

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



N°Attribué par la bibliothèque

--	--	--	--	--	--	--	--	--	--



Année: 2020

Sur la régression relative

Mémoire présenté en vue de l'obtention du diplôme de

Master Académique

Université de Saida - Dr Moulay Tahar

Discipline : Mathématiques

Spécialité : Analyse stochastique, statistique des processus et
applications (ASSPA)

par

Becharef Bakhtia¹

Sous la direction de

Docteur Sâadia Rahmani

Soutenue le 15/09/2020 devant le jury composé de

Mme. Wahiba Benzatout	Université Dr. Moulay Tahar. Saïda	Présidente
Melle. Saâdia Rahmani	Université Dr. Moulay Tahar. Saïda	Encadreur
Mme. Fatima Benziadi	Université Dr. Moulay Tahar. Saïda	Examinaterice
Mme. Rachida Rouane	Université Dr. Moulay Tahar. Saïda	Examinaterice

1. e-mail : bhtia108@gmail.com

Remerciements

Gloire soit rendu au Dieu tout puissant le très miséricordieux pour tous ses bienfaits dont il m'a comblé et de m'avoir donné le courage et la force pour réaliser ce modeste travail.

Les grandes leçons ne sont pas tirées d'un livre mais d'un prof. tel que vous Dr. Sâadia Rahmani, **Merci beaucoup** pour votre incroyable gentillesse, encouragement, encadrement, vos conseils, et votre soutien au cours de la réalisation de ce projet de mémoire. De tout cœur, merci !

Je souhaite aussi exprimer ma profonde gratitude aux professeurs Mme. Benzatout, Mme. Benziadi et Mme. Rouane pour avoir accepter d'être examinateurs et de faire partie de mon jury.

Je tiens à exprimer mes sincères remerciements à mes chers enseignants Mr. Kandouci et Mme. Mokhtari ainsi que tous les docteurs du Laboratoire des Modèles Stochastiques, Statistique et Applications pour leur professionnalisme et leur disponibilité tout ces années de scolarité.

Merci à Mes très chers parents qui ont toujours été là pour moi, grâce à vous j'ai pu surmonter tous les obstacles. je remercie mes sœurs khadidja et Aya, Mon frère Mohamed El Seddik pour leur encouragement, toute ma famille et tous ceux qui ont cru en moi et m'ont encouragée de près où de loin, je suis honorée de vous voir fières de moi.

À mes camarades de la promotion 2015-2016 avec lesquelles j'ai passé une scolarité exceptionnelle, riche d'enseignements, et d'expériences de rencontres, je veux ici exprimer ma sincère amitié. Merci pour tous les bons moments passés ensemble.

Enfin, que toutes les personnes qui ont permis que ce travail de voir le jour soient assurées de ma profonde reconnaissance.

Table des matières

I Présentation générale	6
1 Introduction	7
1.1 Historique	9
1.1.1 La régression classique et relative	9
1.1.2 Les données fonctionnelles	10
1.2 Organisation du mémoire	13
2 Préliminaires	15
2.1 Définitions	15
2.1.1 Le noyau	15
2.1.2 La convergence presque complète	16
2.1.3 La notion de mélange	17
2.1.4 Modélisation mathématique des variables fonctionnelles	17
2.1.5 La semi-métrique	18
2.1.6 Probabilités de petites boules	18
2.1.7 Pondération locale du noyau	19
2.1.8 L'entropie de <i>Kolmogorov</i>	21
2.2 Outils	22
2.2.1 Inégalités exponentielles	22
2.2.2 Théorème centrale limite de <i>Lyapunov</i>	23

II Cas réel	24
3 Régression classique	25
3.1 Modèle et son estimateur	25
3.2 La convergence presque complète	26
3.2.1 Hypothèses et notations	27
3.2.2 Résultat asymptotique	27
3.2.3 Démonstration	27
4 Régression relative	34
4.1 Modèle et son estimateur	34
4.2 La convergence presque complète	37
4.2.1 Hypothèses et notations	38
4.2.2 Résultat asymptotique	38
4.2.3 Démonstration	38
III Régression relative cas fonctionnel	45
5 Cas d'échantillons indépendants et identiquement distribués	46
5.1 Modèle et son estimateur	46
5.2 La convergence presque complète ponctuelle	47
5.2.1 Hypothèses et notations	48
5.2.2 Résultat asymptotique	48
5.2.3 Démonstration	48
5.3 La convergence presque complète uniforme	52
5.3.1 Hypothèses et notations	52
5.3.2 Résultat asymptotique	53
5.3.3 Démonstration	53
5.4 La convergence en moyenne quadratique	57
5.4.1 Hypothèses et notations	57
5.4.2 Résultat asymptotique	58
5.4.3 Démonstration	58

TABLE DES MATIÈRES

5.5	La normalité asymptotique	62
5.5.1	Hypothèses et notations	62
5.5.2	Résultat asymptotique	62
5.5.3	Démonstration	63
6	Cas d'échantillon dépendant	68
6.1	Modèle et son estimateur	68
6.2	la convergence presque complète	68
6.2.1	Hypothèses et notations	68
6.2.2	Résultat asymptotique	69
6.2.3	Démonstration	69
7	Application et conclusion	76
7.1	Application	76
7.2	Conclusion	80
	Bibliographie	81

Première partie

Présentation générale

Chapitre 1

Introduction

La théorie de l'estimation est l'une des branches les plus basiques de la statistique. Cette théorie est habituellement divisée en deux composantes principales, à savoir, l'estimation paramétrique et l'estimation non paramétrique. Le problème de l'estimation non paramétrique consiste, dans la majeure partie des cas, à estimer, à partir des observations, une fonction inconnue, élément d'une certaine classe fonctionnelle. Rappelons qu'une procédure non paramétrique est définie indépendamment de la distribution ou la loi de l'échantillon d'observations. Plus particulièrement, on parle de méthode d'estimation non paramétrique lorsque celle-ci ne se ramène pas à l'estimation d'un nombre fini de paramètres réels associés à la loi de l'échantillon.

Un des problèmes centraux en statistique est celui de l'estimation des caractéristiques fonctionnelles associées à la loi des observations, telles que, par exemple, la fonction de densité, la fonction de répartition ou la fonction de régression. Ce dernier est l'un des outils, les plus utilisés en statistique. Elle est très pratique lorsqu'on s'intéresse à la relation entre une variable réponse Y et une covariable X qui peut être réelle, vectorielle ou fonctionnelle. La régression peut aussi être utilisée pour prédire la valeur de la variable réponse, à partir de valeurs connues d'une ou plusieurs covariables (ou variables explicatives). Les applications de la régression, couvrent la plupart des domaines.

Il y a plusieurs façons pour expliquer cette relation. La modélisation la plus générale est

basée sur la considération suivante :

$$Y = r(X) + \varepsilon.$$

où $r(X) = I\!\!E(Y|X)$; $r(X)$ est inconnu et il faut l'estimer à partir des observations (X_i, Y_i) . Les termes d'erreur ε sont aléatoires ; Ils indiquent qu'il n'existe pas de relation exacte entre la variable réponse Y et la variable explicative X . On suppose, aussi, que $I\!\!E(\varepsilon|X = x) = 0$ et $Var(\varepsilon|X = x) = \sigma^2(x)$.

Le problème consiste donc à déterminer (ou plutôt à estimer) pour chaque réalisation x de la variable X , la valeur de la fonction $r(x)$. Pour caractériser cette fonction, une première approche consiste à utiliser un modèle de régression paramétrique. On suppose que cette fonction peut s'écrire comme une fonction explicite des valeurs de X . Cette dernière peut être linéaire, par exemple

$$r(x) = \alpha + \beta x;$$

et on cherche alors à déterminer les meilleures valeurs des paramètres α et β compte tenu d'un critère, par exemple celui des moindres carrés. Nous nous ramenons alors à l'estimation d'un nombre fini de paramètres. Dans certains cas nous pouvons disposer pour cette estimation d'un échantillon $\{(X_i; Y_i); i = 1, \dots, n\}$ de couples indépendants et ayant chacun la même loi que $(X; Y)$. Souvent, l'utilisation d'un modèle paramétrique n'est pas justifiée, il est alors possible de se suffire de la seule donnée de l'échantillon pour réaliser une estimation. Ce sera à l'aide d'un modèle non paramétrique. Dans ce cas on ne dispose d'aucune forme paramétrique pour r mais seulement d'hypothèses générales de régularité comme la dérivabilité.

La méthode des moindres carrés pondérée est parmi les critères utilisés dans l'estimation statistique pour les modèles de régression, ce critère peut ne pas être adapté à certains situations. En effet, l'étude de la régression par la méthode des erreurs quadratiques moyennes se traduit par le principe que toutes les variables ayant un poids égal, donc la présence de valeurs aberrantes peut conduire à un résultat non pertinent.

Dans ce travail on cherche à estimer l'opérateur $r(x)$ avec un autre critère. L'estimateur

construit est basé sur la minimisation de l'erreur quadratique moyenne relative, cette technique est utile dans l'analyse des données avec des réponses positives comme le cours des actions ou la durée de vie, particulièrement courantes dans les domaines économiques, finance, ou les études biomédicales. Elle est donc la plus appropriée comme mesure de la performance que le critère des moindres carrés dans de nombreuses situations pratiques.

1.1 Historique

1.1.1 La régression classique et relative

Dans le contexte non paramétrique, les premiers résultats ont été obtenus par *Tukey* [28]. Tandis que l'estimation par la méthode du noyau a été utilisée pour la première fois en 1964 séparément par *Nadaraya* et *Watson*. Cette méthode d'estimation a connu un développement continu. En effet, *Devroye* [7] a établi la convergence uniforme presque sûre de cet estimateur. Le taux de convergence optimal pour la régression non paramétrique a été donné par *Stone* ([26], [27]). *Collomb* ([4], [5]) a apporté une contribution déterminante sur ce modèle. Ces travaux se sont focalisés sur l'utilisation de la régression dans la prévision de séries chronologiques. Les premiers résultats asymptotiques sur l'estimation non paramétrique de la fonction de régression sur les processus α -mélangeants ont été élaborés par *Györfi* et al. [13]. Dans ce cadre α -mélangeant, *Vieu* [30] a donné les termes asymptotiquement exacts de l'erreur quadratique de l'estimateur à noyau de la fonction de régression. Nous renvoyons à *Sarda* et *Vieu* [25] pour un large éventail de références. Les premiers résultats conséquents sur la notion de régression relative ont été obtenus en 1989 par *Campbell* et al [2], où la régression relative a été utilisée comme outil de classification. En 1991 *Ruiz Velasco* [24] a évoqué l'efficacité asymptotique de la régression relative logistique dans un contexte paramétrique, en particulier lorsque les variables explicatives sont normalement distribuées. *Park*, *Heungsun*, et *Stefanski* [19] ont utilisé des techniques de régression non paramétrique pour produire un estimateur basé sur l'espérance conditionnelle de la variable réponse inverse. *Jones* et al. [17] ont construit un estimateur consistant pour ce modèle, en utilisant la méthode du noyau. Ils ont établi des propriétés asymptotiques, notamment sa convergence en moyenne quadratique dans le cas où les observations sont indépendantes et identiquement distribuées. Dans un article publié

en 2008 par *Giorgi et al.* [11], une approche portant sur l'estimation de la survie relative des patients atteint d'un cancer colorectal, a permis d'obtenir une estimation flexible et représentant une forme réaliste d'un point de vue épidémiologique.

1.1.2 Les données fonctionnelles

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 illustrants 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 années quatre-vingt, sous l'impulsion, notamment, des travaux de *Grenander* [12] et *Ramsay* [20]. Il a été popularisé par *Ramsay* et *Silverman* [21], puis par les différents ouvrages de *Bosq* [1], *Ramsay* et *Silverman* [22] et , [23] et *Ferraty* et *Vieu* [9]. 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.

Dans le contexte particulier de l'étude de séries temporelles, l'approche introduite par *Bosq* [1] fait apparaître 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. 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.* [16], *Hall et al.* [14].

Exemple sur les données fonctionnelles [3]

Études longitudinales : Il est courant de disposer des données concernant un même phénomène mesuré quantitativement à différents temps de mesure. On peut alors souvent considérer que l'on dispose de courbes aléatoires dépendantes du temps (réalisation de processus à temps continu indexés par le temps).

1. Courbe des températures relevées en un point donné, à différents instants, courbe des cumuls mensuels de précipitations en un point donné. De nombreux exemples sont disponibles dans les données **CanadianWeather** du package **fda**.

On représente à titre d'exemple à la Figure 7.1 l'évolution des températures au cours d'une année (mesures moyennes journalières) dans 4 stations météorologiques canadiennes, et au graphique 1.2 les précipitations journalières dans 3 autres stations.

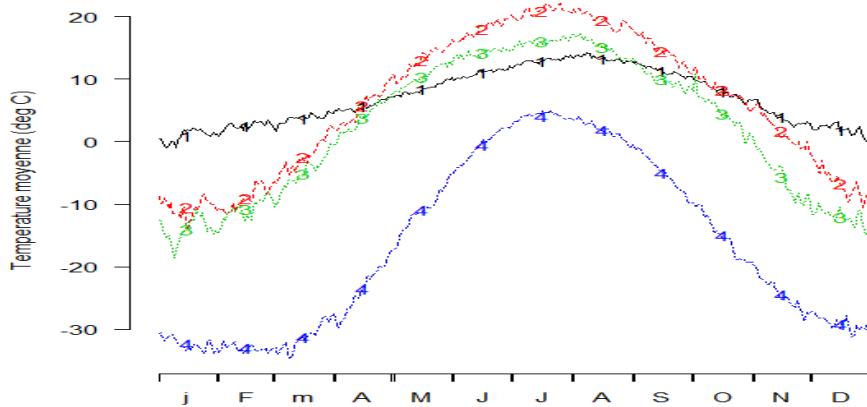


FIGURE 1.1 – Températures moyennes dans les 4 stations.

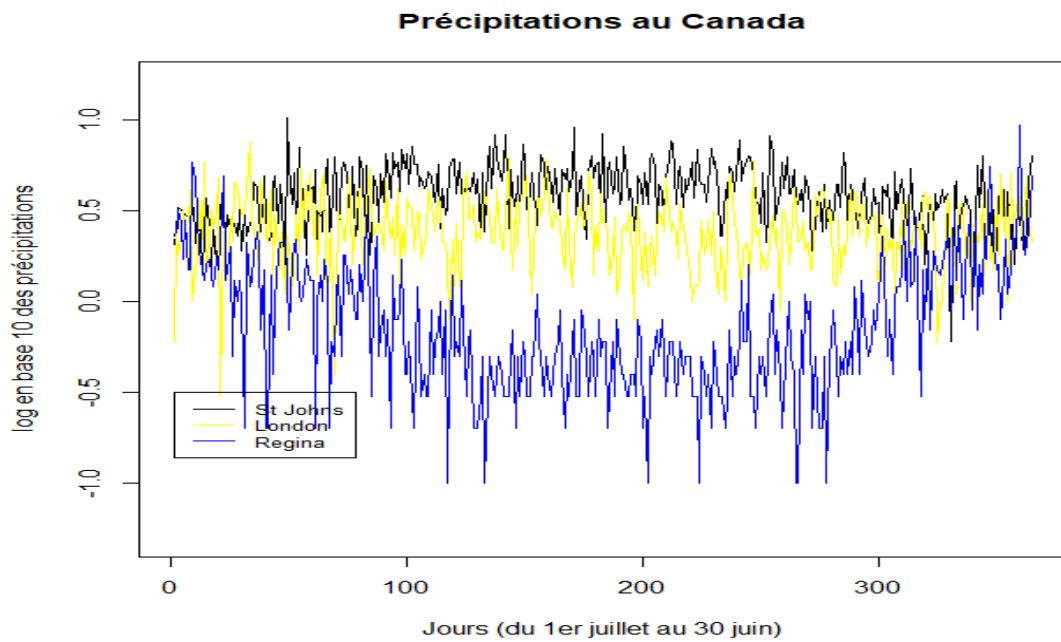


FIGURE 1.2 – Les précipitations journalières dans 3 stations

2. Courbe de croissance d'un individu (ou d'une plante) au cours du temps. On peut citer par exemple, Les données **growth** du package **fda** qui contiennent les tailles de filles et de garçons, mesurées à 31 âges, entre 1 et 18 ans.

Les courbes associées pour 10 filles et 10 garçons de l'échantillon sont représentées à la Figure 1.3.

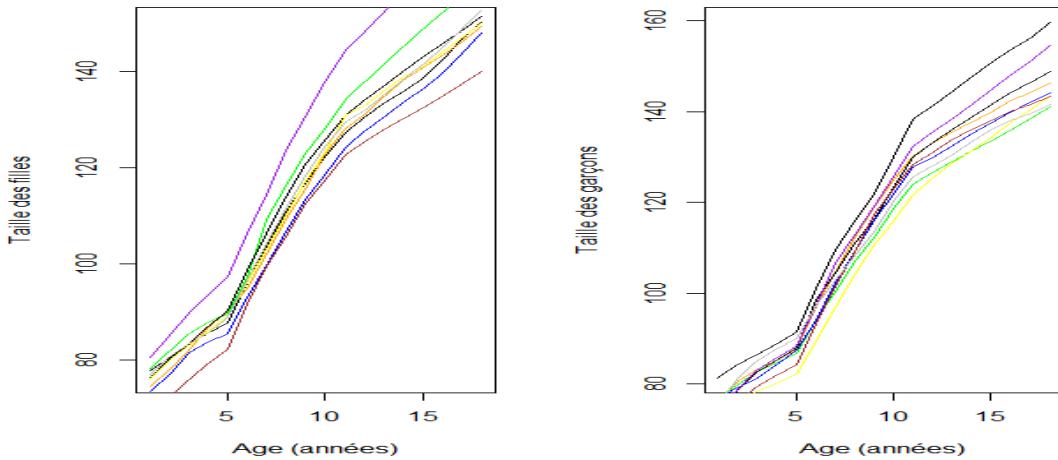


FIGURE 1.3 – Courbe de croissance des filles et des garçons.

1.2 Organisation du mémoire

Notre travail est scindé en trois parties :

Après une partie introductive (comprenant les deux premiers chapitres) dans laquelle nous avons donné un bref historique sur la régression classique, la régression relative et les données fonctionnelles. Nous y énonçons également les définitions et outils probabilistes auxquels nous avons eu recours tout au long de ce mémoire.

Concernant la deuxième partie, nous l'avons consacrée au cadre réel, en s'intéressant uniquement au cas d'observations i.i.d. Dans ce contexte, cette deuxième partie est composée également de deux chapitres. Nous commençons au 3^{ème} chapitre à présenter un estimateur à noyau de la régression classique, et nous établissons sa convergence presque complète. Dans le 4^{ème} chapitre, nous passons à l'étude de l'erreur relative en établissant la même propriété asymptotique que le chapitre précédent.

La troisième partie est consacrée au cadre fonctionnel et est divisée en trois chapitres. Dans le 5^{ème} chapitre, nous nous intéressons au cas où les observations sont indépendantes et identiquement distribuées, la réponse est réelle et la covariable est fonctionnelle. Nous établissons la convergence presque complète ponctuelle et uniforme, la convergence en moyenne quadratique et la normalité asymptotique. Nous généralisons dans le Chapitre

6, le résultat de la convergence presque complète du chapitre précédent, au cas où les données sont α -mélangeantes. Certains résultats théoriques obtenus au chapitre précédent sont illustrés à travers une application sur des données simulées dans le septième chapitre. Ce dernier est achevé par une conclusion générale. Une bibliographie abondante est présentée dans la fin de ce mémoire.

Chapitre 2

Préliminaires

Ce chapitre est destinée à quelques définitions et outils statistiques nécessaires pour l'élaboration des différents résultats présentés dans ce mémoire.

2.1 Définitions

2.1.1 Le noyau

L'étude de la relation entre deux variable aléatoires X et Y est un sujet très important dans l'estimation non paramétrique. Les premiers travaux ont été étendus à la notion de régression par la méthode du noyau K , tel que K est défini par :

Définition 2.1. [9] Une fonction K est définie de \mathbb{R} dans \mathbb{R}^+ tel que $\int_{-\infty}^{+\infty} K(u)du = 1$ est dite Noyau :

- i) De type 0 s'il est de support $[0, 1]$ compact et pour tout $u \in [0, 1]$, $K(u) > 0$.
- ii) De type 1 s'il existe deux constantes réelles $0 < C_1 < C_2 < \infty$ tel que :

$$C_1 \mathbb{1}_{[0,1]} < K < C_2 \mathbb{1}_{[0,1]}^{\text{1}}.$$

- iii) De type 2 si son support est $[0, 1]$, de plus s'il est dérivable et sa dérivée existe sur $[0, 1]$ est satisfaite pour deux constantes réelles $-\infty < C_1 < C_2 < 0$:

$$C_2 < K' < C_1.$$

1. $\mathbb{1}_{(p)}$ la fonction indicatrice qui vaut 1 lorsque p est vrai, 0 sinon.

Exemples 2.1. Plusieurs types de noyaux sont couramment utilisés : rectangulaire, Uniforme, Triangulaire, Quadratique et Gaussien.

- Noyau rectangulaire est défini par :

$$K(u) = \frac{1}{2} \mathbb{1}_{[1,-1]}(u).$$

- Noyau Uniforme est défini par :

$$K(u) = \frac{1}{2} \mathbb{1}_{(|u| \leq 1)}.$$

- Noyau Triangulaire est défini par :

$$K(u) = (1 - |u|) \mathbb{1}_{(|u| \leq 1)}.$$

- Noyau Quadratique est défini par :

$$K(u) = \frac{15}{16} (1 - u^2)^2 \mathbb{1}_{(|u| < 1)}.$$

- Noyau Gaussien est défini par :

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}u^2).$$

2.1.2 La convergence presque complète

Le concept de convergence presque complète a été introduit par *Hsu* et *Robbins* [15]. Elle implique la convergence presque sûre et se prête bien aux calculs faisant intervenir des sommes de variables aléatoires. Malgré cela, elle ne commence à devenir populaire dans la communauté statistique que dans les années 1980 après les travaux de *Collomb* ([4] , [5]). Elle est utilisée surtout en statistique non-paramétrique. Ce critère ainsi que sa vitesse de convergence sont présentés dans les définitions suivantes :

Définition 2.2. [9] On dit que la suite de variables aléatoires réelles $(X_n)_{n \in \mathbb{N}}$ converge presque complètement vers une variable aléatoire X lorsque $n \rightarrow \infty$ si et seulement si :

$$\forall \epsilon > 0 \sum_{n \geq 0} \mathbb{P}(|X_n - X| > \epsilon) < \infty,$$

et cette convergence est notée par :

$$\lim_{n \rightarrow \infty} (X_n)_n = X, \quad p.co.$$

Définition 2.3. [9] On dit que la vitesse de convergence presque complète de la suite de variables aléatoires réelles $(X_n)_{n \in \mathbb{N}}$ vers X est d'ordre u_n (u_n étant une suite numérique déterministe) et on note $X_n = \mathcal{O}_{p.co.}(u_n)$, si :

$$\exists \epsilon_0 > 0, \quad \sum_{n \in \mathbb{N}} \mathbb{P}[|X_n - X| > \epsilon_0 u_n] < \infty.$$

2.1.3 La notion de mélange

De nombreux propriétés statistiques fonctionnelles ont été établies en considérant des échantillons indépendants. Cependant, il est parfois intéressant d'étudier des échantillons dépendants afin de pouvoir répondre à des situations où les données ne sont pas nécessairement indépendantes. Il y a plusieurs types de modélisation de la dépendance au sein d'un échantillon. Nous nous intéresserons dans ce mémoire à des variables α -mélangeantes.

Définition 2.4. [9] Soit $\{\xi_n\}_{n \in \mathbb{Z}}$ une famille de variables aléatoires définie sur l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans un espace probabilisable (Ω', \mathcal{A}') . On note pour $-\infty \leq j \leq k \leq +\infty$, par \mathcal{A}_j^k la σ -algèbre engendrée par les variables aléatoires $(\xi_s, j \leq s \leq k)$. On définit le coefficient α de mélange fort par :

$$\alpha(n) = \sup_k \sup_{A \in \mathcal{A}_{-\infty}^k} \sup_{B \in \mathcal{B}_k^{+\infty}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

La suite $(\xi_n)_{n \in \mathbb{Z}}$ est dite α -mélangeante (ou fortement mélangeante), si :

$$\lim_{n \rightarrow \infty} \alpha(n) = 0. \tag{2.1}$$

Parmi les différents types de décroissance des coefficients de α -mélange on distinguera les cas de décroissance arithmétique et géométrique. Nous nous concentrerons dans ce mémoire sur le premier type.

Définition 2.5. [9] La suite $(\xi_n)_{n \in \mathbb{N}}$ est dite arithmétiquement (algébriquement) α -mélangeante si :

$$\exists C > 0, \quad \alpha(n) \leq Cn^{-a}, \quad \text{avec } a > 0. \tag{2.2}$$

2.1.4 Modélisation mathématique des variables fonctionnelles

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 la troisième partie de ce mémoire à la régression relative dans laquelle la variable réponse est réelle tandis que la variable explicative est fonctionnelle.

Définition 2.6. Une variable aléatoire est dite variable aléatoire fonctionnelle si elle prend ses valeurs dans un espace vectoriel de dimension infinie. Typiquement, il s'agit donc d'une application mesurable $X : \Omega \rightarrow F$. Une donnée fonctionnelle est alors une réalisation de la variable X .

2.1.5 La semi-métrique

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. En plus des métriques disponibles il est assez souvent intéressant de considérer des semi-métriques.

Définition 2.7. [9] d est une semi-métrique sur un espace \mathcal{F} si elle vérifie les deux conditions suivantes :

1. $\forall x \in \mathcal{F} \quad d(x, x) = 0,$
2. $\forall (x, y, z) \in \mathcal{F} \times \mathcal{F} \times \mathcal{F} \quad d(x, y) \leq d(x, z) + d(z, y),$

Les semi-métriques permettent un éventail plus large de topologies possibles que l'on pourra choisir en fonction de la nature des données et du problème considéré.

2.1.6 Probabilités de petites boules

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 non paramétriques en fonction de la dimension. 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 (c'est à dire, 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^d , provenant de la probabilité que la variable explicative appartienne à la boule de centre x et de rayon h . Dans le cas d'une variable explicative fonctionnelle, les résultats asymptotiques sont exprimés à partir de quantités plus générales appelées probabilités de petites boules et définies par :

Définition 2.8. [3] Soit X une variable aléatoire fonctionnelle à valeurs dans un espace de Banach F , on appelle probabilités de petites boules de X les quantités suivantes :

$$\varphi_x(h) = \mathbb{P}(d(x, X) \leq h), \quad x \in \mathcal{F}, h > 0,$$

où $B(x, h) = \{X \in \mathcal{F}, d(x, X) \leq h\}$, avec x est le centre de la boule $B(x, h)$ et h son rayon.

2.1.7 Pondération locale du noyau

Dans le cas des dimensions finies, les techniques de pondération locale sont très appréciées en estimation non paramétriques car elles sont très bien adaptées aux modèles non paramétriques. L'une des approches les plus courantes parmi ces méthodes est certainement celle du noyau.

Cas de variables aléatoires uni-dimensionnées

La pondération locale du noyau est basée sur une fonction de noyau K et une largeur de fenêtre h . Si x est un nombre réel fixé, la pondération locale du noyau transforme n variable aléatoire réels X_1, X_2, \dots, X_n à $\omega_1, \omega_2, \dots, \omega_n$, tel que :

$$\omega_i = \omega_i(x, h, K) = \frac{1}{h} K\left(\frac{x - X_i}{h}\right). \quad (2.3)$$

L'idée principale de la pondération locale autour de x est d'attribuer à chaque variable aléatoire X_i un poids prenant en compte la distance entre x et X_i , plus X_i est éloigné de x , plus la pondération est petite. Pour préciser la notion de pondération locale du noyau, considérons le noyau rectangulaire et réécrivons les ω_i comme suit :

$$\omega_i = \frac{1}{h} \mathbf{1}_{[x-h, x+h]}(X_i).$$

Dans cette situation, la caractéristique locale de la pondération semble évidente puisque les variables aléatoires réelles en dehors de l'intervalle $[x-h, x+h]$ sont ignorées. En outre, la normalisation $\frac{1}{h}$ est proportionnelle à la taille de l'ensemble $[x-h, x+h]$ sur lequel les X_i sont prises en compte. Cette notion de pondération locale n'est pas seulement valable pour le noyau rectangulaire, mais aussi pour tous les noyaux compacts.

Cas de variables aléatoires multi-dimensionnées

Dans des situations multivariées, on observe n vecteurs aléatoires X_1, X_2, \dots, X_n évalués en \mathbb{R}^p . La pondération locale précédente du noyau peut être facilement étendue à cette situation. En effet, il suffit de considérer un noyau multivarié $K^*(u)$ qui sera une fonction de \mathbb{R}^p dans \mathbb{R} . Le premier moyen de le faire est de définir $K^*(u)$ comme un produit de p fonctions de noyaux réels K_1, K_2, \dots, K_p :

$$\forall u = (u_1, u_2, \dots, u_p)^t \in \mathbb{R}^p, \quad K^*(u) = K_1(u_1) \times K_2(u_2) \times \dots \times K_p(u_p).$$

Soit x un vecteur fixé de \mathbb{R}^p . La pondération locale du noyau multivarié consiste à transformer les n vecteurs aléatoires X_1, X_2, \dots, X_n à $\omega_1, \omega_2, \dots, \omega_n$, tel que :

$$\omega_i = \frac{1}{h^p} K^*\left(\frac{x - X_i}{h}\right).$$

Si l'on considère les noyaux compatibles de manière compacte, il apparaît clairement que les ω_i sont des transformations pondérées localement des variables X_i , puisque $\omega_i = 0$ tant que le X_i correspondant est en dehors de certains voisins de x .

Cas de variables aléatoires infini-dimensionnées

Soit X_1, X_2, \dots, X_n n variables aléatoires fonctionnelles évaluées dans \mathcal{F} , et soit x un élément fixé de \mathcal{F} . Une extension fonctionnelle des idées de la pondération locale du noyau multivarié sera la transformation de ces n variables aux n quantités suivantes :

$$\frac{1}{V(h)} K\left(\frac{d(x, X_i)}{h}\right).$$

Où d est une semi-métrique sur \mathcal{F} , K est un noyau réel. Dans cette expression $V(h)$ serait le volume de :

$$B(x, h) = \{x' \in \mathcal{F}, d(x, x') \leq h\}.$$

Qui est la boule par rapport à la topologie induite par d , centrée à x et au rayon h . Cependant, cette approche demande de définir $V(h)$. Il faut avoir une mesure sur \mathcal{F} . C'est la principale différence avec les cas réels et multivariés pour lesquels la mesure de Lebesgue est implicitement utilisée ce qui n'est pas le cas dans l'espace fonctionnel \mathcal{F} . Par conséquent, afin de se libérer d'un choix d'une mesure particulière, nous construisons la normalisation en utilisant directement la distribution de probabilité de la variable aléatoire fonctionnelle, les variables pondérées locales du noyau fonctionnel sont alors définies par :

$$\omega_i = \frac{K\left(\frac{d(x, X_i)}{h}\right)}{\mathbb{E}\left(K\left(\frac{d(x, X_i)}{h}\right)\right)}. \quad (2.4)$$

2.1.8 L'entropie de *Kolmogorov*

L'entropie de *Kolmogorov*² est un outil qui permet de mesurer la complexité des ensembles, dans le sens où une grande entropie signifie qu'une quantité importante d'informations est nécessaire pour décrire l'ensemble.

Définition 2.9. Soit \mathcal{S} un sous- ensemble de l'espace semi-métrique \mathcal{F} , et soit $\epsilon > 0$, un ensemble fini de pointe X_1, X_2, \dots, X_N dans \mathcal{F} est appelé un ϵ -net pour \mathcal{S} si $\mathcal{S} \subset \bigcup_{k=1}^N B(x_k, \epsilon)$, la quantité $\varphi_{\mathcal{S}}(\epsilon) = \log(N_{\epsilon}(\mathcal{S}))$, où $N_{\epsilon}(\mathcal{S})$ est le nombre minimal des boules ouvertes en \mathcal{F} de rayon ϵ qui est nécessaire pour couvrir \mathcal{S} , s'appelle l'entropie de *Kolmogorov* de \mathcal{S} .

Cette notion représente la mesure de la complexité d'un ensemble dans le sens où une entropie élevée signifie que beaucoup d'informations sont nécessaires pour d'écrire un élément avec une précision ϵ . Par conséquent le choix de la structure topologique (d'autre façon, le choix de la semi-métrique), joue un rôle crucial quand on regarde les résultats asymptotiques uniformes sur certains sous-ensemble $\mathcal{S}_{\mathcal{F}}$. Plus particulièrement, la semi-métrique peut également augmenter la concentration de la mesure de probabilité de la variable fonctionnelle X pour minimiser ϵ l'entropie de sous ensemble $\mathcal{S}_{\mathcal{F}}$.

2. Ce concept a été introduit par *Kolmogorov* au milieu des années 1950.

2.2 Outils

2.2.1 Inégalités exponentielles

L'outil que nous allons utiliser de manière déterminante dans les problèmes de convergence presque-complète est l'inégalité exponentielle de *Hoeffding* ci-dessous :

Lemme 2.1. [8] Soit $\Delta_1, \dots, \Delta_n$ des variables aléatoires centrées, indépendantes et de même loi, telle qu'il existe deux réels positifs δ_1 et δ_2 vérifiant :

$$|\Delta_1| \leq \delta_1 \text{ et } \mathbb{E}|\Delta_1|^2 \leq \delta_2$$

Alors, pour tout $\varepsilon \in]0, \delta_2/\delta_1[$ on a :

$$\mathbb{P}\left[n^{-1} \left| \sum_{i=1}^n \Delta_i \right| > \varepsilon\right] \leq 2 \exp\left(\frac{-n\varepsilon^2}{4\delta_2}\right) \quad (2.5)$$

Le lemme suivant donne L'inégalité de *Fuk Nageav*, cette inégalité est en fait une extension au cadre de variables fortement mélangeantes de L'inégalité de *Hoeffding* :

Lemme 2.2. [8] Soit $\{\Delta_i, i \in \mathbb{N}\}$ une famille de variables aléatoires à valeurs dans \mathbb{R} qui vérifient la condition de mélange forte (2.1) avec des coefficients à décroissance algébrique tels que définis en (2.2). On pose

$$s_n^2 = \sum_{i=1}^n \sum_{j=1}^n |Cov(\Delta_i, \Delta_j)|.$$

Si $\|\Delta\|_\infty < \infty$, $\forall i$, alors on a pour tout $\varepsilon > 0$ et pour tout $r > 1$:

$$\mathbb{P}\left[\left| \sum_{k=1}^n \Delta_k \right| > 4\varepsilon\right] \leq 4 \left(1 + \frac{\varepsilon^2}{rs_n^2}\right)^{-\frac{r}{2}} + 2ncr^{-1} \left(\frac{2r}{\varepsilon}\right)^{a+1}. \quad (2.6)$$

Le lemme suivant présente une inégalité de covariance qui est très utile pour le calcul de s_n^2 , définie dans le lemme précédent :

Lemme 2.3. [8] Inégalité de *Dvydov-Rio* : Soit $\{\Delta_i, i \in \mathbb{N}\}$ une famille de variables aléatoires à valeurs dans \mathbb{R} qui vérifient la condition de mélange forte (2.1), et telle que $\|\Delta\|_\infty < \infty$, $\forall i$. On a pour tout $i \neq j$:

$$|Cov(\Delta_i, \Delta_j)| \leq 4\|\Delta_i\|_\infty \|\Delta_j\|_\infty \alpha_{|i-j|}.$$

Le corollaire suivant présente une inégalité exponentielle très utile dans le cas fonctionnel.

Corollaire 2.4. [9]

(i) Si $\forall m \geq 2$, $\exists C_m > 0$, $\mathbb{E}(Z_1^m) \leq C_m a^{2(m-1)}$, on a :

$$\forall \epsilon > 0, \mathbb{P}\left(\left|\sum_{i=1}^n Z_i\right| > \epsilon n\right) \leq 2 \exp\left(\frac{-\epsilon^2 n}{2a^2(1+\epsilon)}\right).$$

(i.i) Supposons que les variables dépendent de n (c'est à dire : $Z_i = Z_{i,n}$), si $\forall m \geq 2$, $\exists C_m > 0$, $\mathbb{E}(Z_1^m) \leq C_m a^{2(m-1)}$ et si $U_n = n^{-1}a_n^2 \log n$ vérifie $\lim_{n \rightarrow \infty} U_n = 0$, nous obtenons :

$$\frac{1}{n} Z_i = \mathcal{O}_{p.co.}(\sqrt{U_n}).$$

2.2.2 Théorème centrale limite de Lyapunov

Dans la littérature, il existe plusieurs versions du théorème central limite. Dans ce travail, nous focalisons sur la version de *Lyapunov*.

Théorème 2.5. [29] Soit X_1, X_2, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées de carré intégrables de moyenne μ_i et de variance σ_i^2 .

Soit $S_n^2 = \sum_{i=1}^n \sigma_i^2$, la condition de Lyapunov suppose l'existence d'une valeur $\delta > 0$ tel que :

$$\lim_{n \rightarrow \infty} \frac{1}{S_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}[|X_i - \mu_i|^{2+\delta}] = 0.$$

Deuxième partie

Cas réel

Chapitre **3**

Régression classique

L'un des modèles le plus fréquemment rencontré en statistique paramétrique ou non paramétrique est le modèle de régression. Le principe de la régression non paramétrique remonte au dix neuvième siècle. La régression non paramétrique est devenue une méthode populaire pour analyser une relation entre une variable dépendante Y et une variable indépendante X . Son objet, est d'estimer cette relation de dépendance sans faire d'hypothèses paramétriques sur la forme de cette dépendance. Dans ce chapitre, nous présentons le modèle et son estimateur à noyau dans la première section. Dans la deuxième section, nous établissons sa convergence presque complète en précisant sa vitesse de convergence.

3.1 Modèle et son estimateur

Considérons n couples de variables aléatoires (i.i.d.) $(X_i, Y_i)_{1 \leq i \leq n}$ ayant la même loi que le couple (X, Y) , où X et Y sont à valeurs dans \mathbb{R} . Comme il a été déjà mentionné dans l'introduction, la méthode la plus communément utilisée pour étudier la relation entre les deux variables aléatoires X et Y est la régression, qui suppose un modèle de la forme :

$$Y = r(X) + \epsilon_i, \quad (3.1)$$

où les erreurs aléatoires ϵ_i sont non corrélées, de moyennes nulles et de variances σ^2 .

On cherchera, dans une famille fixée de fonctions quelle est celle pour laquelle les Y sont les plus proches de $r(X)$. Cette proximité se mesure en général par un risque utilisant l'erreur

quadratique moyenne (MSE en anglais), et on essayera alors de déterminer la fonction $\tilde{r}(X)$ qui rendra cette erreur la plus petite possible, c'est à dire :

$$\mathbb{E}[(Y - \tilde{r}(X))^2] = \min_{\tilde{r}} \mathbb{E}[(Y - r(X))^2].$$

Il est connu que ce minimum est donné par l'espérance conditionnelle :

$$r(X) = \mathbb{E}[Y/X = x],$$

La preuve de cette égalité est trouvée en différenciant l'espérance $\mathbb{E}[(Y - r(X))^2]$ par rapport à $r(X)$, en égalant le résultat à 0, et finalement en isolant $r(X)$, on obtient :

$$\begin{aligned} \frac{\partial}{\partial r(X)} \mathbb{E}[(Y - r(X))^2/X] &= 2 \mathbb{E}[Y - r(X)/X] \\ &= 2 \mathbb{E}[Y/X] - 2 r(X) \\ &= 0, \\ \Rightarrow r(X) &= \mathbb{E}[Y/X]. \end{aligned}$$

L'estimateur de *Nadaraya-Watson* (1964) de la régression est donnée par :

$$\hat{r}(x) = \frac{\hat{g}(x)}{\hat{f}(x)}, \quad (3.2)$$

tel que

$$\hat{g}(x) = \frac{1}{nh} \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n Y_i \omega_i,$$

et

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n \omega_i,$$

où K est le noyau, ω_i est la pondération locale définie par la formule (2.3) et $\hat{f}(x)$ est l'estimateur à noyau de la fonction de densité de x .

3.2 La convergence presque complète

Notre but dans cette section est l'établissement de la convergence presque complète de l'estimateur (3.2).

3.2.1 Hypothèses et notations

Dans toute la suite, C désignera une constante générique positive qui prendra différentes valeurs selon le contexte, mais ne dépendra pas de n . D'autre part, toutes les limites sont obtenues pour $n \rightarrow \infty$; sauf indication.

Nous introduisons les hypothèses suivantes :

- (H.1) Les fonctions $f(x)$ et $g(x)$ sont k -fois continument dérивables et bornées respectivement par M, M' .
- (H.2) La densité $f(x)$ est positive.
- (H.3) Le noyau K est borné, intégrable et à support compact, de plus, il est supposé d'ordre k c'est à dire :

$$\int_{\mathbb{R}} |Z|^j K(z) dz = \begin{cases} 0 & \text{si } j < k \\ C & \text{si } j = k \text{ où } C \text{ est une constante} \end{cases}$$

- (H.4) Le paramètre de lissage h vérifie :

$$\begin{cases} \lim_{n \rightarrow \infty} h = 0, \\ \text{et} \\ \lim_{n \rightarrow \infty} \frac{\log n}{nh} = 0. \end{cases}$$

3.2.2 Résultat asymptotique

Théorème 3.1. *Supposons que les hypothèses précédentes soient réalisées alors on a :*

$$\hat{r}(x) - r(x) = \mathcal{O}(h^k) + \mathcal{O}_{p.co.}\left(\sqrt{\frac{\log n}{nh}}\right).$$

3.2.3 Démonstration

La preuve du résultat précédent est basée sur la décomposition suivante :

$$\begin{aligned} \hat{r}(x) - r(x) &= \hat{r}(x) - \frac{g(x)}{\hat{f}(x)} + \frac{g(x)}{\hat{f}(x)} + r(x) \\ &= \frac{\hat{g}(x) - g(x)}{\hat{f}(x)} + \frac{f(x)r(x) - \hat{f}(x)r(x)}{\hat{f}(x)} \\ &= \frac{\hat{g}(x) - g(x)}{\hat{f}(x)} + \frac{f(x) - \hat{f}(x)}{\hat{f}(x)}r(x). \end{aligned}$$

Par conséquent, le résultat énoncé dans le Théorème 3.1 découle des lemmes suivants :

Lemme 3.2. Sous les hypothèses du Théorème 3.1, on a :

- (i) $\mathbb{E}[\hat{f}(x)] - f(x) = \mathcal{O}(h^k)$,
- (ii) $\mathbb{E}[\hat{g}(x)] - g(x) = \mathcal{O}(h^k)$.

Lemme 3.3. Sous les hypothèses du Théorème 3.1, on a :

$$(i) \quad \hat{f}(x) - \mathbb{E}[\hat{f}(x)] = \mathcal{O}_{p.co.}\left(\sqrt{\frac{\log n}{nh}}\right),$$

$$(ii) \quad \hat{g}(x) - \mathbb{E}[\hat{g}(x)] = \mathcal{O}_{p.co.}\left(\sqrt{\frac{\log n}{nh}}\right).$$

Lemme 3.4. Sous les hypothèses précédentes, on a :

$$\exists \delta > 0, \quad \sum_{i=1}^n \mathbb{P}\left[|\hat{f}(x)| \leq \delta\right] < \infty.$$

Preuve du Lemme 3.2

- Commençons par la partie (i)

Comme les X_i sont équidistribuées, on a ;

$$\begin{aligned} \mathbb{E}[\hat{f}(x)] &= \mathbb{E}\left[\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)\right] \\ &= \frac{1}{h} \mathbb{E}\left[K\left(\frac{x-X}{h}\right)\right] \\ &= \frac{1}{h} \int_{\mathbb{R}} K\left(\frac{x-u}{h}\right) f(u) du. \end{aligned}$$

En effectuant le changement de variable :

$$z = \frac{x-u}{h} \Rightarrow \begin{cases} u = x - hz \\ du = -hdz, \end{cases} \quad (3.3)$$

on obtient :

$$\mathbb{E}[\hat{f}(x)] = \int_{\mathbb{R}} K(z) f(x - hz) dz,$$

puisque f est k fois dérivable, on utilise le développement de Taylor suivant :

$$f(x - hz) = f(x) + \sum_{j=1}^{k-1} \frac{(-1)^j (zh)^j}{j!} f^j(x) + \frac{(-1)^k (zh)^k}{k!} f^k(\theta_x),$$

où θ_x est entre x et $x - hz$. D'autre part, en utilisant la deuxième partie de l'hypothèse (H.3), nous obtenons :

$$\mathbb{E}[\hat{f}(x)] = f(x) + (-1)^k h^k \int_{\mathbb{R}} z^k K(z) \frac{f^k(\theta_x)}{k!} dz,$$

puisque f^k est continue, alors lorsque $h \rightarrow 0$ on a : $f^k(\theta_x) \rightarrow f^k(x)$,
on obtient donc :

$$\begin{aligned} |\mathbb{E}[\hat{f}(x)] - f(x)| &\leq \frac{h^k}{k!} |f^k(x)| \int_{\mathbb{R}} |z^k| K(z) dz \\ &\leq C_1 h^k, \end{aligned}$$

avec $C_1 = \frac{M.C}{k!}$, ce qui donne le résultat de la partie (i). ■

– Pour ce qui concerne la partie (ii), on a :

$$\begin{aligned} \mathbb{E}[\hat{g}(x)] &= \mathbb{E}\left[\frac{1}{nh} \sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h}\right)\right] \\ &= \frac{1}{h} \mathbb{E}\left[Y K\left(\frac{x-X}{h}\right)\right] \\ &= \frac{1}{h} \int_{\mathbb{R}} Y K\left(\frac{x-u}{h}\right) f(u) du. \end{aligned}$$

En conditionnant par rapport à $X = x$, on obtient :

$$\begin{aligned} \mathbb{E}[\hat{g}(x)] &= \frac{1}{h} \int_{\mathbb{R}} \mathbb{E}[Y/X = x] K\left(\frac{x-u}{h}\right) f(u) du \\ &= \frac{1}{h} \int_{\mathbb{R}} r(x) K\left(\frac{x-u}{h}\right) f(u) du. \end{aligned}$$

D'autre part, on a $r(x) = \frac{g(x)}{f(x)} \Rightarrow g(x) = r(x)f(x)$, ce qui conduit à :

$$\mathbb{E}[\hat{g}(x)] = \frac{1}{h} \int_{\mathbb{R}} K\left(\frac{x-u}{h}\right) g(u) du.$$

En effectuant le même changement de variable (3.3), et la symétrie du noyau K nous permet de déduire :

$$\mathbb{E}[\hat{g}(x)] = \int_{\mathbb{R}} K(z) g(x - hz) dz.$$

Puisque g est k fois différentiable, on utilise le développement de Taylor comme précédemment. Ainsi, nous obtenons :

$$\mathbb{E}[\hat{g}(x)] = g(x) + (-1)^k h^k \int_{\mathbb{R}} z^k K(z) \frac{g^k(\theta_x)}{k!} dz,$$

puisque g^k est continue, alors lorsque $h \rightarrow 0$ on a : $g^k(\theta_x) \rightarrow g^k(x)$.

On obtient donc :

$$\begin{aligned} |\mathbb{E}[\hat{g}(x)] - g(x)| &\leq \frac{h^k}{k!} |g^k(x)| \int_{\mathbb{R}} |z^k| K(z) dz \\ &\leq C_2 h^k, \end{aligned}$$

avec $C_2 = \frac{M' \cdot C}{k!}$, et le résultat de la partie (ii) est donc achevé. ■

Preuve du Lemme 3.3

La preuve du Lemme 3.3 nécessite l'utilisation de l'inégalité exponentielle.

– Commençons par la partie (i)

on a :

$$\begin{aligned} \hat{f}(x) - \mathbb{E}[\hat{f}(x)] &= \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) - \mathbb{E}\left[\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)\right] \\ &= \frac{1}{nh} \sum_{i=1}^n \left[K\left(\frac{x - X_i}{h}\right) - \mathbb{E}\left[K\left(\frac{x - X_i}{h}\right)\right] \right], \end{aligned}$$

on pose :

$$\Delta_i = \frac{1}{h} \left[K\left(\frac{x - X_i}{h}\right) - \mathbb{E}\left[K\left(\frac{x - X_i}{h}\right)\right] \right].$$

Pour appliquer L'inégalité (2.5), il faut d'abord vérifier les deux conditions suivantes :

1. $|\Delta_i| < \delta_1$,
2. $\mathbb{E}[\Delta_i^2] < \delta_2$.

Pour la condition 1. et puisque K est borné, on a :

$$\begin{aligned} |\Delta_i| &\leq \frac{1}{h} \left[\left| K\left(\frac{x - X_i}{h}\right) \right| + \left| \mathbb{E}\left[K\left(\frac{x - X_i}{h}\right)\right] \right| \right] \\ &\leq \frac{C}{h} = \delta_1. \end{aligned}$$

Pour la condition 2. on a :

$$\begin{aligned}
 \mathbb{E}[\Delta_i^2] &= \frac{1}{h^2} \mathbb{E} \left[\left[K \left(\frac{x - X_i}{h} \right) - \mathbb{E} \left[K \left(\frac{x - X_i}{h} \right) \right] \right]^2 \right] \\
 &= \frac{1}{h^2} \text{Var} \left[K \left(\frac{x - X_i}{h} \right) \right] \\
 &\leq \frac{1}{h^2} \mathbb{E} \left[K^2 \left(\frac{x - X_i}{h} \right) \right] \\
 &\leq \frac{1}{h^2} \int_{\mathbb{R}} K^2 \left(\frac{x - u}{h} \right) f(u) du.
 \end{aligned}$$

Par un changement de variable, nous obtenons :

$$\mathbb{E}[\Delta_i^2] \leq \frac{1}{h} \int_{\mathbb{R}} K^2(z) f(x - hz) dz,$$

K est de carré intégrable et de plus lorsque $h \rightarrow 0$, $f(x - hz) \rightarrow f(x)$, nous obtenons donc :

$$\mathbb{E}[\Delta_i^2] \leq \frac{M}{h} = \delta_2.$$

Puisque les deux conditions sont vérifiées, on peut appliquer l'inégalité exponentielle (2.5) :

$$\begin{aligned}
 \mathbb{P} \left(\frac{1}{n} \left| \sum_{i=1}^n \Delta_i \right| > \epsilon' \right) &\leq 2 \exp \left(\frac{-n\epsilon'^2}{4\delta_2} \right) \\
 &\leq 2n^{\frac{-\epsilon'^2}{4M}},
 \end{aligned}$$

avec $\epsilon' = \epsilon \sqrt{\frac{\log n}{nh}}$, c'est à dire :

$$\sum_{i=1}^n \mathbb{P} \left(\frac{1}{n} \left| \sum_{i=1}^n \Delta_i \right| > \epsilon' \right) < \infty.$$

Ce qui prouve la partie (i). ■

– Concernant la partie (ii), on a :

$$\begin{aligned}
 \widehat{g}(x) - \mathbb{E}[\widehat{g}(x)] &= \frac{1}{nh} \sum_{i=1}^n Y_i K \left(\frac{x - X_i}{h} \right) - \mathbb{E} \left[\frac{1}{nh} \sum_{i=1}^n Y_i K \left(\frac{x - X_i}{h} \right) \right] \\
 &= \frac{1}{nh} \sum_{i=1}^n \left[Y_i K \left(\frac{x - X_i}{h} \right) - \mathbb{E} \left[Y_i K \left(\frac{x - X_i}{h} \right) \right] \right].
 \end{aligned}$$

On pose :

$$\Delta_i = \frac{1}{h} \left[Y_i K\left(\frac{x - X_i}{h}\right) - \mathbb{E}\left[Y_i K\left(\frac{x - X_i}{h}\right)\right] \right],$$

D'une manière analogue à la partie (i), il faut d'abord vérifier les deux conditions précédentes.

Pour la condition 1. et puisque K est borné, on a :

$$\begin{aligned} |\Delta_i| &\leq \frac{1}{h} \left[\left| Y_i K\left(\frac{x - X_i}{h}\right) \right| + \left| \mathbb{E}\left[Y_i K\left(\frac{x - X_i}{h}\right)\right] \right| \right] \\ &\leq \frac{C}{h} = \delta_1. \end{aligned}$$

Pour la condition 2. on a :

$$\begin{aligned} \mathbb{E}[\Delta_i^2] &= \frac{1}{h^2} \mathbb{E} \left[\left[Y_i K\left(\frac{x - X_i}{h}\right) - \mathbb{E}\left[Y_i K\left(\frac{x - X_i}{h}\right)\right] \right]^2 \right] \\ &= \frac{1}{h^2} \text{Var} \left[Y_i K\left(\frac{x - X_i}{h}\right) \right] \\ &\leq \frac{1}{h^2} \mathbb{E} \left[Y_i^2 K^2 \left(\frac{x - X_i}{h} \right) \right] \\ &\leq \frac{1}{h^2} \int_{\mathbb{R}} Y_i^2 K^2 \left(\frac{x - u}{h} \right) f(u) du, \end{aligned}$$

Ainsi, après un changement de variable, nous aurons :

$$\mathbb{E}[\Delta_i^2] \leq \frac{1}{h} Y_i^2 K^2(z) f(x - hz) dz,$$

K étant de carré intégrable et de plus, lorsque $h \rightarrow 0$, $f(x - hz) \rightarrow f(x)$ nous obtenons donc :

$$\mathbb{E}[\Delta_i^2] \leq \frac{M}{h} = \delta_2.$$

On peut maintenant appliquer l'inégalité exponentielle (2.5) :

$$\begin{aligned} \mathbb{P} \left(\frac{1}{n} \left| \sum_{i=1}^n \Delta_i \right| > \epsilon' \right) &\leq 2 \exp \left(\frac{-n\epsilon'^2}{4\delta_2} \right) \\ &\leq 2n^{\frac{-\epsilon'^2}{4M}}, \end{aligned}$$

avec $\epsilon' = \epsilon \sqrt{\frac{\log n}{nh}}$, c'est à dire :

$$\sum_{i=1}^n \mathbb{P} \left(\frac{1}{n} \left| \sum_{i=1}^n \Delta_i \right| > \epsilon' \right) < \infty.$$

■

Preuve de Lemme 3.4

Nous remarquons que :

$$|\hat{f}(x)| \leq \frac{f(x)}{2} \Rightarrow |\hat{f}(x) - f(x)| > \frac{f(x)}{2},$$

d'où :

$$\mathbb{P}\left[|\hat{f}(x)| \leq \frac{f(x)}{2}\right] \leq \mathbb{P}\left[|\hat{f}(x) - f(x)| > \frac{f(x)}{2}\right].$$

Comme $f(x) > 0$, on prend $\delta = \frac{f(x)}{2}$, nous obtenons :

$$\sum_{i=1}^n \mathbb{P}\left[|\hat{f}(x)| \leq \delta\right] < \infty.$$

Ce qui complète la preuve du Lemme 3.4. ■

Chapitre **4**

Régression relative

Nous présentons dans ce chapitre, la régression relative qui est définie à partir d'une minimisation de l'erreur quadratique moyenne relative. La présentation du modèle ainsi que son estimateur à noyau est explicitement donnée dans la première section. Dans la deuxième section, nous établissons la vitesse de convergence presque complète de l'estimateur construit.

4.1 Modèle et son estimateur

On considère $\{(X_i, Y_i), i = 1, \dots, n\}$ une suite de n -observations réelles indépendantes, de même loi que le couple (X, Y) . Nous nous concentrerons sur l'étude du lien entre une variable explicative X et une réponse Y . La modélisation la plus générale de cette relation est basée sur la considération suivante :

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

où ϵ est une variable aléatoire réelle centée et indépendante de X , telle que :

$$\begin{cases} \mathbb{E}(\epsilon/X) = 0; \\ var(\epsilon/X) = \sigma^2 < \infty. \end{cases}$$

Dans ce chapitre, on obtient l'estimateur de l'opérateur de régression $r(x)$ par la minimisation de l'erreur quadratique moyenne relative, c'est à dire :

$$\min_r \mathbb{E} \left[\left(\frac{Y - r(x)}{Y} \right)^2 \middle| X = x \right], \quad \text{pour } Y > 0. \quad (4.1)$$

(4.1) est une mesure plus significative, lorsqu'on a des valeurs aberrantes. La solution du problème (4.1) peut être explicitement exprimée par le rapport des deux premiers moments inverses conditionnels de Y sachant X .

$$r(x) = \frac{\mathbb{E}[Y^{-1}/X]}{\mathbb{E}[Y^{-2}/X]}, \quad (4.2)$$

En effet

$$\begin{aligned} \frac{\partial}{\partial r(x)} \mathbb{E}\left[\left(\frac{Y - r(x)}{Y}\right)^2 \middle| X = x\right] &= 2 \mathbb{E}\left[\left(\frac{Y - r(x)}{Y^2}\right) \middle| X = x\right] \\ &= 2 \mathbb{E}[Y^{-1}/X = x] - 2 r(x) \mathbb{E}[Y^{-2}/X = x] = 0 \\ \Rightarrow r(x) &= \frac{\mu(-1)}{\mu(-2)}. \end{aligned}$$

On peut démontrer facilement, que (4.2) est le meilleur prédicteur de Y sachant X .

En effet, soit $R(x)$ un prédicteur quelconque de Y sachant X , alors :

$$\begin{aligned} \mathbb{E}\left[\left(\frac{Y - R(x)}{Y}\right)^2 \middle| X = x\right] &= \mathbb{E}\left[\left(\frac{(Y + r(x) - r(x) - R(x))}{Y}\right)^2 \middle| X = x\right] \\ &= \mathbb{E}\left[\left(\left(\frac{Y - r(x)}{Y}\right)^2 + \left(\frac{r(x) - R(x)}{Y}\right)^2 + 2 \frac{(Y - r(x))(r(x) - R(x))}{Y^2}\right) \middle| X\right] \\ &= \mathbb{E}\left[\left(\frac{Y - r(x)}{Y}\right)^2 \middle| X = x\right] + \mathbb{E}\left[\left(\frac{r(x) - R(x)}{Y}\right)^2 \middle| X = x\right] \\ &\quad + 2 \mathbb{E}\left[\frac{(Y - r(x))(r(x) - R(x))}{Y^2} \middle| X = x\right] \\ &= \mathbb{E}\left[\left(\frac{Y - r(x)}{Y}\right)^2 \middle| X = x\right] + \mathbb{E}\left[\left(\frac{r(x) - R(x)}{Y}\right)^2 \middle| X = x\right] \\ &\quad + 2 \mathbb{E}\left[\frac{(Y - r(x))}{Y^2} \middle| X\right](r(x) - R(x)). \end{aligned} \quad (4.3)$$

Remarquons que le troisième terme de (4.3) est nul car :

$$\mathbb{E}\left[\left(\frac{Y - r(x)}{Y^2}\right) \middle| X = x\right] = \mathbb{E}\left[\frac{Y}{Y^2} \middle| X = x\right] - \mathbb{E}\left[\frac{r(x)}{Y^2} \middle| X = x\right]$$

$$\begin{aligned}
 &= \mathbb{E}(Y^{-1}/X = x) - r(x)\mathbb{E}(Y^{-2}/X = x) \\
 &= \mu(-1) - \frac{\mu(-1)}{\mu(-2)}\mu(-2) \\
 &= 0.
 \end{aligned}$$

Donc, on aura :

$$\begin{aligned}
 \mathbb{E}\left[\left(\frac{Y - R(x)}{Y}\right)^2 / X = x\right] &= \mathbb{E}\left[\left(\frac{Y - r(x)}{Y}\right)^2 / X = x\right] + \mathbb{E}\left[\left(\frac{r(x) - R(x)}{Y}\right)^2 / X = x\right] \\
 &= \mathbb{E}\left[\frac{Y^2 - 2Yr(x) + r(x)^2}{Y^2} / X = x\right] + \mathbb{E}\left[\left(\frac{r(x) - R(x)}{Y}\right)^2 / X = x\right] \\
 &= 1 - 2\mathbb{E}(Y^{-1}r(x)/X = x) + \mathbb{E}(Y^{-2}r(x)^2/X = x) \\
 &\quad + \mathbb{E}(Y^{-2}/X = x)\mathbb{E}((r(x) - R(x))^2/X = x) \\
 &= 1 - 2\mu(-1)r(x) + \mu(-2)r(x)^2 + \mu(-2)(r(x) - R(x))^2 \\
 &= 1 - 2\frac{\mu(-1)^2}{\mu(-2)} + \frac{\mu(-1)^2}{\mu(-2)} + \mu(-2)(r(x) - R(x))^2 \\
 &= 1 - \frac{\mu(-1)^2}{\mu(-2)} + \mu(-2)(r(x) - R(x))^2 \\
 &= \frac{\mu(-2) - \mu(-1)^2}{\mu(-2)} + \mu(-2)(r(x) - R(x))^2 \\
 &= \frac{\mathbb{E}(Y^{-2}/X = x) - \mathbb{E}(Y^{-1}/X = x)^2}{\mu(-2)} + \mu(-2)(r(x) - R(x))^2 \\
 &= \frac{\mathbb{E}[(Y^{-1})^2/X = x] - \mathbb{E}(Y^{-1}/X = x)^2}{\mu(-2)} + \mu(-2)(r(x) - R(x))^2 \\
 &= \frac{Var(Y^{-1}/X)}{\mu(-2)} + \mu(-2)(r(x) - R(x))^2. \tag{4.4}
 \end{aligned}$$

Le premier terme de (4.4) ne dépend pas de $R(x)$ et le deuxième terme est minimisé lorsque

$R(x) = r(x)$ p.s, donc $r(x)$ est le meilleur prédicteur de l'erreur quadratique moyenne relative.

Le résultat (4.4) montre également que l'erreur de prédiction quadratique moyenne relative de $r(x)$ est donnée par $\frac{Var(Y^{-1}/X)}{\mu(-2)}$.

Remarque 4.1. Le prédicteur $r(x)$ peut-être exprimer par :

$$r(x) = \frac{\mathbb{E}(Y^{-1}/X)}{Var(Y^{-1}/X) + (\mathbb{E}(Y^{-1}/X))^2}.$$

Cette remarque suggère une méthode permettant d'estimer cette fonction. Donc, cela motive l'utilisation des méthodes de modélisation de la moyenne et de la variance pour ajuster les modèles à la moyenne et à la variance de Y^{-1} en fonction de X .

L'une des méthodes les plus populaires permettant de construire un estimateur de la fonction $r(x)$ est la méthode à noyau. plus précisément, on peut construire des estimateurs à noyau de $E(Y^{-\ell}|X)$ pour $\ell = 1, 2$. Ainsi, l'estimateur de la fonction $r(x)$ est donné par :

$$\hat{r}(x) = \frac{\sum_{i=1}^n Y_i^{-1} K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^n Y_i^{-2} K\left(\frac{x - X_i}{h}\right)}. \quad (4.5)$$

Cet estimateur peut être exprimé par :

$$\hat{r}(x) = \frac{\hat{g}_1(x)}{\hat{g}_2(x)}, \quad (4.6)$$

avec

$$\hat{g}_1(x) = \frac{1}{nh} \sum_{i=1}^n Y_i^{-1} K\left(\frac{x - X_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n Y_i^{-1} \omega_i,$$

et

$$\hat{g}_2(x) = \frac{1}{nh} \sum_{i=1}^n Y_i^{-2} K\left(\frac{x - X_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n Y_i^{-2} \omega_i,$$

où ω_i est définie dans la formule (2.3).

4.2 La convergence presque complète

Pour étudié la convergence presque complète de l'estimateur (4.5), on fixe un point $x \in \mathbb{R}$ et on introduit les hypothèses suivantes :

4.2.1 Hypothèses et notations

- (H.1) La fonction r est deux fois continûment dérivable au voisinage de x .
- (H.2) La densité f de la variable explicative et la fonction $\mathbb{E}(Y^{-2}/X = x)$ sont strictement positives au point x , de plus la variable Y est borné.
- (H.3) Le noyau K est symétrique supposé borné, intégrable et à support compact.
- (H.4) Le paramètre de lissage h est tel que :

$$\begin{cases} \lim_{n \rightarrow \infty} h = 0; \\ \lim_{n \rightarrow \infty} \frac{nh}{\log n} = \infty. \end{cases}$$

4.2.2 Résultat asymptotique

Sous les hypothèses précédentes, on a le résultat suivant :

Théorème 4.1. *Sous les hypothèses (H.1)-(H.4), on a :*

$$\hat{r}(x) - r(x) = \mathcal{O}(h^2) + \mathcal{O}_{p.co}\left(\sqrt{\frac{\log n}{nh}}\right). \quad (4.7)$$

4.2.3 Démonstration

Pour démontrer ce théorème, on pose $r(x) = \frac{g_1(x)}{g_2(x)}$, tel que :

$$\begin{cases} g_1(x) = \mu(-1); \\ g_2(x) = \mu(-2). \end{cases}$$

La démonstration de ce théorème est basée sur la décomposition suivante :

$$\begin{aligned} \hat{r}(x) - r(x) &= \frac{\hat{g}_1(x)}{\hat{g}_2(x)} - \frac{g_1(x)}{g_2(x)} + \frac{g_1(x)}{g_2(x)} - r(x) \\ &= \frac{\hat{g}_1(x)}{\hat{g}_2(x)} - \frac{g_1(x)}{g_2(x)} + \frac{r(x)g_2(x)}{\hat{g}_2(x)} - \frac{r(x)\hat{g}_2(x)}{\hat{g}_2(x)} \\ &= \frac{1}{\hat{g}_2(x)}(\hat{g}_1(x) - g_1(x)) + \frac{r(x)}{\hat{g}_2(x)}(g_2(x) - \hat{g}_2(x)) \\ &= \frac{1}{\hat{g}_2(x)}(\hat{g}_1(x) - \mathbb{E}\hat{g}_1(x) + \mathbb{E}\hat{g}_1(x) - g_1(x)) + \frac{r(x)}{\hat{g}_2(x)}(g_2(x) - \mathbb{E}\hat{g}_2(x) + \mathbb{E}\hat{g}_2(x) - \hat{g}_2(x)). \end{aligned}$$

Ainsi, la preuve du Théorème 4.7 est une conséquence directe des lemmes suivants :

Lemme 4.2. Sous les hypothèses **(H.1)** – **(H.3)**, on a :

$$(i) \quad \mathbb{E}\widehat{g}_1(x) - g_1(x) = \mathcal{O}(h^2).$$

$$(ii) \quad \mathbb{E}\widehat{g}_2(x) - g_2(x) = \mathcal{O}(h^2).$$

Lemme 4.3. Sous les hypothèses **(H.1)** – **(H.4)**, on a :

$$(i) \quad \widehat{g}_1(x) - \mathbb{E}\widehat{g}_1(x) = \mathcal{O}\left(\sqrt{\frac{\log n}{nh}}\right).$$

$$(ii) \quad g_2(x) - \mathbb{E}\widehat{g}_2(x) = \mathcal{O}\left(\sqrt{\frac{\log n}{nh}}\right).$$

Lemme 4.4. Sous les hypothèses du Lemme 4.3, on a :

$$\exists \delta > 0, \quad \text{tel que} \sum_{i=1}^n \mathbb{P}[|\widehat{g}_2(x)| < \delta] < \infty.$$

Démonstration du Lemme (4.2)

– Pour le premier résultat (i), on a

$$\begin{aligned} \mathbb{E}\widehat{g}_1(x) &= \mathbb{E}\left[\frac{1}{nh} \sum_{i=1}^n Y_i^{-1} K\left(\frac{x-X_i}{h}\right)\right] \\ &= \frac{1}{h} \mathbb{E}\left[K\left(\frac{x-X_1}{h}\right) Y_1^{-1}\right] \\ &= \frac{1}{h} \mathbb{E}\left[K\left(\frac{x-X_1}{h}\right) \mathbb{E}(Y_1^{-1} / X = X_1)\right] \\ &= \frac{1}{h} \int_{\mathbb{R}} K\left(\frac{x-u}{h}\right) f(u) \mathbb{E}(Y_1^{-1} / X = X_1) du. \end{aligned}$$

On remplace $g_1(x) = \mathbb{E}(Y_1^{-1} / X = X_1)f(x)$, alors :

$$\mathbb{E}[\widehat{g}_1(x)] = \frac{1}{h} \int_{\mathbb{R}} K\left(\frac{x-u}{h}\right) g_1(u) du.$$

On pose $z = \frac{x-u}{h}$, on obtient donc :

$$\mathbb{E}[\widehat{g}_1(x)] = \int_{\mathbb{R}} K(z) g_1(x - hz) dz. \quad (4.8)$$

Puisque la fonction f est deux fois continûment dérivable au voisinage de x (d'après **(H.1)**), on applique le développement de Taylor d'ordre 2 défini par :

$$g_1(x - hz) = g_1(x) - hzg_1^{(1)}(x) + \frac{(zh)^2}{2}g_1^{(2)}(x) + O(h^2)$$

sur l'équation (4.8), on obtient alors :

$$\mathbb{E}[\hat{g}_1(x)] = \int_{\mathbb{R}} (K(z)g_1(x) - hzg_1^{(1)}(x) + \frac{(zh)^2}{2}g_1^{(2)}(x) + O(h^2))dz.$$

Et par la symétrie du noyau K , on peut écrire :

$$\mathbb{E}[\hat{g}_1(x)] = g_1(x) + \frac{h^2}{2} \int_{\mathbb{R}} (z^2 K(z)g^{(2)}(x))dz + O(h^2).$$

D'où, on conclut que :

$$\mathbb{E}[\hat{g}_1(x)] - g_1(x) = \mathcal{O}(h^2).$$

■

– La preuve du deuxième résultat (ii) est analogue à la preuve de (i) :

En effet,

$$\begin{aligned} \mathbb{E}\hat{g}_2(x) &= \mathbb{E}\left[\frac{1}{nh} \sum_{i=1}^n Y_i^{-2} K\left(\frac{x-X_i}{h}\right)\right] \\ &= \frac{1}{h} \mathbb{E}\left[K\left(\frac{x-X_1}{h}\right) Y_1^{-2}\right] \\ &= \frac{1}{h} \mathbb{E}\left[K\left(\frac{x-X_1}{h}\right) \mathbb{E}(Y_1^{-2} / X = X_1)\right] \\ &= \frac{1}{h} \int_{\mathbb{R}} K\left(\frac{x-u}{h}\right) f(u) \mathbb{E}(Y_1^{-2} / X = X_1) du \\ &= \frac{1}{h} \int_{\mathbb{R}} K\left(\frac{x-u}{h}\right) g_2(u) du. \end{aligned}$$

Puis, en considérant le changement de variables $z = \frac{x-u}{h}$ et on utilisant le développement de Taylor de $g_2(x - hz)$ à l'ordre 2, on obtient :

$$\mathbb{E}[\hat{g}_2(x)] - g_2(x) = \mathcal{O}(h^2).$$

ce qui complète la preuve du lemme (4.2). ■

Démonstration du Lemme (4.3)

– Commençons par la preuve du résultat (i), on a :

$$\begin{aligned}
 \widehat{g}_1(x) - \mathbb{E}\widehat{g}_1(x) &= \frac{1}{nh} \sum_{i=1}^n Y_i^{-1} K\left(\frac{x-X_i}{h}\right) - \mathbb{E}\left[\frac{1}{nh} \sum_{i=1}^n Y_i^{-1} K\left(\frac{x-X_i}{h}\right)\right] \\
 &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h} \left(Y_i^{-1} K\left(\frac{x-X_i}{h}\right) - \mathbb{E}\left[Y_i^{-1} K\left(\frac{x-X_i}{h}\right)\right] \right)
 \end{aligned}$$

Pour appliquer l'inégalité de *Hoeffding* (voir le Lemme 2.6), on pose

$$\Delta_i = \frac{1}{h} \left(Y_i^{-1} K\left(\frac{x-X_i}{h}\right) - \mathbb{E}\left[Y_i^{-1} K\left(\frac{x-X_i}{h}\right)\right] \right).$$

Donc on a :

$$\begin{aligned}
 |\Delta_i| &= \left| \frac{1}{h} \left(Y_i^{-1} K\left(\frac{x-X_i}{h}\right) - \mathbb{E}\left[Y_i^{-1} K\left(\frac{x-X_i}{h}\right)\right] \right) \right| \\
 &\leq \frac{1}{h} \left| \left(Y_i^{-1} K\left(\frac{x-X_i}{h}\right) \right) + \left| \mathbb{E}\left[Y_i^{-1} K\left(\frac{x-X_i}{h}\right)\right] \right| \right| \\
 &\leq \frac{1}{h} (C_1 + C_2) \\
 &\leq \frac{C}{h}.
 \end{aligned}$$

Ce résultat est obtenu par la bornitude du noyau K et Y^{-1} . D'autre part et puisque :

$$\begin{aligned}
 \mathbb{E}(\Delta_i^2) &= \mathbb{E}\left[\left(\frac{1}{h}\left(Y_i^{-1}K\left(\frac{x-X_i}{h}\right) - \mathbb{E}\left[Y_i^{-1}K\left(\frac{x-X_i}{h}\right)\right]\right)\right)^2\right] \\
 &= \frac{1}{h^2}\mathbb{E}\left[\left(Y_i^{-1}K\left(\frac{x-X_i}{h}\right)\right)^2 + \mathbb{E}^2\left[Y_i^{-1}K\left(\frac{x-X_i}{h}\right)\right]\right] \\
 &\quad - 2\left(Y_i^{-1}K\left(\frac{x-X_i}{h}\right)\mathbb{E}\left[Y_i^{-1}K\left(\frac{x-X_i}{h}\right)\right]\right) \\
 &= \frac{1}{h^2}\left[\mathbb{E}\left(Y_i^{-1}K\left(\frac{x-X_i}{h}\right)\right)^2 + \mathbb{E}^2\left[Y_i^{-1}K\left(\frac{x-X_i}{h}\right)\right]\right] \\
 &\quad - 2 + \mathbb{E}^2\left[Y_i^{-1}K\left(\frac{x-X_i}{h}\right)\right] \\
 &= \frac{1}{h^2}\left[\mathbb{E}\left(Y_i^{-1}K\left(\frac{x-X_i}{h}\right)\right)^2 - \mathbb{E}^2\left[Y_i^{-1}K\left(\frac{x-X_i}{h}\right)\right]\right] \\
 &= Var(\nu_i) \\
 &\leq \mathbb{E}(\nu_i^2),
 \end{aligned}$$

tel que $\nu_i = \frac{1}{h}\left[\mathbb{E}\left(Y_i^{-1}K\left(\frac{x-X_i}{h}\right)\right)\right]$.

Alors :

$$\begin{aligned}
 \mathbb{E}(\nu_i^2) &= \mathbb{E}\left[\frac{1}{h^2}\left(Y_i^{-2}K^2\left(\frac{x-X_i}{h}\right)\right)\right] \\
 &= \mathbb{E}\left[\frac{1}{h^2}\left(\mathbb{E}(Y_i^{-2}/X_i)K^2\left(\frac{x-X_i}{h}\right)\right)\right] \\
 &= \frac{1}{h^2}\int \phi(u)f(u))K^2\left(\frac{x-u}{h}\right)du.
 \end{aligned}$$

Avec $\phi(u) = \mathbb{E}(Y_i^{-2}/X_i = u)$.

En utilisant le changement de variable $Z = \frac{x-u}{h}$, on obtient donc :

$$\begin{aligned}\mathbb{E}(\nu_i^2) &= \frac{1}{h} \int \phi(x - hz)f(x - hz)K^2(z)dz \\ &\leq \frac{C}{h} = \delta_2.\end{aligned}$$

Ce dernier résultat est obtenue à cause de la bornitude de ϕ , la continuité de $f(x)$ et aussi car le noyau K est à support compact.

Donc, en appliquant l'inégalité de *Hoeffding*, on aura :

$$\mathbb{P}\left[\frac{1}{n} \left| \sum_{i=1}^n \Delta_i \right| > \epsilon\right] \leq 2 \exp\left(\frac{-n\epsilon^2}{4\delta_2}\right). \quad (4.9)$$

En prenant $\epsilon = \epsilon_0 \sqrt{\frac{\log n}{nh}}$, on obtient

$$\begin{aligned}\mathbb{P}\left[|\mathbb{E}\hat{g}_1(x) - \hat{g}_1(x)| > \epsilon_0 \sqrt{\frac{\log n}{nh}}\right] &\leq 2 \exp\left(\frac{-n\epsilon_0^2 h \log n}{4nhC}\right) \\ &\leq 2 \exp\left(\frac{-\epsilon_0^2 \log n}{4C}\right) \\ &\leq 2 n^{\frac{-\epsilon_0^2}{4C}}.\end{aligned}$$

Ce dernier résultat conduit à :

$$\sum_{i=1}^n \mathbb{P}\left[|\mathbb{E}\hat{g}_1(x) - \hat{g}_1(x)| > \epsilon_0 \sqrt{\frac{\log n}{nh}}\right] < \infty,$$

Sous la condition suffisante suivante $\frac{\epsilon_0^2}{4C} > 1$.

Ce qui prouve le résultat (i) du Lemme (4.3). ■

– La preuve du résultat (ii) est similaire.

En effet :

$$\begin{aligned}\hat{g}_2(x) - \mathbb{E}\hat{g}_2(x) &= \frac{1}{nh} \sum_{i=1}^n Y_i^{-2} K\left(\frac{x - X_i}{h}\right) - \mathbb{E}\left[\frac{1}{nh} \sum_{i=1}^n Y_i^{-2} K\left(\frac{x - X_i}{h}\right)\right] \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h} \left(Y_i^{-2} K\left(\frac{x - X_i}{h}\right) - \mathbb{E}\left[Y_i^{-2} K\left(\frac{x - X_i}{h}\right)\right] \right)\end{aligned}$$

Nous appliquons l'inégalité de précédente (voir Lemme 2.6), avec :

$$\Delta_i = \frac{1}{h} \left(Y_i^{-2} K\left(\frac{x - X_i}{h}\right) - \mathbb{E}\left[Y_i^{-2} K\left(\frac{x - X_i}{h}\right)\right] \right).$$

Puis, nous suivrons les mêmes étapes précédentes, nous obtenons donc :

$$\sum_{i=1}^n \mathbb{P}\left[|\mathbb{E}\widehat{g}_2(x) - \widehat{g}_2(x)| > \epsilon_0 \sqrt{\frac{\log n}{nh}}\right] < \infty.$$

Démonstration du Lemme (4.4)

Commençons par la remarque suivante :

$$\widehat{g}_2(x) \leq \frac{g_2(x)}{2} \Rightarrow |\widehat{g}_2(x) - g_2(x)| \geq \frac{g_2(x)}{2}.$$

Alors, on peut écrire

$$\mathbb{P}\left[\widehat{g}_2(x) \leq \frac{g_2(x)}{2}\right] \Rightarrow \mathbb{P}\left[|\widehat{g}_2(x) - g_2(x)| \geq \frac{g_2(x)}{2}\right].$$

Donc

$$\sum_{i=1}^n \mathbb{P}\left[\widehat{g}_2(x) \leq \frac{g_2(x)}{2}\right] \Rightarrow \sum_{i=1}^n \mathbb{P}\left[|\widehat{g}_2(x) - g_2(x)| \geq \frac{g_2(x)}{2}\right] < \infty.$$

Il suffit de prendre $\delta = \frac{g_2(x)}{2}$. ■

Troisième partie

Régression relative cas fonctionnel

Chapitre 5

Cas d'échantillons indépendants et identiquement distribués

L'étude des 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.

Les résultats que nous énonçons dans ce chapitre, sont liés aux propriétés asymptotiques de l'estimateur à noyau de la régression relative au cas d'une variable explicative fonctionnelle. Nous supposons que l'échantillon que nous étudions est constitué de variables indépendantes et identiquement distribuées. Après la présentation du modèle et son estimateur dans la première section, nous établissons la convergence presque complète (ponctuelle et uniforme) dans la deuxième et la troisième section. La quatrième section est consacrée à l'erreur quadratique moyenne. Nous achevons ce chapitre par un résultat sur la normalité asymptotique.

5.1 Modèle et son estimateur

Soit \mathcal{F} un espace fonctionnel semi-métrique menu de la métrique d . Nous considérons $\{(X_i, Y_i), i = 1, \dots, n\}$ une suite de n -observations indépendantes, de même loi que le couple (X, Y) , à valeurs dans $\mathcal{F} \times \mathbb{R}$. La relation entre X et Y est exprimé par le modèle

suivant :

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

où les erreurs aléatoires ϵ_i sont non corrélées, de moyennes nulles et de variances σ^2 .

Comme au chapitre précédent, $r(X)$ est la solution du problème de minimisation suivant :

$$\min_r \mathbb{E} \left[\left(\frac{Y - r(x)}{Y} \right)^2 \middle| X = x \right], \quad \text{pour } Y > 0. \quad (5.1)$$

avec $r(X) = \frac{\mu(-1)}{\mu(-2)}$,

où : $\mu(-j) = \mathbb{E}(Y^{-j}|X)$, pour $j = 1, 2$.

L'estimateur à noyau K de $r(X)$, où X est une variable aléatoire à valeurs dans \mathcal{F} et $Y \in \mathbb{R}$ est donné par :

$$\hat{r}(X) = \frac{\sum_{i=1}^n Y_i^{-1} K(h^{-1}d(x, X_i))}{\sum_{i=1}^n Y_i^{-2} K(h^{-1}d(x, X_i))}. \quad (5.2)$$

Cet estimateur peut être exprimé par

$$\hat{r}(x) = \frac{\hat{g}_1(x)}{\hat{g}_2(x)},$$

avec

$$\hat{g}_j(x) = \frac{1}{n \mathbb{E}[K(h^{-1}d(x, X_i))]} \sum_{i=1}^n Y_i^{-j} K(h^{-1}d(x, X_i)),$$

C'est à dire :

$$\hat{g}_j(x) = \frac{1}{n} \sum_{i=1}^n Y_i^{-j} \omega_i, \quad \text{pour } j = 1, 2,$$

où ω_i est définie dans la formule (2.4).

5.2 La convergence presque complète ponctuelle

Pour étudier la convergence presque complète, on fixe un point $x \in \mathcal{F}$ tel que \mathcal{N}_x soit un voisinage de ce point, et on introduit les hypothèses suivantes :

5.2.1 Hypothèses et notations

(H.1) $\mathbb{P}(X \in B(x, h)) = \varphi_x(h) > 0$ pour tout $h > 0$; où $B(x, h)$ est la boule fermée, centrée en x et de rayon h , et

$$\begin{cases} \lim_{h \rightarrow 0} \varphi_x(h) = 0, \\ \lim_{n \rightarrow +\infty} \frac{n\varphi_x(h)}{\log n} = +\infty. \end{cases}$$

(H.2) Pour tout $(x_1, x_2) \in \mathcal{N}_x^2$, on a :

$$|g_j(x_1) - g_j(x_2)| \leq Cd^{k_j}(x_1, x_2) \quad j = 1, 2.$$

(H.3) Le noyau K est de type 1.

(H.4) $\mathbb{E}[Y^{-j}/X] < C < +\infty$, $\forall j \geq 2$.

5.2.2 Résultat asymptotique

Théorème 5.1. *Sous les hypothèses précédentes, on a :*

$$|\hat{r}(x) - r(x)| = \mathcal{O}(h^{k_1}) + \mathcal{O}(h^{k_2}) + \mathcal{O}_{p.co}\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right).$$

5.2.3 Démonstration

La démonstration du Théorème (5.1) est basée sur la décomposition suivante :

$$\begin{aligned} \hat{r}(x) - r(x) &= \frac{\hat{g}_1(x)}{\hat{g}_2(x)} - \frac{g_1(x)}{\hat{g}_2(x)} + \frac{g_1(x)}{\hat{g}_2(x)} - r(x) \\ &= \frac{\hat{g}_1(x)}{\hat{g}_2(x)} - \frac{g_1(x)}{g_2(x)} + \frac{r(x)g_2(x)}{\hat{g}_2(x)} - \frac{r(x)\hat{g}_2(x)}{\hat{g}_2(x)} \\ &= \frac{1}{\hat{g}_2(x)}(\hat{g}_1(x) - g_1(x)) + \frac{r(x)}{\hat{g}_2(x)}(g_2(x) - \hat{g}_2(x)) \\ &= \frac{1}{\hat{g}_2(x)}(\hat{g}_1(x) - \mathbb{E}\hat{g}_1(x) + \mathbb{E}\hat{g}_1(x) - g_1(x)) + \frac{r(x)}{\hat{g}_2(x)}(g_2(x) - \mathbb{E}\hat{g}_2(x) + \mathbb{E}\hat{g}_2(x) - \hat{g}_2(x)). \end{aligned} \tag{5.3}$$

Et donc la preuve est une conséquence des lemmes suivants :

Lemme 5.2. *Sous les hypothèses (H.1)-(H.4), on a :*

$$|\mathbb{E}\hat{g}_j(x) - g_j(x)| = \mathcal{O}(h^{k_j}), \quad j = 1, 2.$$

Lemme 5.3. *Sous les hypothèses précédentes, on a :*

$$|\widehat{g}_j(x) - \mathbb{E}\widehat{g}_j(x)| = \mathcal{O}_{p.co}\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right), \quad j = 1, 2.$$

Lemme 5.4. *Sous les hypothèses du Théorème 5.1, on a :*

$$\sum_{i=1}^n \mathbb{P}\left(\widehat{g}_2(x) < \frac{g_2(x)}{2}\right) < \infty.$$

Preuve du Lemme 5.2

Puisque $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ sont indépendants et identiquement distribués, on peut écrire :

$$\begin{aligned} |\mathbb{E}[\widehat{g}_j(x)] - g_j(x)| &= \left| \mathbb{E}\left(\frac{1}{n\mathbb{E}[K(h^{-1}d(x, X_i))]} \sum_{i=1}^n Y^{-j} K(h^{-1}d(x, X_i))\right) - g_j(x) \right| \\ &\leq \left| \frac{1}{\mathbb{E}[K(h^{-1}d(x, X_1))]} \mathbb{E}\left[\mathbb{E}(Y^{-j}/X_1 = x) K(h^{-1}d(x, X_1)) - g_j(x)\right] \right| \\ &\leq \frac{1}{\mathbb{E}[K_1(x)]} [\mathbb{E}[K_1(x)|g_j(X_1) - g_j(x)|]]. \end{aligned}$$

Sous les hypothèses (H.1) et (H.2), nous obtenons :

$$\begin{aligned} |\mathbb{E}[\widehat{g}_j(x)] - g_j(x)| &\leq \frac{1}{\mathbb{E}[K_1]} \mathbb{E}[(K_1(x)) \mathbb{1}_{B(x,h)}) |g_j(X_1) - g_j(x)|] \\ &\leq Ch^{k_j}, \end{aligned}$$

donc :

$$|\mathbb{E}[\widehat{g}_j(x)] - g_j(x)| = \mathcal{O}(h^{k_j}).$$

■

Preuve du Lemme 5.3

On a :

$$\begin{aligned}
 \hat{g}_j(x) - \mathbb{E}(\hat{g}_j(x)) &= \frac{1}{n\mathbb{E}[K(h^{-1}d(x, X_i))]} \sum_{i=1}^n Y_i^{-j} K(h^{-1}d(x, X_i)) \\
 &\quad - \mathbb{E}\left[\frac{1}{n\mathbb{E}[K(h^{-1}d(x, X_i))]} \sum_{i=1}^n Y_i^{-j} K(h^{-1}d(x, X_i))\right] \\
 &= \frac{1}{n\mathbb{E}[K(h^{-1}d(x, X_i))]} \sum_{i=1}^n \left[Y_i^{-j} K(h^{-1}d(x, X_i)) - \mathbb{E}\left(Y_i^{-j} K(h^{-1}d(x, X_i))\right) \right].
 \end{aligned}$$

En posant :

$$\Delta_{i,j} = \frac{1}{\mathbb{E}[K_1]} [K_i Y_i^{-j} - \mathbb{E}[K_i Y_i^{-j}]], \quad \text{pour } j = 1, 2,$$

on aura :

$$\sum_{i=1}^n \Delta_{i,j} = \hat{g}_j(x) - \mathbb{E}[\hat{g}_j(x)].$$

La preuve du Lemme 5.3 est basée sur l'application du Corollaire 2.4. Pour ce faire, nous nous concentrons sur les moments absolus de la variable aléatoire $\Delta_{i,j}$.

En utilisant le binôme de Newton et pour $\ell \leq m$, nous avons d'un coté :

$$\Delta_{i,j}^m = \frac{1}{\mathbb{E}[K_1]^m} [K_i Y_i^{-j} - \mathbb{E}[K_i Y_i^{-j}]]^m,$$

c'est à dire :

$$\mathbb{E}[|\Delta_{i,j}^m|] = \sum_{\ell=0}^m \frac{1}{\mathbb{E}[K_1]^m} \binom{\ell}{m} C' \mathbb{E}[|Y_1^{-j} K_1(x)|]^\ell \mathbb{E}(K_1 Y_1^{-j})^{m-\ell},$$

où $\binom{\ell}{m} = \frac{m!}{\ell!(m-\ell)!}$
et d'autre part, on a :

$$\begin{aligned}
 \mathbb{E}[|Y_1^{-j\ell} K_1^\ell|] &= \mathbb{E}[K_1^\ell \mathbb{E}[|Y_1^{-j\ell}| / X_1]] \\
 &= C \mathbb{E}[K_1^\ell] \\
 &\leq C \varphi_x(h),
 \end{aligned}$$

ce qui implique :

$$\frac{1}{\mathbb{E}^\ell[K_1]} \mathbb{E}[|Y_1^{-j\ell} K_1^\ell|] = \mathcal{O}(\varphi_x(h)^{-\ell+1}),$$

et :

$$\frac{1}{\mathbb{E}[K_1]} \mathbb{E}[|Y_1^{-j} K_1|] \leq C.$$

Alors :

$$\begin{aligned} \mathbb{E}[|\Delta_{i,j}^m|] &\leq C \sum_{\ell=0}^m \frac{1}{(\mathbb{E}[K_1])^k} \mathbb{E}[|Y_1^{-j\ell} K_1^\ell(x)|] \\ &\leq C \max_{\ell=0,\dots,m} \varphi_x^{-\ell+1}(h) \\ &\leq C \varphi_x^{-m+1}(h) \end{aligned}$$

c'est à dire :

$$\mathbb{E}[|\Delta_{i,j}^m|] = \mathcal{O}(\varphi_x^{-m+1}(h)).$$

En appliquant l'inégalité du Corollaire 2.4, on obtient :

$$\begin{aligned} \mathbb{P}\left(n^{-1} \left| \sum_{i=1}^n \Delta_{i,j} \right| > \eta \sqrt{\frac{\log n}{n \varphi_x(h)}}\right) &\leq 2 \exp\left(\frac{-\eta^2 \frac{\log n}{n \varphi_x(h)} n}{\frac{2}{\varphi_x(h)^2}}\right) \\ &\leq 2 \exp(\log n^{\frac{-\eta^2}{2}}) \\ &\leq 2 n^{\frac{-\eta^2}{2}}, \end{aligned}$$

ce qui donne :

$$\mathbb{P}\left(|\hat{g}_j(x) - \mathbb{E}[\hat{g}_j(x)]| > \eta \sqrt{\frac{\log n}{n \varphi_x(h)}}\right) \leq C' n^{C\eta^2},$$

avec $a^2 = \varphi_x(h)^{-1}$.

Pour $j = 1, 2$, nous choisissons η de tel sorte que :

$$\sum_{i=1}^n \mathbb{P}\left(|\hat{g}_j(x) - \mathbb{E}[\hat{g}_j(x)]| > \eta \sqrt{\frac{\log n}{n \varphi_x(h)}}\right) < +\infty.$$

■

Preuve du Lemme 5.4

Afin de prouver le Lemme (5.4), on commence par remarquer que :

$$\hat{g}_2(x) \leq \frac{g_2(x)}{2} \Rightarrow |\hat{g}_2(x) - g_2(x)| \geq \frac{g_2(x)}{2}.$$

Alors, on peut écrire

$$\mathbb{P}\left[\hat{g}_2(x) \leq \frac{g_2(x)}{2}\right] \Rightarrow \mathbb{P}\left[|\hat{g}_2(x) - g_2(x)| \geq \frac{g_2(x)}{2}\right].$$

Donc

$$\sum_{i=1}^n \mathbb{P}\left[\widehat{g}_2(x) \leq \frac{g_2(x)}{2}\right] \Rightarrow \sum_{i=1}^n \mathbb{P}\left[|\widehat{g}_2(x) - g_2(x)| \geq \frac{g_2(x)}{2}\right] < +\infty.$$

ce qui complète la preuve de ce lemme. ■

5.3 La convergence presque complète uniforme

On s'intéresse dans cette section à la version globale de la convergence précédente. Pour cela, on fixe un sous-ensemble $\mathcal{S}_{\mathcal{F}}$ de \mathcal{F} et on note par $\psi_{\mathcal{S}_{\mathcal{F}}}(\cdot)$ la fonction ϵ -entropie de $\mathcal{S}_{\mathcal{F}}$.

5.3.1 Hypothèses et notations

Nous introduisons les hypothèses suivantes :

(U.1) Pour tout $x \in \mathcal{S}_{\mathcal{F}}$ et $h > 0$:

$$0 < C\varphi_x(h) \leq \mathbb{P}(X \in B(x, h)) \leq C'\varphi_x(h) < \infty.$$

(U.2) Il existe $\eta > 0$, tel que :

$$\forall x, x' \in \mathcal{S}_{\mathcal{F}}^\eta, \quad |g_j(x) - g_j(x')| \leq Cd^k(x, x'),$$

tel que $\mathcal{S}_{\mathcal{F}}^\eta = \{x \in \mathcal{F} : \exists x' \in \mathcal{S}_{\mathcal{F}}, \quad d(x, x') \leq \eta\}$.

(U.3) Le noyau K est borné et lipschitzien de support $[0, 1]$.

(U.4) Les fonctions φ_x et $\psi_{\mathcal{S}_{\mathcal{F}}}$ sont :

(U.4.a) Il existe $\eta_0 > 0$ tel que pour tout $\eta < \eta_0$, $\varphi'_x(\eta) < C$, où φ'_x est la première dérivée de la fonction φ_x .

(U.4.b) Pour un entier assez grand n , on a :

$$\frac{(\log n)^2}{n\varphi_x(h)} < \psi_{\mathcal{S}_{\mathcal{F}}}\left(\frac{\log n}{n}\right) < \frac{n\varphi_x(h)}{\log n}.$$

(U.4.c) La fonction ϵ -entropie de kolmogorov satisfaite :

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

(U.5) Pour $j \geq 2$,

$$\mathbb{E}(|Y^{-j}|/X = x) < C < +\infty \quad \forall x \in \mathcal{S}_{\mathcal{F}} \text{ et } \inf_{x \in \mathcal{S}_{\mathcal{F}}} g_2(x) \geq C' > 0.$$

5.3.2 Résultat asymptotique

Théorème 5.5. *Sous les hypothèses (U.1)-(U.5), on a :*

$$\sup_{x \in \mathcal{S}_{\mathcal{F}}} |\hat{r}(x) - r(x)| = \mathcal{O}(h^{k_1}) + \mathcal{O}(h^{k_2}) + \mathcal{O}_{p.co}\left(\sqrt{\frac{\psi_{\mathcal{S}_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\varphi_x(h)}}\right).$$

5.3.3 Démonstration

La preuve du Théorème 5.5 est basée sur la décomposition (5.3) et les lemmes suivants :

Lemme 5.6. *Sous les hypothèses (U.1) et (U.3)-(U.5), on a :*

$$\sup_{x \in \mathcal{S}_{\mathcal{F}}} |\mathbb{E}\hat{g}_j(x) - g_j(x)| = \mathcal{O}(h^{k_j}), \quad j = 1, 2.$$

Lemme 5.7. *Sous les hypothèses (U.1)-(U.4), on a :*

$$\sup_{x \in \mathcal{S}_{\mathcal{F}}} |\hat{g}_j(x) - \mathbb{E}\hat{g}_j(x)| = \mathcal{O}_{p.co}\left(\sqrt{\frac{\psi_{\mathcal{S}_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\varphi_x(h)}}\right), \quad j = 1, 2.$$

Lemme 5.8. *Sous les hypothèses du Lemme 5.7, nous obtenons :*

$$\exists \delta > 0, \quad \sum_{n=1}^{\infty} \mathbb{P}\left(\inf_{x \in \mathcal{S}_{\mathcal{F}}} \hat{g}_2(x) < \delta\right) < \infty.$$

Preuve du Lemme 5.6

La preuve du Lemme 5.6 est similaire à la preuve du Lemme 5.2, tel que :

$$\begin{aligned} |\mathbb{E}[\hat{g}_j(x)] - g_j(x)| &= \left| \mathbb{E}\left(\frac{1}{n\mathbb{E}[K(h^{-1}d(x, X_i))]} \sum_{i=1}^n Y^{-j} K(h^{-1}d(x, X_i))\right) - g_j(x) \right| \\ &\leq \left| \frac{1}{\mathbb{E}[K(h^{-1}d(x, X_1))]} \mathbb{E}\left[\mathbb{E}(Y^{-j}/X_1 = x) K(h^{-1}d(x, X_1)) - g_j(x)\right] \right| \\ &\leq \frac{1}{\mathbb{E}[K_1(x)]} [\mathbb{E}[K_1(x)|g_j(X_1) - g_j(x)|]]. \end{aligned}$$

Sous les hypothèses (U.1) et (U.2), $\forall x \in \mathcal{S}_{\mathcal{F}}$ nous obtenons :

$$\begin{aligned} |\mathbb{E}[\hat{g}_j(x)] - g_j(x)| &\leq \frac{1}{\mathbb{E}[K_1]} \mathbb{E}[(K_1(x)) \mathbb{1}_{B(x,h)} |g_j(X_1) - g_j(x)|] \\ &\leq Ch^{k_j}, \end{aligned}$$

où C ne dépend pas de x , donc :

$$\sup_{x \in \mathcal{S}_{\mathcal{F}}} |\mathbb{E}[\hat{g}_j(x)] - g_j(x)| = \mathcal{O}(h^{k_j}).$$

■

Preuve du Lemme 5.7

Soit x_1, x_2, \dots, x_N un ensemble fini de points de \mathcal{F} tel que :

$$\mathcal{S}_{\mathcal{F}} \subset \bigcup_{k=1}^N B(x_k, \epsilon), \quad \text{avec } \epsilon = \frac{\log n}{n},$$

pour tout $x \in \mathcal{S}_{\mathcal{F}}$, on note :

$$K(x) = \arg \min_{k \in \{1, 2, \dots, N_{\epsilon}(\mathcal{S}_{\mathcal{F}})\}} d(x, x_k), \quad \text{et} \quad K_i(x) = K\left(\frac{d(x, X_i)}{h}\right).$$

La preuve de ce lemme est basée sur la décomposition suivante :

$$\begin{aligned} \sup_{x \in \mathcal{S}_{\mathcal{F}}} |\hat{g}_j(x) - \mathbb{E}[\hat{g}_j(x)]| &\leq \underbrace{\sup_{x \in \mathcal{S}_{\mathcal{F}}} |\hat{g}_j(x) - \hat{g}_j(x_{k(x)})|}_{F_1} + \underbrace{\sup_{x \in \mathcal{S}_{\mathcal{F}}} |\hat{g}_j(x_{k(x)}) - \mathbb{E}[\hat{g}_j(x_{k(x)})]|}_{F_2} \\ &\quad + \underbrace{\sup_{x \in \mathcal{S}_{\mathcal{F}}} |\mathbb{E}[\hat{g}_j(x_{k(x)})] - \mathbb{E}[\hat{g}_j(x)]|}_{F_3}, \end{aligned}$$

donc, il suffit de traiter chaque terme F_i pour $i = 1, 2, 3$:

– Pour le terme F_1 , sous l'hypothèse (U.1) et donc de la condition :

$$C\varphi_x(h) \leq \mathbb{E}[K_1(x)] \leq C'\varphi_x(h),$$

nous obtenons :

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

puisque le noyau K est lipschitzien sur $[0, 1]$ d'après l'hypothèse (U.3) (c'est à dire, le cas où $K(1) = 0$), nous obtenons :

$$F_1 \leq \frac{C}{n} \sum_{i=1}^n Q_i, \tag{5.4}$$

avec :

$$Q_i = \frac{\epsilon Y^{-j}}{h\varphi_x(h)} \sup_{x \in \mathcal{S}_{\mathcal{F}}} \mathbb{1}_{B((x,h) \cup B(x_{k(x)},h))}(X_i).$$

Appliquons l'inégalité exponentielle 2.4 (i.i) sur Q_i .

Nous calculons donc la quantité $\mathbb{E}[|Q_i|^m]$.

Sous l'hypothèse (U.5) ($\mathbb{E}(|Y^{-jm}|/X = x) < C < \infty$), nous écrivons :

$$\mathbb{E}[|Q_i|^m] \leq \frac{C \epsilon^m}{h^m \varphi_x(h)^{m-1}},$$

pour l'application de l'inégalité précédente, nous choisissons $a^2 = \frac{\epsilon}{h\varphi_x(h)}$.

Nous obtenons donc :

$$F_1 = \mathcal{O}_{p.co.} \left(\sqrt{\frac{\psi_{\mathcal{S}_{\mathcal{F}}} \left(\frac{\log n}{n} \right)}{n \varphi_x(h)}} \right).$$

D'autre part, le cas où le noyau K est lipschitzien sur $[0, 1]$ (c'est à dire, le cas où $K(1) > C > 0$) il faut décomposer le terme F_1 en trois termes comme suit :

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

avec :

$$\begin{cases} F_{11} = \frac{1}{\varphi_x(h)} \sum_{i=1}^n |K_i(x) - K_i(x_{k(x)})| Y_i^{-j} \mathbb{1}_{B(x,h) \cap B(x_{k(x)},h)}(X_i); \\ F_{12} = \frac{1}{\varphi_x(h)} \sum_{i=1}^n K_i(x) Y_i^j \mathbb{1}_{B(x,h) \cap \overline{B(x_{k(x)},h)}}(X_i); \\ F_{13} = \frac{1}{\varphi_x(h)} \sum_{i=1}^n K_i(x_{k(x)}) Y_i^{-j} \mathbb{1}_{\overline{B(x,h)} \cap B(x_{k(x)},h)}(X_i). \end{cases}$$

où \overline{A} est le complémentaire de l'ensemble A .

Pour ce qui concerne le terme F_{11} , nous utilisons le fait que K est une fonction lipschitzienne sur $[0, 1]$, et on a :

$$F_{11} \leq \sup_{x \in \mathcal{S}_{\mathcal{F}}} \frac{C}{n} \sum_{i=1}^n Z_{i,j}, \quad \text{avec } Z_{i,j} = \frac{\epsilon}{h\varphi_x(h)} Y_i^{-j} \mathbb{1}_{B(x,h) \cap B(x_{k(x)},h)}(X_i) Y_i^{-j}.$$

Concernant les termes F_{12} et F_{13} , comme K est borné, nous obtenons :

$$F_{12} \leq \frac{C}{n} \sum_{i=1}^n W_{i,j}, \quad \text{avec } W_{i,j} = \frac{1}{\varphi_x(h)} Y_i^{-j} \mathbb{1}_{B(x,h) \cap \overline{B(x_{k(x)},h)}}(X_i),$$

et :

$$F_{13} \leq \frac{C}{n} \sum_{i=1}^n V_{i,j}, \quad \text{avec } V_{i,j} = \frac{1}{\varphi_x(h)} Y_i^{-j} \mathbb{1}_{\overline{B(x,h)} \cap B(x_{k(x)},h)}(X_i),$$

donc, nous remplaçons Δ_i dans l'inégalité exponentielle (voir le Corollaire 2.4) respectivement par $Z_{i,j}$, $W_{i,j}$ et $V_{i,j}$ avec $a^2 = \frac{\epsilon}{h\varphi_x(h)}$, nous obtenons :

$$F_{11} = \mathcal{O}_{p.co.}\left(\sqrt{\frac{\epsilon \log n}{nh\varphi_x(h)}}\right),$$

$$F_{12} = \mathcal{O}\left(\frac{\epsilon}{\varphi_x(h)}\right) + \mathcal{O}_{p.co.}\left(\sqrt{\frac{\epsilon \log n}{n\varphi_x(h)^2}}\right),$$

et

$$F_{13} = \mathcal{O}\left(\frac{\epsilon}{\varphi_x(h)}\right) + \mathcal{O}_{p.co.}\left(\sqrt{\frac{\epsilon \log n}{n\varphi_x(h)^2}}\right).$$

D'autre part, d'après les conditions (U.4.a) et (U.4.b), on a :

$$F_1 = \mathcal{O}_{p.co.}\left(\sqrt{\frac{\psi_{\mathcal{S}_F}(\epsilon)}{n\varphi_x(h)}}\right).$$

En suivant la même démarche, nous pouvons obtenir :

$$F_3 = \mathcal{O}_{p.co.}\left(\sqrt{\frac{\psi_{\mathcal{S}_F}(\epsilon)}{n\varphi_x(h)}}\right).$$

– Pour le terme F_2 , on remarque que pour tout $\eta > 0$:

$$\begin{aligned} \mathbb{P}\left(F_2 > \eta \sqrt{\frac{\psi_{\mathcal{S}_F}(\epsilon)}{n\varphi_x(h)}}\right) &= \mathbb{P}\left(\max_{\{k \in 1, 2, \dots, N\}} |\widehat{g}_j(x_{k(x)}) - \mathbb{E}[\widehat{g}_j(x_{k(x)})]| > \eta \sqrt{\frac{\psi_{\mathcal{S}_F}(\epsilon)}{n\varphi_x(h)}}\right) \\ &\leq N \max_{\{k \in 1, 2, \dots, N\}} \mathbb{P}\left(|\widehat{g}_j(x_{k(x)}) - \mathbb{E}[\widehat{g}_j(x_{k(x)})]| > \eta \sqrt{\frac{\psi_{\mathcal{S}_F}(\epsilon)}{n\varphi_x(h)}}\right), \end{aligned}$$

Appliquons maintenant, l'inégalité exponentielle du Corollaire 2.4, pour :

$$\Delta_{i,j} = \frac{1}{\mathbb{E}[K_1(x_{k(x)})]} [K_i(x_{k(x)}) Y_i^{-j} - \mathbb{E}[K_i(x_{k(x)}) Y_i^{-j}]],$$

puisque $\mathbb{E}[|\Delta_{i,j}|]^m = \mathcal{O}(\varphi_x(h)^{-m+1})$, alors, on peut prendre $a^2 = \frac{1}{\varphi_x(h)}$, par suite, pour tout $\eta > 0$:

$$\begin{aligned} \mathbb{P}\left(|\widehat{g}_j(x_{k(x)}) - \mathbb{E}[\widehat{g}_j(x_{k(x)})]| > \eta \sqrt{\frac{\psi_{\mathcal{S}_F}(\epsilon)}{n\varphi_x(h)}}\right) &= \mathbb{P}\left(\frac{1}{n} \left| \sum_{i=1}^n \Delta_{i,j} \right| > \eta \sqrt{\frac{\psi_{\mathcal{S}_F}(\epsilon)}{n\varphi_x(h)}}\right) \\ &\leq 2 \exp(-C\eta^2 \psi_{\mathcal{S}_F}(\epsilon)), \end{aligned}$$

En utilisant le fait que $\psi_{\mathcal{S}_F}(\epsilon) = \log N$, et en choisissant η de tel sorte que $C\eta^2 = \beta$, et d'après l'hypothèse **(U.4.c)** nous obtenons :

$$N \max_{\{k \in 1, 2, \dots, N\}} \mathbb{P}\left(|\hat{g}_j(x_{k(x)}) - \mathbb{E}[\hat{g}_j(x_{k(x)})]| > \eta \sqrt{\frac{\psi_{\mathcal{S}_F}(\epsilon)}{n\varphi_x(h)}}\right) \leq CN^{1-\beta}.$$

Preuve du Lemme 5.8

Nous remarquons que :

$$\inf_{x \in \mathcal{S}_F} \leq \frac{g_2(x)}{2} \Rightarrow \exists x \in \mathcal{S}_F \quad \text{tel que} \quad g_2(x) - \hat{g}_2(x) \geq \frac{g_2(x)}{2},$$

qui implique aussi que $\sup_{x \in \mathcal{S}_F} |g_2(x) - \hat{g}_2(x)| \geq \frac{g_2(x)}{2}$, on déduit du Lemme 5.6 que :

$$\mathbb{P}\left(\inf_{x \in \mathcal{S}_F} \leq \frac{g_2(x)}{2}\right) \leq \mathbb{P}\left(\sup_{x \in \mathcal{S}_F} |g_2(x) - \hat{g}_2(x)| \geq \frac{g_2(x)}{2}\right),$$

donc :

$$\sum_{i=1}^{\infty} \mathbb{P}\left(\inf_{x \in \mathcal{S}_F} \leq \frac{g_2(x)}{2}\right) < \infty.$$

5.4 La convergence en moyenne quadratique

On remplace **(H.1)**, **(H.3)** et **(H.4)** respectivement par les hypothèses suivantes :

5.4.1 Hypothèses et notations

(M.1) La propriété de concentration **(H.1)** est satisfaite, de plus, il existe une fonction $\mathcal{X}_x(\cdot)$ tel que :

$$\forall s \in [0, 1], \quad \lim_{r \rightarrow 0} \frac{\varphi_x(sr)}{\varphi_x(r)} = \mathcal{X}_x(s).$$

(M.2) La fonction $\Psi_j(\cdot) = \mathbb{E}\left[g_j(X) - g_j(x)/d(x, X) = \cdot\right]$, $j = 1, 2$ est dérivable au point 0.

(M.3) Le noyau K est de type 2.

(M.4) La probabilité de petite boule satisfaite :

$$n\varphi_x(h) \rightarrow \infty.$$

(M.5) Les fonctions $\mathbb{E}[Y^{-j}/X = x]$ sont continues au voisinage de x , pour $j = 1, 2, 3, 4$.

5.4.2 Résultat asymptotique

Théorème 5.9. *Sous les hypothèses (M.1)-(M.5), nous obtenons :*

$$\mathbb{E}[\hat{r}(x) - r(x)]^2 = B_n^2(x)h^2 + \frac{\sigma^2(x)}{n\varphi_x(h)} + o(h) + o\left(\frac{1}{n\varphi_x(h)}\right),$$

où :

$$\sigma^2(x) = \frac{(g_2(x) - 2r(x)\mathbb{E}[Y^{-3}/X=x] + r^2(x)\mathbb{E}[Y^{-4}/X=x])\beta_2}{g_2^2(x)\beta_1^2}$$

et :

$$B_n(x) = \frac{(\Psi'_1(0) - r(x)\Psi'_2(0))\beta_0}{\beta_1 g_2(x)}, \quad (5.5)$$

avec : $\beta_0 = K(1) - \int_0^1 (sK(s))' \mathcal{X}_x(s) ds$ et $\beta_j = K^j(1) - \int_0^1 (K^j)'(s) \mathcal{X}_x(s) ds$ pour $j = 1, 2$.

5.4.3 Démonstration

La preuve du Théorème 5.9 se débute par écrire le biais et la variance sous les formules suivantes :

$$\mathbb{E}[\hat{r}(x)] = \frac{\mathbb{E}[\hat{g}_1(x)]}{\mathbb{E}[\hat{g}_2(x)]} + \mathcal{O}\left(\frac{1}{n\varphi_x(h)}\right),$$

et :

$$\begin{aligned} \text{Var}[\hat{r}(x)] &= \frac{\text{Var}[\hat{g}_1(x)]}{(\mathbb{E}[\hat{g}_2(x)])^2} - 2\frac{\mathbb{E}[\hat{g}_1(x)]\text{Cov}(\hat{g}_1(x), \hat{g}_2(x))}{(\mathbb{E}[\hat{g}_2(x)])^3} \\ &\quad + \frac{[\hat{g}_2(x)](\mathbb{E}[\hat{g}_1(x)])^2}{(\mathbb{E}[\hat{g}_2(x)])^4} + o\left(\frac{1}{n\varphi_x(h)}\right), \end{aligned}$$

Donc, la démonstration est une conséquence des deux lemmes suivants :

Lemme 5.10. *Sous les hypothèses du Théorème 5.9, on a :*

$$\mathbb{E}[\hat{g}_j(x)] = g_j(x) + \Psi'_j(0)\frac{\beta_0}{\beta_1}h + o(h), \quad \text{pour } j = 1, 2.$$

Lemme 5.11. *Sous les hypothèses précédentes, on a pour $j = 1, 2$:*

$$\text{Var}[\hat{g}_j(x)] = \mathbb{E}[Y^{-2j}/X=x]\frac{\beta_2}{\beta_1^2 n\varphi_x(h)} + o\left(\frac{1}{n\varphi_x(h)}\right),$$

et

$$\text{Cov}(\hat{g}_1(x), \hat{g}_2(x)) = \mathbb{E}[Y^{-3}/X=x]\frac{\beta_2}{\beta_1^2 n\varphi_x(h)} + o\left(\frac{1}{n\varphi_x(h)}\right).$$

Preuve de Lemme 5.10

D'après l'équidistribution des observations, nous écrivons pour $j = 1, 2$:

$$\begin{aligned}\mathbb{E}[\widehat{g}_j(x)] &= \mathbb{E}\left[\frac{1}{n\mathbb{E}[k(h^{-1}d(x, X_i))]} \sum_{i=1}^n Y_i^{-j} k(h^{-1}d(x, X_i))\right] \\ &= \frac{1}{\mathbb{E}[K_1]} \mathbb{E}[K_1 \mathbb{E}[Y_1^{-j} / X_1]].\end{aligned}$$

D'autre part nous avons :

$$\begin{aligned}\mathbb{E}[K_1 \mathbb{E}[Y_1^{-j} / X_1]] &= g_j(x) \mathbb{E}[K_1] + \mathbb{E}[K_1 \mathbb{E}[g_j(X_1) - g_j(x)/d(X_1, x)]] \\ &= g_j(x) \mathbb{E}[K_1] + \mathbb{E}[K_1 \Psi_j(d(X_1, x))],\end{aligned}$$

d'où, la définition de Ψ_j , nous permet d'écrire :

$$\mathbb{E}\widehat{g}_j(x) = g_j(x) + \frac{1}{\mathbb{E}[K_1]} \mathbb{E}[K_1 \Psi_j(d(X_1, x))],$$

car $\Psi(0) = 0$, nous obtenons :

$$\mathbb{E}[K_1 \Psi_j(d(X_1, x))] = \Psi'_j(0) \mathbb{E}[d(X_1, x) K_1] + o(\mathbb{E}[d(X_1, x) K_1])$$

c'est à dire :

$$\mathbb{E}\widehat{g}_j(x) = g_j(x) + \frac{1}{\mathbb{E}[K_1]} \Psi'_j(0) \mathbb{E}[d(X_1, x) K_1] + o(\mathbb{E}[d(X_1, x) K_1]) \quad (5.6)$$

D'autre part pour, $j = 1, 2$ et sous l'hypothèse (M.1) :

$$\begin{aligned}\mathbb{E}(K_1^j) &= \int_0^1 K^j(s) dP^{|d(x, X)|h^{-1}}(s) \\ &= \int_0^1 \left[K^j(1) - \int_s^1 (K^j(s))' ds \right] dP^{|d(x, X)|h^{-1}}(s) \\ &= K^j(1) \varphi_x(h) - \int_s^1 (K^j(s))' \varphi_x(sh) ds \\ &= \varphi_x(h) \left[K^j(1) - \int_s^1 (K^j(s))' \mathcal{X}_x(s) ds \right] + o(\varphi_x(h)) \\ &= \varphi_x(h) \beta_j + o(\varphi_x(h)),\end{aligned} \quad (5.7)$$

et :

$$\begin{aligned}
 \mathbb{E}[K_1 d(x, X_1)] &= h \int_0^1 s K(s) dP|d(x, X)|h^{-1}(s) \\
 &= h \int_0^1 \left[K(1) - \int_s^1 (s K(s))' ds \right] dP|d(x, X)|h^{-1}(s) \\
 &= h \left[K(1) \varphi_x(h) - \int_s^1 (s K(s))' \varphi_x(sh) ds \right] \\
 &= h \varphi_x(h) \left[K(1) - \int_s^1 (s K(s))' \mathcal{X}_x(s) ds \right] + o(h \varphi_x(h)) \\
 &= h \varphi_x(h) \beta_0 + o(h \varphi_x(h)).
 \end{aligned} \tag{5.8}$$

On remplace (5.7) pour $j = 1$ et (5.8) dans (5.6), nous obtenons :

$$\mathbb{E}\hat{g}_j(x) = g_j(x) + \frac{h \Psi'_j(0) \varphi_x(h) \left[K(1) - \int_s^1 (s K(s))' \mathcal{X}_x(s) ds \right]}{\varphi_x(h) \left[K(1) - \int_s^1 (K(s))' \mathcal{X}_x(s) ds \right]} + o\left(\frac{h \varphi_x(h)}{\varphi_x(h)}\right),$$

ainsi :

$$\mathbb{E}\hat{g}_j(x) = g_j(x) + h \Psi'_j(0) \left[\frac{K(1) - \int_0^1 (s K(s))' \mathcal{X}_x(s) ds}{K(1) - \int_0^1 K'(s) \mathcal{X}_x(s) ds} \right] + o(h).$$

Ce qui complète la preuve du Lemme 5.10. ■

Preuve du Lemme 5.11

On a pour $j = 1, 2$:

$$\begin{aligned}
 \text{Var}[\hat{g}_j(x)] &= \frac{1}{(n \mathbb{E}[K_1])^2} \sum_{i=1}^n \text{Var}[K_i Y_i^{-j}] \\
 &= \frac{1}{n (\mathbb{E}[K_1])^2} \text{Var}[K_1 Y_1^{-j}] \\
 &= \frac{1}{n (\mathbb{E}[K_1])^2} \left(\mathbb{E}(K_1^2 Y_1^{-2j}) - \mathbb{E}^2(K_1 Y_1^{-j}) \right).
 \end{aligned}$$

On conditionne par rapport X , de plus sous l'hypothèse (M.1) puis (M.4) nous obtenons :

$$\mathbb{E}[K_1^2 Y_1^{-2j}] = \mathbb{E}[Y^{-2j} / X = x] \mathbb{E}(K_1^2),$$

d'après (5.7), nous obtenons :

$$\mathbb{E}[K_1^2 Y_1^{-2j}] = \mathbb{E}[Y^{-2j} / X = x] \left(K^2(1) - \int_0^1 (K^2(s))' \mathcal{X}_x(u) du \right) + o(1),$$

d'autre part et d'après l'hypothèse (M.5), on a :

$$\begin{aligned} \mathbb{E}[K_1 Y_1^{-j}] &= \mathbb{E}[K_1 \mathbb{E}[Y_1^{-j}]] \\ &\leq C \varphi_x(h) \\ &= \mathcal{O}(\varphi_x(h)), \end{aligned}$$

Donc :

$$Var[K_1 Y_1^{-j}] = \mathbb{E}[Y^{-2j} / X = x] \left(K^2(1) - \int_0^1 (K^2(s))' \mathcal{X}_x(u) du \right) + \mathcal{O}(\varphi_x^2(h)),$$

nous pouvons écrire donc :

$$Var[\hat{g}_j(x)] = \frac{\mathbb{E}[Y^{-2j} / X = x] \left(K^2(1) - \int_0^1 (K^2(s))' \mathcal{X}_x(u) du \right)}{n \varphi_x(h) \left(K(1) - \int_0^1 (K(s))' \mathcal{X}_x(s) ds \right)^2} + o\left(\frac{1}{n \varphi_x(h)}\right).$$

Nous faisons les mêmes étapes pour le terme de covariance, c'est à dire :

$$Cov(\hat{g}_1(x), \hat{g}_2(x)) = \frac{1}{n(\mathbb{E}[K_1])^2} Cov(K_1 Y_1^{-2}, K_1 Y_1^{-1}),$$

où :

$$Cov(K_1 Y_1^{-2}, K_1 Y_1^{-1}) = \mathbb{E}[K_1^2 Y_1^{-3}] - \mathbb{E}[K_1 Y_1^{-2}] \mathbb{E}[K_1 Y_1^{-1}].$$

D'après (5.7), nous obtenons :

$$Cov(\hat{g}_1(x), \hat{g}_2(x)) = \frac{\mathbb{E}[Y^{-3} / X = x] \left(K^2(1) - \int_0^1 (K^2(s))' \mathcal{X}_x(u) du \right)}{n \varphi_x(h) \left(K(1) - \int_0^1 (K(s))' \mathcal{X}_x(s) ds \right)^2} + o\left(\frac{1}{n \varphi_x(h)}\right).$$

Ce qui termine la preuve du Lemme 5.11. ■

5.5 La normalité asymptotique

Les premiers travaux s'intéressant à la normalité asymptotique de l'estimateur à noyau (5.2) sont dus à *Masry* [18]. Il a considéré le cas d'un échantillon constitué de variables α -mélangeantes mais il n'a pas donné pas l'expression des termes asymptotiquement dominants du biais et de la variance.

Parallèlement, *Ferraty et al.* [10] ont obtenu l'expression explicite de la loi asymptotique (c'est à dire des termes dominants du biais et de la variance) dans le cas d'un échantillon de variables indépendantes.

Le fait d'avoir explicité les termes dominants du biais 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 quadratique moyenne.

5.5.1 Hypothèses et notations

On garde les mêmes hypothèses de la convergence en moyenne quadratique.

5.5.2 Résultat asymptotique

Théorème 5.12. *Sous les hypothèses (M.1)-(M.5), pour tout $x \in \mathcal{F}$ on a :*

$$\left(\frac{n\varphi_x(h)}{\sigma^2(x)} \right)^{\frac{1}{2}} (\hat{r}(x) - r(x) - B_n(x) - o(h)) \xrightarrow{D} \mathcal{N}(0, 1), \quad \text{quand } n \rightarrow \infty.$$

Où \xrightarrow{D} est la convergence en distribution.

5.5.3 Démonstration

La preuve du Résultat 5.12 est basée sur la décomposition relativement complexe suivante suivante :

$$\begin{aligned}
 \hat{r}(x) - r(x) &= \frac{\hat{g}_1(x)}{\hat{g}_2(x)} - \frac{\mathbb{E}[\hat{g}_1(x)]}{\mathbb{E}[\hat{g}_2(x)]} + \frac{r(x)}{\hat{g}_2} [\mathbb{E}[\hat{g}_2(x)] - \hat{g}_2(x)] + \frac{r(x)}{\hat{g}_2} [\hat{g}_2(x) - \mathbb{E}[\hat{g}_2(x)]] \\
 &\quad - \frac{\mathbb{E}[\hat{g}_1(x)]}{\mathbb{E}[\hat{g}_2(x)]} + \frac{\mathbb{E}[\hat{g}_1(x)]}{\mathbb{E}[\hat{g}_2(x)]} + \frac{\mathbb{E}[\hat{g}_1(x)]}{\hat{g}_2(x)} - \frac{g_1(x)}{g_2(x)} \\
 &= \frac{\hat{g}_1(x)}{\hat{g}_2(x)} - \frac{\mathbb{E}[\hat{g}_1(x)]}{\hat{g}_2(x)} + \frac{r(x)}{\hat{g}_2} [\mathbb{E}[\hat{g}_2(x)] - \hat{g}_2(x)] + \frac{r(x)}{\hat{g}_2} [\hat{g}_2(x) - \mathbb{E}[\hat{g}_2(x)]] \\
 &\quad - \frac{1}{\hat{g}_2(x)} \frac{\mathbb{E}[\hat{g}_1(x)]}{\mathbb{E}[\hat{g}_2(x)]} (\hat{g}_2(x) - \mathbb{E}[\hat{g}_2(x)]) + \frac{\mathbb{E}[\hat{g}_1(x)]}{\mathbb{E}\hat{g}_2(x)} - \frac{g_1(x)}{g_2(x)} \\
 &= \frac{1}{\hat{g}_2(x)} \left(\hat{g}_1(x) - \mathbb{E}[\hat{g}_1(x)] + (\mathbb{E}[\hat{g}_2(x)] - \hat{g}_2(x)) \frac{g_1(x)}{g_2(x)} \right. \\
 &\quad \left. - \left(\frac{\mathbb{E}[\hat{g}_1(x)]}{\mathbb{E}[\hat{g}_2(x)]} - \frac{g_1(x)}{g_2(x)} \right) [\hat{g}_2(x) - \mathbb{E}[\hat{g}_2(x)]] \right) + \frac{\mathbb{E}[\hat{g}_1(x)]}{\mathbb{E}\hat{g}_2(x)} - \frac{g_1(x)}{g_2(x)}.
 \end{aligned}$$

Pour simplifier la preuve, on pose :

$$A_n = \frac{1}{\mathbb{E}[\hat{g}_2(x)] g_2(x)} [\mathbb{E}[\hat{g}_1(x)] g_2(x) - \mathbb{E}[\hat{g}_2(x)] g_1(x)],$$

et :

$$D_n = \frac{1}{g_2(x)} \left[\left[\hat{g}_1(x) - \mathbb{E}[\hat{g}_1(x)] \right] g_2(x) + \left[\mathbb{E}[\hat{g}_2(x)] - \hat{g}_2(x) \right] g_1(x) \right].$$

Donc la décomposition précédente devient :

$$\hat{r}(x) - r(x) = \frac{1}{\hat{g}_2(x)} [D_n - A_n (\hat{g}_2(x) - \mathbb{E}[\hat{g}_2(x)])] + A_n.$$

Par suite, la démonstration du résultat précédent est une conséquence des lemmes suivants :

Lemme 5.13. *Sous les hypothèses du Théorème 5.12, nous obtenons :*

$$\left(\frac{n\varphi_x(h)}{g_2^2(x)\sigma^2(x)} \right)^{\frac{1}{2}} \left(\left[\hat{g}_1(x) - \mathbb{E}[\hat{g}_1(x)] \right] g_2(x) + \left[\mathbb{E}[\hat{g}_2(x)] - \hat{g}_2(x) \right] g_1(x) \right) \xrightarrow{D} \mathcal{N}(0, 1).$$

C'est à dire :

$$\frac{\sqrt{n\varphi_x(h)}}{\sigma(x)} D_n \xrightarrow{D} \mathcal{N}(0, 1).$$

Lemme 5.14. *Sous les hypothèses du Théorème 5.12, nous obtenons :*

$$\widehat{g}_2(x) \xrightarrow{P} g_2(x),$$

et

$$\left(\frac{n\varphi_x(h)}{g_2^2(x)\sigma^2(x)} \right)^{\frac{1}{2}} A_n(\widehat{g}_2(x) - \mathbb{E}[\widehat{g}_2(x)]) \xrightarrow{P} 0.$$

Preuve du Lemme 5.13

Pour la preuve du Lemme 5.13, on a :

$$\begin{aligned} \left[\widehat{g}_1(x) - \mathbb{E}[\widehat{g}_1(x)] \right] g_2(x) + \left[\mathbb{E}[\widehat{g}_2(x)] - \widehat{g}_2(x) \right] g_1(x) &= g_1(x)\widehat{g}_2(x) - g_1(x)\mathbb{E}[\widehat{g}_2(x)] - g_2(x)\widehat{g}_1(x) \\ &\quad + g_2(x)\mathbb{E}[\widehat{g}_1(x)] \\ &= g_1(x)\widehat{g}_2(x) - g_2(x)\widehat{g}_1(x) + g_2(x)\mathbb{E}[\widehat{g}_1(x)] \\ &\quad - g_1(x)\mathbb{E}[\widehat{g}_2(x)], \end{aligned}$$

d'une part, on a :

$$\begin{aligned} g_1(x)\widehat{g}_2(x) - g_2(x)\widehat{g}_1(x) &= \frac{g_1(x)}{n\mathbb{E}[K(h^{-1}d(x, X_i))]} \sum_{i=1}^n Y_i^{-2} K(h^{-1}d(x, X_i)) \\ &\quad - \frac{g_2(x)}{n\mathbb{E}[K(h^{-1}d(x, X_i))]} \sum_{i=1}^n Y_i^{-1} K(h^{-1}d(x, X_i)) \\ &= \sum_{i=1}^n \left[\frac{g_1(x)}{n\mathbb{E}[K(h^{-1}d(x, X_i))]} Y_i^{-2} K(h^{-1}d(x, X_i)) \right. \\ &\quad \left. - \frac{g_2(x)}{n\mathbb{E}[K(h^{-1}d(x, X_i))]} Y_i^{-1} K(h^{-1}d(x, X_i)) \right], \end{aligned}$$

et d'autre part :

$$\begin{aligned}
 g_2(x)\mathbb{E}[\widehat{g}_1(x)] - g_1(x)\mathbb{E}[\widehat{g}_2(x)] &= g_2(x)\mathbb{E}\left[\frac{1}{n\mathbb{E}[K(h^{-1}d(x, X_i))]} \sum_{i=1}^n Y_i^{-1} K(h^{-1}d(x, X_i))\right] \\
 &\quad - g_1(x)\mathbb{E}\left[\frac{1}{n\mathbb{E}[K(h^{-1}d(x, X_i))]} \sum_{i=1}^n Y_i^{-2} K(h^{-1}d(x, X_i))\right] \\
 &= \sum_{i=1}^n \mathbb{E}\left[\frac{g_2(x)}{n\mathbb{E}[K(h^{-1}d(x, X_i))]} Y_i^{-1} K(h^{-1}d(x, X_i))\right. \\
 &\quad \left.- \frac{g_1(x)}{n\mathbb{E}[K(h^{-1}d(x, X_i))]} Y_i^{-2} K(h^{-1}d(x, X_i))\right],
 \end{aligned}$$

On pose :

$$L_i(x) = \frac{\sqrt{n\varphi_x(h)}}{n\mathbb{E}[K_1]} K_i(g_1(x)Y_i^{-2} - g_2(x)Y_i^{-1}),$$

et

$$S_n = \sum_{i=1}^n (L_i(x) - \mathbb{E}[|L_i(x)|]).$$

Nous obtenons donc :

$$\sqrt{n\varphi_x(h)}\sigma^{-1}((\widehat{g}_2(x) - \mathbb{E}[\widehat{g}_2(x)])g_1(x) - (\widehat{g}_1(x) - \mathbb{E}[\widehat{g}_1(x)])g_2(x)) = \frac{S_n}{\sigma}.$$

pour compléter la preuve de ce lemme, il suffit donc de montrer la normalité asymptotique de S_n . Nous appliquons le théorème de Lyaponov 2.5 sur $L_i(x)$. C'est à dire montrons que pour $\delta > 0$:

$$\frac{\sum_{i=1}^n \mathbb{E}[|L_i(x) - \mathbb{E}[|L_i(x)|]|^{2+\delta}]}{\left(Var\left(\sum_{i=1}^n L_i(x)\right)\right)^{\frac{2+\delta}{2}}} \rightarrow 0.$$

on a :

$$\begin{aligned}
 \operatorname{Var}\left(\sum_{i=1}^n L_i(x)\right) &= n\varphi_x(h) \operatorname{Var}[\widehat{g}_1(x)g_2(x) + \widehat{g}_2(x)g_1(x)] \\
 &= n\varphi_x(h)(\operatorname{Var}[\widehat{g}_1(x)]g_2^2(x) + \operatorname{Var}[\widehat{g}_2(x)]\widehat{g}_1^2(x) \\
 &\quad + g_1(x)g_2(x)\operatorname{Cov}(\widehat{g}_1(x), \widehat{g}_2(x))) \\
 &= n\varphi_x(h)\left(\frac{\beta_2}{\beta_1^2 n\varphi_x(h)}(g_2(x)g_1(x)\mathbb{E}[Y^{-3}/X=x] \right. \\
 &\quad \left.+ g_1^2(x)\mathbb{E}[Y^{-4}/X=x]) + o\left(\frac{1}{n\varphi_x(h)}\right)\right),
 \end{aligned}$$

donc :

$$\operatorname{Var}\left(\sum_{i=1}^n L_i(x)\right) = \sigma + o(1),$$

d'autre part nous appliquons les inégalités C_r , pour obtenir :

$$\sum_{i=1}^n \mathbb{E}[|L_i(x) - \mathbb{E}[|L_i(x)|]|^{2+\delta}] \leq C \sum_{i=1}^n \mathbb{E}[|L_i(x)|^{2+\delta}] + C' \sum_{i=1}^n |\mathbb{E}[L_i(x)]|^{2+\delta}.$$

Pour tout $j > 0$, $\mathbb{E}[K_1^j] = \mathcal{O}(\varphi_x(h))$, donc sous l'hypothèse (H.4), on a :

$$\begin{aligned}
 \sum_{i=1}^n \mathbb{E}[|L_i(x)|^{2+\delta}] &= n^{-\frac{\delta}{2}}(\varphi_x(h))^{-1-\frac{\delta}{2}} \mathbb{E}[K_1^{2+\delta}|g_1(x)Y_i^{-2} - g_2(x)Y_i^{-1}|^{2+\delta}] \\
 &\leq n^{-\frac{\delta}{2}}(\varphi_x(h))^{-1-\frac{\delta}{2}} \mathbb{E}[K_1^{2+\delta}[2^{1+\delta}g_1(x)^{2+\delta}\mathbb{E}[|Y_i^{-2(\delta+2)}]/X]] \\
 &\quad + 2^{1+\delta}g_2(x)^{2+\delta}\mathbb{E}[|Y_i^{-(\delta+2)}]/X]] \\
 &\leq C(n\varphi_x(h))^{-\frac{\delta}{2}} \left(\mathbb{E}[K_1^{2+\delta}]\varphi_x(h))^{-1}\right) \rightarrow 0.
 \end{aligned}$$

D'autre part et pour le deuxième terme, on a :

$$\begin{aligned}
 \sum_{i=1}^n |\mathbb{E}[L_i(x)]|^{2+\delta} &\leq n^{\frac{-\delta}{2}}(\varphi_x(h))^{\frac{-(2+\delta)}{2}} \left|\mathbb{E}[K_1|g_1(x)Y_i^{-2} - g_2(x)Y_i^{-1}|]\right|^{2+\delta} \\
 &\leq Cn^{\frac{-\delta}{2}}(\varphi_x(h))^{\frac{-(2+\delta)}{2}} \left|\mathbb{E}[K_1]\right|^{2+\delta} \\
 &\leq Cn^{\frac{-\delta}{2}}(\varphi_x(h))^{\frac{1+\delta}{2}} \rightarrow 0.
 \end{aligned}$$

Ce qui termine la preuve du Lemme 5.13. ■

Preuve du Lemme 5.14

D'après les Lemmes 5.10 et 5.11, nous avons :

$$\mathbb{E}[\hat{g}_2(x) - g_2(x)] \rightarrow 0,$$

et

$$Var[\hat{g}_2(x)] \rightarrow 0,$$

donc :

$$\hat{g}_2(x) - g_2(x) \xrightarrow{P} 0.$$

D'autre part :

$$\mathbb{E} \left[\left(\frac{n\varphi_x(h)}{g_1(x)^2 \sigma^2} \right)^{\frac{1}{2}} A_n (\hat{g}_2(x) - \mathbb{E}[\hat{g}_2(x)]) \right] = 0$$

et

$$Var \left[\left(\frac{n\varphi_x(h)}{g_1(x)^2 \sigma^2} \right)^{\frac{1}{2}} A_n (\hat{g}_2(x) - \mathbb{E}[\hat{g}_2(x)]) \right] = \mathcal{O}(A_n^2) = \mathcal{O}(h^2) \rightarrow 0,$$

ce qui montre que :

$$\left(\frac{n\varphi_x(h)}{g_1(x)^2 \sigma^2} \right)^{\frac{1}{2}} A_n (\hat{g}_2(x) - \mathbb{E}[\hat{g}_2(x)]) \xrightarrow{P} 0.$$

■

Chapitre 6

Cas d'échantillon dépendant

Dans ce chapitre, on garde la structure fonctionnelle de la variable explicative et nous généralisons le résultat de la convergence presque complète du chapitre précédent pour des données qui ne sont pas nécessairement indépendantes. Plus précisément, nous focalisons sur le cas où les observations sont α -mélangeantes.

6.1 Modèle et son estimateur

On garde le même estimateur à noyau (5.2) défini au chapitre 5, la seule différence qui apparaît est le type de dépendante entre les variables.

6.2 la convergence presque complète

on garde les hypothèses (H.1)-(H.4)du chapitre précédent et nous ajoutons les hypothèses suivantes :

6.2.1 Hypothèses et notations

(H.5) $(X_i, Y_i)_{1 \leq i \leq n}$ est une suite algébriquement α -mélangeante, dont le coefficient de mélange vérifie :

$$\exists C, a \in \mathbb{R}_+^*, \alpha(n) \leq Cn^{-a}.$$

(H.6) $0 < \sup_{i \neq j}(\text{IP}((X_i, Y_j) \in B(x, h) \times B(x, h))) = \mathcal{O}\left(\frac{\varphi_x(h)^{\frac{a+1}{a}}}{n^{\frac{1}{a}}}\right).$

(H.7) Il existe $\eta > 0$ tel que $Cn^{\frac{3-a}{a+1}+\eta} \leq \varphi_x(h) \leq C'n^{\frac{1}{1-a}}$, avec $a > \frac{5 + \sqrt{17}}{2}$.

6.2.2 Résultat asymptotique

Théorème 6.1. *Sous les conditions (H.1)-(H.7), on a :*

$$\hat{r}(x) - r(x) = \mathcal{O}(h^k) + \mathcal{O}_{p.co.}\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right).$$

6.2.3 Démonstration

La démonstration du théorème précédent est basée sur la décomposition (5.3) et les lemmes suivants :

Lemme 6.2. *Sous les hypothèses du théorème 6.1, on a :*

- (i) $\mathbb{E}[\hat{g}_1(x)] - g_1(x) = \mathcal{O}(h^k)$,
- (ii) $\mathbb{E}[\hat{g}_1(x)] - \hat{g}_1(x) = \mathcal{O}_{p.co.}\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right)$.

Lemme 6.3. *Sous les mêmes hypothèses, on a :*

- (i) $\mathbb{E}[\hat{g}_2(x)] - g_2(x) = \mathcal{O}(h^k)$,
- (ii) $\mathbb{E}[\hat{g}_2(x)] - \hat{g}_2(x) = \mathcal{O}_{p.co.}\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right)$.

Lemme 6.4. *Sous les hypothèses précédentes, on a :*

$$\exists \delta > 0, \quad \text{tel que} \quad \sum_{i \geq 1} \mathbb{P}[\hat{g}_2(x) < \delta] < \infty.$$

Preuve du Lemme 6.2

Pour la preuve du Lemme 6.2, on a :

$$\hat{g}_1(x) = \frac{1}{n} \sum_{i=1}^n Y_i^{-1} \omega_i,$$

donc, on peut écrire :

$$\begin{aligned} \mathbb{E}\hat{g}_1(x) - g_1(x) &= \mathbb{E}[Y_i^{-1} \omega_i] - g_1(x) \\ &= \mathbb{E}[\mathbb{E}[Y_i^{-1} \omega_i / X_1]] - g_1(x) \\ &= \mathbb{E}[g_1(X_1) \omega_1 - g_1(x)] \end{aligned}$$

en utilisant la condition (H.2), on montre que :

$$\mathbb{E}\hat{g}_1(x) - g_1(x) \leq \mathbb{E}|g_1(X_1) - g_1(x)|\omega_1$$

$$\leq C_1 \mathbb{E}[d^k(x, X_1)\omega_1]$$

comme $\mathbb{E}[\omega_1] = 1$, on a :

$$\mathbb{E}\hat{g}_1(x) - g_1(x) \leq C(h^k),$$

ainsi :

$$\mathbb{E}\hat{g}_1(x) - g_1(x) = \mathcal{O}(h^k).$$

Ce qui démontre le résultat (i).

Pour le deuxième résultat (ii), on applique l'inégalité de *Fuk-Nagaev* 2.2 aux variables Δ_i .

Avec :

$$\begin{aligned} \mathbb{E}[\hat{g}_1(x)] - \hat{g}_1(x) &= \left| \frac{\sum_{i=1}^n Y_i^{-1} K(h^{-1}d(x, X_i)) - \mathbb{E}\left[\frac{\sum_{i=1}^n Y_i^{-1} K(h^{-1}d(x, X_i))}{n \mathbb{E}[K(h^{-1}d(x, X_i))] } \right]}{n \mathbb{E}[K(h^{-1}d(x, X_i))]} \right| \\ &= \frac{1}{n} \left| \sum_{i=1}^n \frac{Y_i^{-1} K(h^{-1}d(x, X_i)) - \mathbb{E}[Y_i^{-1} K(h^{-1}d(x, X_i))]}{\mathbb{E}[K(h^{-1}d(x, X_i))]} \right|, \end{aligned}$$

on pose :

$$\Delta_i = Y_i^{-1} K(h^{-1}d(x, X_i)) - \mathbb{E}[Y_i^{-1} K(h^{-1}d(x, X_i))].$$

Cette inégalité est basée sur le calcul de :

$$\begin{aligned}
 S_n^2 &= \sum_{i=1}^n \sum_{j=1}^n |Cov(\Delta_i, \Delta_j)| \\
 &= \sum_{i=1}^n \left[\sum_{i \neq j}^n Cov(\Delta_i, \Delta_j) + Cov(\Delta_i, \Delta_i) \right] \\
 &= \sum_{i=1}^n \left[\sum_{i \neq j}^n Cov(\Delta_i, \Delta_j) + Var(\Delta_i) \right] \tag{6.1} \\
 &= \sum_{i=1}^n \sum_{i \neq j}^n Cov(\Delta_i, \Delta_j) + \sum_{i=1}^n Var(\Delta_i) \\
 &= S_n^{2*} + n Var\Delta_i.
 \end{aligned}$$

Où :

$$\begin{aligned}
 Cov(\Delta_i, \Delta_j) &= \mathbb{E}(\Delta_i \cdot \Delta_j) \\
 &= \mathbb{E}[(Y^{-j} K_i - \mathbb{E} Y^{-j} K_i)(Y^{-j} K_j - \mathbb{E} Y^{-j} K_j)] \\
 &= \mathbb{E}[Y^{-j} K_i Y^{-j} K_j - Y^{-j} K_i \mathbb{E} Y^{-j} K_j - Y^{-j} K_j \mathbb{E} Y^{-j} K_i + \mathbb{E} Y^{-j} K_i \mathbb{E} Y^{-j} K_j] \\
 &= \mathbb{E}[Y^{-j} K_i Y^{-j} K_j] - \mathbb{E}[Y^{-j} K_i] \mathbb{E} Y^{-j} K_j \\
 &= Cov(Y^{-j} K_i, Y^{-j} K_j),
 \end{aligned}$$

sous les hypothèses **(H.3)** et **(H.5)**, on a :

$$|Cov(\Delta_i, \Delta_j)| \leq C \mathbb{E}(K_i \cdot K_j).$$

Pour le terme de S_n^{2*} de (6.1) on utilise les techniques de *Masry* [18] et on divise cette somme comme suit :

$$S_n^{2*} = \sum_{i=1}^n \sum_{E_1} Cov(\Delta_i, \Delta_j) + \sum_{i=1}^n \sum_{E_2} Cov(\Delta_i, \Delta_j),$$

où :

$$E_1 = \{(i, j) \text{ tel que } 1 \leq |i - j| \leq m_n\},$$

et

$$E_2 = \{(i, j) \text{ tel que } m_n + 1 \leq |i - j| \leq n - 1\},$$

où la suite $m_n \rightarrow +\infty$ quand $n \rightarrow \infty$.

Pour la première somme, on a :

$$\begin{aligned} \sum_{i=1}^n \sum_{E_1} |Cov(\Delta_i(x), \Delta_j(x))| &\leq C \sum_{i=1}^n \sum_{E_1} |\mathbb{E}[K_i(x)K_j(x)] + \mathbb{E}^2[K_1(x)]| \\ &\leq C n m \varphi_x(h) \left(\left(\frac{\varphi_x(h)}{n} \right)^{\frac{1}{a}} + \varphi_x(h) \right). \end{aligned} \quad (6.2)$$

Sur l'ensemble E_2 on utilise l'inégalité de *Davydov-Rio's 2.3*, d'où pour $i \neq j$ on a :

$$|Cov(K_i(x), K_j(x))| \leq C\alpha(|i - j|).$$

Donc :

$$\begin{aligned} \sum_{i=1}^n \sum_{E_2} |Cov(K_i(x), K_j(x))| &\leq \sum_{i=1}^n \sum_{E_2} |\mathbb{E}[K_i(x)K_j(x)]| \\ &\leq C n \alpha(m_n) \\ &\leq C n^2 m_n^{-\alpha}. \end{aligned} \quad (6.3)$$

D'après (6.2) et (6.3) :

$$\begin{aligned} |S_n^{2*}| &\leq C n m \varphi_x(h) \left(\left(\frac{\varphi_x(h)}{n} \right)^{\frac{1}{a}} + \varphi_x(h) \right) + C n^2 m_n^{-\alpha}, \\ \text{ainsi pour } m_n = \left(\frac{\varphi_x(h)}{n} \right)^{\frac{1}{a}}, \text{ on aura :} \\ |S_n^{2*}| &\leq C n \frac{\varphi_x(h)^{1-\frac{1}{a}+\frac{1}{a}}}{n^{\frac{-1}{a}+\frac{1}{a}}} + C n^2 \left(\frac{\varphi_x(h)^{\frac{-1}{a}}}{n^{\frac{-1}{a}}} \right)^{-a} \\ &\leq C n \varphi_x(h), \end{aligned}$$

Donc :

$$S_n^{2*} = \mathcal{O}(n\varphi_x(h)).$$

Pour le terme de la variance, on montre que sous (H.1), on a :

$$Var(\Delta_1(x)) \leq C(\mathbb{E}(K_1)^2 - (\mathbb{E}K_1)^2) \leq (\varphi_x(h) + (\varphi_x(h))^2).$$

Nous obtenons donc :

$$S_n^2 = \mathcal{O}(\varphi_x(h)).$$

Par suite, l'application de l'inégalité de *Fuk-Nagaev* entraîne :

$$\begin{aligned}
 \mathbb{P}[|\mathbb{E}\widehat{g}_1(x) - \widehat{g}_1(x)| > \epsilon] &= \mathbb{P}\left[\left|\frac{1}{n\varphi_x(h)} \sum_{i=1}^n \Delta_i\right| > \epsilon\right] \\
 &= \mathbb{P}\left[\left|\sum_{i=1}^n \Delta_i\right| > \frac{4(\epsilon n\varphi_x(h))}{4}\right] \\
 &\leq 4\left(1 + \frac{\epsilon^2 n^2 \varphi_x^2(h)}{n\varphi_x(h) 16r}\right)^{\frac{-r}{2}} + 2ncr^{-1}\left(\frac{8r}{\epsilon n\varphi_x(h)}\right)^{a+1} \\
 \text{pour } \epsilon = \epsilon_0 \sqrt{\frac{\log n}{n\varphi_x(h)}} \text{, on a :} \\
 \mathbb{P}\left[|\mathbb{E}\widehat{g}_1(x) - \widehat{g}_1(x)| > \epsilon_0 \sqrt{\frac{\log n}{n\varphi_x(h)}}\right] &\leq 4\left(1 + \frac{\epsilon_0^2 (\log n) n^2 \varphi_x^2(h)}{n^2 (\varphi_x(h))^2 16r}\right)^{\frac{-r}{2}} + 2ncr^{-1}\left(\frac{8r(n\varphi_x(h))^{\frac{1}{2}}}{\epsilon_0 n\varphi_x(h) (\log n)^{\frac{-1}{2}}}\right) \\
 &\leq 4\left(1 + \frac{\epsilon_0^2 \log n}{16r}\right)^{\frac{-r}{2}} + 2ncr^{-1}\left(\frac{8r^{a+1}}{\epsilon_0}\right)(n\varphi_x(h) \log n)^{-(\frac{a+1}{2})},
 \end{aligned}$$

prenons $r = C(\log(n))^2$. Nous obtenons alors :

$$A_1 = \left(1 + \frac{\epsilon_0^2 \log n}{16r}\right)^{\frac{-r}{2}} \leq c \exp \frac{\epsilon_0^2 \log n}{32},$$

d'autre part :

$$A_2 \leq c\epsilon^{-(a+1)} n^{-(\frac{a+1}{2})+1+ab} (\varphi_x(h))^{(\frac{a+1}{2})},$$

sous la condition (H.7) sur la fonction $\varphi_x(h)$, on déduit :

$$A_2 \leq cn^{-1-\frac{a}{2}[-2b+p\theta]},$$

et avec un choix convenable de ϵ_0 , on obtient :

$$\sum_{i=1}^n \mathbb{P}\left[|\mathbb{E}\widehat{g}_1(x) - \widehat{g}_1(x)| > \epsilon_0 \sqrt{\frac{\log n}{n\varphi_x(h)}}\right] < \infty.$$

■

Preuve du Lemme 6.3

Pour la preuve de la partie (i) du Lemme 6.3, on a :

$$\widehat{g}_2(x) = \frac{1}{n} \sum_{i=1}^n Y_i^{-2} \omega_i,$$

on peut écrire :

$$\begin{aligned}\mathbb{E}\hat{g}_2(x) - g_2(x) &= \mathbb{E}[Y_i^{-2}\omega_1] - g_2(x) \\ &= \mathbb{E}[\mathbb{E}[Y_i^{-2}\omega_1/X_1]] - g_2(x) \\ &= \mathbb{E}[g_2(X_1)\omega_1 - g_2(x)].\end{aligned}$$

De même que la preuve du Lemme 6.2, la condition (**H.2**) nous permet d'écrire :

$$\begin{aligned}|\mathbb{E}\hat{g}_2(x) - g_2(x)| &\leq \mathbb{E}|g_2(X_1) - g_2(x)|\omega_1 \\ &\leq Ch^k.\end{aligned}$$

Donc :

$$\mathbb{E}\hat{g}_2(x) - g_2(x) = \mathcal{O}(h^k).$$

Pour la partie (ii), on applique l'inégalité exponentielle (2.5) aux variables aléatoires $Z_i = Y_i^{-2}\omega_i - \mathbb{E}[Y_i^{-2}\omega_i]$, on montre qu'il existe ϵ_0 tel que :

$$\sum_{n \in \mathbb{N}^*} \mathbb{P}\left[\frac{1}{n} \left| \sum_{i=1}^n Y_i^{-2}\omega_i - \mathbb{E}[Y_i^{-2}\omega_i] \right| > \epsilon_0 \sqrt{\frac{\log n}{n\varphi_x(h)}}\right] < \infty.$$

De même, comme K et Y_i^{-2} sont bornés alors :

$$|\omega_i| \leq \frac{C}{\varphi_x(h)}$$

donc :

$$Var(\omega_i^2) \leq \frac{C'}{\varphi_x(h)}$$

ainsi :

$$\begin{aligned}\mathbb{P}\left[|\mathbb{E}\hat{g}_2(x) - \hat{g}_2(x)| > \epsilon_0 \sqrt{\frac{\log n}{n\varphi_x(h)}}\right] &\leq 2 \exp\left(\frac{-n\epsilon_0^2\varphi_x(h)\log n}{4nC'\varphi_x(h)}\right) \\ &\leq 2 \exp\left(\frac{-\epsilon_0^2 \log n}{4C'}\right) \\ &\leq 2 n^{\frac{-\epsilon_0^2}{4C'}}.\end{aligned}$$

Donc :

$$\sum_{i=1}^n \mathbb{P}\left[|\mathbb{E}\hat{g}_2(x) - \hat{g}_2(x)| > \epsilon_0 \sqrt{\frac{\log n}{n\varphi_x(h)}}\right] \leq \sum_i^n 2n^{\frac{-\epsilon_0^2}{4C'}}.$$

il suffit de choisir $\frac{-\epsilon_0^2}{4C'} > 1$. Nous obtenons donc :

$$\sum_{i=1}^n \mathbb{P} \left[|\mathbb{E}\widehat{g}_2(x) - \widehat{g}_2(x)| > \epsilon_0 \sqrt{\frac{\log n}{n\varphi_x(h)}} \right] < \infty.$$

Ce qui complète la preuve du Lemme 6.3. ■

Preuve du Lemme 6.4

La démonstration du Lemme 6.4 est similaire à la preuve du Lemme 5.4 du Chapitre 5. ■

Chapitre 7

Application et conclusion

La dernière partie de ce mémoire est consacrée à une application sur des données simulées dont l'objectif principal est la comparaison entre l'estimateur à noyau de la régression relative et celui de la régression classique. Nous terminons cette partie par une conclusion générale.

7.1 Application

Si les recherches théoriques des statisticiens se focalisent de plus en plus sur le traitement des données fonctionnelles, c'est essentiellement parce que de plus en plus de domaines y ont recours. Dans cette application, nous présentons un exemple dans lequel nous montrons la supériorité de l'estimateur à noyau de la régression relative (5.2) sur celui de la régression classique défini par :

$$\tilde{r}(x) = \frac{\sum_{i=1}^n Y_i K(h^{-1}d(x, X_i))}{\sum_{i=1}^n K(h^{-1}d(x, X_i))}, \quad (7.1)$$

en testant la sensibilité de notre procédure avec la présence de valeurs aberrantes .

On considère les courbes explicatives suivantes :

$$X_i(t) = a_i \sin(4(b_i - t)) + b_i + \eta_{i,t}, \quad \text{pour } t \in [0, 1[\quad \text{et} \quad i = 1, 2, \dots, 300,$$

où $a_i \sim \mathcal{N}(5, 2)$, $b_i \sim \mathcal{N}(0, 0.1)$ et $\eta_{i,t} \sim \mathcal{N}(0, 0.2)$.

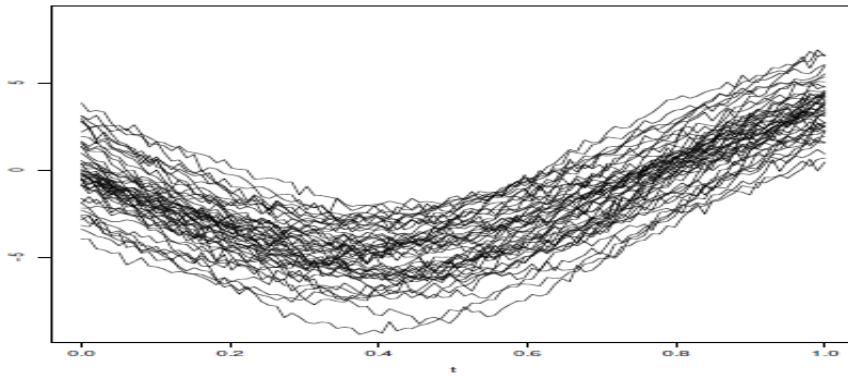


FIGURE 7.1 – Les données fonctionnelles.

Les courbes $X_i(t)$ sont discrétisées sur la même grille générées à partir de 100 mesures équidistantes dans l'intervalle $[0, 1[$, on définit l'opérateur de régression par :

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

Ainsi, la réponse scalaire Y est définie par :

$$Y = r(X) + \epsilon, \quad \text{où } \epsilon \sim \mathcal{N}(0, 1).$$

Pour réaliser notre objectif, nous comparons sur un échantillon fini, les comportements de l'estimateur à noyau de la régression classique (7.1) avec celui de la régression relative (5.2) dans les deux cas :

1. L'absence des valeurs aberrantes.
2. les données sont affectées par certaines valeurs aberrantes

Dans cet exemple d'application, nous utilisons la règle de MAD-Médian pour détecter les valeurs aberrantes, c'est à dire on dit que Y_i est une valeur aberrante si :

$$\frac{|Y_i - M|}{MAD * 0.6745} > C,$$

où M est la médiane de l'échantillon, et le MAD est la médiane de l'écart absolu donné par :

$$MAD = \text{median}(|Y_1 - M|, |Y_2 - M|, \dots, |Y_n - M|),$$

et C est considéré comme $\sqrt{\chi^2_{.975}}$ (La racine carré du quantile de la distribution (Chi-deux) avec un degré de liberté).

L'application de la méthode de MAD-Médian sur cette exemple de données identifie 23 valeurs aberrantes. Pour la première comparaison, nous supprimons toutes les observations des valeurs aberrantes détectées à partir des données d'origine. Ainsi, notre premier exemple est réalisé par des données ne contenant pas de valeurs aberrantes. De plus, nous avons divisé au hasard ces données en deux sous échantillons de tel sorte que le premier appelé échantillon d'apprentissage I_1 contenant les 200 premières observations, par contre le second appelé échantillon de test I_2 contenant 50 observations. Ensuite, la performance des deux estimateurs est décrit par l'erreur quadratique moyenne, et l'erreur quadratique moyenne relative suivantes :

$$MSE = \frac{1}{50} \sum_{i \in I_2} (Y_i - \hat{\theta}(X_i))^2,$$

et

$$RMSE = \frac{1}{Var(Y)} \sum_{i \in I_2} (Y_i - \hat{\theta}(X_i))^2,$$

où $\hat{\theta}$ signifie les deux estimateur étudiés (l'estimateur à noyau de la régression classique (7.1) et relative (5.2)).

Méthode	MSE	RMSE
ERC	2.05	0.0263
ERR	2.09	0.0268

TABLE 7.1 – Les valeurs de MSE et RMSE pour les données simulées.

Où ERC (estimateur de la régression classique) et ERR (estimateur de la régression relative). Comme dans toutes les méthodes de lissage, le choix du paramètre de lissage joue un rôle crucial. Dans cette illustration, nous utilisons la procédure de validation croisée, où h est choisie selon le principe suivant :

$$h_{opt} = \arg \min_h CV(h) = \arg \min_h \sum_{j=1}^n (Y_j - \hat{\theta}^{-j}(X_j)).$$

Où $\hat{\theta}^{-j}$ est l'estimateur de la régression classique (respectivement relative) privé de la j ème observation.

Le deuxième paramètre crucial dans cette étude est la semi-métrique dont son choix est basé sur la forme des courbes X_i , et il est claire que la semi métrique induite par L'ACP est bien adapté à cet ensemble de données. Enfin, pour les deux estimateurs (classique (7.1), relative (5.2)) nous avons utilisé un noyau quadratique.

Les résultats obtenue sont présentés dans la figure 7.2, il est clair qu'il n'y a pas de différence significative entre les deux méthodes (classique, relative).

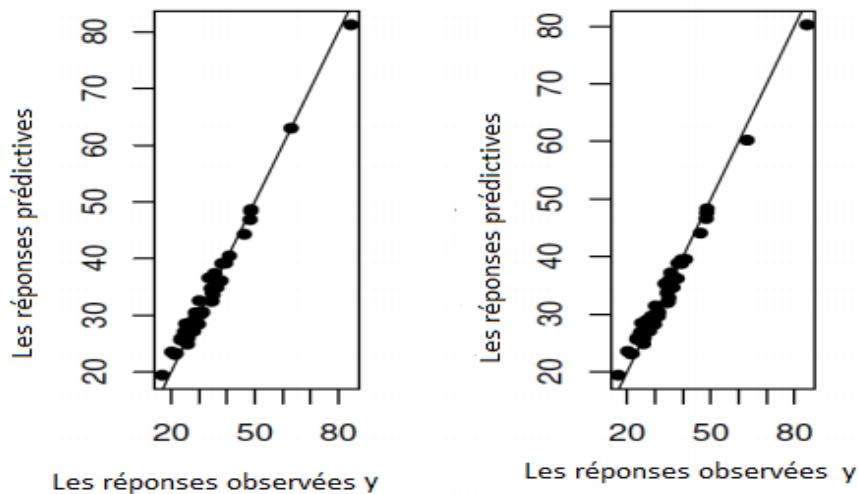


FIGURE 7.2 – Les résultats de prédiction, à gauche par rapport à MSE et à droite par rapport à RSME.

Dans la deuxième partie de cette application, nous nous concentrerons sur la comparaison des deux modèles en présence de valeurs aberrantes. Pour cela, nous avons introduit des valeurs aberrantes en multipliant certaines valeurs de Y dans l'échantillon d'apprentissage par 10. De plus, les deux estimateurs sont obtenus en gardant les mêmes conditions précédentes (les choix de paramètre de lissage, semi métrique et noyau). Enfin, nous rapportons les résultats obtenue dans le tableau 7.2 (respectivement 7.3), où nous calculons les valeurs du MSE et RMSE respectivement en fonction du nombre des valeurs aberrantes introduites.

les statistiques	50	40	30	20	10	0
ERC	3412	2202.0	1178.0	557.9	138.8	3.047
ERR	5.035	4.62	4.086	3.945	3.549	3.56

TABLE 7.2 – Les valeurs de MSE en fonction du nombre de valeurs aberrantes introduites.

les statistiques	50	40	30	20	10	0
ERC	21.080	13.96	7.379	3.379	0.8862	0.0186
ERR	0.03104	0.02843	0.02559	0.0238	0.02265	0.02177

TABLE 7.3 – Les valeurs de RMSE en fonction du nombre de valeurs aberrantes introduites.

Rappelons que, dans le premier cas, les deux estimateurs sont équivalents. Cependant, quand il y a des valeurs aberrantes l'estimateur à noyau de la régression relative est mieux que celui de la régression classique. En effet, la méthode classique de noyau est très sensible à la présence de valeurs aberrantes. Les valeurs de MSE et RMSE dans la méthode du noyau augmentent sensiblement par rapport au nombre de valeurs aberrants, alors que ces erreurs restent très faibles dans le cas d'estimation par la méthode de l'erreur relative.

7.2 Conclusion

L'idée générale de ce mémoire, est la modélisation de la co-variabilité entre une variable explicative (réelle ou fonctionnelle) et une variable réponse scalaire par minimisation de l'erreur moyenne quadratique relative. la principale caractéristique de cette fonction de perte est qu'elle prend en compte la forme de poids de chaque observation, contrairement à la fonction de perte de l'erreur moyenne quadratique où tous les observations ont le même poids.

Dans un premier temps, nous avons considéré une suite d'observations réelles et nous avons construit un estimateur à noyau pour la fonction de régression classique (cette estimation a été faite à partir de l'erreur quadratique moyenne), puis, la fonction de régression relative (l'estimation de l'erreur quadratique moyenne relative). Nous avons étudié la convergence presque complète de ces deux estimateurs en précisant leurs vitesses de convergence. Nous

nous sommes concentré sur le cas où les observations sont indépendantes et identiquement distribuées.

Dans un second temps, nous avons généralisé les résultats précédents au cas des observations qui ne sont pas comme généralement en statistique, des réalisations de variables aléatoires réelles ou vectorielles (vecteurs aléatoires), mais des fonctions aléatoires : courbes, images, etc... Il s'agit de données de dimension infinie, c'est-à-dire rentrant dans le champ de la "très grande dimension". Par contre, la variable réponse est toujours réelle. Cette généralisation a été étudiée dans les deux cas : des observations i.i.d., dans laquelle nous avons étudié la convergence presque complète ponctuelle et uniforme, la convergence en moyenne quadratique et la normalité asymptotique du même estimateur considéré. Le deuxième cas étudié était le cas dépendant avec un résultat de convergence presque complète.

D'un point de vue théorique, les résultats asymptotiques sont optimales tout comme le cas de la méthode classique. Par contre, d'un point de vue pratique, la méthode de régression relative possède plus d'avantages que la méthode de régression classique. Nous avons justifié cette supériorité à travers une application sur des données simulées.

Bibliographie

- [1] Bosq, D. (2000). *Linear Processes in Function Spaces : Theory and applications*. Lecture Notes in Statistics, **149**, Springer.
- [2] Campbell, M. Karen, Donner, Allan. (1989). Classification efficiency of multinomial logistic regression relative to ordinal logistic regression. *J. Amer. Statist. Assoc.* **406**, 587-591.
- [3] Chagny, G. (2018). Statistique pour données fonctionnelles. *Polyycopié*. Université Paris Dauphine.
- [4] Collomb, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau. *Z. Wahrsch. Verw. Gebiete*, **66**, 441-460.
- [5] Collomb, G. (1985). Nonparametric regression. *Statistics*, **16**, 309-324.
- [6] Demongeot, J., Hamie, A., Laksaci, A., and Rachdi, M. (2016). Relative-error prediction in nonparametric functional statistics : Theory and practice. *Journal of Multivariate Analysis*, **146**, 261-268.
- [7] Devroye, L. (1978). The uniform convergence of nearest neighbor regression function estimators and their application in optimization. *Transactions on Information Theory* **24**, 142-51.
- [8] Ferraty, F. et Vieu, P. (2002). Statistique fonctionnelle : Modèles Non-Paramétriques de Régression, Notes de cours de DEA.
- [9] Ferraty, F., P. Vieu. (2006). *Non parametric functional Data analysis, Theory and practice*, Springer-Verlag.

- [10] Ferraty, F.,Mas, A. and Vieu, P. (2007) Advances on nonparametric regression for fonctionnal data. *ANZ Journal of Statistics* **49** 267-286.
- [11] Giorgi, R. , Belot, Au. Gaudart, Je. , Launoy, Guy. (2008). The performance of multiple imputation for missing covariate data within the context of regression relative survival analysis. *Stat. Med.*, **30**, 6310-6331.
- [12] Grenander, U. (1981). *Abstract inference. Wiley Series in Probability and Mathematical Statistics*. John Wiley et Sons, Inc., New York.
- [13] Gyôrfi, L., Hardie, W., Sarda, P. et Vieu, P., (1989). *Nonparametric curve estimation from time séries, Lecture Notes in Statistics*, 60, Springer-Verlag.
- [14] Hall, P. et Vial, C. (2006b). Assessing the finite dimensionality of functional data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68**, 689–705.
- [15] Hsu, L. et Robbins, H. (1947). Complete convergence and the law of large numbers. *Proceedings of the national academy of sciences*, **33**, 25-31.
- [16] James, G.M., Hastie, T.J. and Sugar, C.A. (2000). Principal component models for sparse functional data *Biometrika*, **87**, 587-602.
- [17] Jones, M. C. Park, H., Shin, K., Vines, S. K. and Jeong, S. (2008). Relative error prediction via kernel regression smoothers. *J. Statist. Plann. Inference* **10**, 2887-2898.
- [18] Masry, E. (2005). Nonparametric regression estimation for dependent functional data : asymptotic normality Stochastic Process. Appl. **115**, 155-177.
- [19] Park, H. ,Stefanski, L. A. (1998). Relative-error prediction. *Statist. Probab.Lett.* **40**, 227-236.
- [20] Ramsay, J.O. (1982). When the data are functions. *Psychometrika* **47** 37,9- 396.
- [21] Ramsay, J. et Silverman, B. (1997). *Functional Data Analysis* Springer-Verlag, New York
- [22] Ramsay, J. et Silverman, B. (2002). *Applied functional data analysis : Methods and case studies* Springer-Verlag, New York
- [23] Ramsay, J. et Silverman, B. (2005). *Functional Data Analysis (Second Edition)* Springer-Verlag, New York.

- [24] Ruiz-Velasco, S. (1991). Asymptotic efficiency of logistic regression relative to linear discriminant analysis. *Biometrika* **78**, 235-243.
- [25] Sarda, P., and Vieu, P. (2000). *Kernel regression. Smoothing and regression : approaches, computation, and application*. Wiley Series in Probability and Statistics.
- [26] Stone, C.J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8**, 1348-1360.
- [27] Stone, C.J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10**, 1040-1053.
- [28] Tukey, J. W. (1961). *Curves as parameters, and touch estimation*. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability.
- [29] Vander Vaart, A.W. (2000). Asymptotic Statistics. *Cambridge series on statistical and probabilistic mathematics*.
- [30] Vieu, Ph. (1991). Quadratic errors for nonparametric estimates under dependence. *J. Multivariate Anal.* **39**, 324-347.