statist. Soc. Ser. B* **55** (4) 757-796.

[234] He, G., Müller, H.G. et Wang, J.L. (2000). Extending correlation and regression from multivariate to functional data. *Asymptotics in Statistics and Probability*, Ed. Puri, M.L., VSP International Science Publishers, 301-315.

[235] He, G., Müller, H.-G. et Wang, J.-L. (2003) Functional canonical analysis for square integrable stochastic processes. *J. Multivariate Anal.* **85** (1) 54-77.

[236] He, G., Müller, H.-G. et Wang, J.-L. (2004) Methods of canonical analysis for functional data. (English summary) Contemporary data analysis : theory and methods. *J. Statist. Plann. Inference* **122** (1-2) 141-159.

[237] Heckman, N.E. et Zamar, R.H. (2000) Comparing the shapes of regression functions. (English summary) *Biometrika* **87** (1) 135-144.

[238] Hedli-Griche, S. (2008) Estimation de l'opérateur de régression pour des données fonctionnelles et des erreurs corrélées. *PhD Thesis.*

[239] Helland, I.S. (1990) Partial least squares regression and statistical models. *Scandinavian Journal of Statistics* **17** 97-114.

[240] Henderson, B. (2006) Exploring between site differences in water quality trends : a functional data analysis approach. *Environmetrics* **17** 65-80.

[241] Hitchcock, D.B., Casella, G. et Booth, J.G. (2006) Improved estimation of dissimilarities by presmoothing functional data. *J. Amer. Statist. Assoc.* **101** (473) 211-222.

[242] Hjellvik, V. and Tjøstheim, D. (1995) Nonparametric tests of linearity for time series. *Biometrika* **82** 351-368

[243] Hjellvik, V. and Tjøstheim, D. (1996) Nonparametric statistics for testing linearity and serial independence. *J. Nonparametr. Statist.* **6** (2-3) 223-251

[244] Hjellvik, V., Yao, Q. and Tjøstheim, D. (1998) Linearity testing using local polynomial approximation. *J. Statist. Plann. Infer.* **68** 295-321.

[245] Hlubinka, D. et Prchal, L. (2007) Changes in atmospheric radiation from the statistical point of view. *Comp. Stat. & Data Anal.* **51** (10) 4926-4941.

[246] Hoeffding, W. (1963) Probability inequalities for sums of bounded random variables. *Journal of American Statistic Association* **58**, 15-30.

[247] Hoerl, A.E. et Kennard, R.W. (1980) Ridge regression : advances, algorithms and applications. *American Journal of Mathematical Management ScWatson, Geoffrey S. Smooth regression analysis. Sankhyā Ser. A 26 1964 359–372iences* **1** 5-83.

[248] Hoover, D.R., Rice, J.A., Wu, C.O. et Yang, L.-P. (1998) Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85** (4) 809-822.

[249] Horowitz, J.L. and Spokoiny, V.G. (2001) An adaptative, rate-optimal test of a parametric mean-regression model against a nonparametric alternative *Econometrica* **69** 599-632.

[250] Hyndman, R.J. et Ullah, S. (2007) Robust forecasting of mortality and fertility rates : A functional data approach *Comp. Stat. & Data Anal.* **51** (10) 4942-4957.

[251] Hyvärinen, A., Karhunen, J. et Oja, E. (2001) *Independent Component Analysis.* Wiley Interscience.

[252] Ibragimov, I.A. (1962) Some limit theorems for stationary processes *Theor. Prob. Appl* **7** 349-382.

[253] Jacob, P. et Oliveira, P. E. (1995) A general approach to nonparametric histogram estimation. *Statistics* **27** (1-2)73-92.

[254] James, G.M. (2002) Generalized linear models with functional predictors. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** (3) 411-432.

[255] James, G.M. et Hastie, T.J. (2001) Functional linear discriminant analysis for irregularly sampled curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63** (3) 533-550.

[256] James, G.M., Hastie, T.J. and Sugar, C.A. (2000) Principal component models for sparse functional data *Biometrika* **87** 587-602.Watson, Geoffrey S. Smooth regression analysis. Sankhyā Ser. A 26 1964 359–372

[257] James, G.M. et Silverman, B.W. (2005) Functional adaptive model estimation. (English summary) *J. Amer. Statist. Assoc.* **100** (470) 565-576.

[258] James, G.M. et Sood, A. (2006) Performing hypothesis tests on the shape of functional data. (English summary) *Comp. Stat. & Data Anal.* **50** (7) 1774-1792.

[259] James, G.M. et Sugar, C.A. (2003) Clustering for sparsely sampled functional data. *J. Amer. Statist. Assoc.* **98** (462) 397-408.

[260] Jank W. et Shmueli, G. (2006) Functional1 Data Analysis in Electronic Commerce Research *Statistical Science* **21** (2) 155-166.

[261] Kashimov, S. (1993) The central limit theorem for generalized U-statistics for weakly dependent vectors. *Theory Proba. Appl.* **38** 456-469.

[262] Kawasaki, Y. et Ando, T. (2004) Functional data analysis of the dynamics of yield curves. *COMPSTAT 2004—Proceedings in Computational Statistics* 1309–1316, Physica, Heidelberg.

[263] Kemperman, J.H.B. (1987) The median of a finite measure on a Banach space. *Statistical data analysis based on the L_1-norm and related methods* (Neuchâtel, 1987), 217–230, North-Holland, Amsterdam.

[264] Kim, T.Y. (1993) A note on moment bounds for strong mixing sequences*Statistics and Probability Letters*, **16 N°2**, 163-168.

[265] Cox, D.D. and Kim, T.Y. (1995) Moment bounds for mixing random variables useful in nonparametric function estimation *Stochastic Processes and their Applications*, **56**, 151-158.

[266] Klokov, S.A. et Veretennikov, A.Y. (2004) Sub-exponential mixing rate for a class of Markov chains. *Math. Commun.* **9** (1) 9-26.

[267] Kneip, A. et Engel, J. (1995) Model estimation in nonlinear regression under shape invariance. *Ann. Statist.* **23** (2) 551-570.

[268] Kneip, A. et Gasser, T. (1988) Convergence and consistency results for self-modeling nonlinear regression. *Ann. Statist.* **16** (1) 82-112.

[269] Kneip, A. et Gasser, T. (1992) Statistical tools to analyze data representing a sample of curves. *Ann. Statist.* **20** (3) 1266-1305.

[270] Kneip, A., Li, X., MacGibbon, K.B. et Ramsay, J.O. (2000) Curve registration by local regression. (English, French summary) *Canad. J. Statist.* **28** (1) 19-29.

[271] Kneip, A., Sickles, R.C. et Song, W. (2004) Functional data analysis and mixed effect models. *COMPSTAT 2004—Proceedings in Computational Statistics* 315-326, Physica, Heidelberg.

[272] Kneip, A. et Utikal, K.J. (2001) Inference for density families using functional principal component analysis. With comments and a rejoinder by the authors. *J. Amer. Statist. Assoc.* **96** (454) 519-542.

[273] Kulasekara, K.B. (2001) Variable selection by stepwise slicing in nonparametric regression. *Statist. Probab. Lett.* **51** (4) 327-336.

[274] Laksaci, A. (2007) Erreur quadratique de l'estimateur à noyau de la densité conditionnelle à variable explicative fonctionnelle. (French) [Quadratic error of the kernel estimator of conditional density when the regressor is functional] *C. R. Math. Acad. Sci. Paris* **345** (3) 171-175.

[275] Lavergne, P. and Vuong, Q. (1996) Nonparametric selection of regressor : the nonnested case. *Econometrica* **64** 207-219

[276] Lavergne, P. and Patilea, V. (2007) Un test sur la significativité des variables explicatives en régression non-paramétrique *JDS Angers 2007*

[277] Leng X. et Müller, H.G. (2006) Classification using functional data analysis for temporal gene expression data. *Bioinformatics* **22** 68-76.

[278] Laukaitis, A. Functional data analysis for cash flow and transactions intensity continuous-time prediction using Hilbert-valued autoregressive processes *European Journal of Operational Research* **185** (3) 1607-1614.

[279] Laukaitis, A. et Rackauskas, A. (2002) Functional data analysis of payment systems *Nonlinear Analysis : Modelling and Control* **7** (2) 53-68.

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

[281] Lee, S.-Y., Zhang, W. et Song, X.-Y. (2002) Estimating the covariance function with functional data. *British J. Math. Statist. Psych.* **55** (2) 247-261.

[282] Leurgans, S. E., Moyeed, R. A. et Silverman, B. W. (1993) Canonical correlation analysis when the data are curves. *J. Roy. Statist. Soc. Ser. B* **55** (3) 725-740

[283] Li, K.-C. (1991) Sliced inverse regression for dimension reduction. With discussion and a rejoinder by the author. *J. Amer. Statist. Assoc.* **86** (414) 316-342.

[284] Li, K.-C., Aragon, Y., Shedden, K. et Thomas Agnan, C. (2003) Dimension reduction for multivariate response data. (English summary) *J. Amer. Statist. Assoc.* **98** (461) 99-109.

[285] Li, W.V. and Shao, Q.M. (2001) Gaussian processes : inequalities, small ball probabilities and applications. In :C.R. Rao and D. Shanbhag (eds.) *Stochastic processes, Theory and Methods.* Handbook of Statitics, **19**, North-Holland, Amsterdam.

[286] Liebscher, E. (1996) Central limit theorems for sums of α-mixing random variables. *Stochastics Stochastics Rep.* **59** (3-4) 241-258.

[287] Liebscher, E (2001) Central limit theorems for α-mixing triangular arrays with applications to nonparametric statistics. *Math. Methods Statist.* **10** (2) 194-214.

[288] Lifshits, M.A., Linde, W. et Shi, Z. (2006) Small deviations of Riemann-Liouville processes in L_q-spaces with respect to fractal measures. *Proc. London Math. Soc.* (3) **92** (1) 224-250.

[289] Lifshits, M.A. et Simon, T. (2005) Small deviations for fractional stable processes. *Ann. Inst. H. Poincaré Probab. Statist.* **41** 725-752.

[290] Labbas, A. et Mourid, T. (2003) Estimation et prévision d'un processus autorégressif Banach. (French) [Estimation and prediction for a Banach autoregressive process] *Ann. I.S.U.P.* **47** (3) 19-38.

[291] Locantore, N., Marron,J.S., Simpson, D.G., Tripoli, N., Zhang, J.T. and Cohen, K.L. (1999) Robust component analusis for functional data (with discussion) **8** 1-74.

[292] López-Pintado, S. et Romo, J. (2005) A Half-Graph Depth For Functional Data *Statistics and Econometrics Working Papers* **ws051603** Universidad Carlos III, Departamento de Estadística y Econometría.

[293] López-Pintado, S. et Romo, J. (2006a) On the concept of depth for functional data. *Statistics and Econometrics Working Papers* **ws061230** Universidad Carlos III, Departamento de Estadística y Econometría.

[294] López-Pintado, S. et Romo, J. (2006b) Depth-based classification for functional data. (English summary) Data depth : robust multivariate analysis, computational geometry and applications, 103-119, *DIMACS Ser. Discrete Math. Theoret. Comput. Sci., 72, Amer. Math. Soc.*, Providence, RI.

[295] López-Pintado, S. et Romo, J. (2007a) Depth-based inference for functional data. *Comp. Stat. & Data Anal.* **51** (10) 4957-4968.

[296] López-Pintado, S. et Romo, J. (2007b) Functional analysis via extensions of the band depth *IMS Lecture Notes, Monograph Series Complex Datasets and Inverse Problems : Tomography, Networks and Beyond* **54** 103-120.

[297] Lucero, J.C. (1999) Computation of the harmonics-to-noise ratio of a voice signal using a functional data analysis algorithm. *J. Sound Vibration* **222** (3) 512-520.

[298] Malfait, N. et Ramsay, J.O. (2003) The historical functional linear model. (English, French summary) *Canad. J. Statist.* **31** (2) 115-128.

[299] Mammen, E. (1993) Bootstrap and wild bootstrap for high-dimensional linear models. *Ann. Statist.* **21** (1) 255-285.

[300] Mandelbrot, B.B. et Van Ness, J.W. (1968) Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Review* **10** (4) 422-437.

[301] Manté, C., Yao, A.-F. et Degiovanni, C. (2007) Principal component analysis of measures, with special emphasis on grain-size curves *Comp. Stat. & Data Anal.* **51** (10) 4969-4984.

[302] Marion, J.-M. et Pumo, B. (2004) Comparaison des modèles ARH(1) et ARHD(1) sur des données physiologiques. (French. English, French summary) [Comparison of the ARH(1) and ARHD(1) models on physiological data] *Ann. I.S.U.P.* **48** (3) 29-38.

[303] Martens, H. et Næs, T. (1989) *Multivariate calibration.* John Wiley & Sons, Ltd., Chichester.

[304] Mas, A. (1999) Normalité asymptotique de l'estimateur empirique1 de l'opérateur d'autocorrélation d'un processus ARH(1). (French) [Asymptotic normality for the empirical estimator of the autocorrelation operator of an ARH(1) process] *C. R. Acad. Sci. Paris Sér. I Math.* **329** (10) 899-902.

[305] Mas, A. (2000) *Estimation d'opérateurs de corrélation de processus fonctionnels : lois limites, tests, déviations modérées* Thèse de doctorat de l'université de Paris 6.

[306] Mas, A. (2002) Weak convergence for the covariance operators of a Hilbertian linear process. *Stochastic Process. Appl.* **99** (1) 117-135.

[307] Mas, A. (2004a) Un nouveau TCL pour les opérateurs de covariance dans le modèle ARH(1). (French) [A new central limit theorem for covariance operators in the ARH(1) model] *Ann. I.S.U.P.* **48** (3) 49-61.

[308] Mas, A. (2004b) Consistance du prédicteur dans le modèle ARH(1) : le cas compact. (French) [Consistency of the predictor in the ARH(1) model : the compact case] *Ann. I.S.U.P.* **48** (3) 39-48.

[309] Mas, A. (2007a) Testing for the mean of random curves : a penalization approach *Statistical Inference for Stochastic Processes* **10** (2) 147-163.

[310] Mas, A. (2007b) Weak convergence in the functional autoregressive model. *J. Multivariate Anal.* **98** (6) 1231-1261.

[311] Mas, A. et Menneteau, L. (2003) Large and moderate deviations for infinite-dimensional autoregressive processes. *J. Multivariate Anal.* **87** (2) 241-260.

[312] Mas, A. et Pumo, B. (2007) The ARHD model. *J. Statist. Plann. Inference* **137** (2) 538-553.

[313] Mas, A. et Pumo, B. (2007) Functional linear regression with derivatives, *su1bmitted*

[314] Masry, E. (2005) Nonparametric regression estimation for dependent functional data : asymptotic normality *Stochastic Process. Appl.* **115** (1) 155-177.

[315] Massy, W.F. (1965) Principal Components Regression in Exploratory Statistical Research *Journal of the American Statistical Association* **60** (309) 234-256.

[316] Menneteau, L. (2005) Some laws of the iterated logarithm in Hilbertian autoregressive models. *J. Multivariate Anal.* **92** (2) 405-425.

[317] Meiring, W. (2005) Mid-Latitude Stratospheric Ozone Variability : a Functional Data Analysis Study of Evidence of the Quasi-Biennial Oscillation, Time Trends and Solar Cycle in Ozoneonde Observations *Technical report* **403** Statistic and Applied Probability, University of California, Santa Barbara.

[318] Merlevède, F. (1996) *Processus linéaires hilbertiens : inversibilité, théorèmes limites. Estimation et prévision.* Thèse de doctorat de l'université de Paris 6.

[319] Merlevède, F. (1997) Résultats de convergence presque sûre pour l'estimation et la prévision des processus linéaires hilbertiens. (French) [Results of almost sure convergence for the estimation and the prediction of Hilbertian linear processes] *C. R. Acad. Sci. Paris Sér. I Math.* **324** (5) 573-576.

[320] Merlevède, F., Peligrad, M. et Utev, S. (1997) Sharp conditions for the CLT of linear processes in a Hilbert space. *J. Theoret. Probab.* **10** (3) 681-693.

[321] Mizuto, M. (2004) Clustering methods for functional data : k-means, single linkage a1nd moving clustering. *COMPSTAT 2004 — Procedings in Computational Statistics.* 1503-1510 Physica, Heidelberg.

[322] Mokkadem, A. (1990) Propriétés de mélange des processus autorégressifs polynomiaux. (French) [Mixing properties of polynomial autoregressive processes] *Ann. Inst. H. Poincaré Probab. Statist.* **26** (2) 219-260.

[323] Mourid, T. (1993) Processus autorégressifs banachiques d'ordre supérieur. (French) [Higher-order Banach-space-valued autoregressive processes] *C. R. Acad. Sci. Paris Sér. I Math.* **317** (12) 1167-1172.

[324] Mourid, T. (2002) Estimation and prediction of functional autoregressive processes. *Statistics* **36** (2) 125-138.

[325] Müller, H.-G. (2005) Functional modelling and classification of longitudinal data. With discussions by Ivar Heuch, Rima Izem, and James O. Ramsay and a rejoinder by the author. *Scand. J. Statist.* **32** (2) 223-246.

[326] Müller, H.G., Sen, R. et Stadtmüller, U. (2008) Functional data Analysis for Volatility Process. *soumis*

[327] Müller, H.-G., et Stadtmüller, U. (2005) Generalized functional linear models. (English summary) *Ann. Statist.* **33** (2) 774-805.

[328] Müller, H.G. et Zhang, Y. (2005) Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories. *Biometrics* **61** (4) 1064-1075.

[329] Nadaraya, E.A (1964) On estimating regression. *theor. Prob. Appl.* **10** 186-196.

[330] Nerini, D., Ghattas, B. (2007) Classifying densities using funcional regression trees : Applications in oceanology *Comp. Stat. & Data Anal.* **51** (10) 4984-4993.

[331] Ocaña, F.A., Aguilera, A.M. and Valderama M. (1999) Functional principal components analysis by choice of norm *J. Mult. Analysis* **71** (2) 262-276

[332] Opgen-Rhein, R. et Strimmer, K. (2006) Infering gene dependency networks from genomic longitudinal data : a functional data approach. *REVSTAT- Statistical Journal* **4** (1) 53-65.

[333] Ormoneit, D., Black, M.J., Hastie, T., Kjellström, H (2005) Representing cyclic human motion using functional analysis. *Image Vision Comput.* **23** (14) 1264-1276.

[334] Ortega-Moreno, M., Valderrama, M.J. et Ruiz-Molina, J.C. A state space model for non-stationary functional data. *COMPSTAT 2002 (Berlin)* 135-140 Physica, Heidelberg.

[335] Palma, W. (2007) *Long-memory time series. Theory and methods.* Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ. xviii+285 pp.

[336] Pardo-Fernández, J.C. (2005) *Tests in nonparametric regression based on the error distribution* PhD Thesis.

[337] Pardo-Fernández, J.C. (2007) Comparison of error distributions in nonparametric regression. *Statist. Probab. Lett.* **77** (3) 350-356.

[338] Pezzulli, S. and Silverman, B. (1993) Some properties of smoothed principal component analysis for functional data *Computational Statistics*, **8** 1-16.

[339] Piccioni, M., Scarlatti, S., Trouvé, A., (1998) A variational problem arising from speech recognition. (English summary) *SIAM J. Appl. Math.* **58** (3) 753-771 (electronic).

[340] Politis, D.N. et Romano, J.P. (1994) The stationary bootstrap. *J. Amer. Statist. Assoc.* **89** (428) 1303-1313.

[341] Pousse, A. et Téchené, J.-J. (1997) Quelques propriétés extrémales des valeurs singulières d'un opérateur compact et leurs applications aux analyses factorielles d'une probabilité ou d'une fonction aléatoire. I. Quelques propriétés extrémales des valeurs singulières d'un opérateur compact. (French. English summary) *Probab. Math. Statist.* **17** (2) *Acta Univ. Wratislav.* **2029** 197-221.

[342] Pousse, A. et Téchené, J.-J. (1998) Quelques propriétés extrémales des valeurs singulières d'un opérateur compact et leurs applications aux analyses factorielles d'une probabilité ou d'une fonction aléatoire. II. Critères d'analyses factorielles linéaires d'une probabilité ou d'une fonction aléatoire. (French. English summary)

[343] Prchal, L. and Sarda, P. (2008) Spline estimator for the functional linear regression with functional response. (submitted) *Probab. Math. Statist.* **18** (1) *Acta Univ. Wratislav.* **2045** 1-18.

[344] Preda, C. (1999) Analyse factorielle d'un processus : problèmes d'approximation et de régression (in french), PhD Lille I, 1999.

[345] Preda, C. (2007) Regression models for functional data by reproducing kernel Hilbert spaces methods. *J. Statist. Plann. Inference* **137** (3) 829-840.

[346] Preda, C. et Saporta, G. (2002) Régression PLS sur un processus stochastique. *Revue de Statistique Appliquée* **50** (2) 27-45.

[347] Preda, C. ; Saporta, G. (2004) PLS approach for clusterwise linear regression on functional data. *Classification, clustering, and data mining applications* 167-176, Stud. Classification Data Anal. Knowledge Organ., Springer, Berlin.

[348] Preda, C. et Saporta, G. (2005a) PLS regression on a stochastic process. *Comput. Statist. Data Anal.* **48** (1) 149-158.

[349] Preda, C. et Saporta, G. (2005b) Clusterwise PLS regression on a stochastic process. *Comput. Statist. Data Anal.* **49** (1) 99-108.

[350] Preda, C. et Saporta, G. (2007) PCR and PLS for Clusterwise Regression on Functional Data *Selected Contributions in Data Analysis and Classification* 589-598, Springer, Berlin, Heidelberg

[351] Preda, C., Saporta, G. et Lévéder, C. (2007) PLS classification of functional data . *Computational Statistics***22** (2) 223-235.

[352] Pumo, B. (1992) *Estimation et prévision de processus autorégressifs fonctionnels. Application aux processus à temps contiWatson, Geoffrey S. Smooth regression analysis. Sankhyā Ser. A 26 1964 359–372nu.* Thèse de doctorat de l'Université de Paris 6.

[353] Pumo, B. (1995) Les processus autorégressifs à valeurs dans $C_{[0,\delta]}$. Estimation de processus discrétisés. (French. English, French summary) [Estimation and prediction of $C_{[0,\delta]}$-valued autoregressive processes] *C. R. Acad. Sci. Paris Sér. I Math.* **320** (4) 497-500.

[354] Pumo, B. (1998) Prediction of continuous time processes by C[0, 1]-valued autoregressive processes. *Statist. Inf. Stochast. Processes* **3** 297-309.

[355] Rachdi, M. et Vieu, P. (2007) Nonparametric regression for functional data : automatic smoothing parameter selection. *J. Statist. Plann. Inference* **137** (9) 2784-2801.

[356] Rachedi, F. (2004) Vitesse de convergence de l'estimateur crible d'un ARB(1). (French) [Rate of convergence of the sieve estimator of an ARB(1) process] *Ann. I.S.U.P.* **48** (3) 87-96.

[357] Rachedi, F. (2005) Vitesse de convergence en norme p-intégrale et normalité asymptotique de l'estimateur crible de l'opérateur d'un ARB(1). (French) [Rate of convergence in p-integral norm and asymptotic normality of the sieve estimator of the operator in an ARB(1) process] *C. R. Math. Acad. Sci. Paris* **341** (6) 369-374.

[358] Rachedi, F. et Mourid, T. (2003) Estimateur crible de l'opérateur d'un processus ARB(1). (French. English, French summary) [Sieve estimator of the operator of an ARB(1) process] *C. R. Math. Acad. Sci. Paris* **336** (7) 605-610.

[359] Ramsay, J.O. (1982) When the data are functions. *Psychometrika* **47** (4) 379-396.

[360] Ramsay, J.O. (2000a) Differential equation models for statistical functions. *Canad. J. Statist.* **28** (2) 225-240.

[361] Ramsay, J.O. (2000b) Functional components of variation in handwriting. *Journal of the American Statistical Association* **95** 9-15.

[362] Ramsay, J.O., Bock, R. et Gasser, T. (1995) Comparison of height acceleration curves in the Fels, Zurich, and Berkeley growth data, *Annals of Human Biology* **22** 413-426

[363] Ramsay, J. and Dalzell, C. (1991) Some tools for functional data analysis *J. R. Statist. Soc. B.* **53** 539-572.

[364] Ramsay, J.O. et Li, X. (1998) Curve registration. (English summary) *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60** (2) 351-363.

[365] Ramsay, J.O., Munhall, K.G., Gracco V.L. and Ostry D.J. (1996) Functional data analysis of lip motion. *J Acoust Soc Am* **99** 3178 ?3727.

[366] Ramsay, J. and Silverman, B. (1997) *Functional Data Analysis* Springer-Verlag, New York

[367] Ramsay, J. and Silverman, B. (2002) *Applied functional data analysis : Methods and case studies* Spinger-Verlag, New York

[368] Ramsay, J. and Silverman, B. (2005) *Functional Data Analysis (Second Edition)* Spinger-Verlag, New York

[369] Rao, C. R. (1958) Some statistical methods for comparison of growth curves. *Biometrics* **14** 1-17

[370] Ratcliffe, S.J., Heller, G.Z. et Leader, L.R. (2002a) Functional data analysis with application to periodically stimulated foetal heart rate data. I : Functional regression. *Stat Med.* **21** (8) 1103-1114.

[371] Ratcliffe, S.J., Heller, G.Z. et Leader, L.R. (2002b) Functional data analysis with application to periodically stimulated fetal heart rate data : II : functional logistic regression. *Stat Med.* **21** (8) 1115-1127.

[372] Raz, J. (1990) Testing for no effect when estimating a smooth function by nonparametric regression : a randomization approach *J. Amer. Statist. Assoc* **85** (409) 132-138

[373] Reddy, S.K.et Dass, M. (2006) Modeling on-line art auction dynamics using functional data analysis. (English summary) *Statist. Sci.* **21** (2) 179-193.

[374] Rhomari, N. (2002) Approximation et inégalités exponentielles pour les sommes de vecteurs aléatoires dépendants. (French) [Approximation and exponential inequalities for sums of dependent random vectors] *C. R. Math. Acad. Sci. Paris* **334** (2) 149-154.

[375] Rice, J.A. et Silverman, B.W. (1991) Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53** (1) 233-243.

[376] Rio, E. (1995) About the Lindeberg method for strongly mixing sequences. *ESAIM Probab. Statist.* **1** 35-61 (electronic).

[377] Rio, E (2000) *Théorie asymptotique des processus aléatoires faiblement dépendants.* (French) [Asymptotic theory of weakly dependent random processes] Mathématiques & Applications (Berlin) [Mathematics & Applications] **31** Springer-Verlag, Berlin, x+169 pp.

[378] Romain, Y. (2002) Perturbation of functional tensors with applications to covariance operators *Statist. and Proba. Lett.* **58** 253-264.

[379] Rønn, B.B. (2001) Nonparametric maximum likelihood estimation for shifted curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63** (2) 243-259.

[380] Rosenblatt, M. (1956) A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U. S. A.* **42** 43-47.

[381] Rossi, F. et Conan-Guez, B. (2005a) Estimation consistante des paramètres d'un modèle non linéaire pour des données fonctionnelles discrétisées aléatoirement. *Comptes rendus de l'Académie des Sciences - Série I* **340** (2) 167-170.

[382] Rossi, F. et Conan-Guez, B. (2005b) Functional Multi-Layer Perceptron : a Nonlinear Tool for Functional Data Analysis. *Neural Networks* **18** (1) 45-60.

[383] Rossi, F. et Conan-Guez, B. (2005c) Un modèle neuronal pour la régression et la discrimination sur données fonctionnelles. *Revue de Statistique Appliquée* **LIII** (4) 5-30.

[384] Rossi, F. et Conan-Guez, B. (2006) Theoretical Properties of Projection Based Multilayer Perceptrons with Functional Inputs. *Neural Processing Letters* **23** (1) 55-70.

[385] Rossi, F., Delannay, N., Conan-Guez, B. et Verleysen, M. (2005) Representation of Functional Data in Neural Networks *Neurocomputing* **64** 183-210.

[386] Rossi, F., Conan-Guez, B. et El Golli, A. (2004) Clustering Functional Data with the SOM algorithm. *In Proceedings of ESANN 2004* 305-312, Bruges, Belgium.

[387] Rossi, F. et Villa, N. (2006) Support Vector Machine for Functional Data Classification. *Neurocomputing* **69** (7-9) 730-742.

[388] Rossi, N., Wang, X. et Ramsay, J.O. (2002) Nonparametric item response function estimates with the EM algorithm. *Journal of the Behavioral and Educational Sciences* **27** 291-317.

[389] Ruiz-Medina, M.D., Salmerón, R. et Angulo, J.M. (2007) Kalman filtering from POP-based diagonalization of ARH(1) *Computational Statistics & Data Analysis* **51** (10) 4994-5008.

[390] Sarda, P. et Vieu, P. (1999) Smoothing kernel regression. In *Smoothing and Regression. Approaches. Computations and Applications* Ed. M. Schimek. Wiley, New-York.

[391] Shmileva, E. (2006) Small ball probabilities for jump Lévy processes from the Wiener domain of attraction. *Statist. Probab. Lett. 76* (17) 1873-1881.

[392] Shen, Q. et Faraway, J. (2004) An F test for linear models with functional responses. (English summary) *Statist. Sinica* **14** (4) 1239-1257.

[393] Silverman, B.W. (1995) Incorporatig effects into functional principal component analysis *J.R. Stat. Soc. B.* **57** 673-689

[394] Silverman, B.W. (1996) Smoothed functional principal component analysis by choice of norm *Ann. Stat.* **24** 1-24

[395] Song, J.J., Lee, H.J., Morris, J.S. et Kang, S. (2007) Clustering of time-course gene expression data using functional data analysis *Computational Biology and Chemistry* **31** (4) 265-274.

[396] Spitzner, D., Marron, J.S. et Essick, G.K. (2003) Mixed-model functional ANOVA for studying human tactile perception. (English summary) *J. Amer. Statist. Assoc.* **98** (462) 263-272.

[397] Stone, C.J. (1980) Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** (6) 1348-1360.

[398] Stone, C.J. (1982) Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** (4) 1040-1053.

[399] Stone, C.J. (1983) Optimal uniform rate of convergence for nonparametric estimators of a density function or its derivatives. *Recent advances in statistics* 393-406, Academic Press, New York.

[400] Stone, C.J. (1985) Additive regression and other nonparametric models. *Ann. Statist.* **13** (2) 689-705.

[401] Stute, W. (1997) Nonparametric model checks for regression. (English summary) *Ann. Statist.* **25** (2) 613-641.

[402] Stute, W. and González-Manteiga, W. (1996) NN goodness-of-fit tests for linear models *J. Statist. Plann. Inference* **53** (1) 75-92.

[403] Stute, W., González Manteiga, W., and Presedo Quindimil, M. (1998a) Bootstrap approximations in model checks for regression. *J. Amer. Statist. Assoc.* **93** (441) 141-149.

[404] Stute, W., Thies, S. and Zhu, L.-X. (1998b) Model checks for regression : an innovation process approach. *Ann. Statist.* **26** (5) 1916-1934.

[405] Stute, W., and Zhu, L.-X. (2005) Nonparametric checks for single-index models. *Ann. Statist.* **33** (3) 1048-1083.

[406] Tarpey, T et Kinateder, K.K.J. (2003) Clustering functional data. *J. Classification* **20** (1) 93-114.

[407] Tarpey, T., Petkova, E. et Ogden, R.T.Profiling placebo responders by self-consistent partitioning of functional data. *J.A.S.A.* **98** (464) 850-858.

[408] Tukey, J.W. (1961) Curves as parameters, and touch estimation. *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.* **I** 681-694. Univ. California Press, Berkeley, Calif.

[409] Tucker, L.R. (1958) Determination of parameters of functional equations by factor analysis. *Psychometrica* **23** 19-23.

[410] Turbillon, C. (2007) *Estimation et prévision des processus Moyenne Mobile fonctionnels* Thèse de doctorat de l'université de Paris 6.

[411] Valderrama, M.J., Ocaña, F.A. et Aguilera, A.M. (2002) Forecasting PC-ARIMA models for functional data. *COMPSTAT* (Berlin) 25-36.

[412] Valderrama, M.J., Ortega-Moreno, M., González, P. et Aguilera, A.M. (2003) Derivation of a state-space model by functional data analysis. *Comput. Statist.* **18** (4) 533-546.

[413] Vidakovic, B. (2001) Wavelet-Based Functional Data Analysis : Theory, Applications and Ramifications *Proceedings PSFVIP-3* F3399.

[414] Viele, K. (2001) Evaluating fit in functional data analysis using model embeddings. (English, French summary) *Canad. J. Statist.* **29** (1) 51-66.

[415] Villa, N. et Rossi, F. (2006a) SVM fonctionnels par interpolation spline. In *Actes des 38ièmes Journées de Statistique de la SFDS* Clamart, France.

[416] Villa, N. et Rossi, F. (2006b) Un résultat de consistance pour des SVM fonctionnels par interpolation spline. *Comptes Rendus Mathématiques* **343** (8) 555-560.

[417] Vimond, M. (2005) Empirical statistical inference for homothetic shifted regression models *Publication du LSP* **LSP-2005-06**.

[418] Vimond, M. Efficient estimation for homothetic shifted regression models *Publication LSP* **LSP-2006-01**.

[419] Vimond, M. (2006b) Goodness of fit test for homothetic shifted regression models *Publication LSP* **LSP-2006-02**.

[420] Volkonskii, V. A. et Rozanov, Y. A. (1959) Some limit theorems for random functions. I, *Theory Probab. Appl.* **4** 178-197.

[421] Wand, M.P. (1990) On exact $\mathbb{L}^1$ rates in nomparametric kernel regression. *Scand.J.Statist* **17 N°3** , 251-256.

[422] Wang, K. et Gasser, T. (1997) Alignment of curves by dynamic time warping. *Ann. Statist.* **25** (3) 1251-1276.

[423] Wang, K. et Gasser, T. (1998) Asymptotic and bootstrap confidence bounds for the structural average of curves. *Ann. Statist.* **26** (3) 972-991.

[424] Wang, K. et Gasser, T. (1999) Synchronizing sample curves nonparametrically. *Ann. Statist.* **27** (2) 439-460.

[425] Wang, S., Jank W. et Shmueli, G. (2007) Explaining and Forecasting Online Auction Prices and their Dynamics using Functional Data Analysis *Journal of Business and Economic Statistics* accepté.

[426] Watson, G.S. (1964) Smooth regression analysis. *Sankhyā Ser. A* **26** 359-372.

[427] Whang, Y.J. and Andrews, D.W.K. (1993) Tests of specification for parametric and semiparametric models. *J. Econometrics* **57** (1-3), 277-318.

[428] Withers, C.S. (1981) Conditions for linear processes to be strong-mixing. *Z. Wahrsch. Verw. Gebiete* **57** (4) 477-480.

[429] Wold, H. (1975) Soft modelling by latent variables : the non-linear iterative partial least squares (NIPALS) approach. *Perspectives in probability and statistics (papers in honour of M. S. Bartlett on the occasion of his 65th birthday)*, 117-142. Applied Probability Trust, Univ. Sheffield, Sheffield.

[430] Wu, C.O., Chiang, C.-T. et Hoover, D.R. (1998) Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J. Amer. Statist. Assoc.* **93** (444) 1388-1402.

[431] Yao, F., Müller, H.-G. et Wang, J.-L. (2005a) Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** (470) 577-590.

[432] Yao, F., Müller, H.-G. et Wang, J.-L. (2005b) Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33** (6) 2873-2903.

[433] Yao, F. and Lee, T.C.M. (2006) Penalised spline models for functional principal component analysis *J.R. Stat. Soc. B* **68** (1) 3-25.

[434] Yokoyama, R. (1980) Moments bounds for stationary mixing sequences *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **52**, 45-57.

[435] Yoshihara, K. (1976) Limiting behavior of U-statistics for stationary, absolutely regular processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **35** (3) 237-252.

[436] Yoshihara, K. (1994) *Weakly dependent stochastic sequences and their applications. Vol. IV. Curve estimation based on weakly dependent data.* Sanseido Co., Ltd., Chiyoda-ku. viii+336 pp.

[437] Yoshihara, K. (2004) *Weakly dependent stochastic sequences and their applications. Vol. XIV. Recent topics on weak and strong limit theorems.* Sanseido Co., Ltd., Chiyoda-ku. vi+410 pp.

[438] Yu, Y. et Lambert, D. (1999) Fitting trees to functional data : with an application to time-of-day patterns *J. Comput. Graph. Statist.* **8** 749-762.

[439] Zhang, C. et Dette, H. (2004) A power comparison between nonparametric regression tests. *Statist. Probab. Lett.* **66** (3) 289-301.

[440] Zhang, J.-T. et Chen, J. Statistical inferences for functional data. *Ann. Statist.* **35** (3) 1052-1079.

[441] Zhao, X., Marron, J.S. et Wells, M.T. (2004) The functional data analysis view of longitudinal data. (English summary) *Statist. Sinica* **14** (3) 789-808.