

N° d'ordre :

REPUBLIQUE ALGERIENNE DEMOCRATIQUE & POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR & DE LA RECHERCHE
SCIENTIFIQUE



UNIVERSITÉ DJILLALI LIABES
FACULTÉ DES SCIENCES EXACTES
SIDI BEL ABBES

THESE DE DOCTORAT DE 3^{ème} CYCLE

Présentée par
MECHAB WARDIA CHAHINEZ

Domaine : Mathématiques-Informatiques (MI)

Filière : Mathématiques

*Intitulé de la formation : Statistiques Mathématiques appliquées à
l'économie et à la finance.*

Analyse non paramétrique par régression relative.

Soutenue le : 31/01/2017

Devant le jury composé de :

Président : GHERIBALLAH ABDELKADER, Professeur, Université de S.B.A.

Examineurs :

- ATTOUCHE MOHAMMED KADI, Professeur, Université de S.B.A.
- BENCHIKH TEWFIK, Maître de Conférences (A), Université de S.B.A.
- MADANI FETHI, Maître de Conférences (A), Université de Saida.

Directeur de thèse : LAKSACI ALI, Professeur, Université de S.B.A.

Co-Directeur de thèse :

Année universitaire 2016/2017.

Remerciements

Je suis particulièrement reconnaissante à mon encadreur **le Professeur Ali laksaci** qui m'a remis le pied à l'étrier de la recherche. Je voudrais lui exprimer ma gratitude pour sa grande disponibilité, pour le temps qu'il a passé à m'aider afin d'améliorer et compléter mes travaux, pour ses conseils avisés qui m'ont permis de mener à bien cette thèse.

Je tiens bien entendu à exprimer ma sincère reconnaissance à Monsieur le **Professeur Abdelkader Gheribellah** qui m'a fait l'honneur d'accepter la présidence du jury de cette thèse.

Je suis très honorée de la présence de Messieurs **les Professeurs Fethi Madani, Attouch Mohammed Kadi et Benchikh Tawfik** qui ont accepté de s'intéresser à ce travail et de le juger.

Finalement, je voudrais terminer avec ces quelques lignes, sans doute trop courtes, pour exprimer toute ma gratitude à ma famille et plus particulièrement à mon très cher père qui, m'a transmis la passion des mathématiques. Je lui suis infiniment redevable pour avoir eu confiance en moi, même lorsque l'avenir paraissait incertain. Je tiens également à exprimer toute mon affection à ma très chère mère qui m'a soutenu au cours de ces années et a su être à l'écoute de mes soucis et de mes difficultés. Je tiens enfin à remercier tous les membres de ma famille et toutes les personnes proches qui m'ont apporté leur soutien durant ces années. Je vous dédie ce mémoire de thèse.

Table des matières

1	Introduction	5
1.1	Notions de mélange	11
1.2	Notions d'association	12
1.2.1	Cas de variables aléatoires Hilbertiennes	19
1.3	Quelques inégalités utiles	23
1.3.1	Inégalité de Markov	23
1.3.2	Inégalité de Hölder	23
1.3.3	Inégalité de covariance	23
1.3.4	Inégalité de type Fuk-Nagaev	23
1.3.5	Inégalité de Kallabis et Neumman(2006) [60]	25
2	Régression non-paramétrique relative	39
2.1	Présentation du modèle	39
2.2	Convergence presque complète de l'estimateur	41
2.2.1	Notations et hypothèses	41
2.3	Démonstration du théorème	41
2.3.1	Démonstration des Lemmes	42
3	Régression relative pour variables associées	47
3.1	Introduction	47
3.2	Convergence presque complète	48

3.3	La normalité asymptotique	50
3.4	Démonstration des résultats préliminaires	52
4	Régression relative fonctionnelle	69
4.1	Présentation du modèle et son estimateur	69
4.2	Convergence de l'estimateur	71
4.3	Démonstration du théorème	72
5	Application sur des données réelles	79

Chapitre 1

Introduction

La précision de l'estimateur obtenu par les méthodes paramétriques est parfois illusoire dans beaucoup d'applications pratiques à cause de la déviation du modèle choisi par rapport à la réalité. De plus, les calculs peuvent se révéler pénibles ou même inextricables lorsque les conditions de réalisation des tests ne sont pas vérifiées, ce qui rend difficile d'estimer convenablement le modèle paramétrique. Ces considérations ont amené à l'utilisation d'autres méthodes qui permettent de s'affranchir de ces conditions ainsi que de toute autres hypothèses restrictives sur la loi de l'échantillon. Il s'agit des méthodes dites non-paramétriques qui restent valides quelle que soit la distribution des données. Aussi, l'utilisation de l'approche non paramétrique offre une très grande flexibilité de modélisation pour les applications réelles, en apportant un regard nouveau sur les techniques d'estimation. Depuis, plusieurs problématiques de recherche ont été abordées sur le plan théorique et algorithmique dans le but d'apporter un développement pour les méthodes de régression et afin d'élargir le spectre d'application de ces méthodes.

Comme la notion de régression relative est plus récente que les autres on dispose d'un plus petit nombre de résultats. Les premiers résultats conséquents ont été obtenus en 1989 par Campbell et al. dans [25], où la régression relative a été utilisée comme outil de classification. En 1991, dans [99], Ruiz Velasco a évoqué l'efficacité asymptotique de la régression relative logistique dans un

contexte paramétrique, en particulier lorsque les variables explicatives sont normalement distribuées. Dans [78], les auteurs ont utilisé des techniques de régression non paramétrique pour produire un estimateur basé sur l'espérance conditionnelle de la variable réponse inverse. Dans [59], Jones et al. 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. (voir [45]), 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 plausible d'un point de vue épidémiologique.

À l'ère du tout numérique, le développement des systèmes informatiques et la puissance de calcul des machines sont à l'origine de l'accroissement constant de la masse de données à traiter. Cette inflation ; toujours plus importante ; révèle à chaque fois les limites des techniques d'analyses utilisées, ce qui pose un réel défi aux statisticiens pour la construction et la mise au point de nouvelles méthodes statistiques, adaptées à cette profusion de données. C'est l'une des raisons pour lesquelles ; durant les vingt dernières années ; l'analyse des données fonctionnelles est devenue l'un des outils les plus utilisés pour étudier les données dans leurs propres dimensions, apportant entre autre plusieurs éléments de réponse au problème. En effet l'intérêt est double : En pratique, il arrive souvent de récolter des données pour lesquelles les instants d'observation sont différents pour chaque individu, ce qui rend difficile l'approche par les méthodes classiques de la statistique multivariée (par exemple lorsque certaines données sont manquantes). D'autre part, la variabilité temporelle est extrêmement fréquente dans les données réelles (en particulier quand on possède des observations à haute résolution d'un phénomène quelconque), la statistique fonctionnelle permet alors d'éviter la simplification de certaines observations en les remplaçant par exemple par une moyenne .

Les domaines d'utilisation des méthodes fonctionnelles sont très variés et ca-

ractérisés par les quantités ou dimensions des données, de la météorologie[106] au traitement des images satellites[27] en passant par la génétique (voir [2], [62] et [102]). Afin de mettre en relief le large potentiel applicatif portant sur l'étude de données, considérées comme des fonctions (ou courbes), nous présentons quelques exemples typiques provenant de différentes disciplines où des échantillons de données fonctionnelles ont été utilisés.

- Dans [87], l'aspect fonctionnel des données est pris en compte par l'étude et la comparaison de courbes de croissance. L'auteur envisage l'analyse en composantes principales (ACP) et l'analyse factorielle de données fonctionnelles dans des situations où les courbes de croissances deviennent presque linéaires par la transformation de l'axe du temps. Bien d'autres travaux ont été consacrés à l'étude fonctionnelle de courbes de croissance comme dans [86], [43], [56] et [84].
- Dans le domaine de la médecine, grâce aux progrès effectués au niveau des appareils et des méthodes de mesure, une grande palette de méthodes liées à la statistique fonctionnelle est mise en œuvre dans l'étude de certains phénomènes comme dans l'évolution de certains cancers (voir [26] et [38]), l'étude des déformations de la cornée [65] ou l'effet placebo [105].
- La climatologie est un autre domaine qui a historiquement contribué au développement de certains outils/modèles statistiques propres aux données fonctionnelles, pour étudier les variations de la température mensuelle moyenne au cours de l'année en différents points du pays. On retrouve cette approche dans l'étude du phénomène El niño ; proposée dans [80] ; où l'évolution de diverses mesures correspondent à des individus pour lesquels la série temporelle est étudiée dans une modélisation auto-regressive fonctionnelle.
- Les méthodes géophysiques permettent de visualiser et d'explorer des données fonctionnelles. On étudie par exemple l'évolution de diverses mesures géophysiques d'une zone donnée pendant une année. On peut évoquer notamment le travail de Dabo-Niang et al. [29] qui s'intéresse

à déterminer la nature du sol de différents endroits de l'Amazonie en classant les courbes recueillies par les satellites.

Ces différents champs de recherche ont trouvé un réel écho au sein du cercle statistique où un grand nombre de travaux ; aussi bien théoriques que pratiques ; ont été produits. On pourrait citer à cet égard les travaux de Grenander [46], Dauxois et al. [30] et Ramsay [85] dans lesquels on retrouve l'étude des variables aléatoires fonctionnelles. Les ouvrages de Ramsay et Silverman [57], Ramsay et Silverman [83], Ramsay et Silverman [84] et Ferraty et Vieu [6] présentent un large éventail de méthodes statistiques développées pour le traitement de divers problèmes d'estimation dans lesquels interviennent les variables statistiques fonctionnelles. Dans [8], D. Bosq propose une approche faisant apparaître une suite de données fonctionnelles dépendantes qui modélisent la série chronologique observée. Le processus discret est alors présenté comme un processus à temps continu qu'il découpe en un échantillon de courbes successives. On pourra se référer à [11], comme l'un des principaux travaux théoriques utilisé dans l'étude des estimateurs issus de modèles à variables fonctionnelles. Pour les données longitudinales, issues de mesures répétées, en quelques points choisis, d'un phénomène temporel, les approches utilisées sont différentes de la méthode de variables aléatoires fonctionnelles. Une riche littérature existe dans ce domaine et on peut citer en particulier [31] et [48], pour une comparaison des deux méthodes. Cependant, certains outils de la statistique fonctionnelle s'adaptent facilement à l'étude de ce type de données (voir [31], [48], [55], [56] et [73]).

Dans de nombreux champs d'applications, notamment en sismologie (voir les données étudiées dans Estévez-Pérez et al [44]), en économétrie (voir les articles traités par Engle et Russell [82] et Nielson et Linton [53]) ou encore dans le domaine de la santé publique, la condition d'indépendance entre les variables observées est souvent subjective, voir erronée. De par la probable existence de corrélation entre les données d'observations, il est très peu réaliste de supposer leur indépendance. Toutefois, il serait intéressant de modéliser l'indépendance mais aussi d'introduire une structure probabiliste

qui permettra de contrôler la dépendance au sein des variables de l'échantillon statistique ce qui reflétera l'évolution exacte d'un phénomène aléatoire. C'est pour cette raison que, depuis quelques années, la notion de dépendance a été largement utilisée et a fait l'objet d'une intense activité de recherche. On peut consulter par exemple l'article de Rosenblatt [92] qui fut l'un des premiers qui ont traité ce sujet et qui a amené certains auteurs, dont Ibragimov [51] fut l'un des premiers à introduire la condition de mélange fort à son étude sur les champs aléatoires stationnaires avec des densités spectrales rationnelles. Pour ce qui nous concerne ici, nous nous tiendrons particulièrement à la notion de mélange fort appelée aussi α -mélange qui ; son nom ne l'indique pas ; est la conception de dépendance la plus faible. Ce choix a été motivé du fait que ce type de variables soit assez général et très présent dans la pratique statistique (voir Doukhan [36]) et des nombreux résultats dont on dispose et qui nous permettront d'étudier ce type de modélisation. On considère des échantillons de variables α -mélangeantes, condition introduite par Rosenblatt [92], qui est l'une des plus générales et notamment vérifiée par différents types de processus que l'on retrouvera par exemple dans Bradley [13].

Signalons que de nombreux auteurs se sont intéressés à ce type de dépendance. En 1971, sous une séquence de variables aléatoires fortement mélangeantes, Welsch [107] suggère des conditions suffisantes pour la convergence faible dans un espace de Skorohod. À l'idée d'aborder le concept du mélange fort, Rosenblatt et Murray [93] ont établi un théorème de la limite centrale avec une ergodicité uniforme pour des processus stationnaires. Bien d'autres auteurs se sont aussi penchés sur cette notion de mélange fort, comme Rao [88], Bradley et al. [15], Bradley [12] et Rosenblatt [94] qui avait étudié la normalité asymptotique pour l'estimation de la densité spectrale sous la condition de mélange fort. Sur les propriétés de base des conditions de l' α -mélange, en 1986 Bradley et Richard [14] tentent de compléter leur travail et on peut citer d'autres travaux sur le thème comme Bulinskii et Doukhan [21] et Louani et Ould-Saïd [66].

En 2000, dans [71], Merlevède et Peligrad répondent à une conjecture de Bradley en 1997 [16] et ont pu améliorer tous les théorèmes de la limite centrale antérieurs pour ce type de dépendance. Par la suite, dans une note à l'académie des sciences de Paris [40], Ferraty et al. publient une étude d'un modèle de régression non-paramétrique pour des variables aléatoires fonctionnelles fortement mélangeantes. Dans [69], Masry évoque la normalité asymptotique de l'estimation de la fonction de régression fonctionnelle pour les processus fortement mélangeants. Les résultats établis dans l'article de Delsol, [32], sont liés aux propriétés asymptotiques de l'estimateur à noyau généralisé pour une variable explicative fonctionnelle dans le cas d'un échantillon alpha-mélangeant.

Les travaux réalisés lors de la thèse seront regroupés en quatre chapitres ;

- Dans le deuxième chapitre, nous traitons le cas où les observations sont indépendantes identiquement distribuées. Comme résultats asymptotiques nous démontrons la convergence presque complète de notre estimateur. La vitesse de convergence garde sa forme usuelle, où la régularité du modèle est bien exploitée dans la partie biais, ainsi que le problème du fléau de la dimension est également présent dans la partie dispersion.
- Le troisième chapitre est consacré au cas où les variables sont quasi-associées. Sous des conditions de régularité classiques, nous démontrons la convergence presque complète ainsi que la normalité asymptotique de notre estimateur à noyau construit à partir de la minimisation du carré moyen de l'erreur relative.
- Le cas où la variable explicative est de type fonctionnelle est traité dans le quatrième chapitre. Nous établissons les propriétés asymptotiques de notre estimateur dans le cas où les observations $(X_i, Y_i)_{i=1,2,\dots,n}$ sont fortement mélangeants.
- Au dernier chapitre, nous présentons un exemple qui nous permettra à la fois de donner une meilleure visualisation mais aussi d'illustrer comment interpréter la qualité des prévisions de notre estimateur comparée

à celle de la régression classique du noyau. Nous terminons cette thèse par une conclusion générale qui synthétise le travail réalisé ; ainsi, un bilan sera fait sur les objectifs fixés. Nous donnons aussi quelques perspectives de recherche sur l'ensemble des chapitres abordés afin d'améliorer le modèle de régression relative .

Nous rappelons brièvement la définition liée à l' α - mélange ainsi que les outils fondamentaux qui interviendront dans notre étude pour le cas d'une famille de variables aléatoires algébriquement α - mélangeantes, à savoir l'inégalité de covariance.

1.1 Notions de mélange

Définition 1.1.1 *On appelle coefficient de mélange fort entre deux tribus \mathcal{A} et \mathcal{B} , le nombre réel*

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{(A,B) \in (\mathcal{A}, \mathcal{B})} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

On montre que : $0 \leq \alpha(\mathcal{A}, \mathcal{B}) \leq \frac{1}{4}$.

Définition 1.1.2 (Rosenblatt (1956)[92]) *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathfrak{F}, \mathbb{P})$. On appelle coefficients de mélange fort, les réels $\alpha(n)$ définis par :*

$$\alpha(n) = \sup\{\alpha(\mathfrak{F}_1^k, \mathfrak{F}_{n+k}^\infty), k \in \mathbb{N}^*\}.$$

Où \mathfrak{F}_j^l désigne la tribu engendrée par les variables $(X_i, j < i < l)$.

Proposition 1.1.1 (Billingsley (1968)[6]) *Si $(X_n)_{n \geq 1}$ est une suite α -mélangeante, alors $(f(X_n))_{n \geq 1}$ est aussi α -mélangeante pour toute fonction mesurable f .*

Définition 1.1.3 *On dit qu'une famille $\{X_i, i \in \mathbb{N}^*\}$ est fortement mélangée ou alpha-mélangée, si l'on a :*

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

On distingue deux types de mélange fort :

Définition 1.1.4 *On dit que la famille $\{X_i, i \in \mathbb{N}^*\}$ est algébriquement alpha-mélangée, s'il existe deux constantes positives non nulles C et a telles que les coefficients de mélange vérifient*

$$\alpha(n) \leq Cn^{-a}.$$

Définition 1.1.5 *S'il existe deux constantes $s \in \mathbb{R}_+^*$ et $t \in]0, 1[$ telles que les coefficients de mélange vérifient*

$$\alpha(n) \leq st^n,$$

On dit que la famille $\{X_i, i \in \mathbb{N}^\}$ est géométriquement alpha-mélangée.*

1.2 Notions d'association

Il n'est pas rare de relever une absence de dépendance dans l'enceinte d'un échantillon, ce que nous qualifierons d'indépendance. Lehmann [61] compte parmi les premiers à s'être essentiellement intéressé par la distribution des paires de variables aléatoires et introduit une notion de dépendance bien adaptée à ce cadre, remarquant immédiatement des conséquences sur la covariance entre les variables aléatoires. ce qui était donc naturel que les notions de dépendance positives soient apparues dans un contexte bivarié.

Définition 1.2.1 (Lehmann(1966)[61]) *Deux variables aléatoires X et Y sont dites PQD (Positive Quadrant Dependence) si, pour tous les nombres réels x et y*

$$\mathbb{P}(X \leq x, Y \leq y) \geq \mathbb{P}(X \leq x) \cdot \mathbb{P}(Y \leq y).$$

Plus tard, inspiré de la dépendance positive définie par Lehmann, cette idée a été étendue aux distributions multivariées par Esary, Porschan et Walkup en 1967 [39], qui avaient pour but de tenter une formalisation de ce genre de notions de dépendance et d'explorer quelques-unes de ses propriétés, en essayant d'étendre le champ d'application du comportement au-delà des distributions particulières considérées, en proposant à travers l'association une notion plus forte.

Définition 1.2.2 *Soit f une application de \mathbb{R}^n dans \mathbb{R} . On dit que f est croissante si pour tout $i = 1, \dots, n$, l'application*

$$t \mapsto f(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_n)$$

est croissante sur \mathbb{R} .

On considère $X = \{X_t, t \in T\}$ une famille de variables aléatoires réelles indexée par un ensemble d'indices T définies sur un espace de probabilité $(\Omega, \mathfrak{F}, \mathbb{P})$.

Définition 1.2.3 (Esary et al.(1967) [39]) *On dit que cette famille est associée ; ou positivement dépendante ; si pour des sous-ensembles finis $I, J \subset T$ et toutes fonctions f et g croissantes $f : \mathbb{R}^{|I|} \rightarrow \mathbb{R}$ et $g : \mathbb{R}^{|J|} \rightarrow \mathbb{R}$,*

$$(1.1) \quad \text{Cov}(f(X_s, s \in I), g(X_t, t \in J)) \geq 0,$$

à chaque fois que la covariance existe, avec $|I|$ et $|J|$ les cardinaux de I et J .

Dans des études de problèmes de physique statistique et des problèmes de la fiabilité et la durabilité de systèmes, dont le problème de la percolation ou d'autres problèmes liés à la théorie des statistiques, quelques idées de dépendance positive ont été explorées. La structure d'association a été décrite à travers les inégalités FKG qui portent désormais le nom de leurs auteurs Fortuin, Kasteleyn et Ginibre [42], dans l'étude des propriétés de l'Hamiltonien

du modèle d'Ising décrivant l'interaction entre deux corps ferromagnétiques. Nous allons résumer dans ce qui suit les principales propriétés de telle dépendance .

Propriété 1.2.1 (Esary et al. (1967)[39] et Suquet et al. (2001)[22])

1. *Tout sous ensemble d'un ensemble fini de variables aléatoires réelles associées est associé.*
2. *Si deux ensembles de variables associées sont disjoints, leur union donne un ensemble associé.*
3. *Une famille infinie de variables aléatoires réelles est associée si toute sous-famille finie est associée.*
4. *Tout singleton formé d'une variable aléatoire réelle X est associé.*
5. *Si $X = (X_1, \dots, X_n)$ est associé et si f_1, \dots, f_k sont des fonctions toutes croissantes de \mathbb{R}^n dans \mathbb{R} (resp. toutes décroissantes), alors le vecteur $Y = (f_1(X), \dots, f_k(X))$ sera associé.*
6. *Si $X^{(k)} = (X_1^{(k)}, \dots, X_n^{(k)})$ est associé pour tout k et si $X^{(k)}$ converge en loi vers $X = (X_1, \dots, X_n)$ lorsque k tend vers ∞ , alors X est associé.*

Voici quelques exemples importants de vecteurs associés :

Exemple 1.2.1 (Processus gaussien, (Pitt (1982)[81])) :

Un champ aléatoire gaussien (X_1, \dots, X_n) est associé si et seulement si sa fonction de covariance $Cov(X_i, X_j) \geq 0$ pour tout i différent de j .

Exemple 1.2.2 (Processus linéaire (Doukhan et Louhichi (1999)[37]))

Soit $(\varepsilon_i)_{i \in \mathbb{Z}}$ une suite de variables aléatoires indépendantes prenant les valeurs $\left\{ \frac{-1}{2}, \frac{1}{2} \right\}$ avec équiprobabilité et $(X_n)_{n \in \mathbb{Z}}$ une suite aléatoire vérifiant l'équation d'auto-régression suivante :

$$X_n = \frac{1}{2}X_{n-1} + \varepsilon_n, n \in \mathbb{Z},$$

alors pour tout $n \in \mathbb{Z}$, $X_n = \sum_{i=0}^{\infty} 2^{-i} \varepsilon_{n-i}$ presque sûrement ainsi le processus linéaire $(X_n)_{n \in \mathbb{Z}}$ est associé.

Exemple 1.2.3 (Suquet(1994)[104])

- Tout vecteur aléatoire à composantes indépendantes est associé.
- Pour X associé, notons $S_i = \sum_{k=1}^i X_k$. Alors, le vecteur des sommes partielles (S_1, \dots, S_n) est associé.
- Le vecteur aléatoire $X = (X_1, \dots, X_n)$ est associé si et seulement s'il vérifie $Cov(f(X), g(X)) \geq 0$, pour f et g une paire de fonctions croissantes de \mathbb{R}^n dans $\{0, 1\}$.

À partir de 1980, un intérêt particulier est porté à l'étude de plusieurs aspects probabilistes et statistiques, des variables associées. Après avoir développé une inégalité pour les fonctions caractéristiques, qui a longtemps servi dans le cadre de la normalité asymptotique pour le cas associé, Newman établit dans [74] le théorème central limite (TCL) pour des variables strictement stationnaires. En 1981, Newman et Wright généralisent, dans [76], le TCL dans un cas fonctionnel avant que Cox et Grimmett (1984)[28] ne remplacent la condition de stationnarité par des conditions sur les moments des variables associées .

Diverses modifications et extensions ont été apportées à la définition d'association. Par exemple en 1984, Neumann dans son étude [75] étend la définition 1.2.3; de la faible association ou association positive; en imposant la condition (3.13) seulement pour les ensembles I et J tels que $I \cap J = \emptyset$. Il établit la loi forte des grands nombres pour des suites strictement stationnaires sous l'hypothèse que

$$\frac{1}{n} \sum_{i=1}^n Cov(X_1, X_i) \rightarrow 0.$$

Sous certaines hypothèses techniques sur le noyau et la fenêtre, Roussas étudie dans [98] la consistance uniforme forte des estimateurs à noyau des dérivées d'ordre r de f pour des suites strictement stationnaires dans lequel il

obtient des taux de convergence.

Dans le cadre des données complètes (non censurées), supposons que $\{X_n, n \geq 1\}$ est un processus strictement stationnaire de variables aléatoires associées. Par une estimation non-paramétrique de type noyau de la densité marginale f , Bagai et Prakasa Rao [5] montrent que dans pareil cas, l'estimateur défini par :

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right),$$

est fortement consistant aussi bien ponctuellement qu'uniformément sous un ensemble de conditions. En se basant sur l'estimateur de Kaplan-Meier, Cai et Roussas [24] se proposent d'étudier le comportement asymptotique de l'estimateur de la fonction de répartition $\tilde{F}_n(\cdot)$ dans le cas d'un modèle non censuré en présence de variables aléatoires associées. Sous des variables bornées et sous certaines conditions sur la structure de covariance des variables satisfaisant une inégalité exponentielle du type Bernstein-Hoeffding et par application de la loi forte des grands nombres, Ioannides et Roussas [52] obtiennent une vitesse de convergence à l'ordre de $(\frac{n}{\log^2 n})^{\frac{1}{3}}$. L'approche introduite par Ioannides et Roussas [52] a été améliorée par Oliveira [77] en utilisant des variables associées non bornées mais le taux révélé à la fin était plutôt lent (taux d'ordre $(\frac{n}{\log^5 n})^{\frac{1}{3}}$), ce qui indique que l'inégalité exponentielle utilisée était loin d'être la forme optimale. On pourra citer aussi les travaux de Li et Wang [63] qui développent la notion de LPQD (Linear Positive Quadrant Dependence), de Lin et Li [64] et celui de Xing et Yang [109].

Certaines notions de dépendance positives ont été considérées en essayant de répondre aux difficultés spécifiques, ou d'étendre le champ d'application de certains des résultats ou même essayer de caractériser l'indépendance des covariances des variables aléatoires. Quelques notions sont restées concentrées sur les distributions bivariées, tandis que d'autres ont essayé de faire face à des distributions multivariées étendant les propriétés les plus pertinentes. Ceci est le cas de Masry [67].

Mais, bien que les premiers résultats étaient un peu moins encourageants à

fournir de meilleurs taux de convergence de l'estimateur à noyau de la densité, Henriques et Oliveira [49] de leur côté proposent un taux de décroissance exponentielle.

En 2007, Douge améliore sensiblement le résultat de Masry [67] et établit une nouvelle inégalité exponentielle qui conduit à un taux de convergence uniforme presque sûre de f_n sous la condition de décroissance géométrique des covariances.

La dépendance négative a été naturellement introduite en inversant les inégalités dans les définitions de dépendance positive ci-dessus. L'apparition de ce type de dépendance est arrivé dans Alam et Saxena [1] et de là, avec une image homologue à celle de (3.13), un article fut publié en 1983 par Joag-Dev et Proschan [58] visant à développer la notion d'association négative qui en plus des propriétés fondamentales des variables NA, montrent que certaines distributions multivariées (les loi gaussiennes, les lois de Dirichlet, les lois multinomiales,...) possèdent la propriété de NA.

Définition 1.2.4 (Joag-Dev et Proschan(1983)[58]) *La famille $\{X_1, X_2, \dots, X_n\}$ est dite négativement associée (NA) si, pour des sous-ensembles finis $I, J \subset T$ et toute fonction f et g non décroissantes $f : \mathbb{R}^{|I|} \rightarrow \mathbb{R}$, $g : \mathbb{R}^{|J|} \rightarrow \mathbb{R}$,*

$$(1.2) \quad \text{Cov}(f(X_s, s \in I), g(X_t, t \in J)) \leq 0,$$

à chaque fois que la covariance existe.

Depuis, l'écoulement de la littérature a finalement connu son essor dans la construction d'un corps plein de résultats à partir des prévisions probabilistes avec une application plus statistique à l'esprit ; à cet égard, nous citons les travaux de Newman [75], Matula [70], Roussas (voir [96], [97]) et Bulinski [17].

N'empêche que la dépendance négative attire sans doute moins d'attention dans les années précédentes, par rapport à l'intérêt porté en association positive, mais il y a, bien sûr, quelques fonctionnalités spécifiques qui finalement

intéressent certains auteurs, on cite les contributions de Su et al. [103] qui à la base des inégalités pour moments, un principe d'invariance faible est prouvé sous une séquence de variables négativement associées, Zhang [110], Baek et al. [4], Jabbari et Azarnoosh [54], Xing et al. [108], Doosti et Dewan [34], Mi-Hwa Ko [72] mais aussi Guang et Hui Cai [47] qui proposent un principe d'invariance fort pour les variables négativement associées.

Afin de compléter l'éventail de nos résultats, nous allons nous étendre au cadre des variables aléatoires **quasi-associées**.

Une inégalité de covariance a été établie par Birkel (voir [7]) a donné naissance à une autre notion de dépendance faible appelée la "quasi-association".

Définition 1.2.5 *Un ensemble à valeurs aléatoires réelles $X = \{X_t, t \in T\}$ avec $\mathbb{E}[X_t^2] < \infty$ est dit quasi-associé si, pour tous les sous-ensembles disjoints finis $I, J \subset T$ et pour toute fonction lipschitzienne bornée f et g respectivement définies sur $\mathbb{R}^{|I|}$ et $\mathbb{R}^{|J|}$,*

$$(1.3) \text{Cov}(f(X_s, s \in I), g(X_t, t \in J)) \leq \text{lip}(f)\text{lip}(g) \sum_{s \in I} \sum_{t \in J} |\text{Cov}(X_s, X_t)|.$$

Avec

$$(1.4) \quad \text{lip}(f) = \sup_{x \neq y} \frac{|f(x) - f(y)|}{\|x - y\|_1} < \infty,$$

$$(1.5) \quad \|x\|_1 = \sum_{s=1}^m |x_s| \text{ pour } x = (x_1, \dots, x_m) \in \mathbb{R}^m.$$

Toutes les normes sur \mathbb{R}^m sont équivalentes, donc le choix de la norme $\|\cdot\|_1$ dans (1.5) dépend des raisons de commodité seulement .

En effet, à l'usage de fonctions lisses, l'analogue de (1.3) a été présenté dans Birkel [7] pour des variables aléatoires associées (le lecteur pourra aussi consulter Roussas [95], Peligrad and Shao [79] et l'article de Bulinski [17].)

Dans Bulinski et Shabanovich [19], il a été montré que toute séquence de variables aléatoires négativement ou positivement associées avec des seconds moments finis satisfaisant l'inégalité (1.3) est un champ quasi-associé .

Soit maintenant $X = \{X_t, t \in T\}$ un champ aléatoire à valeurs dans \mathbb{R}^k .

les concepts mentionnés ci-dessus seront généralisés pour des familles de variables aléatoires à valeurs vectorielles (voir : Burton et al .[23], Bulinski[18] , et Bulinski Shahkin [20]).

Définition 1.2.6 (Bulinski(2000)[18]) *Un champ aléatoire X à valeurs dans \mathbb{R}^k est dit quasi-associé si, pour des sous ensembles finis disjoints $I, J \subset T$ et pour toutes fonctions lipschitziennes bornées f et g , respectivement définies sur $\mathbb{R}^{|I|}$ et $\mathbb{R}^{|J|}$,*

$$(1.6) \quad \text{Cov}(f(X_s, s \in I), g(X_t, t \in J)) \leq \text{lip}(f)\text{lip}(g) \sum_{s \in I} \sum_{t \in J} \sum_{r,q=1}^k |\text{Cov}(X_{s,r}, X_{t,q})|$$

où $X_{s,r}$ désigne la r -ième composante du vecteur X_s .

Remarque 1.2.1 *En 2002, Shashkin [100] a montré que le champ aléatoire gaussien $X = \{X_t, t \in T\}$ à valeur dans \mathbb{R}^k dont la fonction de covariance prend des valeurs positives et négatives est quasi-associé.*

1.2.1 Cas de variables aléatoires Hilbertiennes

Dans la plupart des problèmes liés au domaine de l'apprentissage statistique, on considère la variable aléatoire X à valeurs dans \mathbb{R}^d . Mais au cours de ces dernières années, la quantité et la nature des données à exploiter, ont conduit à l'émergence de nouveaux outils statistiques qui ont également tiré profit des capacités croissantes de calcul. Par ailleurs, les propriétés particulières des espaces de Hilbert et leur riches structures géométriques, par rapport à d'autres espaces fonctionnels possibles, ont suscité un grand intérêt au traitement et à l'étude de problèmes avec des variables aléatoires hilbertiennes et en particulier à valeurs dans des espaces de Hilbert séparables (c'est à dire admettant un sous ensemble dénombrable dense).

Nous allons maintenant définir la quasi-association pour des variables aléatoires à valeurs dans un espace de Hilbert séparable.

Soit $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ un espace de Hilbert séparable muni d'une base orthonormée $\{e_k, k \geq 1\}$.

Définition 1.2.7 *Une famille $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires à valeurs dans \mathbb{H} est dite quasi-associée par rapport à la base orthonormale $\{e_k, k \geq 1\}$ si, pour tout $d \geq 1$, la séquence des vecteurs aléatoires $\{(\langle X_n, e_1 \rangle, \dots, \langle X_n, e_d \rangle)\}$ est quasi-associée.*

En dépit de la diversité des méthodes d'estimation non-paramétrique, des avancées récentes en traitements de variables aléatoires hilbertiennes font apparaître que les techniques basées sur l'estimation par la méthode du noyau sont facilement transposables au cadre des processus quasi-associés Hilbertiens. La combinaison des avantages respectifs de chaque méthode permet d'établir de très intéressantes avancées dans l'étude de ce type de données, comme dans [35] et [3]. Dans [35], L. Douge définit les notions d'association et de quasi-association pour des suites de variables aléatoires à valeurs dans un espace de Hilbert séparable et établit pour ces suites une inégalité exponentielle et un théorème central limite. Dans [3], en utilisant l'approche semi-paramétrique de l'indice fonctionnelle pour la densité conditionnelle, des résultats sur la convergence presque complète de l'estimateur à noyau ainsi que la normalité asymptotique ; pour des variables quasi associées hilbertiennes par rapport à la structure en un seul indice ; sont établis.

Dans ce qui suit on va donner quelques exemples de processus (classiques) qui vérifient la propriété d'association.

Exemple 1.2.4 (Processus autorégressifs hilbertiens d'ordre 1)

On considère $(X_n, n \in \mathbb{Z})$ une suite de variables aléatoires à valeurs dans un espace de hilbert séparable \mathbb{H} , muni de son produit scalaire $\langle \cdot, \cdot \rangle$ et de sa norme $\| \cdot \|$. Soit ρ un opérateur linéaire borné sur \mathbb{H} tel que :

$$\exists j_0 \geq 1; \| \rho^{j_0} \|_{\mathcal{L}} < 1.$$

où \mathcal{L} est l'espace des applications linéaires continues de \mathbb{H} dans \mathbb{H} , muni de la norme :

$$\forall l \in \mathcal{L}, \quad \|l\|_{\mathcal{L}} = \sup_{\|x\| \leq 1} \|l(x)\|.$$

Il s'agit de modéliser des processus à temps continu par des processus à temps discret à valeurs dans un espace fonctionnel de Hilbert. Un tel modèle est l'unique solution stationnaire de l'équation :

$$(1.7) \quad X_n = \rho(X_{n-1}) + \epsilon_n, n \in \mathbb{Z}.$$

où la suite $(\epsilon_n, n \in \mathbb{Z})$ est un \mathbb{H} -bruit blanc fort¹.

Si on suppose qu'il existe une base orthonormale $(v_k, k \geq 0)$ de \mathbb{H} telle que

$$\forall k \geq 0, \quad \rho(v_k) = \alpha_k v_k, \quad \text{avec } \alpha_k \geq 0,$$

et pour tout $n \in \mathbb{Z}$, $(\langle \epsilon_n, v_k \rangle, k \geq 0)$ une suite de variables aléatoires associées, alors $(X_n, n \in \mathbb{Z})$ est aussi une suite associée.

Exemple 1.2.5 (Processus Linéaire hilbertien)

Soient $(\epsilon_n)_{n \in \mathbb{Z}}$ une suite de variables aléatoires hilbertiennes et $(\rho_i)_i$ une famille d'opérateurs linéaires symétriques.

On définit $(X_n)_{n \in \mathbb{Z}}$ le processus linéaire à valeurs dans un espace de hilbert \mathbb{H} par :

$$\forall n \in \mathbb{Z}, \quad X_n = \sum_{i=0}^{\infty} \rho_i(\epsilon_{n-i}).$$

On suppose qu'il existe une base orthonormale $(v_k, k \geq 0)$ de \mathbb{H} telle que

$$\forall i \geq 0, \forall k \geq 1, \quad \rho_i(v_k) = \alpha_{i,k} v_k, \quad \text{où } \alpha_{i,k} \geq 0,$$

alors si $(\epsilon_n)_{n \in \mathbb{Z}}$ est associée par rapport à la base $(v_k, k \geq 0)$, $(X_n)_{n \in \mathbb{Z}}$ est associé par rapport à la même base.

1. On dit que $(\epsilon_n, n \in \mathbb{Z})$ est un \mathbb{H} -bruit blanc fort si, les ϵ_n sont des variables aléatoires à valeurs dans \mathbb{H} , indépendantes et de même loi, telles que : $\forall n \in \mathbb{Z}, \quad \left(0 < \mathbb{E} \|\epsilon_n\|^2 < +\infty\right) \wedge \left(\mathbb{E}[\epsilon_n] = 0\right)$.

Exemple 1.2.6 (Processus Gaussien) :

Soit $\xi = \{\xi(t), t \geq 0\}$ un processus à temps continu gaussien centré à accroissement stationnaire et $\{\varphi_i, i \geq 0\}$ un ensemble complet orthonormé de fonctions de $\mathbb{L}^2([0, 1])$.

Par la représentation de Karhunen-Loève et d'après le lemme de Mercer, on a :

$$\forall k \in \mathbb{N}, \quad \xi_{t+k} = \sum_{i=0}^{\infty} \eta_i^k \varphi_i(t), \quad \text{avec } \eta_i^k = \int_0^1 \xi_{t+k} \varphi_i(t) dt.$$

On considère $(X_k)_{k \in \mathbb{N}}$ la suite de variables aléatoires à valeurs dans l'espace de Hilbert $\mathbb{L}^2([0, 1])$, muni du produit scalaire

$$\langle X_n, \varphi_i \rangle = \int_0^1 X_n(t) \varphi_i(t) dt = \eta_n^i$$

où $(\eta_k^i, i \geq 0, k \in \mathbb{N})$ est une suite de variables aléatoires gaussiennes et $(\varphi_i, i \geq 0)$ une base orthonormale de $\mathbb{L}^2([0, 1])$, telle que :

$$\forall k \in \mathbb{N}, \quad \forall t \in]0, 1[, \quad X_k(t) = \xi_{t+k}.$$

Alors $(X_k)_{k \in \mathbb{N}}$ est une suite de variables aléatoires quasi-associées à valeurs dans $\mathbb{L}^2([0, 1])$.

1.3 Quelques inégalités utiles

1.3.1 Inégalité de Markov

Soit X une variable aléatoire réelle finie. On suppose que X est presque sûrement positive ou nulle, alors, pour tout réel a strictement positif, on a

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

1.3.2 Inégalité de Hölder

Soit $1 < p, q < \infty$, tels que $\frac{1}{p} + \frac{1}{q} = 1$ et $X \in \mathfrak{L}^p(\Omega, \mathcal{A}, \mathbb{P})$, $Y \in \mathfrak{L}^q(\Omega, \mathcal{A}, \mathbb{P})$, alors on a

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{\frac{1}{p}} \mathbb{E}[|Y|^q]^{\frac{1}{q}}.$$

1.3.3 Inégalité de covariance

Soit $(X_n)_{n \in \mathbb{Z}}$ une famille de variables aléatoires α -mélangeantes. Etant donnée la variable aléatoire Γ (resp. Γ') qui, pour tout $k \in \mathbb{Z}$, est $\sigma(X_i, -\infty \leq i \leq k)$ -mesurable (resp. $\sigma(X_i, n+k \leq i \leq +\infty)$ -mesurable).

Proposition 1.3.1 *Si Γ et Γ' sont bornées, alors :*

$$(1.8) \quad \exists C > 0, \quad \text{Cov}(\Gamma, \Gamma') \leq C\alpha(n).$$

Soient p, q et r trois réels positifs tels que $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$ et $\mathbb{E}[\Gamma^p] < \infty$ (resp. $\mathbb{E}[\Gamma'^q] < \infty$), alors :

$$(1.9) \quad \exists C > 0, \quad \text{Cov}(\Gamma, \Gamma') \leq C (\mathbb{E}[\Gamma^p])^{\frac{1}{p}} (\mathbb{E}[\Gamma'^q])^{\frac{1}{q}} (\alpha(n))^{\frac{1}{r}}.$$

1.3.4 Inégalité de type Fuk-Nagaev

L'inégalité exponentielle un autre outil que nous avons utilisé de manière déterminante dans notre problème de convergence presque complète et que

l'on peut rencontrer sous l'appellation de "inégalité de type Fuk-Nagaev" dans Rio [91]. Nous nous en tiendrons ici à la version la plus explicite que celles qui sont effectivement disponibles dans la littérature mais, qui est largement suffisante dans notre contexte et que l'on trouve sous cette forme dans le livre de Ferraty et Vieu [6].

Soit $\{X_i, i \in \mathbb{N}^*\}$ une famille de variables aléatoires identiquement distribuées et qui vérifie la condition de mélange fort avec des coefficients à décroissance algébrique.

On pose

$$s_n^2 = \sum_{i=1}^n \sum_{j=1}^n |\text{Cov}(X_i, X_j)|.$$

Proposition 1.3.2 *S'il existe $p > 2$ et $M > 0$ tel que $(\forall t > M, \mathbb{P}(|X_1| > t) \leq t^{-p})$, alors :*

$$(1.10) \quad \forall \varepsilon > 0, \forall r \geq 1, \quad \mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| > \varepsilon\right) \leq C \left\{ \left(1 + \frac{\varepsilon^2}{r s_n^2}\right)^{\frac{-r}{2}} + n r^{-1} \left(\frac{r}{\varepsilon}\right)^{\frac{p(a+1)}{(a+p)}} \right\}.$$

Cette inégalité est une des extensions de l'inégalité de Bernstein suivante, qu'on peut trouver sous diverses formes dans Hoeffding [50].

Théorème 1.3.1 (Rio([90])) *Soit $\{X_1, \dots, X_n\}$ une famille de variables aléatoires réelles indépendantes telle que $\mathbb{E}[X_i] = 0$. Supposons qu'il existe des constantes positives V et K telles que, pour tout t positif,*

$$\log \mathbb{E}[\exp(tS_n)] \leq \frac{Vt^2}{2(1 - Kt)}.$$

avec $(S_n)_n$ une martingale adaptée à la filtration $(\mathcal{F}_n)_n$ telle que $S_0 = 0$, alors :

$$\mathbb{P}(S_n^* \geq t) \leq \exp\left(-\frac{t^2}{2(V + Kt)}\right),$$

où $S_n^* = \max(0, S_1, \dots, S_n)$.

Le développement des résultats de l'estimation non paramétrique sous dépendance s'est fait en parallèle avec celui de telles inégalités. Depuis, la puissance de ces inégalités a été améliorée dans Bosq [9], jusqu'à l'inégalité de Rio [91]. Le lecteur intéressé trouvera dans Rio [89] mais aussi dans Bosq [10] un ensemble de plusieurs versions de cette inégalité, dont certaines sont sous des énoncés plus généraux.

1.3.5 Inégalité de Kallabis et Neumman(2006) [60]

Soient X_1, \dots, X_n des variables aléatoires réelles telles que $\mathbb{E}[X_i] = 0$ et $\mathbb{P}(|X_i| \leq M) = 1$, pour $i = 1, \dots, n$ et $M < \infty$. Soit $\sigma_n^2 = \text{Var}(X_1 + \dots + X_n)$. Supposons qu'il existe $K < \infty$ et $\beta > 0$ tels que pour $1 \leq s_1 \leq \dots \leq s_u$ et $1 \leq t_1 \leq \dots \leq t_v$, on a l'inégalité suivante :

$$|\text{Cov}(X_{s_1} \dots X_{s_u}, X_{t_1} \dots X_{t_v})| \leq K^2 M^{u+v-2} v \exp\{-\beta(t_1 - s_u)\},$$

alors, pour tout t positif,

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left(-\frac{\frac{t^2}{2}}{A_n + B_n^{\frac{1}{3}} t^{\frac{5}{3}}}\right),$$

où $A_n \leq \sigma_n^2$ et $B_n = \left(\frac{2(K \vee M)}{1 - \exp(-\beta)}\right) \left(\frac{16nk^2}{9A_n(1 - \exp(-\beta))} \vee 1\right)$.

Bibliographie

- [1] Alam, K., Saxena, K. M. L. (1981). Positive dependence in multivariate distributions. *Comm. Statist. A* 10 : 1183-1196.
- [2] Araki, Y., Konishi, S. et Imoto, S. Functional discriminant analysis for microarray gene expression data via radial basis function networks. (English summary) *COMPSTAT 2004-Proceedings in Computational Statistics*, Physica, Heidelberg 613-620.
- [3] Attaoui, S., Laksaci, A. & Ould Saïd, E. (2012). Some asymptotics results on the nonparametric conditional density estimate in the single index for quasiassociated Hilbertian processes. Preprint, LMPA, No 476, June 2012. Univ. du Littoral Côte d'Opale. Submitted.
- [4] Baek, J. I., Park, S. T., Chung, S. M., Seo, H. Y. (2005). On the almost sure convergence of weighted sums of negatively associated random variables. *Commun. Korean Math. Soc.* 20 : 539-546.
- [5] Bagai, I., Prakasa Rao, B. L. S. (1995). Kernel-type density and failure rate estimation for associated sequences. *Ann. Inst. Statist. Math.* 47 : 253-266.
- [6] Billingsley, Patrick. *Convergence of probability measures*. John Wiley & Sons, Inc., New York-London-Sydney 1968
- [7] Birkel, T. (1988) On the convergence rate in the central limit theorem for associated processes. *Ann. Probab.*, 16., 1685-1698.

- [8] Bosq, D. (1991) Modelization, nonparametric estimation and prediction for continuous time processes. In Nonparametric functional estimation and related topics (Spetses, 1990), 509-529, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. 335, Kluwer Acad. Publ., Dordrecht.
- [9] Bosq, D. (1993). Bernstein type large deviations inequalities for partial sums of strong mixing processes. *Statistics*, 24, 59-70.
- [10] Bosq, D. (1996). Limits theorems for Banach-valued autoregressive processes. Application to real continuous time processes. *Bull. Belg. Math. Soc. Simon. Stevin.* 3(5) : 537-555.
- [11] Bosq, D. (2000) Linear Processes in Function Spaces : Theory and Applications Lecture Notes in Statistics 149 Springer-Verlag, New York.
- [12] Bradley, R. C. Counterexamples to the central limit theorem under strong mixing conditions. *Limit theorems in probability and statistics, Vol. I, II (Veszprém, 1982)*, 153–172, *Colloq. Math. Soc. János Bolyai*, 36, North-Holland, Amsterdam, 1984.
- [13] Bradley, R.C. (2005) Basic properties of strong mixing conditions. A survey and some open questions. Update of, and a supplement to, the 1986 original. *Probab. Surv.* 2 107-144
- [14] Bradley, Richard C. Basic properties of strong mixing conditions. *Dependence in probability and statistics (Oberwolfach, 1985)*, 165–192, *Progr. Probab. Statist.*, 11, Birkhäuser Boston, Boston, MA, 1986.
- [15] Bradley, Richard C., Jr. On the strong mixing and weak Bernoulli conditions. *Z. Wahrsch. Verw. Gebiete* 51 (1980), no. 1, 49–54.
- [16] Bradley, Richard C. On quantiles and the central limit question for strongly mixing sequences. Dedicated to Murray Rosenblatt. *J. Theoret. Probab.* 10 (1997), no. 2, 507–555. MR1455156
- [17] Bulinski, A.V. (1996) On the convergence rates in the central limit theorem for positively and negatively dependent random fields. In : Ibragimov, I.A., Zaitsev, A.Yu. (Eds.), *Proceedings of the Kolmogorov semester in*

- the Euler Math. Inst., March 1993, Probability Theory and Mathematical Statistics. Gordon and Breach, London, 3-14.
- [18] Bulinski, A.V. (2000) Asymptotic normality of quasi-associated vector random fields. *Obozr. Prom. i Prikl. Mat.*, 7, 482-483 (in Russian).
- [19] Bulinski, A.V., Shabanovich, E. (1998) Asymptotical behaviour of some functionals of positively and negatively dependent random fields. *Fundam. Appl. Math.*, 4, 479-492 (in Russian).
- [20] Bulinski, A.V., Shashkin, A.P. (2003) : Rates in the CLT for vector-valued random fields. Prépublications no. 3/24 de Mathématiques de l'Université Paris-X Nanterre. 16 pp.
- [21] Bulinskii, Alexander ; Doukhan, Paul. Inégalités de mélange fort utilisant des normes d'Orlicz. (French) [Strong mixing moment inequalities using Orlicz norms] *C. R. Acad. Sci. Paris Sér. I Math.* 305 (1987), no. 19, 827–830.
- [22] Bulinski, Alexander ; Suquet, Charles. Normal approximation for quasi-associated random fields. *Statist. Probab. Lett.* 54 (2001), no. 2, 215–226.
- [23] Burton, R., Dabrowski, A.R., Dehling, H. (1986) An invariance principle for quasiassociated random vectors. *Stochastic Process. Appl.*, 23, 301-306.
- [24] Cai, Z., Roussas, G. G. (1998). Kaplan-Meier estimator under association. *J. Multivariate Anal.* 67 : 318-348.
- [25] Campbell, M. Karen ; Donner, Allan. Classification efficiency of multinomial logistic regression relative to ordinal logistic regression. *J. Amer. Statist. Assoc.* 84 (1989), no. 406, 587-591.
- [26] Cao, J. et Ramsay, J.O. (2007) Parameter cascades and profiling in functional data analysis *Comp. Stat.* 22 (3) 335-351.
- [27] Cardot, H., Crambes, C. and Sarda, P. (2006) Ozone pollution forecasting using conditional mean and conditional quantiles with functional covariates. *Statistical methods for biostatistics and related fields*, Hårdle, Mori and Vieu (Eds.), Springer.

- [28] Cox, J. T., Grimmett, G. (1984). Central limit theorem for associated random variables and the percolation model. *Ann. Probab.* 12 : 514-528.
- [29] Dabo-Niang, S. (2004b) Kernel density estimator in an infinite-dimensional space with a rate of convergence in the case of diffusion process. *Appl. Math. Lett.* 17 (4) 381-386.
- [30] Dauxois, J., Pousse, A. and Romain, Y. (1982) Asymptotic theory for the principal component analysis of a vector random function : some applications to statistical inference *J. Multivariate Anal.* 12 (1) 136-154.
- [31] Davidian, M., Lin, X. et Wang, J.-L. (2004) Introduction [Emerging issues in longitudinal and functional data analysis] *Statistica Sinica* 14 (3) 613-614.
- [32] Delsol, Laurent. Nonparametric methods for α -mixing functional random variables. *The Oxford handbook of functional data analysis*, 130–165, Oxford Univ. Press, Oxford, 2011.
- [33] Demichev, V. P. A functional central limit theorem for the volumes of excursion sets of quasiassociated random fields. (Russian) *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* 412 (2013), *Veroyatnost i Statistika.* 19, 109-120, 275 ; translation in *J. Math. Sci. (N. Y.)* 204 (2015), no. 1, 69-77
- [34] Doosti, H., Dewan, I. (2010). Wavelet linear density estimation for associated stratified size-biased sample.
- [35] Douge, L. (2010). Théorèmes limites pour des variables quasi-associées hilbertiennes. *Ann. I.S.U.P.*, 54, 51-60.
- [36] Doukhan, P. (1994). *Mixing : Properties and Examples*. Lecture Notes in Statistics., Springer-Verlag, Berlin.
- [37] Doukhan, Paul ; Louhichi, Sana. A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* 84 (1999), no. 2, 313-342.

- [38] Erbas, Bircan ; Hyndman, Rob J. ; Gertig, Dorota M. Forecasting age-specific breast cancer mortality using functional data models. *Stat. Med.* 26 (2007), no. 2, 458–470.
- [39] Esary, J. D. ; Proschan, F. ; Walkup, D. W. Association of random variables, with applications. *Ann. Math. Statist.* 38 1967 1466–1474.
- [40] Ferraty, Frédéric ; Goia, Aldo ; Vieu, Philippe. Régression non-paramétrique pour des variables aléatoires fonctionnelles mélangeantes. (French) [Nonparametric regression for mixing functional random variables] *C. R. Math. Acad. Sci. Paris* 334 (2002), no. 3, 217–220.
- [41] Ferraty, F., Vieu, P. (2006). *Nonparametric functional data analysis. Theory and practice*
- [42] Fortuin, C. M. ; Kasteleyn, P. W. ; Ginibre, J. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.* 22 (1971), 89–103.
- [43] Gasser, T., Hall, P. et Presnell, B. (1998) Nonparametric estimation of the mode of a distribution of random curves. *J. R. Statomptes. Soc. Ser. B Stat. Methodol.* 60 (4) 681-691.
- [44] G. Estévez-Pérez, A. Quintela-del-Rio and P. Vieu, Convergence rate for cross-validatory bandwidth in kernel hazard estimation from dependent samples. *J. Statist. Plann. Inference* 104 (2002), 1-30.
- [45] Giorgi, Roch ; Belot, Aurélien ; Gaudart, Jean ; Launoy, Guy. The performance of multiple imputation for missing covariate data within the context of regression relative survival analysis. *Stat. Med.* 27 (2008), no. 30, 6310-6331.
- [46] Grenander, U. (1981) *Abstract inference*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.
- [47] Cai, Guang-hui. A strong invariance principle for negatively associated random fields." *Czechoslovak Mathematical Journal* 61.1 (2011) : 27-40.
- [48] Hall, P., Müller, H.G. et Wang, J.L. (2006) Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* 34 (3) 1493-1517.

- [49] Henriques, C., Oliveira, P. E. (2005). Exponential rates for kernel density estimation under association. *Statist. Neerlandica* 59 (4) : 448-466.
- [50] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58, 15-30.
- [51] Ibragimov, I. A. On the spectrum of stationary Gaussian sequences satisfying a strong mixing condition. II. Sufficient conditions. Rate mixing. (Russian) *Teor. Veroyatnost. i Primenen* 15 1970 24-37. Welsch, Roy E. A weak convergence theorem for order statistics from strong-mixing processes. *Ann. Math. Statist.* 42 (1971), 1637-1646.
- [52] Ioannides, D. A., Roussas, G. G. (1999). Exponential inequality for associated random variables. *Statist. Probab. Lett.* 42 : 423-431.
- [53] J.P. Nielson and O. Linton, Kernel estimation in a nonparametric marker dependent hazard model. *Ann. Statist.* 23 (1995), 1735-1748.
- [54] Jabbari, H., Azarnoosh, H. (2006). Almost sure convergence rates for the estimation of a covariance operator for negatively associated samples. *JIRSS* (2006) Vol. 5, Nos. 1-2, pp 53-67
- [55] James, G.M., Hastie, T.J. and Sugar, C.A. (2000) Principal component models for sparse functional data *Biometrika* 87 587-602. Watson, Geoffrey S. Smooth regression analysis. *Sankhya Ser. A* 26 1964 359-372.
- [56] James, G.M. et Sugar, C.A. (2003) Clustering for sparsely sampled functional data. *J. Amer. Statist. Assoc.* 98 (462) 397-408.
- [57] Jim Ramsay et Bernard Silverman. *Functional Data Analysis*. Springer Series in Statistics. Springer Verlag, June 1997.
- [58] Joag-Dev, Kumar ; Proschan, Frank. Negative association of random variables, with applications. *Ann. Statist.* 11 (1983), no. 1, 286-295.
- [59] Jones, M. C. ; Park, Heungsun ; Shin, Key-Il ; Vines, S. K. ; Jeong, Seok-Oh. Relative error prediction via kernel regression smoothers. *J. Statist. Plann. Inference* 138 (2008), no. 10, 2887-2898.
- [60] Kallabis, R. S. et Neumann, M. H. (2006). An exponential inequality under weak dependence. *Bernoulli*, 12(2), 333-350.

- [61] Lehmann, E. L. Some concepts of dependence. *Ann. Math. Statist.* 37 1966 1137–1153.
- [62] Leng X. et Müller, H.G. (2006) Classification using functional data analysis for temporal gene expression data. *Bioinformatics* 22 68-76.
- [63] Li, Y. X., Wang, J. F. (2008). The law of the iterated logarithm for positively dependent random variables. *J. Math. Anal. Appl.* 33 : 259-265
- [64] Lin, Z., Li, D. (2007). Asymptotic normality for L1-norm kernel estimator of conditional median under association dependence. *J. Multivariate Anal.* 98 : 1214-1230.
- [65] Locantore, N., Marron, J.S., Simpson, D.G., Tripoli, N., Zhang, J.T. and Cohen, K.L. (1999) Robust component analysis for functional data (with discussion) 8 1-74.
- [66] Louani, Djamel ; Ould-Saïd, Elias. Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis. *J. Nonparametr. Statist.* 11 (1999), no. 4, 413–442.
- [67] Masry, E. (2002). Multivariate probability density estimation for associated processes : strong consistency and rates. *Statist. Probab. Lett.* 58 : 205-219
- [68] Masry, Elias. Local polynomial fitting under association. *J. Multivariate Anal.* 86 (2003), no. 2, 330-359.
- [69] Masry, Elias. Nonparametric regression estimation for dependent functional data : asymptotic normality. *Stochastic Process. Appl.* 115 (2005), no. 1, 155–177.
- [70] Matula, P. (1992). A note on the almost sure convergence of sums of negatively dependent random variables. *Statist. Probab. Lett.* 15 : 209-213.
- [71] Merlevède, Florence ; Peligrad, Magda. The functional central limit theorem under the strong mixing condition. *Ann. Probab.* 28 (2000), no. 3, 1336–1352.

- [72] Mi-Hwa Ko. On the marcinkiewicz zygmund laws of large numbers for negatively dependent random fields. *Honam Mathematical J.* 33 (2011), No. 2, pp. 163-171.
- [73] Müller, H.-G. (2005) Functional modelling and classification of longitudinal data. With discussions by Ivar Heuch, Rima Izem, and James O. Ramsay and a rejoinder by the author. *Scand. J. Statist.* 32 (2) 223-246.
- [74] Newman, C. M. (1980). Normal fluctuations and the FKG inequalities. *Comm. Math. Phys.* 74 : 119-128.
- [75] Newman, C. M. (1984). Asymptotic independence and limit theorems for positively and negatively dependent random variables, in :Tong, Y. L. (Ed.) *Inequalities in statistics and probability*, IMS Lecture Notes-Monograph Ser., vol. 5 Inst. Math. Statist., Hayward, CA : 127-140.
- [76] Newman, C. M. , Wright, A. L. (1981). An invariance principle for certain dependent sequences. *Ann. Probab.* 9 : 671-675.
- [77] Oliveira, P. E. (2005). An exponential inequality for associated variables. *Statist. Probab. Lett.* 73(2) : 189-197.
- [78] Park, Heungsun ; Stefanski, L. A. Relative-error prediction. *Statist. Probab. Lett.* 40 (1998), no. 3, 227-236.
- [79] Peligrad, M., Shao, Q.-M. (1995) Estimation of variance of partial sums of an associated sequence of random variables. *Stochastic proc. Appl.*, 56, 307-319.
- [80] Philippe Besse, Hervé Cardot, et David Stephenson. Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics*, 4 :673-688, 2000.
- [81] Pitt, L. D. (1982). Positively correlated normal variables are associated. *Ann. Probab.* 10 (3) : 496-499.
- [82] R.F. Engle and J. Russell, Autoregressive conditional duration : A new model for irregularly spaced transaction data. *Econometrica* 66 (1988), 1127-1162.

- [83] Ramsay, J. and Silverman, B. (2002) Applied functional data analysis : Methods and case studies Springer-Verlag, New York.
- [84] Ramsay, J. and Silverman, B. (2005) Functional Data Analysis (Second Edition) Springer-Verlag, New York.
- [85] Ramsay, J.O. (1982) When the data are functions. *Psychometrika* 47 (4) 379- 396.
- [86] Ramsay, J.O., Bock, R. et Gasser, T. (1995) Comparison of height acceleration curves in the Fels, Zurich, and Berkeley growth data, *Annals of Human Biology* 22 413-426.
- [87] Rao, C. R. (1958) Some statistical methods for comparison of growth curves. *Biometrics* 14 1-17.
- [88] Rao, Madiraju Sudhakara. Weak convergence of empirical process for non-stationary strong mixing sequences and their applications to asymptotic normality and efficiency of linear rank statistics. Thesis (Ph.D.)-University of Alberta (Canada). ProQuest LLC, Ann Arbor, MI, 1974.
- [89] Rio, E. (1994). Covariance inequalities for strongly mixing processes. *Ann. de l.I.H.P. Probab..Statist.*, 29, 587.597.
- [90] Rio, E. Inégalités exponentielles et inégalités de concentration, Notes de cours provisoires, version 2, www.math.u-bordeaux1.fr/imb/IMG/pdf/Rio1.pdf
- [91] Rio, E. (2000). Théorie asymptotique des processus aléatoires faiblement dépendants, *Mathématiques et applications*. Springer. New-York.
- [92] ROSENBLATT, M. (1956) A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U. S. A.* 42 43-47.
- [93] Rosenblatt, Murray. Uniform ergodicity and strong mixing. *Z. Wahrscheinlichkeits theorie und Verw. Gebiete* 24 (1972), 79-84.
- [94] Rosenblatt, M. A two-dimensional smoothing spline and a regression problem. *Limit theorems in probability and statistics, Vol. I, II (Veszprém, 1982)*, 915–931, *Colloq. Math. Soc. János Bolyai*, 36, North-Holland, Amsterdam, 1984. MR0807590

- [95] Roussas, G.G. (1994) Asymptotic normality of random fields of positively and negatively associated processes. *J. Multivar. Anal.*, 50, 152-173.
- [96] Roussas, G. G. (1994). Asymptotic normality of a smooth estimate of random fields of positively or negatively associated processes. *J. Multivariate Anal.* 50 : 152-173.
- [97] Roussas, G. G. (2000). Asymptotic normality of the kernel estimate of a probability density function under association. *Statist. Probab. Lett.* 50 : 1-12.
- [98] Roussas, George G. Kernel estimates under association : strong uniform consistency. *Statist. Probab. Lett.* 12 (1991), no. 5, 393-403.
- [99] Ruiz-Velasco, S. Asymptotic efficiency of logistic regression relative to linear discriminant analysis. *Biometrika* 78 (1991), no. 2, 235-243.
- [100] Shashkin, A. P. Quasi-associatedness of a Gaussian system of random vectors. (Russian) *Uspekhi Mat. Nauk* 57 (2002), no. 6(348), 199-200 ; translation in *Russian Math. Surveys* 57 (2002), no. 6, 1243-1244.
- [101] Shelkovich, V.M., Associated and quasi associated homogeneous distributions (generalized functions)," *J. Math. Anal. Appl.* 338 (2008), no. 1, 48-70.
- [102] Song, J.J., Lee, H.J., Morris, J.S. et Kang, S. (2007) Clustering of time-course gene expression data using functional data analysis *Computational Biology and Chemistry* 31 (4) 265-274.
- [103] Su, C., Zhao, L., Wang, Y. (1997). Moment inequalities and weak convergence for negatively associated sequences. *Sci. China (A)* 40 : 172-182.
- [104] Suquet.Ch. Introduction à l'association.PUB. IRMA, LILLE 1994 Vol. 34, No I.
- [105] Tarpey, T., Petkova, E. et Ogden, R.T.Profiling placebo responders by selfconsistent partitioning of functional data. *J.A.S.A.* 98 (464) 850-858.
- [106] Vidakovic, B. (2001) Wavelet-Based Functional Data Analysis : Theory, Applications and Ramifications *Proceedings PSFVIP-3* F3399.

- [107] Welsch, Roy E. Limit laws for extreme order statistics from strong-mixing processes. *Ann. Math. Statist.* 43 (1972), 439-446.
- [108] Xing, G. On the exponential inequalities for strictly stationary and negatively associated random variables *J. Statist. Plann. Inference*, 139 (2009), pp. 3453-3460.
- [109] Xing, G., Yang, S. (2010b). Some exponential inequalities for positively associated random variables and rates of convergence of the Strong Law of Large Numbers. *J. Theo. Probab.* 23 : 169-192.
- [110] Zhang, L. (2001). A law of the iterated logarithm for negatively associated random fields, (Technical report).

Chapitre 2

Régression non-paramétrique relative

Ce chapitre est divisé en trois sections. Dans la première section on présente notre modèle et son estimateur et les deux sections suivantes seront consacrées au résultat et sa démonstration.

2.1 Présentation du modèle

On considère $\{(X_i, Y_i), i = 1, \dots, n\}$ une suite de n -observations indépendantes, de même loi que le couple (X, Y) . On note f la densité de la variable explicative X par rapport à la mesure de Lebesgue sur \mathbb{R} . Dans tout ce chapitre, le modèle central auquel on s'intéresse est le modèle non-paramétrique que l'on écrit en toute généralité

$$(2.1) \quad Y = r(X) + \varepsilon$$

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

$$\mathbb{E}(\varepsilon/X) = 0.$$

La fonction r est estimée comme solution du problème d'optimisation suivant :

$$\min_r \mathbb{E}[(Y - r(x))^2 | X = x]$$

et donc, la solution de ce problème est explicitement donnée par $r(x) = \mathbb{E}[Y|X = x]$.

Dans le présent travail, on se propose d'estimer la fonction r en considérant le critère d'optimisation suivant :

$$\min_r E\left[\left(\frac{Y - r(x)}{Y}\right)^2 \mid X = x\right].$$

La solution de ce problème est aussi facile à expliciter, en utilisant la dérivée par rapport à r . Ainsi, on trouve comme solution

$$r(x) = \frac{E[Y^{-1}|X = x]}{E[Y^{-2}|X = x]}.$$

Par ailleurs, on suppose que les fonction $E[Y^{-1}|X = x]$ et $E[Y^{-2}|X = x]$ sont des fonctions de classe $C^2(\mathbb{R})$.

En utilisant la méthode du noyau, on estime la fonction r par

$$(2.2) \quad \hat{r}(x) = \frac{\sum_{i=1}^n Y_i^{-1} K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^n Y_i^{-2} K\left(\frac{x - X_i}{h}\right)}$$

où K est une fonction de \mathbb{R} dans \mathbb{R} et $h = h_n$ un paramètre réel strictement positif. Cet estimateur peut être exprimé par

$$(2.3) \quad \hat{r}(x) = \frac{\hat{g}_1(x)}{\hat{g}_2(x)}$$

avec

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

et

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

2.2 Convergence presque complète de l'estimateur

On fixe un point $x \in \mathbb{R}$ et on introduit les conditions suivantes :

2.2.1 Notations et hypothèses

- (H1) La fonction f est 2-fois continûment dérivable au voisinage de x .
- (H2) La densité f de la variable explicative et la fonction $\mathbb{E}[Y^{-2}|X = x]$ sont strictements positives au point x .
- (H3) La variable réponse Y est telle que : $|Y| > M > 0$
- (H4) Le noyau K est symétrique supposé borné intégrable et à support compact.
- (H5) Le paramètre de lissage h est tel que :

$$\lim_{n \rightarrow \infty} \frac{nh}{\log n} = \infty.$$

On établit le résultat suivant :

Théorème 2.2.1 *Sous les hypothèses (H1)-(H5), on a*

$$(2.4) \quad \hat{r}(x) - r(x) = O(h^2) + O\left(\sqrt{\frac{\log n}{nh}}\right), \text{ p.co.}$$

2.3 Démonstration du théorème

Notre objectif est de démontrer que $\hat{r}(x) = \frac{\sum_{i=1}^n Y_i^{-1} K\left(\frac{d(x, X_i)}{h}\right)}{\sum_{i=1}^n Y_i^{-2} K\left(\frac{d(x, X_i)}{h}\right)}$ converge presque complètement vers $r(x) = \frac{g_1(x)}{g_2(x)}$.

On notera :

$$g_1(x) = \mathbb{E}[Y^{-1}|X = x]f(x) \quad \text{et} \quad g_2(x) = \mathbb{E}[Y^{-2}|X = x]f(x).$$

La démonstration de ce résultat sera basée sur la décomposition suivante :

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

Ainsi, le théorème est une conséquence directe des lemmes suivants :

Lemme 2.3.1 *Sous les hypothèses (H1)-(H4), on a*

$$(2.5) \quad \mathbb{E}\hat{g}_1(x) - g_1(x) = O(h^2)$$

$$(2.6) \quad \mathbb{E}\hat{g}_2(x) - g_2(x) = O(h^2)$$

Lemme 2.3.2 *Sous les hypothèses (H1)-(H5), on a*

$$(2.7) \quad \hat{g}_1(x) - \mathbb{E}\hat{g}_1(x) = O\left(\sqrt{\frac{\log n}{nh}}\right)$$

$$(2.8) \quad \hat{g}_2(x) - \mathbb{E}\hat{g}_2(x) = O\left(\sqrt{\frac{\log n}{nh}}\right)$$

Lemme 2.3.3 *Sous les hypothèses du lemme 2.3.2*

$$(2.9) \quad \exists \delta > 0, \quad \text{tel que} \quad \sum_{i=1}^n P[|\hat{g}_2(x)| < \delta] < \infty$$

2.3.1 Démonstration des Lemmes

Démonstration du lemme (2.3.1)

Pour le premier résultat, on a

$$\begin{aligned}\mathbb{E}[\hat{g}_1(x)] &= \mathbb{E}\left[\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) Y_i^{-1}\right] \\ &= \frac{1}{h} \mathbb{E}\left[K\left(\frac{x - X_1}{h}\right) Y_1^{-1}\right] \\ &= \frac{1}{h} \mathbb{E}\left[K\left(\frac{x - X_1}{h}\right) E[Y_1^{-1} | X = X_1]\right] \\ &= \frac{1}{h} \int_{\mathbb{R}} K\left(\frac{x - z}{h}\right) f(z) \mathbb{E}[Y^{-1} | X = z] dz\end{aligned}$$

Donc

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

En utilisant le changement de variables $t = \frac{x-z}{h}$, on obtient

$$\mathbb{E} [\hat{g}_1(x)] = \int_{\mathbb{R}} K(t) g_1(x - ht) dt.$$

Par suite, l'application du développement de Taylor de g_1 à l'ordre 2 nous permet d'écrire :

$$g_1(x - th) = g_1(x) - thg_1'(x) + \frac{(th)^2}{2} g_1^{(2)}(x) + o(h^2).$$

Sachant que K est un noyau symétrique, alors

$$E [\hat{g}_1(x)] = g_1(x) + \frac{h^2}{2} \int t^2 K(t) g_1^{(2)}(x) + O(h^2).$$

D'où on conclut que

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

La démonstration du résultat (2.6) est similaire à celle qui précède. De la même manière ;

$$\begin{aligned} E [\hat{g}_2(x)] &= \frac{1}{h} \mathbb{E} \left[K \left(\frac{x - X_1}{h} \right) \mathbb{E} [Y_1^{-2} | X = X_1] \right] \\ &= \frac{1}{h} \int K \left(\frac{x-z}{h} \right) f(z) \mathbb{E} [Y^{-2} | X = z] dz \\ &= \frac{1}{h} \int K \left(\frac{x-z}{h} \right) g_2(z) dz. \end{aligned}$$

En considérant le changement de variables $t = \frac{x-z}{h_n}$ et en utilisant le développement de Taylor de g_2 à l'ordre 2, on montre que

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

ce qui achève la preuve du lemme.

Démonstration du lemme (2.3.2)

On a

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

Appliquons l'inégalité de Hoeffding :

Posons

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

Comme K et Y^{-1} sont supposées bornées, alors il existe une constante C telle que

$$|\Delta_i| \leq \frac{C}{h_n}.$$

Sachant que :

$$\mathbb{E}[\Delta_i^2] = \text{Var}[\mu_i] \leq \mathbb{E}[\mu_i^2],$$

alors :

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

où :

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

En utilisant le changement de variables $t = \frac{x - z}{h_n}$, on aura :

$$\mathbb{E}(\mu_i^2) = \frac{1}{h} \int \phi(x - zh) f(x - zh) K^2(z) dz.$$

Puisque la fonction ϕ est bornée, la densité f continue et le noyau K est à support compact, alors, il existe une constante C telle que :

$$\mathbb{E}(\mu_i^2) \leq \frac{C}{h}.$$

Donc en appliquant l'inégalité de Hoeffding, on aura :
pour tout $\varepsilon > 0$ et pour tout $\delta_2 > 1$,

$$(2.10) \quad P \left[\frac{1}{n} \left| \sum_{i=1}^n \Delta_i \right| > \varepsilon \right] \leq 2 \exp \left(\frac{-n\varepsilon^2}{4\delta_2} \right).$$

En prenant $\varepsilon = \varepsilon_0 \sqrt{\frac{\log n}{nh}}$, on obtient

$$\begin{aligned} \mathbb{P} \left[|\mathbb{E}\hat{g}_1(x) - \hat{g}_1(x)| > \varepsilon_0 \sqrt{\frac{\log n}{nh}} \right] &\leq 2 \exp \left(\frac{-n\varepsilon_0^2 h \log n}{4nhC} \right) \\ &\leq 2 \exp \left(\frac{\varepsilon_0^2 \log n}{4C} \right) \\ &\leq 2n^{-\frac{\varepsilon_0^2}{4C}}. \end{aligned}$$

Par suite, il suffit de choisir $\frac{\varepsilon_0^2}{4C} > 1$ pour que

$$\sum_{i=1}^n P \left[|\mathbb{E}\hat{g}_1(x) - \hat{g}_1(x)| > \varepsilon_0 \sqrt{\frac{\log n}{nh}} \right] \leq \sum_{i=1}^n 2n^{-\frac{\varepsilon_0^2}{4C}} < \infty.$$

- En utilisant les mêmes calculs que précédent on montre que :

$$\hat{g}_2(x) - \mathbb{E}\hat{g}_2(x) = O \left(\sqrt{\frac{\log n}{nh}} \right)$$

Démonstration du lemme (2.3.3)

Le lemme (2.3.2) entraîne en particulier la convergence presque complète de $\hat{g}_2(x)$ vers $g_2(x)$.

Ainsi, pour tout $\varepsilon > 0$ on a

$$\sum_{i=1}^n P [|\hat{g}_2(x) - g_2(x)| > \varepsilon] < \infty.$$

En remarquant 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

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

Donc

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

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

Régression relative pour variables associées

3.1 Introduction

Dans une vaste panoplie de sujets à la pointe des recherches dans divers domaines tels que la finance, la biologie et la médecine, de nombreux mathématiciens et statisticiens s'appliquent assidûment aux mécanismes de la dépendance entre deux ou plusieurs variables aléatoires. Cependant on distingue deux types de dépendance : le mélange et l'association. Dans le cadre d'une estimation non-paramétrique, nous évoquons une notion plus large que celle de l'association négative et positive appelée quasi-association. Ce terme qui fut utilisé comme objet en 1967 dans le travail de Esary, Proschan et Walkup [39] qui contient la définition de la quasi association et qui avait pour intérêt initial d'apporter une application de la fiabilité en statistiques . En 1971, Fortuyn, Kastelyn et Ginibre écrivirent un article [42] qui se voulait une application du concept de La notion de dépendance positive ou association à la mécanique statistique ainsi qu'au modèle mathématique de la percolation. On retrouve aussi des résultats à la base des variables négativement associées dans l'article de Joag Dev, Kumar et Subhash en 1983 [58]

où ils développent des propriétés et proposent des applications statistiques. Par ailleurs en 1999, Doukhan et Louhichi ont établi une estimation de la densité pour une suite de variables faiblement dépendantes, ce qui a permis d'adopter une approche unifiée pour les processus négativement associés et les processus gaussiens et dans un cas particulier de la dépendance faible introduit dans [37]. Bulinski et Suquet (2001) ont réalisé une approximation normale pour des champs aléatoires quasi-associés. [68] un autre article dans lequel on a inclu le thème et pour lequel on a étudié la normalité asymptotique de la densité d'une variable vectorielle sous des données fonctionnelles dépendantes. D'autres travaux ont traité le thème de la quasi-association, comme dans l'article de Shelkovich en 2008 [101], on cite aussi l'article de Lahcen Douge [35] en 2010 ainsi que celui de Demichev en 2013 [33].

L'objectif de notre travail porte essentiellement sur l'estimation non paramétrique sous des données quasi-associées, en supposant que la variable explicative soit vectorielle. Dans un le paragraphe suivant , nous présentons les notations et les hypothèses de base sous lesquelles notre résultat sera interprété. Ensuite en deuxième partie, nous aborderons les propriétés asymptotiques de l'estimateur construit. Plus précisément, nous étudierons la vitesse de convergence presque complète et la normalité asymptotique de l'estimateur. Les résultats obtenus de ce chapitre ont fait l'objet d'un article accepté pour publication dans *Metron* 74, 75–97 (2016).

Passons maintenant au développement du contenu de ce chapitre :

3.2 Convergence presque complète

Soit S un compact de \mathbb{R}^d et f (resp. $f_{i,j}$) la densité de probabilité de la variable aléatoire X (resp. la densité conjointe de (X_i, X_j)).

Nous désignons par C et C' des constantes strictement positives et considérons $\lambda_k := \sup_{s \geq k} \sum_{|i-j| \geq s} \lambda_{i,j}$, avec

$$\lambda_{i,j} = \sum_{k=1}^d \sum_{l=1}^d |Cov(X_i^k, X_j^l)| + \sum_{k=1}^d |Cov(X_i^k, Y_j)| + \sum_{l=1}^d |Cov(Y_i, X_j^l)| + |Cov(Y_i, Y_j)|.$$

où X_i^k représente la k -ième composante de X_i .

Notons K le noyau et h_n Une suite de réels positifs qui converge vers zéro lorsque n tend vers l'infini .

À cet effet, nous traitons la convergence presque complète uniforme du $\tilde{r}(x)$ vers $r(x)$ sur le compact S en posant les hypothèses suivantes :

(H1) La densité f de la variable X est strictement positive de class C^2 sur S et verifie :

$$\sup_{|i-j| \geq 1} \|f_{(X_i, X_j)}\|_{\infty} < \infty.$$

(H2) Les fonctions $r_l(\cdot) = E[Y^{-\gamma} | X = \cdot]$ avec $(l = 1, 2)$ sont de classe C^2 sur S .

(H3) Le processus $\{(X_i, Y_i), i \in \mathbb{N}\}$ est quasi-associié avec un coefficient de covariance λ_k qui satisfait

$$\exists a > 0, \exists C > 0 / \lambda_k \leq C \exp\{-ak\}.$$

(H4) La fonction K est une fonction Lipchitzienne, symétrique d'ordre 1 et à support compact.

(H5) $\exists \gamma \in (0, 1)$ et $\xi_1, \xi_2 > 0$ tels que

$$\frac{(\log n)^{\frac{1}{d}}}{n^{\frac{(1-\gamma-\xi_2)}{d}}} \leq h \leq \frac{C}{(\log n)^{\frac{1+\xi_1}{d}}}.$$

(H6) Soient Y^{-l} avec $l = 1, 2$, les moments inverse d'ordre $l = 1; 2$ de la variable réponse tels que

$$\mathbb{E}[\exp(|Y^{-l}|)] \leq C \quad \text{et} \quad \mathbb{E}[|Y_i^{-l} Y_j^{-l}| | X_i, X_j] \leq C' \quad \text{pour} \quad i \neq j.$$

On obtient le théorème suivant

Théorème 3.2.1 *Sous les conditions (H1)-(H6) et si $\inf_{x \in S} g_2(x) > 0$, on a*

$$(3.1) \quad \sup_{x \in S} |\tilde{r}(x) - r(x)| = O(h^2) + O_{a.co.} \left(\sqrt{\frac{\log n}{n^{1-\gamma} h^d}} \right) \quad a.co.$$

Démonstration du Théorème 3.2.1 On écrit

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

où

$$\tilde{g}_l(x) = \frac{1}{nh^d} \sum_{i=1}^n Y_i^{-l} K(h^{-1}(x - X_i)) \quad \text{pour } l = 1, 2.$$

On utilise la décomposition :

$$(3.2) \quad \tilde{r}(x) - r(x) = \frac{1}{\tilde{g}_2(x)} \left[\tilde{g}_1(x) - g_1(x) \right] + [g_2(x) - \tilde{g}_2(x)] \frac{r(x)}{\tilde{g}_2(x)}$$

où

$$g_1(x) = r_1(x)f(x) \quad \text{and} \quad g_2(x) = r_2(x)f(x).$$

Ainsi, le théorème 3.2.1 est une conséquence des résultats intermédiaire suivants (cf. Lemmes 3.2.1 and 3.2.2).

Lemme 3.2.1 *Sous les conditions (H1), (H2), (H4) et (H5), on a, pour $l = 1, 2$:*

$$|E\tilde{g}_l(x) - g_l(x)| = O(h^2).$$

Lemme 3.2.2 *Sous les conditions (H3)-(H6), on a, pour $\gamma = 1, 2$:*

$$|\tilde{g}_l(x) - E\tilde{g}_l(x)| = O_{a.co.} \left(\sqrt{\frac{\log n}{n^{1-\gamma}h^d}} \right).$$

Corollaire 3.2.1 *Sous les conditions du théorème 3.2.1, on obtient :*

$$\sum_n \mathbb{P} \left(\inf_{x \in S} |\tilde{g}_2(x)| \leq \frac{\inf_{x \in S} g_2(x)}{2} \right) < \infty.$$

3.3 La normalité asymptotique

Cette section est consacrée à la normalité asymptotique de l'estimateur $\tilde{r}(x)$. À cet objectif, nous maintenons les conditions adoptées dans la précédente section et on remplace l'hypothèse (H5) par la suivante :

(H5') $\exists \gamma \in (0, 1)$ et $\xi_1, \xi_2 > 0$ tels que

$$\frac{1}{n^{2(1/2-\gamma/9-\xi_2)/d}} \leq h \leq \frac{C}{n^{(1+\xi_1)/(d+4)}}.$$

On obtient le résultat suivant :

Théorème 3.3.1 *Sous les conditions (H1)-(H4), (H5') et (H6) on a pour tout $x \in \mathcal{A}$:*

$$(3.3) \quad \left(\frac{nh^d}{\sigma^2(x)} \right)^{1/2} (\tilde{r}(x) - r(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{quand } n \rightarrow \infty.$$

où

$$\sigma^2 = \frac{(g_2(x) - 2r(x)E[Y^{-3}|X=x] + r^2(x)E[Y^{-4}|X=x]) \int_{\mathbb{R}^d} K^2(z)dz}{g_2^2(x)}$$

$$\mathcal{A} = \left\{ x \in S, (g_2(x) - 2r(x)E[Y^{-3}|X=x] + r^2(x)E[Y^{-4}|X=x]) g_2^2(x) \neq 0 \right\}$$

et $\xrightarrow{\mathcal{D}}$ désigne la convergence en loi.

Démonstration du théorème 3.3.1. Il est clair que

$$(3.4) \quad \tilde{r}(x) - r(x) = \frac{1}{\tilde{g}_2(x)} [B_n + V_n (\tilde{g}_2(x) - E\tilde{g}_2(x))] + V_n$$

où

$$V_n = \frac{1}{E\tilde{g}_2(x)g_2(x)} \left[[E\tilde{g}_1(x)]g_2(x) - [E\tilde{g}_2(x)]g_1(x) \right]$$

et

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

Alors

$$(3.5) \quad \tilde{r}(x) - r(x) - V_n = \frac{1}{\tilde{g}_2(x)} [B_n + V_n (\tilde{g}_2(x) - E\tilde{g}_2(x))].$$

Donc, le théorème 3.3.1 est une conséquence des résultats suivants.

Lemme 3.3.1 *Sous les hypothèses du théorème 3.3.1, on a*

$$(3.6) \quad \left(\frac{nh^d}{g_2^2(x)\sigma^2} \right)^{1/2} (B_n - E[B_n]) \rightarrow \mathcal{N}(0, 1).$$

Lemme 3.3.2 *Sous les hypothèses du théorème 3.3.1, on a :*

$\tilde{g}_2(x) \rightarrow g_2(x)$, en probabilité ,

$$\left(\frac{nh^d}{g_2(x)^2 \sigma^2(x)} \right)^{1/2} V_n \rightarrow 0,$$

et

$$\left(\frac{nh^d}{g_2(x)^2 \sigma^2(x)} \right)^{1/2} V_n (\tilde{g}_2(x) - E\tilde{g}_2(x)) \rightarrow 0, \text{ en probabilité.}$$

3.4 Démonstration des résultats préliminaires

Démonstration du Lemme 3.2.1. Par des arguments standards on obtient pour $l = 1, 2$,

$$\begin{aligned} \mathbb{E}[\tilde{g}_l(x)] &= \frac{1}{h^d} \int_{\mathbb{R}^d} \mathbb{E}[Y^{-l} | X = u] K\left(\frac{x-u}{h}\right) f(u) du \\ &= \frac{1}{h^d} \int_{\mathbb{R}^d} g_l(u) K\left(\frac{x-u}{h}\right) du \\ &= \int_{\mathbb{R}^d} g_l(x-hz) K(z) dz. \end{aligned}$$

Comme les deux fonctions f et r_l sont de classe \mathcal{C}^2 , on utilise le développement du Taylor sous (H4)

$$|\mathbb{E}[\tilde{g}_l(x) - g_l(x)]| \leq Ch^2.$$

Le dernier résultat complète la démonstration de ce lemme. ■

Démonstration du Lemme 3.2.2. L'objectif est de démontrer que

$$(3.7) \quad \sum_{n>0} \mathbb{P} \left(\sup_{x \in S} |\tilde{g}_l^*(x) - \mathbb{E}[\tilde{g}_l^*(x)]| > \varepsilon \sqrt{\frac{\log n}{n^{1-\gamma} h^d}} \right) < \infty.$$

Pour cela, on utilise les techniques de troncature en introduisant la variable aléatoire suivante :

$$\tilde{g}_l^*(x) = \frac{1}{nh^d} \sum_{i=1}^n k\left(\frac{x-X_i}{h_n}\right) Y^{-2} \mathbb{1}_{|Y^{-2}| < \mu_n}$$

avec $\mu_n = n^{\frac{\gamma}{18}}$.

La démonstration de (3.7) sera basée sur les résultats suivants.

$$(3.8) \quad \sup_{x \in S} |\mathbb{E}[\tilde{g}_l^*(x)] - \mathbb{E}[\tilde{g}_l(x)]| = O\left(\sqrt{\frac{\log n}{n^{1-\gamma}h^d}}\right),$$

$$(3.9) \quad \sup_{x \in S} |\tilde{g}_l^*(x) - \tilde{g}_l(x)| = O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\gamma}h^d}}\right)$$

et

$$(3.10) \quad \sup_{x \in S} |\tilde{g}_l^*(x) - \mathbb{E}[\tilde{g}_l^*(x)]| = O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\gamma}h^d}}\right).$$

Commençons par le résultat (3.10) :

Soit \mathbf{S} un compact de \mathbb{R}^d . Supposons que pour $d_n \leq n^\beta$ et $\tau_n \leq d_n^{-1}$,

$$\mathbf{S} \subset \bigcup_{j=1}^{d_n} B(x_k, \tau_n),$$

où $\beta = \frac{\delta(d+2)}{2} + \frac{1}{2} + \frac{\gamma}{18}$ et $\delta \geq (9 - \gamma - \xi_2)/(d+2)$.

Pour tout $x \in S$, on pose :

$$k(x) = \arg \min_{k \in \{1, \dots, d_n\}} \|x - x_k\|.$$

En introduisant la variable aléatoire $\tilde{g}_l^*(x_{k(x)})$, on considère la décomposition suivante :

pour $l = 1, 2$,

$$\begin{aligned} \sup_{x \in S} \left| \tilde{g}_l^*(x) - \mathbb{E}[\tilde{g}_l^*(x)] \right| &\leq \underbrace{\sup_{x \in S} |\tilde{g}_l^*(x) - \tilde{g}_l^*(x_{k(x)})|}_{T_1} \\ &+ \underbrace{\sup_{x \in S} |\tilde{g}_l^*(x_{k(x)}) - \mathbb{E}[\tilde{g}_l^*(x_{k(x)})]|}_{T_2} \\ &+ \underbrace{\sup_{x \in S} \left| \mathbb{E}[\tilde{g}_l^*(x_{k(x)})] - \mathbb{E}[\tilde{g}_l^*(x)] \right|}_{T_3}. \end{aligned}$$

• Pour T_2 , on utilise l'inégalité de Bernstein sur les variables associées (voir Kallabis and Neumann (2006)). En effet,

$$\begin{aligned}
|\tilde{g}_l^*(x_{k(x)}) - \mathbb{E}[\tilde{g}_l^*(x_{k(x)})]| &= \left| \frac{1}{nh_n^d} \sum_{i=1}^n k\left(\frac{x_{k(x)} - X_i}{h_n}\right) Y_i^{-l} \mathbb{I}_{|Y_i^{-l}| < \mu_n} - \right. \\
&\quad \left. - \mathbb{E} \left[\frac{1}{nh_n^d} \sum_{i=1}^n k\left(\frac{x_{k(x)} - X_i}{h_n}\right) Y_i^{-l} \mathbb{I}_{|Y_i^{-l}| < \mu_n} \right] \right| \\
&= \left| \sum_{i=1}^n \frac{1}{nh_n^d} \left(k\left(\frac{x_{k(x)} - X_i}{h_n}\right) Y_i^{-l} \mathbb{I}_{|Y_i^{-l}| < \mu_n} - \right. \right. \\
&\quad \left. \left. - \mathbb{E} \left[k\left(\frac{x_{k(x)} - X_i}{h_n}\right) Y_i^{-l} \mathbb{I}_{|Y_i^{-l}| < \mu_n} \right] \right) \right| \\
&= \left| \sum_{i=1}^n \Delta_i \right|
\end{aligned}$$

Pour $u \in \mathbb{R}^d$, $v \in \mathbb{R}$, on pose

$$\chi_l(u, v) = k\left(\frac{x_{k(x)} - u}{h_n}\right) Y_v^{-l} \mathbb{I}_{|Y_v^{-l}| < \mu_n} - \mathbb{E} \left[k\left(\frac{x_{k(x)} - u}{h_n}\right) Y_v^{-l} \mathbb{I}_{|Y_v^{-l}| < \mu_n} \right].$$

Donc

$$\Delta_i = \frac{1}{nh_n^d} \chi_l(X_i, Y_i)$$

Par suite, la fonction χ satisfait :

$$\|\chi_l\|_\infty \leq C\mu_n^l \|K\|_\infty \quad \text{et} \quad \text{Lip}(\chi_l) \leq C\mu_n^{l+1} h^{-1} \text{Lip}(K).$$

Par conséquent, pour les deux cas $l = 1, 2$, on aura

$$\|\chi_l\|_\infty \leq C\mu_n^3 \|K\|_\infty \quad \text{et} \quad \text{Lip}(\chi_l) \leq C\mu_n^3 h^{-1} \text{Lip}(K).$$

Notons que pour appliquer l'inégalité de Kallabis et Neumann, nous avons besoin de l'évaluation asymptotique des deux quantités $\text{Var}(\sum_{i=1}^n \Delta_i)$ et $\text{Cov}(\Delta_{s_1} \dots \Delta_{s_u}, \Delta_{t_1} \dots \Delta_{t_v})$, pour $(s_1, \dots, s_u) \in \mathbb{N}^u$ et $(t_1, \dots, t_v) \in \mathbb{N}^v$.

On commence par le terme de la variance :

$$(3.11) \quad \text{Var} \left(\sum_{i=1}^n \Delta_i \right) = n \text{Var}(\Delta_1) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \text{Cov}(\Delta_i, \Delta_j).$$

Sous (H6), on a

$$\begin{aligned}
\text{Var}(\Delta_1) &\leq \frac{1}{n^2 h^{2d}} \mathbb{E}[|Y_1^{-l} K_1(x_k)|^2] \leq C' \frac{1}{n^2 h^{2d}} \mathbb{E}[|K_1(x_k)|^2] \\
&\leq C' n^{-2} h^{-d} \int_{\mathbb{R}^d} K^2(u) f(x_k - hu) du \\
(3.12) \quad &= O(n^{-2} h^{-d}).
\end{aligned}$$

Maintenant, nous allons évaluer le comportement asymptotique de la somme du terme à droite de (3.11). Pour cela on utilise les technique de Masry (1986). En effet, on considère la décomposition suivante

$$\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \text{Cov}(\Delta_i, \Delta_j) = \sum_{i=1}^n \sum_{\substack{j=1 \\ 0 < |i-j| \leq m_n}}^n \text{Cov}(\Delta_i, \Delta_j) + \sum_{i=1}^n \sum_{\substack{j=1 \\ |i-j| > m_n}}^n \text{Cov}(\Delta_i, \Delta_j)$$

où (m_n) est une suite d'entiers qui tend vers l'infini lorsque n tend vers l'infini. En ce qui concerne le cas $|i-j| \leq m_n$, on utilise (H1) et (H6) pour arriver à

$$\begin{aligned}
\mathbb{E}[|\Delta_i \Delta_j|] &\leq C \frac{1}{n^2 h^{2d}} (\mathbb{E}[|Y_i^{-l} K_i(x_k) Y_j^{-l} K_j(x_k)|] + (\mathbb{E}[|Y^{-l} K_1(x_k)|])^2) \\
&\leq C \frac{1}{n^2 h^{2d}} (\mathbb{E}[K_i(x_k) K_j(x_k)] + (\mathbb{E}[K_1(x_k)])^2) \\
&\leq \frac{C}{n^2} \left\{ \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} k(u) k(v) f(X_i, X_j)(x_{k(x)} - h_n u, x_{k(x)} - h_n v) dudv + \right. \\
&\quad \left. + \left(\int_{\mathbb{R}^d} k(u) f(x_{k(x)} - h_n u) du \right)^2 \right\} \\
&= O(n^{-2}).
\end{aligned}$$

Ainsi, Ainsi, sur l'ensemble $E_1 = \{(i, j) / |i-j| \leq m_n\}$ nous obtenons

$$\begin{aligned}
\sum_{i=1}^n \sum_{\substack{j=1 \\ 0 < |i-j| \leq m_n}}^n \text{Cov}(\Delta_i, \Delta_j) &\leq n m_n (\mathbb{E}[\Delta_i \Delta_j]). \\
&\leq C n^{-1} m_n.
\end{aligned}$$

D'autre part, pour le deuxième cas où $|i - j| > m_n$, on utilise la condition sur la quasi-association de séquence (X_i, Y_i) et (H3), pour obtenir dans les deux cas $l = 1, 2$

$$\begin{aligned} \sum_{i=1}^n \sum_{\substack{j=1 \\ |i-j| > m_n}}^n \text{Cov}(\Delta_i, \Delta_j) &\leq \mu_n^6 n^{-2} h^{-2(d+1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ |i-j| > m_n}}^n \lambda_{i,j} \\ &\leq \mu_n^6 n^{-1} h^{-2(d+1)} \lambda_{m_n} \\ &\leq \mu_n^6 n^{-1} h^{-2(d+1)} e^{-am_n}. \end{aligned}$$

Donc,

$$\sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \text{Cov}(\Delta_i, \Delta_j) \leq C (n^{-1} m_n + \mu_n^6 n^{-1} h^{-2(d+1)} e^{-am_n}).$$

Au final, pour $m_n = \frac{1}{a} \log(a \mu_n^6 h^{-2(d+1)})$ et sous (H5), il résulte que

$$(3.13) \quad nh^d \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \text{Cov}(\Delta_i, \Delta_j) \rightarrow 0.$$

En combinant (3.12) et (3.13), on aura :

$$\text{Var} \left(\sum_{i=1}^n \Delta_i \right) = O \left(\frac{1}{nh^d} \right).$$

Nous abordons le terme de covariance (3.11), pour tout $(s_1, \dots, s_u) \in \mathbb{N}^u$ et $(t_1, \dots, t_v) \in \mathbb{N}^v$. Pour cela on distingue les trois cas suivants :

– Pour $t_1 = s_u$, on a

$$(3.14) \quad \begin{aligned} |\text{Cov}(\Delta_{s_1} \dots \Delta_{s_u}, \Delta_{t_1} \dots \Delta_{t_v})| &\leq \left(\frac{C \mu_n^3 \|K\|_\infty}{nh^d} \right)^{u+v} \mathbb{E} |K_1^2(x_k)| \\ &\leq h^d \left(\frac{C \mu_n^3}{nh^d} \right)^{u+v}. \end{aligned}$$

- Pour le cas $t_1 > s_u$, on utilise la propriété de la quasi-association pour montrer que

$$\begin{aligned}
|Cov(\Delta_{s_1} \dots \Delta_{s_u}, \Delta_{t_1} \dots \Delta_{t_v})| &\leq \left(\frac{\mu_n^3 h^{-1} \text{Lip}(K)}{nh^d} \right)^2 \left(\frac{2\mu_n^3 \|K\|_\infty}{nh^d} \right)^{u+v-2} \sum_{i=1}^u \sum_{j=1}^v \lambda_{s_i, t_j} \\
&\leq Ch^{-2} \left(\frac{C\mu_n^3}{nh^d} \right)^{u+v} v \lambda_{t_1 - s_u} \\
(3.15) \quad &\leq Ch^{-2} \left(\frac{C\mu_n^3}{nh^d} \right)^{u+v} v e^{-a(t_1 - s_u)}.
\end{aligned}$$

D'autre part, on a

$$\begin{aligned}
|Cov(\Delta_{s_1} \dots \Delta_{s_u}, \Delta_{t_1} \dots \Delta_{t_v})| &\leq \left(\frac{2C\mu_n^3 \|K\|_\infty}{nh^d} \right)^{u+v-2} (|\mathbb{E} \Delta_{s_u} \Delta_{t_1}|) \\
(3.16) \quad &\leq \left(\frac{C\mu_n^3}{nh^d} \right)^{u+v} h^{2d}.
\end{aligned}$$

En combinant (3.15) à la puissance $\frac{d}{2d+2}$ et (3.16) à la puissance $\frac{d+2}{2d+2}$, on trouve la majoration finale

$$|Cov(\Delta_{s_1}, \dots, \Delta_{s_u}, \Delta_{t_1}, \dots, \Delta_{t_v})| \leq h^d \left(\frac{C\mu_n^3}{nh^d} \right)^{u+v} v e^{-\frac{ad}{2d+2}(t_1 - s_u)}.$$

En appliquant le théorème de Kallabis et Neumann pour $K_n = \frac{C\mu_n^3}{n\sqrt{h_n^d}}$,

$M_n = \frac{C\mu_n^3}{nh_n^d}$ et $A_n = \text{Var}(\sum_{i=1}^n \Delta_i) = \mathcal{O}\left(\frac{1}{nh_n^d}\right)$, nous obtenons

$$\begin{aligned}
\mathbb{P}\left(\left|\tilde{g}_l^*(x_{k(x)}) - \mathbb{E}[\tilde{g}_l^*(x_{k(x)})]\right| > \eta\sqrt{\frac{\log n}{n^{1-\gamma}h^d}}\right) &= \mathbb{P}\left(\left|\sum_{i=1}^n \Delta_i\right| > \eta\sqrt{\frac{\log n}{n^{1-\gamma}h^d}}\right) \\
&\leq \exp\left\{-\frac{\frac{t^2}{2}}{A_n + B_n^{\frac{1}{3}}t^{\frac{5}{3}}}\right\} \\
&\leq \exp\left\{-\frac{\frac{-(\eta\sqrt{\frac{\log n}{n^{1-\gamma}h^d}})^2}{2}}{\text{Var}(\sum_{i=1}^n \Delta_i) + \left[\left(\frac{16nK_n}{9A_n(1-\exp\{-\beta\})} \vee 1\right) \left(\frac{2(K_n \vee M_n)}{1-\exp\{-\beta\}}\right)\right]^{\frac{1}{3}} \left(\eta\sqrt{\frac{\log n}{n^{1-\gamma}h^d}}\right)^{\frac{5}{3}}}\right\} \\
&\leq \exp\left\{-\frac{\eta^2 \log n / (2n^{1-\gamma}h^d)}{\left(\text{Var}(\sum_{i=1}^n \Delta_i) + C\mu_n^3(nh^d)^{-\frac{1}{3}} \left(\frac{\log n}{n^{1-\gamma}h^d}\right)^{\frac{5}{6}}\right)}\right\} \\
&\leq \exp\left\{-\frac{\eta^2 \log n}{Cn^{-\gamma} + \mu_n^3 n^{-\gamma/6} \left(\frac{\log^5 n}{nh^d}\right)^{\frac{1}{6}}}\right\} \\
&\leq C' \exp\{-C\eta^2 \log n\}.
\end{aligned}$$

En utilisant le fait que, $d_n \leq n^\beta$, on aura

$$\mathbb{P}\left(\max_{1 \leq k \leq d_n} |\tilde{g}_l^*(x_k) - \mathbb{E}[\tilde{g}_l^*(x_k)]| > \eta\sqrt{\frac{\log n}{nh^d}}\right) \leq C'n^{\beta-C\eta^2}.$$

Le choix convenable de η achèvera la démonstration de la partie T_2 .

• Nous traitons maintenant les termes T_1 et T_3 : Pour cela, on utilise la condition de Lipschitz sur le noyau K . On aura alors ; pour tout $x \in S$,

$$\begin{aligned}
|\tilde{g}_l^*(x) - \tilde{g}_l^*(x_{k(x)})| &= \frac{1}{nh^d} \left| \sum_{i=1}^n Y_i^{-l} K_i(x) \mathbb{1}_{|Y_i^{-1}| < \mu_n} - \sum_{i=1}^n Y_i^{-l} K_i(x_{k(x)}) \mathbb{1}_{|Y_i^{-1}| < \mu_n} \right| \\
&\leq \frac{C}{nh^{d+1}} \|x - x_{k(x)}\| \sum_{i=1}^n Y_i^{-l} \mathbb{1}_{|Y_i^{-1}| < \mu_n} \\
&\leq \frac{C\tau_n}{\mu_n^l nh^{d+1}} \\
&\leq \frac{C\tau_n}{\mu_n nh^{d+1}}.
\end{aligned}$$

comme $\tau_n = O(n^{-\beta})$, il en résulte

$$(3.17) \quad \frac{C\tau_n}{\mu_n h^{d+1}} = O\left(\sqrt{\frac{\log n}{nh^d}}\right) \quad a.co.$$

donc

$$(3.18) \quad T_1 = \sup_{x \in S} |\tilde{g}_N^*(x) - \tilde{g}_N^*(x_{k(x)})| = O\left(\sqrt{\frac{\log n}{n^{1-\gamma}h^d}}\right)$$

et

$$(3.19) \quad T_3 = \sup_{x \in S} |\mathbb{E}[\tilde{g}_l^*(x_{k(x)})] - \mathbb{E}[\tilde{g}_l^*(x)]| = O\left(\sqrt{\frac{\log n}{n^{1-\gamma}h^d}}\right).$$

En ce qui concerne (3.8) on a

$$\begin{aligned} \sup_{x \in S} |\mathbb{E}[\tilde{g}_l(x)] - \mathbb{E}[\tilde{g}_l^*(x)]| &= \frac{1}{nh^d} \left| \mathbb{E} \left[\sum_{i=1}^n Y_i^{-l} \mathbb{1}_{\{|Y_i^{-1}| > \mu_n\}} K_i \right] \right| \\ &\leq h^{-d} \mathbb{E} \left[|Y_1^{-l}| \mathbb{1}_{\{|Y_1^{-1}| > \mu_n\}} K_1 \right] \\ &\leq h^{-d} \mathbb{E} \left[\exp(|Y_1^{-l}|/4) \mathbb{1}_{\{|Y_1^{-1}| > \mu_n\}} K_1 \right]. \end{aligned}$$

Utilisons l'inégalité de Holder,

$$\begin{aligned} \sup_{x \in S} |\mathbb{E}[\tilde{g}_l(x)] - \mathbb{E}[\tilde{g}_l^*(x)]| &\leq h^{-d} \left(\mathbb{E} \left[\exp(|Y_1^{-l}|/2) \mathbb{1}_{\{|Y_1^{-1}| > \mu_n\}} \right] \right)^{\frac{1}{2}} (\mathbb{E}(K_1^2))^{\frac{1}{2}} \\ &\leq h^{-d} \exp(-\mu_n^l/4) (\mathbb{E}[\exp(|Y^{-l}|)])^{\frac{1}{2}} (\mathbb{E}(K_1^2))^{\frac{1}{2}} \\ &\leq Ch^{\frac{-d}{2}} \exp(-\mu_n^l/4). \end{aligned}$$

Comme $\mu_n = n^{\gamma/18}$ alors on a

$$\sup_{x \in S} |\mathbb{E}[\tilde{g}_l(x_k)] - \mathbb{E}[\tilde{g}_l^*(x_k)]| = o\left(\left(\frac{\log n}{n^{1-\gamma}h^d}\right)^{1/2}\right).$$

L'assertion (3.9) est une conséquence de l'inégalité de Markov. En effet, pour tout $\epsilon > 0$

$$\mathbb{P} \left[\sup_{x \in S} |\tilde{g}_l(x) - \tilde{g}_l^*(x)| > \epsilon \right] = \mathbb{P} \left(\frac{1}{nh^d} \sum_{i=1}^n Y_i^{-l} \mathbb{1}_{|Y_i^{-1}| > \mu_n} K_i > \epsilon \right)$$

$$\begin{aligned}
&\leq n\mathbb{P}(|Y^{-l}| > \mu_n) \\
&\leq n \exp(-\mu_n^l) \mathbb{E}(\exp(|Y^{-1}|)) \\
&\leq Cn \exp(-\mu_n^l).
\end{aligned}$$

Alors,

$$\sum_{n \geq 1} \mathbb{P} \left(\sup_{x \in S} |\tilde{g}_l^*(x) - \tilde{g}_l(x)| > \varepsilon \sqrt{\frac{\log n}{n^{1-\gamma} h^d}} \right) \leq C \sum_{n \geq 1} n \exp(-\mu_n).$$

La définition de μ_n complète la démonstration T_2 . ■

Démonstration du corollaire 3.2.1. Il est clair que $\forall x \in S$, on a

$$\inf_{x \in S} \tilde{g}_2(x) < \frac{\inf_{x \in S} g_2(x)}{2} \Rightarrow \sup_{x \in S} |g_2(x) - \tilde{g}_2(x)| > \frac{\inf_{x \in S} g_2(x)}{2}.$$

Alors,

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

Utilisons le résultat du Lemme 3.2.1 et le Lemme 3.2.2 pour achever la démonstration du corollaire. ■

Démonstration du Lemme 3.3.1.

De même que le Lemme 3.2.1, on peut écrire

$$B_n = B_n - B_n^* + B_n^*$$

où

$$B_n^* = \frac{1}{g_2(x)} \left[\left[\tilde{g}_1^*(x) - E\tilde{g}_1^*(x) \right] g_2(x) + \left[E\tilde{g}_2^*(x) - \tilde{g}_2^*(x) \right] g_1(x) \right].$$

En utilisant les arguments de la preuve du lemme 3.2.1, on obtient pour un point fixé $x \in \mathbb{R}^d$

$$\left(\frac{nh^d}{(g_2(x))^2 \sigma^2(x, \theta_x)} \right)^{1/2} |\tilde{g}_l(x) - \tilde{g}_l^*(x)| = o_p(1).$$

Comme $\mathbb{E}[B_n] = \mathbb{E}[B_n^*] = 0$, alors il suffit de montrer la normalité asymptotique de

$$\left(\frac{nh^d}{(g_2(x))^2 \sigma^2(x, \theta_x)} \right)^{1/2} |B_n^* - \mathbb{E}[B_n^*]|.$$

Pour cela, on pose pour tout $i = 1, \dots, n$, $Z_{ni} = \sqrt{nh^d} \Delta_i$ où

$$\Delta_i = \frac{1}{g_2(x)nh^d} \left(K_i(Y_i^{-1}g_2(x) - Y_i^{-2}g_1(x)) \mathbb{1}_{|Y_i^{-l}| < \mu_n} - \mathbb{E} \left[K_i(Y_i^{-1}g_2(x) - Y_i^{-2}g_1(x)) \mathbb{1}_{|Y_i^{-l}| < \mu_n} \right] \right),$$

Donc,

$$S_n := \sum_{i=1}^n Z_{ni} = \sqrt{nh^d} (B_n^* - \mathbb{E}[B_n^*]).$$

Ainsi, il suffit de vérifier que

$$(3.20) \quad S_n \rightarrow \mathcal{N}(0, \sigma_1(x)).$$

avec $\sigma_1^2(x) = (g_2(x))^2 \sigma^2$. Pour ce faire on utilise les techniques de Doob (1953)(page 228-231). En effet, on considère les deux suites $p = p_n$ et $q = q_n$ telles que

$$p = o(n^{1/2} \mu_n^{-2} h^{d/2}) \quad \text{et} \quad q = O(p^{1-\varsigma}), \quad \text{pour} \quad \varsigma \in (0, 1)$$

et on pose

$$S_n = T_n + T'_n + \zeta_k \text{ avec } T_n = \sum_{j=1}^k \eta_j, \text{ et } T'_n = \sum_{j=1}^k \xi_j$$

où

$$\eta_j := \sum_{i \in I_j} Z_{ni}, \quad \xi_j := \sum_{i \in J_j} Z_{ni}, \quad \zeta_k := \sum_{i=k(p+q)+1}^n Z_{ni}.$$

avec

$$\begin{aligned} I_j &= \{(j-1)(p+q) + 1, \dots, (j-1)(p+q) + p\}, \\ J_j &= \{(j-1)(p+q) + p + 1, \dots, j(p+q)\}. \end{aligned}$$

Il est clair que $k = \left\lfloor \frac{n}{p+q} \right\rfloor$, (où $\lfloor \cdot \rfloor$ désigne la partie entière), on a $\frac{kq}{n} \rightarrow 0$, et $\frac{kp}{n} \rightarrow 1$, $\frac{q}{n} \rightarrow 0$, ce qui implique que $\frac{p}{n} \rightarrow 0$ quand $n \rightarrow \infty$.

Maintenant, pour démontrer notre résultat, il suffit de prouver

$$(3.21) \quad \mathbb{E}(T'_n)^2 + \mathbb{E}(\zeta_k)^2 \rightarrow 0$$

et

$$(3.22) \quad T_n \rightarrow \mathcal{N}(0, \sigma_1^2(x)).$$

Pour (3.21), on utilise la stationnarité des observations

$$(3.23) \quad \mathbb{E}(T'_n)^2 = k \text{Var}(\xi_1) + 2 \sum_{1 \leq i < j \leq k} |\text{Cov}(\xi_i, \xi_j)|$$

et

$$(3.24) \quad k \text{Var}(\xi_1) \leq qk \text{Var}(Z_{n1}) + 2k \sum_{1 \leq i < j \leq q} |\text{Cov}(Z_{ni}, Z_{nj})|.$$

Le premier terme dans la partie à droite de (3.24) est obtenue par les mêmes arguments de (3.12) combinés avec le fait que $\frac{kq}{n} \rightarrow 0$. En effet,

$$(3.25) \quad k \text{Var}(Z_{n1}) = h^d n k q \text{Var}(\Lambda_1) = O\left(\frac{kq}{n}\right) \rightarrow 0, \quad \text{quand } n \rightarrow \infty.$$

Le deuxième terme est

$$k \sum_{1 \leq i < j \leq q} |\text{Cov}(Z_{ni}, Z_{nj})| = knh^d \sum_{1 \leq i < j \leq q} |\text{Cov}(\Lambda_i, \Lambda_j)|.$$

et de même que pour (3.13) on montre que

$$\sum_{1 \leq i < j \leq q} |\text{Cov}(\Lambda_i, \Lambda_j)| = o\left(\frac{q}{n^2 h^d}\right).$$

Donc,

$$(3.26) \quad k \sum_{1 \leq i < j \leq q} |\text{Cov}(Z_{ni}, Z_{nj})| = o\left(\frac{kq}{n}\right) \rightarrow 0, \quad \text{quand } n \rightarrow \infty.$$

La stationnarité des variables permet d'évaluer le deuxième terme de la partie à droite de (3.23) comme suit

$$\begin{aligned}
\sum_{1 \leq i < j \leq k} |Cov(\xi_i, \xi_j)| &= \sum_{l=1}^{k-1} (k-l) |Cov(\xi_1, \xi_{l+1})| \\
&\leq k \sum_{l=1}^{k-1} |Cov(\xi_1, \xi_{l+1})| \\
&\leq k \sum_{l=1}^{k-1} \sum_{(i,j) \in J_1 \times J_{l+1}} Cov(Z_{ni}, Z_{nj}).
\end{aligned}$$

Comme pour tout $(i, j) \in J_1 \times J_j$, on a $|i - j| \geq p + 1 > p$, alors

$$\begin{aligned}
\sum_{1 \leq i < j \leq k} |Cov(\xi_i, \xi_j)| &\leq k \frac{C\mu_n^4}{nh^{d+2}} \sum_{i=1}^p \sum_{j=2p+q+1, |i-j|>p}^{k(p+q)} \lambda_{i,j} \\
&\leq \frac{Ckp\mu_n^4}{nh^{d+2}} \lambda_p \\
&\leq \frac{Ckp\mu_n^4}{nh^{d+2}} e^{-ap} \rightarrow 0.
\end{aligned}$$

Finalement, ce dernier résultat combiné avec (3.24), (3.25), et (3.26) permet d'écrire

$$\mathbb{E}(T'_1)^2 \rightarrow 0 \text{ quand } n \rightarrow \infty.$$

Et comme $(n - k(p + q)) \leq p$, alors

$$\begin{aligned}
\mathbb{E}(\zeta_k)^2 &\leq (n - k(p + q))Var(Z_{n1}) + 2 \sum_{1 \leq i < j \leq n} |Cov(Z_{ni}, Z_{nj})| \\
&\leq pVar(Z_{n1}) + 2 \sum_{1 \leq i < j \leq n} |Cov(Z_{ni}, Z_{nj})| \\
&\leq pn h^d Var(\Lambda_1) + \underbrace{nh^d \sum_{1 \leq i < j \leq n} |Cov(\Lambda_i, \Lambda_j)|}_{o(1)} \\
&\leq \frac{Cp}{n} + o(1).
\end{aligned}$$

Alors,

$$\mathbb{E}(\zeta_k)^2 \rightarrow 0, \quad \text{quand } n \rightarrow \infty.$$

La démonstration de (3.22) est conséquence des assertions suivantes

$$(3.27) \quad \left| \mathbb{E} \left(e^{it \sum_{j=1}^k \eta_j} \right) - \prod_{j=1}^k \mathbb{E} \left(e^{it \eta_j} \right) \right| \rightarrow 0,$$

et

$$(3.28) \quad k \text{Var}(\eta_1) \rightarrow \sigma_1^2(x), \quad k \mathbb{E}(\eta_1^2 \mathbb{I}_{\{\eta_1 > \epsilon \sigma_1(x)\}}) \rightarrow 0.$$

Démonstration de (3.27) :

On a

$$\begin{aligned} & \left| \mathbb{E} \left(e^{it \sum_{j=1}^k \eta_j} \right) - \prod_{j=1}^k \mathbb{E} \left(e^{it \eta_j} \right) \right| \leq \\ (3.29) \quad & \left| \mathbb{E} \left(e^{it \sum_{j=1}^k \eta_j} \right) - \mathbb{E} \left(e^{it \sum_{j=1}^{k-1} \eta_j} \right) \mathbb{E} \left(e^{it \eta_k} \right) \right| + \left| \mathbb{E} \left(e^{it \sum_{j=1}^{k-1} \eta_j} \right) - \prod_{j=1}^{k-1} \mathbb{E} \left(e^{it \eta_j} \right) \right| \\ & = \left| \text{Cov} \left(e^{it \sum_{j=1}^{k-1} \eta_j}, e^{it \eta_k} \right) \right| + \left| \mathbb{E} \left(e^{it \sum_{j=1}^{k-1} \eta_j} \right) - \prod_{j=1}^{k-1} \mathbb{E} \left(e^{it \eta_j} \right) \right| \end{aligned}$$

et successivement, on a

$$\begin{aligned} & \left| \mathbb{E} \left(e^{it \sum_{j=1}^k \eta_j} \right) - \prod_{j=1}^k \mathbb{E} \left(e^{it \eta_j} \right) \right| \leq \left| \text{Cov} \left(e^{it \sum_{j=1}^{k-1} \eta_j}, e^{it \eta_k} \right) \right| \\ (3.30) \quad & + \left| \text{Cov} \left(e^{it \sum_{j=1}^{k-2} \eta_j}, e^{it \eta_{k-1}} \right) \right| + \dots + \left| \text{Cov} \left(e^{it \eta_2}, e^{it \eta_1} \right) \right|. \end{aligned}$$

La quasi-association des observations donne

$$\left| \text{Cov} \left(e^{it \eta_2}, e^{it \eta_1} \right) \right| \leq \frac{Ct^2 \mu_n^4}{nh^{d+2}} \sum_{i \in I_1} \sum_{j \in I_2} \lambda_{i,j}$$

En appliquant cette inégalité à chaque terme de la partie à droite de (3.30)

afin d'obtenir

$$\left| \mathbb{E} \left(e^{it \sum_{j=1}^k \eta_j} \right) - \prod_{j=1}^k \mathbb{E} \left(e^{it \eta_j} \right) \right| \leq \frac{Ct^2 \mu_n^4}{nh^{d+2}}$$

$$\left[\sum_{i \in I_1} \sum_{j \in I_2} \lambda_{i,j} + \sum_{i \in I_1 \cup I_2} \sum_{j \in I_3} \lambda_{i,j} + \dots + \sum_{i \in I_1 \cup \dots \cup I_{k-1}} \sum_{j \in I_k} \lambda_{i,j} \right].$$

Il est clair que pour tout $2 \leq l \leq k-1$, $(i, j) \in I_l \times I_{l+1}$, on a $|i-j| \geq q+1 > q$, alors

$$\sum_{i \in I_1 \cup \dots \cup I_{l-1}} \sum_{j \in I_l} \lambda_{i,j} \leq p \lambda_q.$$

Donc, l'inégalité (3.29) devient

$$\left| \mathbb{E} \left(e^{it \sum_{j=1}^k \eta_j} \right) - \prod_{j=1}^k \mathbb{E} \left(e^{it \eta_j} \right) \right| \leq \frac{Ct^2 \mu_n^4}{nh^{d+2}} k p \lambda_q$$

$$\leq \frac{Ct^2 \mu_n^4}{nh^{d+2}} k p e^{-aq} \rightarrow 0.$$

En ce qui concerne (3.28), on utilise les mêmes arguments que dans (3.23), pour écrire

$$\begin{aligned} \lim_{n \rightarrow \infty} k \text{Var}(\eta_1) &= \lim_{n \rightarrow \infty} k p \text{Var}(Z_{n1}) \\ &= \lim_{n \rightarrow \infty} k p n h^d \text{Var}(\Lambda_1). \end{aligned}$$

Il est évident qu'on peut remarquer que

$$\begin{aligned} \text{Var}(\Lambda_1) &= \frac{1}{n^2 g_2^2(x) h^{2d}} \left\{ \text{Var} \left(K \left(h^{-1}(x - X_i) \right) \left[Y_i^{-1} g_2(x) + Y_i^{-2} g_1(x) \right] \right) \right\} \\ &\quad - \frac{1}{n^2 g_2^2(x) h^{2d}} \left\{ \mathbb{E} \left[K^2 \left(h^{-1}(x - X_i) \right) \left[Y_i^{-1} g_2(x) + Y_i^{-2} g_1(x) \right] \mathbb{1}_{|Y_i^{-1}| > \mu_n} \right] \right\} \\ &\quad - \frac{1}{n^2 g_2^2(x)} \left(\frac{1}{h^d} \mathbb{E} \left[K \left(h^{-1}(x - X_i) \right) \left[Y_i^{-1} g_2(x) + Y_i^{-2} g_1(x) \right] \mathbb{1}_{|Y_i^{-1}| > \mu_n} \right] \right)^2. \end{aligned}$$

De même que (??) et le Lemme 3.2.1 on obtient

$$(3.31) \quad \text{Var}(\Lambda_1) = \frac{\sigma_1^2(x)}{n^2 h^d} + o\left(\frac{1}{n^2 h^d}\right).$$

Donc,

$$k \text{Var}(\eta_1) = \frac{kp\sigma_1^2(x)}{n} + o\left(\frac{kp}{n}\right) \rightarrow \sigma_1^2(x).$$

Pour la deuxième partie de (3.28), on utilise le fait que $|\eta_1| \leq Cp|Z_{n1}| \leq \frac{C\mu_n^2 p}{\sqrt{nh^d}}$ et l'inégalité de Tchebychev pour arriver à

$$\begin{aligned} k\mathbb{E}(\eta_1^2 \mathbb{I}_{\{\eta_1 > \epsilon\sigma_1(x)\}}) &\leq \frac{C\mu_n^4 p^2 k}{nh^d} \mathbb{E}(\eta_1 > \epsilon\sigma_1(x)) \\ &\leq \frac{C\mu_n^4 p^2 k}{nh^d} \frac{\text{Var}(\eta_1)}{\epsilon^2 \sigma_1^2(x)} \\ &= O\left(\frac{\mu_n^4 p^2}{nh^d}\right) \end{aligned}$$

Ce que achèvera cette démonstration. ■

Démonstration du Lemme 3.3.2. Pour la première limite, on a d'après le Lemme (3.2.1)

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

et par les mêmes arguments que ceux utilisés dans (3.11), on montre que

$$\text{Var}[\tilde{g}_2(x)] \rightarrow 0$$

d'où

$$\tilde{g}_2(x) - g_2(x) \rightarrow 0 \quad \text{en probabilité .}$$

Par ailleurs, il est claire que la deuxième limite est une conséquence de cette dernière convergence. Ainsi, il suffit de traiter

$$\text{Var}[(\tilde{g}_2(x) - E\tilde{g}_2(x))] = \text{Var}[\tilde{g}_2(x)] \rightarrow 0$$

et

$$V_n = O(h^p) \quad (\text{voir , Lemme (3.2.1)).}$$

La condition (H2') permet de déduire que

$$\left(\frac{n\phi_x(h)}{g_2(x)^2\sigma^2} \right)^{1/2} A_n (\tilde{g}_2(x) - E\tilde{g}_2(x)) \rightarrow 0 \quad \text{en probabilité.}$$

■

Références

- Bulinski A. & Suquet, C. (2001). Normal approximation for quasi-associated random fields. *Statist. Probab. Lett.* **54**, 215–226.
- Chatfield, C., (2007). The joys of consulting. *Significance* **4**, 33–36.
- Chen, K., Guo, S., Lin, Y., and Ying, Z. (2010). Least absolute relative error estimation. *J. Am. Statist. Assoc.* **105**, 1104–1112.
- Douge, L. (2010) Théorèmes limites pour des variables quasi-associées hilbertiennes. *Ann. I.S.U.P.* **54**, 51–60.
- Esary, J., Proschan, F. & Walkup, D. (1967). Association of random variables with applications. *Ann. Math. Statist.* **38**, 1466–1476.
- Barlow, R.E. & Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing. Probability Models*. Holt, Rinehart & Winston, New York.
- Gramaglia, M., Trullols-Cruces, O., Naboulsi, D., Fiore, M. and Calderon, M. *Vehicular Networks on Two Madrid Highways*, [Technical Report] 2014, <https://hal.inria.fr/hal-00959837v1>.
- Jong-Dev, K., & Proschan, F. (1983). *Negative association of random variables, with applications*. *Ann. Statist.* **11**, 286–295.
- Lin, Zhengyan; Li, Degui. (2007). Asymptotic normality for L_1 -norm kernel estimator of conditional median under association dependence. *J. Multivariate Anal.* **98**, 1214–1230
- Kallabis, R. S & Neumann, M. H. (2006). An exponential inequality under weak dependence. *Bernoulli*, **12**, 333–350.

Khoshgoftaar, T.M., Bhattacharyya, B.B., Richardson, G.D., (1992). Predicting software errors, during development, using nonlinear regression models : a comparative study. *IEEE Trans. Reliab.* **41**, 390-395.

Narula, S.C., Wellington, J.F., (1977). Prediction, linear regression and the minimum sum of relative errors. *Technometrics.* **19**, 185-190.

Jones, M. C., Park, H., Shin, K. Vines, S. K. and Jeong, S.O., (2008) Relative error prediction via kernel regression smoothers. *Journal of Statistical Planning and Inference* **138**, 2887-2898

Shen, V.Y., Yu, T., Thebaut, S.M., (1985). Identifying error-prone software: an empirical study. *IEEE Trans. Software Eng.* **11**, 317-324.

Wilcox, R. (2005). *Introduction to robust estimation and hypothesis testing*. Academic Press.

Yang, Ying and Ye, Fei. (2013) General relative error criterion and M-estimation. *Front. Math. China* **8**, 695-715.

Chapitre 4

Régression relative fonctionnelle

L'objectif de ce chapitre tient essentiellement à introduire le modèle de régression relative pour l'estimation non paramétrique d'une variable réponse réelle, conditionnellement à une variable explicative fonctionnelle.

L'idée est d'étudier la convergence presque complète de l'estimateur construit à partir de la méthode du noyau, dans un cas où la variable explicative est de type fonctionnelle. On s'est basé sur des conditions de dépendance standard. Ce chapitre comporte 3 parties :

- la première section comprend une présentation de notre modèle ainsi que son estimateur.
- Sous des hypothèses appropriées, on présente dans la deuxième section le principal résultat de la convergence presque complète de l'estimateur.
- Dans la troisième section, on donne la démonstration de nos résultats.

4.1 Présentation du modèle et son estimateur

Soit (X, Y) un couple des variables aléatoires défini sur l'espace de probabilité (Ω, \mathcal{A}, P) à valeurs dans $\mathcal{F} \times R$, où \mathcal{F} est un espace fonctionnel semi-métrique de métrique d . La relation entre X et Y est exprimé par le

modèle de régression suivant

$$(4.1) \quad Y = r(X) + \epsilon$$

où ϵ est une variable aléatoire réelle centrée et indépendante de X et r une fonction de \mathcal{F} à valeurs dans \mathbb{R} . Comme au chapitre précédent, la régression relative est le cas où la fonction r est solution du problème d'optimisation suivant :

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

La solution de ce problème peut facilement être explicitée ; en utilisant la dérivée par rapport à r de la fonction $E \left[\left(\frac{Y - r}{Y} \right)^2 \middle| X = x \right]$ on trouve comme solution :

$$r(x) = \frac{E[Y^{-1}|X = x]}{E[Y^{-2}|X = x]}.$$

Par ailleurs, on suppose que les fonctions $E[Y^{-1}|X = x]$ et $E[Y^{-2}|X = x]$ sont des fonctions de classe $C^2(\mathbb{R})$.

En utilisant la méthode du noyau, on estime la fonction r par

$$(4.2) \quad \hat{r}(x) = \frac{\sum_{i=1}^n Y_i^{-1} K \left(\frac{d(x, X_i)}{h} \right)}{\sum_{i=1}^n Y_i^{-2} K \left(\frac{d(x, X_i)}{h} \right)}$$

où K est une fonction de \mathbb{R} dans \mathbb{R} et $h = h_n$ un paramètre réel strictement positif.

Pour simplifier la notation on écrit :

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

avec

$$\hat{g}_1(x) = \frac{1}{n\mathbf{E} \left[K \left(\frac{d(x, X_i)}{h_n} \right) \right]} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{h_n} \right) Y_i^{-1}$$

et

$$\widehat{g}_2(x) = \frac{1}{n\mathbf{E} \left[K \left(\frac{d(x, X_i)}{h_n} \right) \right]} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{h_n} \right) Y_i^{-2}.$$

4.2 Convergence de l'estimateur

Pour étudier la convergence presque complète de l'estimateur on introduit les hypothèses suivantes :

(H1) $P(X \in B(x, h)) = \phi_x(h) > 0$ pour tout $h > 0$; où $B(x, h)$ est la boule fermée, centrée en x et de rayon h et $\lim_{h \rightarrow 0} \phi_x(h) = 0$.

(H2) Pour $k > 0$ et $C > 0$, il existe un voisinage N_x de x tel que

$$\forall x_1, x_2 \in N_x, \quad |g(x_1) - g(x_2)| \leq C d^k(x_1, x_2),$$

où g désigne $E[Y^{-1}|X = x]$ ou bien $E[Y^{-2}|X = x]$.

(H3) Le noyau K est une fonction à support compact dans $[0, 1]$ vérifiant

$$0 < C_3 < K(t) < C_4 < \infty.$$

(H4) Le paramètre de lissage h_n est tel que : $\lim_{n \rightarrow \infty} h_n = 0$ et $\lim_{n \rightarrow \infty} \log n / n \phi_x(h_n) = 0$.

(H5) La variable réponse Y est telle que : $|Y| > C > 0$.

(H6) $(X_i, Y_i)_{1 \leq i \leq n}$ est une suite algébriquement α -mélangeante, dont le coefficient de mélange vérifie

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

(H7) $0 < \sup_{i \neq j} \{ \mathbb{P}((X_i, X_j) \in B(x, h) \times B(x, h)) \} = O \left(\frac{(\phi_x(h))^{(a+1)/a}}{n^{1/a}} \right)$.

(H8) Il existe $\eta > 0$, tel que, $C n^{\frac{3-a}{a+1} + \eta} \leq \phi_x(h) \leq C' n^{\frac{1}{1-a}}$, avec $a > \frac{5 + \sqrt{17}}{2}$.

On démontre le résultat suivant

Théorème 4.2.1 *Sous les conditions (H1)-(H8), on a*

$$(4.3) \quad \widehat{r}(x) - r(x) = O(h^k) + O \left(\sqrt{\frac{\log n}{n \phi_x(h)}} \right), \text{ p.co.}$$

4.3 Démonstration du théorème

De même que le cas précédent, nous avons

$$(4.4) \quad \widehat{r}(x) = \frac{\widehat{g}_1(x)}{\widehat{g}_2(x)}$$

et

$$r(x) = \frac{E[Y^{-1}|X = x]}{E[Y^{-2}|X = x]} := \frac{g_1(x)}{g_2(x)}.$$

On utilise la décomposition suivante :

$$(4.5) \quad \widehat{r}(x) - r(x) = \frac{\widehat{g}_1(x) - g_1(x)}{\widehat{g}_2(x)} + (g_2(x) - \widehat{g}_2(x)) \frac{r(x)}{\widehat{g}_2(x)}.$$

D'où, le théorème est une conséquence directe des lemmes suivants :

Lemme 4.3.1 *Sous les conditions du théorème précédent on a*

$$(4.6) \quad \widehat{g}_1(x) - g_1(x) = O\left(h^k + \sqrt{\frac{\log n}{n\varphi_x(h)}}\right)$$

Lemme 4.3.2 *Sous les conditions précédentes, on a*

$$(4.7) \quad \widehat{g}_2(x) - g_2(x) = O\left(h^k + \sqrt{\frac{\log n}{n\varphi_x(h)}}\right)$$

Lemme 4.3.3 *Sous les conditions précédentes, on a*

$$(4.8) \quad \exists \delta > 0, \quad \text{tel que} \quad \sum P[|\widehat{g}_2(x)| < \delta] < \infty.$$

Démonstration du lemme (4.3.1)

On a

$$\widehat{g}_1(x) - g_1(x) = \widehat{g}_1(x) - \mathbf{E}\widehat{g}_1(x) + \mathbf{E}\widehat{g}_1(x) - g_1(x)$$

Donc, il suffit de montrer les deux résultats suivants :

$$(4.9) \quad \mathbf{E}\widehat{g}_1(x) - g_1(x) = O(h^k)$$

$$(4.10) \quad \mathbf{E}\widehat{g}_1(x) - \widehat{g}_1(x) = O\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right)$$

Pour (4.16), on a :

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

Posons

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

On peut écrire

$$\begin{aligned} \mathbf{E}\widehat{g}_1(x) - g_1(x) &= \mathbf{E}[Y_1^{-1}\Delta_1] - g_1(x) \\ &= \mathbf{E}[\mathbf{E}[Y_1^{-1}\Delta_1|X_1]] - g_1(x) \\ &= \mathbf{E}[g_1(X_1)\Delta_1 - g_1(x)]. \end{aligned}$$

En utilisant la condition de Lipschitz pour r , on montre que :

$$\begin{aligned} |\mathbf{E}\widehat{g}_1(x) - g_1(x)| &\leq \mathbf{E}|g_1(X_1) - g_1(x)| \Delta_1 \\ &\leq C_1 \mathbf{E}[d^k(x, X_1)\Delta_1] \end{aligned}$$

comme $\mathbf{E}[\Delta_1] = 1$, on a

$$|\mathbf{E}\widehat{g}_1(x) - g_1(x)| \leq C(h^k),$$

Ainsi,

$$\mathbf{E}\widehat{g}_1(x) - g_1(x) = O(h^k).$$

On traite maintenant (4.17). Pour cela, on utilise l'inégalité de Fuk-Nagaev aux variables

$$\Delta_i = Y_i^{-1} K\left(\frac{d(x, X_i)}{h}\right) - \mathbf{E}Y_i^{-1} K\left(\frac{d(x, X_i)}{h}\right).$$

Cette inégalité est basée sur le calcul de

$$\begin{aligned}
S_n^2 &= \sum_{i=1}^n \sum_{j=1}^n |\text{cov}(\Delta_i, \Delta_j)| \\
&= \sum_{i=1}^n \left[\sum_{i \neq j} \text{cov}(\Delta_i, \Delta_j) + \text{cov}(\Delta_i, \Delta_j) \right] \\
&= \sum_{i=1}^n \left[\sum_{i \neq j} \text{cov}(\Delta_i, \Delta_j) + \text{var} \Delta_i \right] \\
&= S_n^{2*} + n \text{var} \Delta_1.
\end{aligned}$$

Pour S_n^{2*} , on utilise les techniques de Masry (1986) et on divise cette somme comme suit

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

et

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

où $m_n \rightarrow \infty$, quand $n \rightarrow \infty$. Soit $J_{1,n}$ (resp. $J_{2,n}$) la somme sur l'ensemble E_1 (resp. sur E_2). Donc,

$$J_{1,n} = \sum_{E_1} |\text{Cov}(\Delta_i(x), \Delta_j(x))| \leq C \sum_{E_1} |E[K_i(x)K_j(x)] + E^2[K_1(x)]|.$$

Sous (H1), (H6) et (H7) on a

$$J_{1,n} \leq C n m_n \phi_x(h) \left(\left(\frac{\phi_x(h)}{n} \right)^{1/a} + \phi_x(h) \right).$$

Sur l'ensemble E_2 , on utilise l'inégalité de Davydov-Rio's (Rio, 2000, p. 87). D'où, pour tout $i \neq j$, on a

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

donc,

$$J_{2,n} = \sum_{E_2} |\text{Cov}(K_i(x), K_j(x))| \leq n^2 m_n^{-a}.$$

Ainsi, pour $m_n = \left(\frac{\phi_x(h)}{n}\right)^{-1/a}$ on aura :

$$(4.11) \quad S_n^{2*} = J_{1,n} + J_{2,n} = O(n\phi_x(h)).$$

Pour le terme variance, on montre que sous (H1), on a

$$(4.12) \quad \text{Var}(\Delta_1(x)) \leq C(EK_1^2 + (EK_1)^2) \leq C(\phi_x(h) + (\phi_x(h))^2).$$

Finalement, de (4.11) et (4.12) on obtient

$$(4.13) \quad S_n^2(x) = O(n\phi_x(h)).$$

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

$$\begin{aligned} P[\mathbb{E}\hat{g}_1(x) - \hat{g}_1(x) > \varepsilon] &= P\left[\left|\frac{1}{n\varphi_x(h)} \sum_{i=1}^n \Delta_i\right| > \varepsilon\right] \\ &= P\left[\left|\sum_{i=1}^n \Delta_i\right| > \frac{4(\varepsilon n\varphi_x(h))}{4}\right] \\ &\leq 4 \left(1 + \frac{\varepsilon^2 n^2 \varphi_x^2(h)}{n\varphi_x(h)16r}\right)^{\frac{-r}{2}} + 2ncr^{-1} \left(\frac{8r}{\varepsilon n\varphi_x(h)}\right)^{a+1}. \end{aligned}$$

Pour $\varepsilon = \varepsilon_0 \sqrt{\frac{\log n}{n\varphi_x(h)}}$

$$\begin{aligned} P\left[\mathbb{E}\hat{g}_1(x) - \hat{g}_1(x) > \varepsilon_0 \sqrt{\frac{\log n}{n\varphi_x(h)}}\right] \\ \leq 4 \left(1 + \frac{\varepsilon_0^2 (\log n) n^2 (\varphi_x)^2(h)}{n^2 (\varphi_x(h))^2 16r}\right)^{\frac{-r}{2}} + 2ncr^{-1} \left(\frac{8r (n\varphi_x(h))^{\frac{1}{2}}}{\varepsilon_0 n\varphi_x(h) (\log n)^{\frac{-1}{2}}}\right) \\ \leq 4 \left(1 + \frac{\varepsilon_0^2 \log n}{16r}\right)^{\frac{-r}{2}} + 2ncr^{-1} \left(\frac{8r^{a+1}}{\varepsilon_0}\right) (n\varphi_x(h) \log n)^{-\left(\frac{a+1}{2}\right)}. \end{aligned}$$

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

$$(4.14) \quad A_1 = \left(1 + \frac{\varepsilon_0^2 \log n}{16r}\right)^{\frac{-r}{2}} \leq c \exp \frac{-\varepsilon_0^2 \log n}{32}.$$

D'autre part,

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

Sous la condition (H8) sur la fonction $\phi_x(h)$, on aboutit à

$$(4.15) \quad A_2 \leq cn^{-1-\frac{a}{2}[-2b+p\theta]}$$

Par suite, par les résultats (4.14) et (4.15) et sous un choix convenable de ε_0 , on aura

$$\sum_{n=1} P \left[\mathbf{E} \widehat{g}_1(x) - \widehat{g}_2(x) > \varepsilon_0 \sqrt{\frac{\log n}{n\varphi_x(h)}} \right] < \infty$$

Démonstration du lemme (4.3.2) De même que le lemme précédent, on a

$$\widehat{g}_2(x) - g_2(x) = \widehat{g}_2(x) - \mathbf{E}\widehat{g}_2(x) + \mathbf{E}\widehat{g}_2(x) - g_2(x).$$

Donc, il suffit de montrer les deux résultats suivants :

$$(4.16) \quad \mathbf{E}\widehat{g}_2(x) - g_2(x) = O(h^k)$$

$$(4.17) \quad \mathbf{E}\widehat{g}_2(x) - \widehat{g}_2(x) = O\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right)$$

Pour (4.16) on a

$$\mathbf{E}\widehat{g}_2(x) - g_2(x) = \mathbf{E}[g_2(X_1)\Delta_1 - g_2(x)].$$

La condition de Lipschitz nous permet d'écrire

$$\begin{aligned} |\mathbf{E}\widehat{g}_2(x) - g_2(x)| &\leq \mathbf{E} |g_2(X_1) - g_2(x)| \Delta_1 \\ &\leq Ch^k. \end{aligned}$$

D'où,

$$\mathbf{E}\widehat{g}_1(x) - g_1(x) = O(h^k).$$

Passons maintenant à la démonstration du résultat (4.17).

Pour cela, on garde les mêmes notations du lemme précédent et en appliquant

l'inégalité de Fuk-Nagaev aux variables $Z_i = Y_i \Delta_i - \mathbf{E} [Y_i^{-2} \Delta_i]$, on montre qu'il existe ε_0 tel que :

$$\sum_{n \in \mathbf{N}^*} \mathbf{P} \left(\frac{1}{n} \left| \sum_{i=1}^n (\mathbf{E}(Y_i^{-2} \Delta_i) - Y_i^{-2} \Delta_i) \right| > \varepsilon_0 \left(\sqrt{\frac{\log n}{n \varphi_x(h)}} \right) \right) < \infty.$$

De même comme K et Y^{-2} sont bornées alors :

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

Donc

$$\text{Var}(\Delta_i^2) \leq C' \varphi_x^{-1}(h).$$

Ainsi, on peut appliquer l'inégalité de de Fuk-Nagaev aux variables pour lesquelles on a pour tout $\varepsilon > 0$ et pour tout $\delta_2 > 1$:

$$\begin{aligned} P \left[\left| \mathbb{E} \hat{g}_2(x) - \hat{g}_2(x) \right| > \varepsilon_0 \sqrt{\frac{\log n}{\varphi_x(h)}} \right] &\leq 2 \exp \left(\frac{-n \varepsilon_0^2 \varphi_x(h) \log n}{4n C' \varphi_x(h)} \right) \\ &\leq 2 \exp \left(\frac{\varepsilon_0^2 \log n}{4C} \right) \\ &\leq 2n^{-\frac{\varepsilon_0^2}{4C}}. \end{aligned}$$

Donc

$$\sum_{i=1}^n P \left[\left| \mathbb{E} \hat{g}_2(x) - \hat{g}_2(x) \right| > \varepsilon_0 \sqrt{\frac{\log n}{n \varphi_x(h)}} \right] \leq \sum_{i=1}^n 2n^{-\frac{\varepsilon_0^2}{4C}}.$$

Pour que la série converge, il suffit de choisir $\frac{\varepsilon_0^2}{4C} > 1$.

Enfin pour achever la démonstration du théorème, il nous suffit de montrer qu'il existe un réel ε strictement positif tel que

$$(4.18) \quad \sum_{n>0} \mathbb{P} (|\hat{g}_2(x)| < \varepsilon) < \infty.$$

Il est clair que :

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

Donc, on peut écrire

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

Ainsi

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

Il suffit donc de prendre $\delta = \frac{g_2(x)}{2}$ pour obtenir (4.3.3).

En conclusion, le résultat du théorème se déduit des résultats intermédiaires précédents. \square

Chapitre 5

Application sur des données réelles

La congestion routière ; ces véhicules plus nombreux à chaque instant qu'à l'instant précédent est un phénomène assez répandu qui se manifeste sur la majorité des routes et autoroutes en raison de divers facteurs mais qui très souvent engendre des impacts cruciaux sur la prospérité socio-économique suscitant, ces dernières années notamment l'intérêt d'un nombre élevé de travaux de simulation et d'études statistiques ayant trait aux faits de la gestion du trafic routier. Pour avoir une idée claire du comportement de notre méthode d'estimation d'un point de vue pratique, nous effectuons une prévision qui fournit le nombre maximum de véhicules transitant par une autoroute 20 secondes à l'avance sur une longueur de voie de 200 m. Afin d'atteindre l'objectif ci-dessus, nous nous sommes servis d'un ensemble de données indiquant les traces de mobilité véhiculaire sur deux autoroutes unidirectionnelles à Madrid et ce à différentes heures de la journée. Les données analysées ont été recueillies sur plusieurs jours de Mai 2010 (Vendredi 7, Lundi 10, Mardi 11, Mercredi 12), sachant que pour chaque jour, deux captures ont été enregistrées : une de 8h30 à 9h00 et l'autre de 11h30 à 12h00. Cependant, nous continuons à œuvrer sur la question de fournir un estimateur rigoureux d'où un modèle à bon sens prédictif en simulant une parcelle de 200m et on exploite les paramètres d'intérêts qui représentent le temps signalé par les boucles d'induction lors du passage d'un véhicule ainsi que la position de

la trace de mobilité du véhicule enregistrée à chaque seconde constatée lors d'une durée de 30 minutes. Dans cette étude applicative nous nous concentrons sur la connectivité des véhicules sur le réseau. Cette dernière mesure le nombre de voitures sur une distance R mètres. Ainsi nous analysons une série chronologique Z_t donnant le nombre de véhicules dans une distance de R mètres à un instant t .

Les données considérées sont données par le schéma suivant :

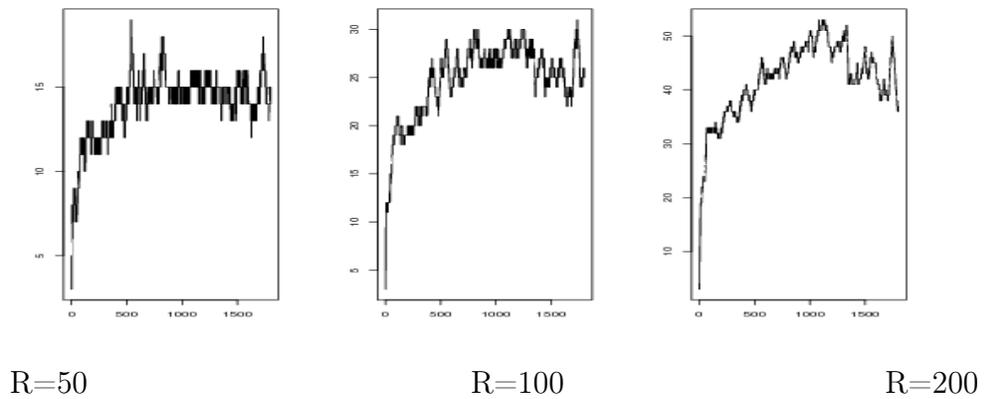


Fig. 1 Case1 : M40 highway from 8 :30 a.m. to 9 :00 a.m.

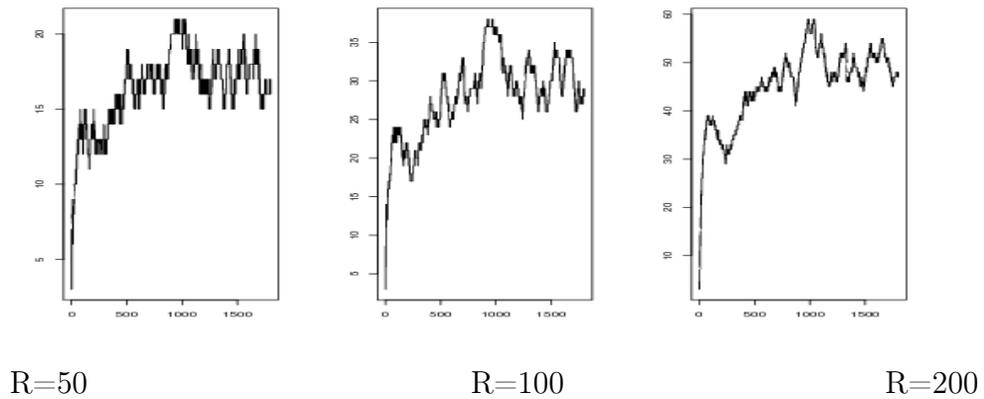


Fig. 2. Case2 : A06 highway from 8 :30 a.m. to 9 :00 a.m.

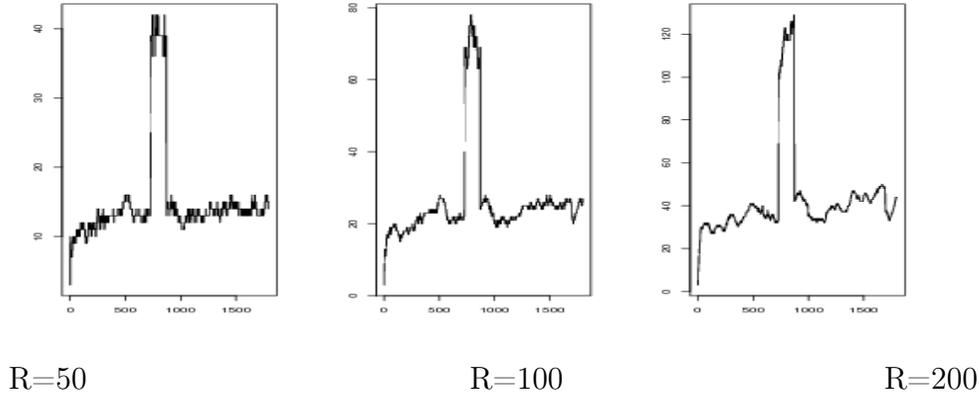


Fig. 3. Case3 : M40 highway from 11 :30 a.m. to 12 :00 a.m.

Nous avons utilisé le détecteur MAD-Median développé par (Wilcox (2005)) pour déterminer les observations aberrantes ; ce modèle considère une observation Y_i comme aberrante si :

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

où M est la medianne, MAD définie par

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

et C est tel que $\sqrt{\chi^2_{.975}}$.

Le tableau suivant donne le résultat du test :

Cases	R=50	R=100	R=200
Case 1	%5	% 10	% 3
Case 2	%10	%5	%8
Case 3	%17	%10	%8

Table 4 : The percentage of outliers

Comme toute autre méthode d'estimation, cette pratique exige deux éléments le noyau K et le paramètre de lissage h . Même si le choix du noyau s'avère

être anodin sur le comportement de l'estimateur, ceci n'est pas tout à fait pareil pour l'ordre de grandeur du paramètre de lissage qui régit le niveau de lissage des courbes estimées, et donc le critère de choix de la fenêtre de lissage optimale reste la question centrale. On convient encore que pour bien tester la validité du modèle en question, on a utilisé la méthode de validation croisée. Sur ce fait, nous comparons notre modèle à la régression classique (C.R.) en calculant l'erreur absolue qui sera donnée par

$$AE = |\text{true value} - \text{predicted value}|$$

<i>Cases</i>	Com. range R	AE(C.R.)	AE(R.E.E.)
<i>Case1</i>	R=50	2.1	3.2
	R=100	8.9	4.5
	R=200	1.4	1.9
<i>Case2</i>	R=50	9.8	4.6
	R=100	2.3	3.7
	R=200	7.4	5.1
<i>Case3</i>	R=50	12.8	5.2
	R=100	10.2	4.1
	R=200	11.7	5.5

Tableau *AE*-erreur

Nous constatons de l'application que l'estimateur fournit un bon rapprochement des données. En effet, les résultats obtenus montrent la supériorité de notre approche sur la régression classique ce qui a permis de mettre en évidence les bonnes propriétés statistiques auxquelles répond la régression relative à la présence des valeurs aberrantes.

Bibliographie

- [1] Dabo-Niang. S. , Rhomari, N. .– *Estimation non paramétrique de la régression avec variable explicative dans un espace métrique et classification.*
- [2] Ferraty, F., Mas, A. and Vieu, P. (2006). Advances on nonparametric regression for functional variables. *Austr. New-Zeal. J. Stat.*, **49**, 1-20.
- [3] Ferraty, F, Laksaci, A. & Vieu, P, (2006)– *Estimating some characteristics of the conditional distribution in nonparametric functional models.* *Statistic. Inf. for Stoch. Proc.* **9**, No. 1, 47-76.
- [4] Ferraty, F. et Vieu, P.– *Statistique fonctionnelle : Modèles Non-Paramétriques de Régression*, Notes de cours de DEA (2002/2003)
- [5] Ferraty, F.. *Modélisation statistique pour variables aléatoires fonctionnelles : Théorie et Application*, Habilitation à diriger des recherches, Université Paul Sabatier Toulouse III, 16 juin 2003.
- [6] Ferraty, F., P. Vieu. *Non parametric functional Data analysis, Theory and practice. 2006, Springer-Verlag.*,
- [7] Ferraty, F., Tadj, A., Laksaci, A. and Vieu, P. (2008). Rate of uniform consistency for nonparametric estimates with functional variables. (Preprint)

- [8] Ferraty, F., Vieu, Ph., *Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés*. C. R. Acad. Sci., Paris, Sér. I, Math. **330**, No.2, (2000), 139-142.
- [9] Györfi L., Härdle W., Sarda P., Vieu P., *Nonparametric Curve Estimation for Time Series*, Springer-Verlag, Berlin, 1989.
- [10] Masry, E. (2005)– *Nonparametric regression estimation for dependent functional data : Asymptotic normality*.
Stoch. Proc. and their Appl., **115**, 155-177.
- [11] Nadaraya, E.A. On estimating regression Teor. Veroyatn. Primen. Teor. Veroyatn. Primen. 9, (1964), 157-159.
- [12] Roussas, George G. Consistent regression estimation with fixed design points under dependence conditions. Stat. Probab. Lett. 8, No.1, (1989), 41-50.
- [13] Sarda, P., Vieu, Ph., (2000), *Kernel regression.*, Schimek, Michael G. (ed.), Smoothing and regression. Approaches, computation and application. Chichester : Wiley. Wiley Series in Probability and Statistics. 43-70.
- [14] Stone, C., J. Consistent nonparametric regression. Discussion. Ann. Stat. 5, (1977), 595-645.
- [15] Tukey, J, W. Curves as parameters, and touch estimation Proc. 4th Berkeley Symp. Math. Stat. Probab. 1, (1961), 681-694.
- [16] Vieu, Ph., (1991), Quadratic errors for nonparametric estimates under dependence. *J. Multivariate Anal.* **39**, No.2, 324-347.
- [17] Watson, G.S. Smooth regression analysis Sankhya, Ser. A 26, (1964), 359-372.