

**Modèles de régression pour
variables aléatoires
uni, multi et ∞ -dimensionnées**

FRÉDÉRIC FERRATY ET PHILIPPE VIEU

Chapitre 1

Présentation générale

1.1 Introduction

Après un premier chapitre destiné à clarifier le vocabulaire, nous commençons par rappeler des éléments de base concernant l'estimation de la régression univariée par technique de noyau. Il s'agit là de choses bien connues, et le lecteur initié trouvera ce Chapitre 2 relativement aisé tandis que le néophyte pourra s'y familiariser avec ces techniques basiques de statistique non-paramétrique. Les chapitres 3 et 4 présenteront des modèles de régression concernant respectivement la statistique des séries temporelles et la statistique non-paramétrique multidimensionnelle. Il s'agit là de choses plus récentes en estimation fonctionnelle mais déjà relativement bien connues. Enfin, les deux derniers Chapitres (5 et 6) aborderont des aspects très récents (et pour certains nouveaux) de la Statistique Fonctionnelle puisqu'ils étudient des modèles de régression pour variables fonctionnelles ; le premier est un modèle de régression non-paramétrique, faisant ainsi le lien avec les chapitres précédents, alors que le second est un modèle de régression paramétrique (linéaire). Quant à la structure de cet ouvrage, elle peut être vue comme simplement une succession de chapitres avec une progression croissante de la difficulté des problèmes abordés. On peut aussi faire une présentation dichotomique de ce document. En effet, parmi les modèles de régression que nous exposons, on peut extraire deux familles selon le type de régresseur que l'on considère. La première famille de modèles de régression correspond à celle dont les régresseurs sont à valeurs dans un espace de dimension fini (variables aléatoires réelles ou vectorielles) alors que la seconde correspond à celle dont les régresseurs sont à valeurs dans un espace de dimension infinie (variables aléatoires fonctionnelles). Ainsi, ce découpage induit naturellement deux parties.

La première partie s'intéresse au cadre classique d'estimation non-paramétrique

d'une fonction de régression à partir de régresseurs réels (Chapitre 2) ou vectoriels (Chapitre 4), tandis qu'une extension au cadre de séries temporelles est présentée au Chapitre 3. L'estimation non-paramétrique, et en particulier l'estimation de régression, constitue un champ de recherche important de la Statistique depuis deux ou trois décennies. Pour s'en convaincre le lecteur pourra consulter les revues bibliographiques de Collomb (1981) et (1985) qui dès le début des années 80 font déjà état de développements nombreux et variés sur ce thème, puis les ouvrages de Härdle (1990) et (1991) et enfin pour terminer l'ouvrage collectif de Schimek (2000) qui dresse un peu le bilan actuel des diverses connaissances en la matière. D'autres ouvrages généraux sur le thème incluent ceux de Bosq et Lecoutre (1987) et Wand et Jones (1995). Ce champ de recherches est potentiellement porteur à la fois de développements théoriques intéressants et de multiples possibilités d'applications. Pour ce qui nous concerne ici nous nous en tiendrons à quelques considérations théoriques. Par ailleurs, vu l'abondance de la littérature, il était impossible dans les Chapitres 2 et 4 de prétendre à une quelconque exhaustivité et c'est la raison pour laquelle nous avons accompagné notre propos de très nombreuses références bibliographiques. Ces références ont été choisies en privilégiant trois types de travaux, les travaux précurseurs, les ouvrages synthétiques et ceux contenant les résultats les plus récents. Nous espérons que ces choix permettront au lecteur de se faire une idée et de remonter à la plus grande partie des travaux existant sur le thème. En fait ces Chapitres 2 et 4 s'adressent essentiellement à un public de néophytes désireux de se familiariser avec les techniques de base en estimation non paramétrique, et nous avons souhaité mettre en évidence dans notre propos deux choses essentielles en ce domaine : le rôle du paramètre de lissage et l'influence de la dimension. Par ailleurs dans le Chapitre 3, nous avons souhaité accorder une place importante à un domaine d'actualité de la Statistique non-paramétrique qui est celui de la prévision de séries chronologiques. Il s'agit d'un domaine sur lequel les premiers résultats conséquents furent établis au début des années 80 (Collomb, 1983 et 1984, et Robinson, 1983), et qui a connu depuis des développements continus (citons par exemple les ouvrages généraux de Györfi *et al.*, 1989, Yoshihara, 1994, Härdle *et al.*, 1997 et celui de Bosq, 1998). Pour ce Chapitre 3, nous avons opté pour une démarche différente de celle des deux précédents en ce sens que nous avons cherché à établir nos résultats sous une formulation des plus récentes qui soit (en évitant toutefois certains raffinements de détails qui pourraient nuire à la clarté du propos). Ainsi ce Chapitre 3 contient essentiellement des résultats asymptotiques qui, bien qu'étant de nature classique (et en particulier de même nature que certains résultats dans Bosq, 1998) sont établis sous des formulations originales. Ce chapitre aussi accorde une large place à la bibliographie qui, bien qu'étant là aussi choisie de manière nécessairement sélective, pourra permettre au lecteur de remonter à la plupart des travaux sur le domaine.

La seconde partie s'intéresse au traitement des variables aléatoires fonctionnelles. Ce domaine de recherche de la Statistique connaît actuellement un franc succès auprès de la communauté internationale des statisticiens si on en veut pour preuve l'ampleur récente des publications scientifiques concernant ce sujet. C'est le cas notamment lorsqu'on s'intéresse à des techniques d'estimation lorsque les données sont qualifiées de courbes (Kneip et Gasser, 1992, Leurgans et *al.*, 1993, Ramsay et Li, 1996, Rice et Silverman, 1991) ou bien de données longitudinales (Besse et *al.*, 1997, Boularan et *al.*, 1995, Fan et Zhang, 1999, Fan et Zhang, 2000, Hoover et *al.*, 1998, Luo, 98, Luo et *al.*, 1998, Nunez-Anton et *al.*, 1999) ou encore de données fonctionnelles (Besse et Cardot, 1996, Besse et *al.*, 1999, Bosq, 1991, Cardot et *al.*, 1999, Cardot et *al.*, 2000, Ferraty et Vieu, 2000, Ramsay et Dalzell, 1991, Ramsay et Silverman, 1997). Il existe deux principales raisons à l'engouement suscité par le traitement des variables fonctionnelles : d'une part, il permet à la fois d'utiliser ou de développer des outils théoriques performants et d'autre part, il offre un énorme potentiel en terme d'applications (imagerie, agro-industrie, reconnaissance de forme, géophysique, économétrie,...). Ainsi, cette thématique couvre toutes les sensibilités émanant de la communauté statisticienne : des plus appliquées aux plus théoriques, sans prédominance de l'une sur l'autre.

Cette partie se décompose en deux chapitres qui s'insèrent pleinement dans ce cadre puisqu'ils s'intéressent aux variables aléatoires fonctionnelles. Seulement, il est illusoire de traiter globalement un sujet aussi vaste. C'est pourquoi nous nous limitons volontairement aux modèles de régression qui, rappelons-le, sont le fil conducteur de ce document. Même dans un contexte de régression, nous ne pouvons prétendre à l'exhaustivité. Nous avons donc décidé de nous restreindre à l'étude de deux modèles de régression pour variable aléatoire fonctionnelle (v.a.f.) : l'un non-paramétrique, étudié au chapitre 5, l'autre paramétrique, faisant l'objet du chapitre 6, ceci en anticipant sur la terminologie définie au paragraphe suivant. Plusieurs motivations sont à l'origine de ce choix. Premièrement, le modèle de régression non-paramétrique nous permet de faire une transition avec le Chapitre 4 puisqu'il généralise des résultats obtenus dans le cadre de la régression non-paramétrique pour variable vectorielle. Deuxièmement, l'étude du modèle de régression paramétrique fait appel à des outils théoriques différents de ceux employés dans le modèle de régression non-paramétrique. Enfin, la juxtaposition de ces deux modèles ouvre de nombreuses perspectives de recherche. L'objectif de cette partie est donc double : familiariser le lecteur avec la modélisation statistique de v.a.f. et donner une boîte à outils suffisamment étendue afin d'appréhender dans les meilleures conditions les développements théoriques que nécessite un tel contexte.

1.2 Quelques modèles de régression

1.2.1 Le cadre général

L'objectif de ce chapitre introductif est d'introduire des modèles généraux et de clarifier certains points de vocabulaires que nous utiliserons abondamment dans cet ouvrage. Nous nous plaçons dans le cadre de l'estimation d'une fonction de régression, que nous noterons r . La variable aléatoire à expliquer sera notée Y , tandis que la variable aléatoire explicative sera notée X . Toutes les variables considérées seront supposées être définies sur un même espace probabilisé (Ω, \mathcal{A}, P) . Les divers modèles de régression que nous abordons peuvent tous être écrits sous la forme :

$$Y = r(X) + \epsilon, \quad (1.1)$$

où ϵ est une variable aléatoire centrée et indépendante de X . Dans tout ce qui suit, nous nous limiterons à une variable aléatoire Y réelle, et nos différents modèles se distingueront selon la nature de la variable X et selon la nature de la relation fonctionnelle r liant Y à X .

1.2.2 Modèles pour variables réelles

Considérons tout d'abord le cas où la variable X est réelle. La littérature statistique dans ce domaine peut se diviser entre modèles paramétriques et modèles non-paramétriques, avec tout l'éventail d'adjectifs intermédiaires que l'on peut rencontrer, semi-paramétrique, semi-non-paramétrique, ... Il nous a semblé utile de préciser dès le début de cet ouvrage quelques points de vocabulaire. Nous adopterons la démarche qui consiste à caractériser un modèle de régression par une hypothèse du type

$$r \in \mathcal{C}, \quad (1.2)$$

où \mathcal{C} est une classe de fonctions, avec ou sans restriction supplémentaire sur la loi de Y , X ou ϵ .

Définition 1.2.1 *Nous dirons que le modèle défini par (1.1) et (1.2) est un modèle paramétrique pour variable réelle lorsque la classe \mathcal{C} est indexable par un nombre fini de paramètres réels. Par opposition nous parlerons de modèle non-paramétrique pour variable réelle lorsque \mathcal{C} est un espace de dimension infinie.*

De manière naturelle, le premier exemple qui vient à l'esprit, est le modèle linéaire qui, que l'on fasse une hypothèse de loi sur ϵ

$$Y = aX + b + \epsilon, a, b \in \mathbb{R}, \epsilon \text{ suivant une loi } \mathcal{N}(0, \sigma^2), \quad (1.3)$$

ou non

$$Y = aX + b + \epsilon, (a, b) \in \mathbb{R}^2, \quad (1.4)$$

est clairement de nature paramétrique au vu de la définition précédente.

Un exemple classique de modèle non-paramétrique est celui qui consiste à poser

$$\mathcal{C} = \{\text{fonctions k-fois continûment dérivables}\}. \quad (1.5)$$

Notons que la littérature fait parfois état d'autres définitions qui ne sont pas toujours équivalentes à celle que nous adoptons ici (Cf. Exercice 1.1). Pour ce qui nous concerne, tout au long de ce document nous en tiendrons à cette définition.

1.2.3 Modèles pour variables vectorielles

Les définitions précédentes s'étendent naturellement au cas vectoriel quand $X \in \mathbb{R}^p$, r étant dans ce cas là une fonction de \mathbb{R}^p dans \mathbb{R} . Ainsi, un modèle de régression se construit à partir d'une hypothèse de type

$$r \in \mathcal{C}. \quad (1.6)$$

Définition 1.2.2 *Nous dirons que le modèle défini par (1.1) et (1.6) est un modèle paramétrique pour variable vectorielle lorsque la classe \mathcal{C} est indexable par un nombre fini de paramètres vectoriels de \mathbb{R}^p . Par opposition nous parlerons de modèle non-paramétrique pour variable vectorielle lorsque \mathcal{C} est un espace de dimension infinie.*

Les modèles linéaires,

$$Y = {}^tAX + b + \epsilon, A \in \mathbb{R}^p, b \in \mathbb{R}, \quad (1.7)$$

avec ou sans hypothèse sur ϵ , entrent clairement dans la catégorie des modèles paramétriques, tandis que les modèles ne faisant que des hypothèses de régularité sur r du type

$$\mathcal{C} = \{\text{fonctions k-fois continûment différentiables}\}, \quad (1.8)$$

sont de nature non-paramétrique. Un autre exemple de modèle non-paramétrique est le modèle additif

$$\mathcal{C} = \{r, r(x^1, \dots, x^p) = \mu + r^1(x^1) + \dots + r^p(x^p), \mu \in \mathbb{R}, r^j \in \mathcal{C}^0\}, \quad (1.9)$$

$$\mathcal{C}^0 = \{\phi, \int \phi(t)dt = 0\}. \quad (1.10)$$

1.2.4 Modèles pour variables fonctionnelles

Avant d'aller plus loin, qu'appelle-t-on précisément variable aléatoire fonctionnelle ? On appelle variable aléatoire fonctionnelle (v.a.f.), notée X , toute application mesurable définie sur un espace probabilisé (Ω, \mathcal{A}, P) et à valeurs dans un espace probabilisable (E, \mathcal{B}_E) où E est un espace de dimension infinie, appelé génériquement espace fonctionnel, \mathcal{B}_E étant la tribu des boréliens de E (tribu engendrée par les ouverts de E). Au delà de l'espace fonctionnel E dans lequel X prend ses valeurs, la terminologie *fonctionnelle* fait essentiellement référence à la dimension infinie de E . En effet, la principale difficulté dans l'étude des modèles pour v.a.f. provient justement de cette dimension infinie. L'idée consiste alors à généraliser les modèles classiques pour variable aléatoire vectorielle en se plaçant dans un cadre purement fonctionnel. Le cas où l'espace E est de dimension finie se ramène naturellement au cadre vectoriel évoqué ci-dessus. Par conséquent, nous nous limiterons de manière implicite à des espaces E de dimension infinie, et nous parlerons d'espace fonctionnel. La fonction r est maintenant une fonction de E dans \mathbb{R} , et nos modèles statistiques s'écrivent comme précédemment

$$r \in \mathcal{C}. \quad (1.11)$$

Définition 1.2.3 *Nous dirons que le modèle défini par (1.1) et (1.11) est un modèle paramétrique pour variable fonctionnelle lorsque la classe \mathcal{C} est indexable par un nombre fini de paramètres de E . Par opposition nous parlerons de modèle non-paramétrique pour variable fonctionnelle lorsque la classe \mathcal{C} n'est pas indexable par un nombre fini de paramètres de E .*

Prenons par exemple le cas où E est un espace de Hilbert et notons $\langle \cdot, \cdot \rangle$ son produit scalaire. Le cadre le plus classique est le modèle linéaire fonctionnel, pour lequel on fait l'hypothèse :

$$\mathcal{C} = \{\text{opérateurs linéaires continus}\}. \quad (1.12)$$

Ce modèle est un modèle paramétrique fonctionnel au sens de la définition précédente, puisque d'après le Théorème de Riesz tout opérateur Φ dans \mathcal{C} admet un représentant unique ϕ dans E défini par

$$\Phi(x) = \langle \phi, x \rangle, \quad \forall x \in E.$$

D'ailleurs, on trouvera parfois dans la littérature le modèle linéaire fonctionnel écrit non pas à partir de (1.1), (1.11) et (1.12) mais plutôt en faisant apparaître de manière explicite ce représentant de r dans E (Voir Exercice 1.2).

1.2.5 Exercices

Exercice 1.1 *On se place dans le cadre où X est une variable réelle. Comme alternative à la Définition 1.2.1, on peut parfois rencontrer dans la littérature la définition suivante : “Un modèle de régression est dit paramétrique (resp. non-paramétrique) si l’ensemble \mathcal{P} des lois de probabilités sur \mathbb{R}^2 auquel celle du couple (X, Y) est supposée appartenir, est indexable par un nombre fini de paramètres réels (resp. est un ensemble de dimension infinie)”. Montrer, en construisant un modèle de régression particulier, que ces deux définitions ne sont pas équivalentes.*

Exercice 1.2 *Soit $E = L^2([0, 1])$ muni du produit scalaire usuel*

$$\langle f, g \rangle = \int_0^1 f(t)g(t)dt,$$

écrire le modèle linéaire fonctionnel sous la forme

$$Y = \int_0^1 r^*(t)X(t)dt + \epsilon, \quad r^* \in E. \quad (1.13)$$

Chapitre 2

Régression non-paramétrique réelle

2.1 Le modèle non-paramétrique

Nous nous plaçons dans ce chapitre dans le cadre de l'estimation de la fonction de régression

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

d'une variable réelle Y sur une variable réelle X , et nous supposons que nous disposons pour cette estimation d'un échantillon $\{(X_i, Y_i), i = 1, \dots, n\}$ de couples indépendants et ayant chacun même loi que (X, Y) . Le modèle qui va nous intéresser est un modèle non-paramétrique, en ce sens que la seule condition que nous ferons sur la fonction r est une condition de régularité

$$r \text{ est } k \text{ fois continûment dérivable,} \tag{2.1}$$

k étant un entier positif ou nul (le cas $k = 0$ correspondant évidemment à l'hypothèse simple de continuité de r). Ce problème a été abondamment étudié dans la littérature, et notre objectif ici n'est pas de présenter une discussion exhaustive des différentes méthodes d'estimation existantes. Nous souhaitons d'avantage nous attarder sur la nature même du modèle défini par (2.1). Bien sûr, comprendre ce modèle statistique nécessite de connaître un estimateur qui, sous ce modèle, possède de bonnes propriétés mathématiques. C'est la raison pour laquelle nous allons nous

limiter à l'étude des estimateurs de type noyau de convolution, puisque la simplicité de leur construction et leur facilité d'utilisation vont de pair avec leur bonnes propriétés asymptotiques. En particulier, ces estimateurs nous permettront de mettre en évidence le problème essentiel lié à ces modèles : la question du choix du paramètre de lissage.

Ce chapitre s'organise de la manière suivante. La méthode du noyau sera rapidement présentée dans le Paragraphe 2.2. Nous donnerons ensuite au Paragraphes 2.3 et 2.4 quelques propriétés asymptotiques de ces estimateurs. Ces quelques propriétés ont été choisies parmi l'abondante littérature existante sur le sujet, de manière à répondre à un objectif double : donner au lecteur les éléments de base en la matière et mettre en valeur le modèle non-paramétrique lui-même, ses caractéristiques, ses avantages et ses limites. Sans sacrifier à la rigueur mathématique nous avons décidé de présenter certains résultats sous une forme légèrement plus restrictive que celle disponible dans la littérature, cela afin d'éviter au maximum que certains aspects techniques non fondamentaux ne viennent nuire à la clarté des points essentiels que nous souhaitons mettre en évidence. Le Paragraphe 2.5 est consacré au problème du choix du paramètre de lissage. La nécessité évidente d'avoir à faire des choix drastiques dans la présentation de ce chapitre, nous a conduit à consacrer un Paragraphe 2.6 relativement conséquent aux compléments bibliographiques. Nous espérons que ces compléments bibliographiques permettront au lecteur de remonter aux développements les plus récents en matière d'estimation par noyau, à ceux relatifs aux autres techniques non-paramétriques de régression et à ceux relatifs aux autres problèmes d'estimation fonctionnelle qui font appel à des techniques similaires à celles présentées ici en régression. Pour terminer, le Paragraphe 2.7 propose quelques exercices qui doivent à la fois permettre de se familiariser avec les techniques de type noyau, et permettre d'établir certaines extensions des résultats présentés dans les paragraphes précédents.

2.2 La méthode du noyau

Les estimateurs de type noyau, introduits indépendamment par Nadaraya (1964) et Watson (1964), sont une des techniques les plus populaires d'estimation sous des modèles de régression de type (2.1). Pour comprendre les idées qui ont amené à l'introduction de ces estimateurs, peut-être faut-il remonter au régressogramme de Tukey (1961) défini de la manière suivante :

$$\hat{r}_{reg}(x) = \frac{\sum_{i=1}^n Y_i \mathbf{1}(X_i \in B_j)}{\sum_{i=1}^n \mathbf{1}(X_i \in B_j)}, \quad \forall x \in B_j, \quad (2.2)$$

où B_j , $j = 1, \dots, J$ est une partition du support de X fixée a priori. Comme son cousin l'histogramme en estimation de densité, cet estimateur primitif présente comme inconvénient d'avoir à choisir à la fois la finesse de la discrétisation (*i.e.* le nombre J de découpages) et la position exacte des bornes des intervalles B_j .

Afin de résoudre ce second problème, un nouvel estimateur peut être construit en remplaçant la discrétisation a priori en intervalles B_j par un seul intervalle mais qui varie de manière continue. Concrètement, cela donne l'estimateur de la fenêtre mobile défini de la manière suivante :

$$\hat{r}_{FM}(x) = \frac{\sum_{i=1}^n Y_i \mathbf{I}(X_i \in [x - h; x + h])}{\sum_{i=1}^n \mathbf{I}(X_i \in [x - h; x + h])}, \quad \forall x, \quad (2.3)$$

où h est un paramètre réel strictement positif.

L'estimateur précédent présente encore le désavantage d'être discontinu par nature. Ainsi sa généralisation naturelle est l'estimateur à noyau, appelé aussi estimateur de Nadaraya-Watson, défini de la manière suivante :

$$\hat{r}_{NW}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)}, \quad \forall x. \quad (2.4)$$

Dans cette définition K est une fonction de \mathbb{R} dans \mathbb{R} (dont nous verrons qu'elle n'est pas nécessairement positive), et h est un paramètre réel strictement positif (dont nous verrons qu'il sera intéressant de le faire dépendre de n). Cet estimateur sera abondamment étudié dans cet ouvrage. Il faut noter que tous les résultats que nous obtiendrons pour cet estimateur \hat{r}_{NW} resteront valables pour l'estimateur de la fenêtre mobile \hat{r}_{FM} puisque ce dernier est un estimateur à noyau particulier correspondant au cas où K est le noyau uniforme

$$K(t) = \mathbf{I}(t \in [-1; +1]).$$

Notons que dans toutes ces définitions nous adoptons implicitement la convention $0/0 = 0$.

2.3 Convergence presque complète

2.3.1 Résultats sous hypothèse de dérivabilité

Nous allons commencer par donner un résultat de convergence presque complète sous le modèle non-paramétrique (2.1). Cette notion de convergence presque complète entraîne à la fois la convergence presque sûre et la convergence en probabilité. Le

lecteur non familier avec cette notion peut se reporter à l'Exercice 2.3. Dans un premier temps nous nous plaçons en un point fixé x . Le modèle (2.1) est renforcé par la condition

$$r \text{ et } f \text{ sont } k \text{ fois continûment dérivables autour de } x, \quad (2.5)$$

et nous supposons en outre que

$$f(x) > 0, \quad (2.6)$$

f désignant la densité de X par rapport à la mesure de Lebesgue sur \mathbb{R} (densité supposée exister). Nous verrons que nous aurons besoin des conditions suivantes sur le paramètre de lissage $h = h(n)$

$$\lim_{n \rightarrow \infty} h = 0 \text{ et } \lim_{n \rightarrow \infty} \frac{nh}{\log n} = \infty. \quad (2.7)$$

Concernant la pondération K , nous supposerons que

$$K \text{ est borné, intégrable et à support compact.} \quad (2.8)$$

Nous verrons que lorsque $k > 0$ il sera intéressant de rajouter l'hypothèse que K est un noyau d'ordre k au sens de Gasser et Müller (1979), c'est à dire qu'il vérifie :

$$\int t^j K(t) dt = 0, \forall j = 1, \dots, k-1 \text{ et } 0 < \left| \int t^k K(t) dt \right| < \infty. \quad (2.9)$$

Notons que cette condition d'ordre est, dès que $k > 2$, incompatible avec une hypothèse de positivité du noyau (Cf. Exercice 2.4). Afin de ne pas nuire à l'exposé des points essentiels que nous souhaitons mettre en valeur ici, nous faisons l'hypothèse suivante

$$|Y| < M < \infty, \quad (2.10)$$

hypothèse qui pourrait être allégée (au moyen par exemple d'une technique de troncature du type de celle introduite par Mack et Silverman (1982) et reprise entre autres dans Györfi *et al.*, 1989, ou Bosq, 1996), mais au prix d'un accroissement sensible de la lourdeur des démonstrations. Dans la forme simplifiée sous laquelle nous le présentons ci-dessous, ce résultat est issu de Sarda et Vieu (2000), mais les idées qui servent de base à cette démonstration reviennent à Collomb (1976) et (1984).

Théorème 2.3.1 **Vitesse de convergence presque complète ponctuelle sous condition de dérivabilité.** *Considérons le modèle (2.5) avec $k > 0$ et supposons que les conditions (2.6)-(2.10) soient réalisées. On a*

$$\hat{r}_{NW}(x) - r(x) = O(h^k) + O\left(\sqrt{\frac{\log n}{nh}}\right), \text{ p.co.} \quad (2.11)$$

Preuve du Théorème 2.3.1 - Dans toute cette preuve C désigne une constante générique. Par ailleurs, vu la définition de \hat{r}_{NW} on peut sans aucune perte de généralité considérer que

$$\int_{\mathbb{R}} K(t)dt = 1.$$

Pour mener à bien cette démonstration on utilise la décomposition suivante (où l'on note $rf = g$) :

$$\hat{r}_{NW}(x) = \frac{\hat{g}(x)}{\hat{f}(x)}, \quad (2.12)$$

avec

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

et

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

et on écrit que

$$\hat{r}_{NW}(x) - r(x) = \frac{\hat{g}(x) - g(x)}{\hat{f}(x)} + \left(f(x) - \hat{f}(x)\right) \frac{r(x)}{\hat{f}(x)}. \quad (2.13)$$

Le résultat énoncé va découler des 5 résultats suivants :

$$E\hat{g}(x) - g(x) = O(h^k), \quad (2.14)$$

$$E\hat{f}(x) - f(x) = O(h^k), \quad (2.15)$$

$$E\hat{g}(x) - \hat{g}(x) = O\left(\sqrt{\frac{\log n}{nh}}\right), \text{ p.co.}, \quad (2.16)$$

$$E\hat{f}(x) - \hat{f}(x) = O\left(\sqrt{\frac{\log n}{nh}}\right), \text{ p.co.}, \quad (2.17)$$

et

$$\exists \delta > 0, \sum_{n=1}^{\infty} P[\hat{f}(x) \leq \delta] < \infty. \quad (2.18)$$

- Preuve de (2.14). Par équidistribution des (X_i, Y_i) , on a que

$$E\hat{g}(x) = h^{-1} EYK\left(\frac{x - X}{h}\right).$$

En conditionnant par rapport à X on arrive à

$$E\hat{g}(x) = h^{-1} \int r(u)K\left(\frac{x - u}{h}\right)f(u)du.$$

Le calcul de cette intégrale se fait en posant $z = (x - u)/h$ pour arriver à

$$E\hat{g}(x) = \int g(x - zh)K(z)dz.$$

Il suffit de développer la fonction g au voisinage de x , ce qui est possible au vu de la condition (2.5). Ceci s'écrit, pour θ_z entre x et $x + zh$:

$$g(x - zh) = g(x) + \sum_{j=1}^{k-1} \frac{(-1)^j (zh)^j}{j!} g^{(j)}(x) + \frac{(-1)^k (zh)^k}{k!} g^{(k)}(\theta_z).$$

On utilise alors la condition (2.9) sur l'ordre du noyau K , et on arrive à :

$$E\hat{g}(x) = g(x) + (-1)^k h^k \frac{\int z^k K(z) g^{(k)}(\theta_z) dz}{k!}.$$

La continuité de $g^{(k)}$ (Cf. (2.5)) et la compacité du support de K , assurent la convergence uniforme en z de $g^{(k)}(\theta_z)$ vers $g^{(k)}(x)$. On arrive finalement à

$$E\hat{g}(x) = g(x) + (-1)^k h^k \int z^k K(z) dz \frac{g^{(k)}(x)}{k!} + o(h^k). \quad (2.19)$$

Ceci achève la preuve de (2.14).

- Preuve de (2.15). La démonstration est similaire à celle qui précède. Nous renvoyons à l'Exercice 2.1 qui établit un résultat plus général que le résultat (2.15) que nous recherchons, à savoir :

$$E\hat{f}(x) - f(x) = (-1)^k h^k \int z^k K(z) dz \frac{f^{(k)}(x)}{k!} + o(h^k). \quad (2.20)$$

- Preuve de (2.16). L'idée essentielle est d'utiliser une inégalité exponentielle de type Bernstein. La version que nous allons utiliser est celle que nous rappelons dans le Lemme 2.3.1 ci-dessous. Dans Hoeffding (1963) le lecteur trouvera plusieurs versions d'inégalités de ce type, sous plusieurs jeux différents de conditions. Pour ce qui nous concerne, nous aurons uniquement besoin de la version simplifiée ci-dessous, version dont la preuve est donnée dans l'article de Hoeffding déjà cité.

Lemme 2.3.1 *Soit $\Delta_1, \dots, \Delta_n$ des v.a.r. centrées, indépendantes et de même loi, telles qu'il existe deux réels positifs d et δ^2 vérifiant :*

$$|\Delta_1| \leq d \text{ et } E\Delta_1^2 \leq \delta^2.$$

Alors, pour tout $\epsilon \in]0, \frac{\delta^2}{d}[$ on a

$$P \left[n^{-1} \left| \sum_{i=1}^n \Delta_i \right| > \epsilon \right] \leq 2e^{-\frac{n\epsilon^2}{4\delta^2}}. \quad (2.21)$$

Pour appliquer ce lemme aux variables

$$\Delta_i = h^{-1} \left(Y_i K\left(\frac{x - X_i}{h}\right) - EY_i K\left(\frac{x - X_i}{h}\right) \right)$$

il faut tout d'abord trouver des majorants pour $|\Delta_i|$ et pour $E\Delta_i^2$. Les conditions (2.8) et (2.10) permettent aisément de construire la première de ces bornes, et on a l'existence d'une constante finie C telle que

$$|\Delta_i| \leq \frac{C}{h}. \quad (2.22)$$

Calculons maintenant le moment d'ordre 2 de ces variables. On a en posant

$$\Gamma_i = \frac{1}{h} Y_i K\left(\frac{x - X_i}{h}\right)$$

et en conditionnant par rapport à X

$$E\Gamma_i^2 = h^{-1} E \left(h^{-1} Y_i^2 K^2\left(\frac{x - X_i}{h}\right) \right) = h^{-2} \int \phi(u) f(u) K^2\left(\frac{x - u}{h}\right) du,$$

où la fonction ϕ est définie par

$$\phi(u) = E(Y^2 | X = u),$$

l'existence de ce moment conditionnel étant assurée par la condition (2.10). On effectue le changement de variable $z = (x - u)/h$ pour arriver à

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

Puisque ϕ est bornée (cf. (2.10)), et puisque f l'est aussi car continue (Cf. (2.5)) sur le support compact de K (Cf. (2.8)), on a l'existence d'une constante finie C telle que :

$$E\Gamma_i^2 \leq \frac{C}{h}, \quad (2.23)$$

et de manière évidente on a l'existence d'une constante C telle que :

$$E\Delta_i^2 \leq \frac{C}{h}. \quad (2.24)$$

Nous sommes donc en mesure d'appliquer le Lemme 2.3.1. Ainsi, il découle de (2.21), (2.22) et (2.24) que pour ϵ suffisamment petit on a

$$P[|E\hat{g}(x) - \hat{g}(x)| > \epsilon] \leq 2e^{-\frac{n\epsilon^2 h}{4C}}. \quad (2.25)$$

Si l'on applique ce résultat à

$$\epsilon = \epsilon_0 \sqrt{\frac{\log n}{nh}},$$

on arrive finalement à l'existence d'une constante C telle que pour tout ϵ_0

$$P \left[|E\hat{g}(x) - \hat{g}(x)| > \epsilon_0 \sqrt{\frac{\log n}{nh}} \right] \leq 2e^{-C\epsilon_0^2 \log n}. \quad (2.26)$$

Pour ϵ_0 bien choisi, le terme de droite dans (2.26) est celui d'une série convergente. Ainsi (2.26) achève la preuve de (2.16).

- Preuve de (2.17). Il suffit de reprendre le calcul précédent mais dans le cas particulier où la variable Y est la variable certaine toujours égale à 1, et le résultat (2.26) s'écrit alors

$$P \left[|E\hat{f}(x) - \hat{f}(x)| > \epsilon_0 \sqrt{\frac{\log n}{nh}} \right] \leq 2e^{-C\epsilon_0^2 \log n}, \quad (2.27)$$

d'où l'on obtient le résultat recherché (2.17).

- Preuve de (2.18). Les résultats (2.15) et (2.17) entraînent en particulier la convergence presque complète de $\hat{f}(x)$ vers $f(x)$. Ainsi pour tout $\epsilon > 0$ on a

$$\sum_{n=1}^{\infty} P \left[\left| \hat{f}(x) - f(x) \right| > \epsilon \right] < \infty.$$

En remarquant que

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

on peut écrire

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

Il suffit maintenant de poser $\delta = \epsilon = f(x)/2$ (ce qui est toujours possible grâce à la condition (2.6)) pour obtenir (2.18). \square

Nous allons maintenant établir une version uniforme du résultat précédent. La démonstration de ces théorèmes s'inspire de Collomb (1976) et (1984). Nous avons bien évidemment besoin d'une version uniforme des hypothèses (2.6) et (2.5), et pour cela nous nous plaçons sur un compact S de \mathbb{R} , tel qu'il existe $\theta > 0$ pour lequel on a

$$r \text{ et } f \text{ sont } k \text{ fois continûment dérivables autour de } S, \quad (2.28)$$

et

$$\inf_{x \in S} f(x) > \theta. \quad (2.29)$$

Pour le reste, les conditions dont nous aurons besoin sont les mêmes que celles nécessitées pour l'obtention des résultats ponctuels ci-dessus, auxquelles s'ajoute la restriction suivante de type Lipschitz sur le noyau

$$\exists \beta > 0, \exists C < \infty, \forall x \in S, \forall y \in S, |K(x) - K(y)| \leq C|x - y|^\beta. \quad (2.30)$$

Théorème 2.3.2 Vitesse de convergence presque complète uniforme sous condition de dérivabilité. *Considérons le modèle (2.28) avec $k > 0$ et supposons que les conditions, (2.7), (2.8), (2.9), (2.10), (2.29) et (2.30) soient réalisées. On a*

$$\sup_{x \in S} |\hat{r}_{NW}(x) - r(x)| = O(h^k) + O\left(\sqrt{\frac{\log n}{nh}}\right), \text{ p.co.} \quad (2.31)$$

Preuve du Théorème 2.3.2 - Comme précédemment on peut supposer que le noyau K est d'intégrale égale à 1, et nous noterons tout au long de la preuve C une constante générique. Le découpage que nous allons faire est similaire à la formule (2.13) utilisée dans le cadre ponctuel, et on écrit que

$$\begin{aligned} \sup_{x \in S} |\hat{r}_{NW}(x) - r(x)| &\leq \frac{\sup_{x \in S} |\hat{g}(x) - g(x)|}{\inf_{x \in S} |\hat{f}(x)|} \\ &+ \sup_{x \in S} |f(x) - \hat{f}(x)| \frac{\sup_{x \in S} |r(x)|}{\inf_{x \in S} |\hat{f}(x)|}. \end{aligned} \quad (2.32)$$

Vu que r est bornée (car continue sur le compact S), le résultat énoncé (2.31) va découler des 5 résultats suivants :

$$\sup_{x \in S} |E\hat{g}(x) - g(x)| = O(h^k), \quad (2.33)$$

$$\sup_{x \in S} |E\hat{f}(x) - f(x)| = O(h^k), \quad (2.34)$$

$$\sup_{x \in S} |E\hat{g}(x) - \hat{g}(x)| = O\left(\sqrt{\frac{\log n}{nh}}\right), \text{ p.co.}, \quad (2.35)$$

$$\sup_{x \in S} |E\hat{f}(x) - \hat{f}(x)| = O\left(\sqrt{\frac{\log n}{nh}}\right), \text{ p.co.}, \quad (2.36)$$

et

$$\exists \delta > 0, \sum_{n=1}^{\infty} P[\inf_{x \in S} |\hat{f}(x)| \leq \delta] < \infty. \quad (2.37)$$

- Preuve de (2.33) et (2.34). Il suffit de remarquer que toutes les approximations en $O(h^k)$ faites au cours de la preuve de (2.14) et (2.15) pendant la démonstration du Théorème 2.3.1 sont uniformes en $x \in S$, cela grâce à la condition de régularité (2.28) et à la compacité du support de K (Cf. (2.8)).

- Preuve de (2.35). Pour démontrer ce résultat nous allons partir du résultat ponctuel (2.25), établi au cours de la preuve du Théorème 2.3.1. Rappelons que ce résultat pouvait s'énoncer de la manière suivante : pour tout $\epsilon > 0$, il existe une constante $C > 0$ indépendante de x telle que

$$P[|E\hat{g}(x) - \hat{g}(x)| > \epsilon] \leq C e^{-\frac{n\epsilon^2 h}{C}}. \quad (2.38)$$

L'idée est de recouvrir le compact S par un nombre fini d'intervalles B_k de longueurs égales. On note τ_n le nombre de ces intervalles, l_n la longueur de chacun d'entre eux et t_k leurs centres. On a :

$$S \subset \cup_{k=1}^{\tau_n} B_k \text{ où } B_k =]t_k - l_n; t_k + l_n].$$

Pour tout x dans S , on notera $B(x)$ (respectivement $t(x)$) le seul parmi ces intervalles (respectivement le centre de cet unique intervalle) contenant x . On utilise maintenant la décomposition suivante :

$$\begin{aligned} P \left[\sup_{x \in S} |E\hat{g}(x) - \hat{g}(x)| > \epsilon \right] &\leq P \left[\max_{k=1, \dots, \tau_n} |E\hat{g}(t_k) - \hat{g}(t_k)| > \frac{\epsilon}{2} \right] + \\ P \left[\sup_{x \in S} |E\hat{g}(x) - \hat{g}(x) - E\hat{g}(t(x)) + \hat{g}(t(x))| > \frac{\epsilon}{2} \right]. \end{aligned} \quad (2.39)$$

Pour ce qui concerne le premier terme du membre de droite de (2.39), on a en appliquant (2.38) :

$$\begin{aligned} P \left[\max_{k=1, \dots, \tau_n} |E\hat{g}(t_k) - \hat{g}(t_k)| > \frac{\epsilon}{2} \right] &\leq \sum_{k=1}^{\tau_n} P \left[|E\hat{g}(t_k) - \hat{g}(t_k)| > \frac{\epsilon}{2} \right] \\ &\leq C \sum_{k=1}^{\tau_n} e^{-\frac{n\epsilon^2 h}{C}} \leq C \frac{e^{-\frac{n\epsilon^2 h}{C}}}{l_n}. \end{aligned} \quad (2.40)$$

Venons en maintenant au second terme du membre de droite de (2.39). En utilisant tout d'abord la condition (2.10) pour borner Y puis la condition de Lipschitz sur K (Cf. (2.30)) on arrive à :

$$\begin{aligned} |\hat{g}(x) - \hat{g}(t(x))| &\leq C \frac{1}{nh} \sum_{i=1}^n \left| K\left(\frac{x - X_i}{h}\right) - K\left(\frac{t(x) - X_i}{h}\right) \right| \\ &\leq C \frac{1}{nh} \sum_{i=1}^n \left| \frac{x - X_i}{h} - \frac{t(x) - X_i}{h} \right|^\beta \\ &\leq C \frac{1}{h^{1+\beta}} |x - t(x)|^\beta \leq C \left(\frac{l_n^\beta}{h^{1+\beta}}\right). \end{aligned} \quad (2.41)$$

Nous allons maintenant appliquer ce qui précède à

$$\epsilon = \epsilon_0 \sqrt{\frac{\log n}{nh}},$$

et nous arrivons ainsi, en utilisant (2.39)-(2.41) à :

$$\begin{aligned} P \left[\sup_{x \in S} |E\hat{g}(x) - \hat{g}(x)| > \epsilon_0 \sqrt{\frac{\log n}{nh}} \right] &\leq C \frac{e^{-\frac{\epsilon_0^2 \log n}{C}}}{l_n} \\ &+ P \left[C \left(\frac{l_n^\beta}{h^{1+\beta}}\right) > \frac{\epsilon_0 \sqrt{\log n}}{2\sqrt{nh}} \right]. \end{aligned} \quad (2.42)$$

On peut toujours choisir, à cause des hypothèses sur h (Cf. (2.7)), une suite l_n telle que

$$l_n = n^{-a}, \quad a > 0 \text{ et } l_n^\beta = o\left(h^{1+\beta} \sqrt{\frac{\log n}{nh}}\right),$$

de sorte que le deuxième terme de la partie droite de (2.42) soit nul pour n suffisamment grand. On arrive alors à

$$P \left[\sup_{x \in S} |E\hat{g}(x) - \hat{g}(x)| > \epsilon_0 \sqrt{\frac{\log n}{nh}} \right] \leq Cn^{-\frac{\epsilon_0^2}{c} + a}. \quad (2.43)$$

On peut toujours trouver alors une valeur de ϵ_0 telle que la partie droite de (2.43) soit le terme général d'une série convergente. Ceci achève la preuve de (2.35).

- Preuve de (2.36). Il suffit de reprendre le calcul précédent mais dans le cas particulier où la variable Y est la variable certaine toujours égale à 1, et le résultat (2.43) s'écrit alors

$$P \left[\sup_{x \in S} |E\hat{f}(x) - \hat{f}(x)| > \epsilon_0 \sqrt{\frac{\log n}{nh}} \right] \leq Cn^{-\frac{\epsilon_0^2}{c} + a}, \quad (2.44)$$

d'où, par un bon choix de ϵ_0 , on obtient le résultat recherché (2.36).

- Preuve de (2.37). Les résultats (2.34) et (2.36) entraînent en particulier la convergence presque complète uniforme sur S de $\hat{f}(x)$ vers $f(x)$. Ainsi pour tout $\epsilon > 0$ on a

$$\sum_{n=1}^{\infty} P \left[\sup_{x \in S} |\hat{f}(x) - f(x)| > \epsilon \right] < \infty.$$

De plus, $\inf_{x \in S} |\hat{f}(x)| \leq \frac{\theta}{2}$ (où θ est défini par (2.29)) implique que $\sup_{x \in S} |\hat{f}(x) - f(x)| > \frac{\theta}{2}$. Ainsi, on en déduit que

$$P \left[\inf_{x \in S} |\hat{f}(x)| \leq \frac{\theta}{2} \right] \leq P \left[\sup_{x \in S} |\hat{f}(x) - f(x)| > \frac{\theta}{2} \right].$$

Il suffit maintenant d'utiliser la convergence presque complète uniforme sur S de $\hat{f}(x)$ vers $f(x)$ et de poser $\delta = \epsilon = \theta/2$ pour obtenir (2.37). \square

2.3.2 Résultats sous hypothèses de continuité

Le prochain résultat est établi sous un modèle de régression plus général que celui du Théorème 2.3.1, puisque l'hypothèse d'existence de dérivées continues pour les fonctions que l'on estime (*i.e.* pour f et r) est remplacée par la condition de continuité. Le gain obtenu ainsi en terme de généralité du modèle statistique est à mettre en balance avec la perte des vitesses de convergence.

Théorème 2.3.3 **Convergence presque complète ponctuelle sous condition de continuité.** *Considérons le modèle (2.5) avec $k = 0$ et supposons que les conditions (2.6), (2.7), (2.8) et (2.10) soient réalisées. On a*

$$\hat{r}_{NW}(x) \rightarrow r(x), \text{ p.co.} \quad (2.45)$$

Preuve du Théorème 2.3.3 - Comme précédemment on peut toujours supposer que K est unitaire en ce sens que

$$\int_{\mathbb{R}} K(t) dt = 1.$$

La démonstration se fait suivant les mêmes lignes que la précédente. Au vu de la décomposition (2.13), le Théorème 2.3.3 sera prouvé dès que seront établis les 5 résultats suivants :

$$E\hat{g}(x) - g(x) \rightarrow 0, \quad (2.46)$$

$$E\hat{f}(x) - f(x) \rightarrow 0, \quad (2.47)$$

$$E\hat{g}(x) - \hat{g}(x) \rightarrow 0, \text{ p.co.}, \quad (2.48)$$

$$E\hat{f}(x) - \hat{f}(x) \rightarrow 0, \text{ p.co.}, \quad (2.49)$$

et

$$\exists \delta > 0, \sum_{n=1}^{\infty} P[\hat{f}(x) \leq \delta] < \infty. \quad (2.50)$$

- Preuve de (2.46). Par équidistribution des (X_i, Y_i) , on a que

$$E\hat{g}(x) = h^{-1} EY K\left(\frac{x - X}{h}\right).$$

En conditionnant par rapport à X on arrive à

$$E\hat{g}(x) = h^{-1} \int r(u)K\left(\frac{x-u}{h}\right)f(u)du.$$

Le calcul de cette intégrale se fait en posant $z = (x - u)/h$ pour arriver à

$$E\hat{g}(x) = \int g(x - zh)K(z)dz. \quad (2.51)$$

La continuité uniforme de g sur le support compact de K amène que

$$g(x - zh) \rightarrow g(x), \quad (2.52)$$

uniformément en z , ce qui achève la preuve de (2.46).

- Preuve de (2.47). La démonstration est similaire à celle qui précède, et est laissée à titre d'exercice (Cf. Exercice 2.1).

- Preuves de (2.48) et (2.49). On remarque qu'aucun des calculs effectués lors de la démonstration de (2.16) et (2.17) au cours de la preuve du Théorème 2.3.1 ne fait intervenir les conditions (2.5) et (2.9). Ainsi (2.16) et (2.17) restent valables sous les hypothèses qui sont les nôtres dans ce théorème, et il suffit d'utiliser la dernière partie de (2.7) pour conclure.

- Preuve de (2.50). La convergence presque complète de $\hat{f}(x)$ vers $f(x)$, qui découle de (2.47) et (2.49), amène que pour tout $\epsilon > 0$ on a

$$\sum_{n=1}^{\infty} P \left[\left| \hat{f}(x) - f(x) \right| > \epsilon \right] < \infty.$$

Il suffit alors de donner les mêmes arguments que ceux avancés pour montrer (2.18).
□

Le prochain résultat est une version uniforme du théorème précédent. Il est établi sous un modèle de régression plus général que celui du Théorème 2.3.2, puisque l'hypothèse d'existence de dérivées continues pour les fonctions que l'on estime (*i.e.* pour f et r) est remplacée par la condition de continuité. Le gain obtenu ainsi en terme de généralité du modèle statistique s'accompagne de la perte des vitesses de convergence.

Théorème 2.3.4 Convergence uniforme presque complète sous condition de continuité. *Considérons le modèle (2.28) avec $k = 0$ et supposons que les conditions (2.7), (2.8), (2.10), (2.29) et (2.30) soient réalisées. On a*

$$\sup_{x \in S} |\hat{r}_{NW}(x) - r(x)| \rightarrow 0, \text{ p.co.} \quad (2.53)$$

Preuve du Théorème 2.3.4 - Comme précédemment on peut toujours supposer que K est unitaire en ce sens que

$$\int_{\mathbb{R}} K(t) dt = 1.$$

La démonstration se fait suivant les mêmes lignes que la précédente, et C dénote toujours une constante générique. Au vu de la décomposition (2.32), le Théorème 2.3.4 sera prouvé dès que seront établis les 5 résultats suivants :

$$\sup_{x \in S} |E\hat{g}(x) - g(x)| \rightarrow 0, \quad (2.54)$$

$$\sup_{x \in S} |E\hat{f}(x) - f(x)| \rightarrow 0, \quad (2.55)$$

$$\sup_{x \in S} |E\hat{g}(x) - \hat{g}(x)| \rightarrow 0, \text{ p.co.}, \quad (2.56)$$

$$\sup_{x \in S} |E\hat{f}(x) - \hat{f}(x)| \rightarrow 0, \text{ p.co.}, \quad (2.57)$$

et

$$\exists \delta > 0, \sum_{n=1}^{\infty} P[\inf_{x \in S} |\hat{f}(x)| \leq \delta] < \infty. \quad (2.58)$$

- Preuve de (2.54). Il suffit de noter que (2.51) est vraie pour tout x dans S , et de remarquer que la compacité de S et l'hypothèse (2.28) assurent que la convergence dans (2.52) est uniforme pour x dans S .

- Preuve de (2.55). Les mêmes arguments que pour prouver (2.54) précédemment peuvent être invoqués.

- Preuves de (2.56) et (2.57). On remarque qu'aucun des calculs effectués lors de la démonstration de (2.35) et (2.36) au cours de la preuve du Théorème 2.3.2 ne fait intervenir les conditions $k > 0$ et (2.9). Ainsi (2.35) et (2.36) restent valables sous les hypothèses qui sont les nôtres dans ce théorème, et il suffit d'utiliser la dernière

partie de (2.7) pour conclure.

- Preuve de (2.58). La convergence presque complète uniforme sur S de $\hat{f}(x)$ vers $f(x)$, qui découle de (2.55) et (2.57), amène que pour tout $\epsilon > 0$ on a

$$\sum_{n=1}^{\infty} P \left[\inf_{x \in S} |\hat{f}(x) - f(x)| > \epsilon \right] < \infty.$$

Il suffit maintenant d'utiliser les mêmes arguments développés pour montrer (2.37) puis de poser $\delta = \epsilon = \theta/2$ pour obtenir (2.58). \square

2.3.3 Résultats sous hypothèse de type Lipschitz

Mentionnons, pour terminer cet exposé des résultats de convergence presque complète, que d'autres modèles non-paramétriques, caractérisables par d'autres conditions de régularité sur r et f , peuvent être étudiés par des techniques similaires. C'est le cas en particulier lorsque les conditions (2.28) ou (2.5) sont remplacées par des conditions de type Lipschitz. Nous donnons uniquement l'énoncé de ces résultats. Les preuves sont très voisines de celles des théorèmes précédents, et par conséquent laissées à titre d'exercice (Cf. Exercice 2.5). Les conditions de Lipschitz sont soit des conditions ponctuelles en x fixé du type

$$\exists \beta > 0, \exists C < \infty, \exists \epsilon > 0, \forall y \in]x - \epsilon, x + \epsilon[, |\phi(x) - \phi(y)| \leq C|x - y|^\beta, \quad (2.59)$$

soit des conditions uniformes sur un compact S du type

$$\exists \beta > 0, \exists C < \infty, \forall x \in S, \forall y \in S, |\phi(x) - \phi(y)| \leq C|x - y|^\beta, \quad (2.60)$$

où ϕ désigne indifféremment f ou r .

Théorème 2.3.5 Vitesse de convergence presque complète ponctuelle sous condition de Lipschitz. *Considérons le modèle (2.59) et supposons que les conditions (2.6), (2.7), (2.8) et (2.10) soient réalisées. On a*

$$\hat{r}_{NW}(x) - r(x) = O(h^\beta) + O\left(\sqrt{\frac{\log n}{nh}}\right), \text{ p.co.} \quad (2.61)$$

Théorème 2.3.6 Vitesse de convergence presque complète uniforme sous condition de Lipschitz. *Considérons le modèle (2.60) et supposons que les conditions (2.7), (2.8), (2.10), (2.29) et (2.30) soient réalisées. On a*

$$\sup_{x \in S} |\hat{r}_{NW}(x) - r(x)| = O(h^\beta) + O\left(\sqrt{\frac{\log n}{nh}}\right). \text{ p.co.} \quad (2.62)$$

2.4 Convergence en moyenne quadratique

2.4.1 Erreur quadratique moyenne ponctuelle

Nous allons maintenant nous intéresser à des résultats asymptotiques en termes de convergence quadratique. La structure de ce paragraphe est similaire à celle du précédent. Les hypothèses dont nous aurons besoin sont sensiblement les mêmes que pour les résultats de convergence presque complète, à quelques exceptions près décrites ci-dessous. Tout d'abord, la condition sur le paramètre de lissage peut être rendue légèrement moins restrictive en ce sens qu'on peut remplacer (2.7) par :

$$\lim_{n \rightarrow \infty} h = 0 \text{ et } \lim_{n \rightarrow \infty} nh = \infty. \quad (2.63)$$

Par ailleurs, afin de pouvoir spécifier les constantes intervenant dans nos développements asymptotiques nous aurons besoin parfois de rajouter la condition suivante sur la loi de (X, Y) :

$$\phi(u) = E(Y^2 | X = u) \text{ est continue au point } x. \quad (2.64)$$

Pour des raisons techniques qui seront explicitées ultérieurement, nous aurons besoin de nous restreindre à des noyaux positifs. Or (*Cf.* Exercice 2.4), cette condition est incompatible avec un noyau d'ordre supérieur à 2. Par conséquent, nous en resterons à un modèle de type (2.5) avec $k = 2$, et les conditions sur le noyau K deviennent alors :

$$K \text{ est borné, intégrable, positif, symétrique et à support compact.} \quad (2.65)$$

Le premier résultat établit la vitesse de convergence en moyenne quadratique pour l'estimateur à noyau de la régression sous hypothèses de dérivabilité, en un point x fixé. La démonstration que nous proposons pour ce résultat s'inspire largement de celle de Collomb (1976).

Théorème 2.4.1 Convergence en moyenne quadratique ponctuelle sous condition de dérivabilité. *Considérons le modèle (2.5) avec $k = 2$ et supposons que les conditions (2.6), (2.10), (2.63), (2.64) et (2.65) soient réalisées. On a*

$$E[\hat{r}_{NW}(x) - r(x)]^2 = B^2(x)h^4 + V(x)\frac{1}{nh} + o(h^4 + \frac{1}{nh}), \quad (2.66)$$

où

$$B(x) = \frac{\int t^2 K(t) dt}{2} \frac{(g^{(2)}(x) - r(x)f^{(2)}(x))}{f(x)}, \quad (2.67)$$

et

$$V(x) = \int K^2(t) dt \frac{(\phi(x) - r^2(x))}{f(x)}. \quad (2.68)$$

Preuve du Théorème 2.4.1 - Il suffit de montrer que :

$$E\hat{r}_{NW}(x) - r(x) = B(x)h^2 + o(h^2), \quad (2.69)$$

et

$$Var\hat{r}_{NW}(x) = V(x)\frac{1}{nh} + o(\frac{1}{nh}). \quad (2.70)$$

- Preuve de (2.69). L'idée principale de cette preuve, dont la difficulté réside en la nécessité de calculer l'espérance de quotients aléatoires, est d'utiliser le développement usuel de $1/z$:

$$\frac{1}{z} = 1 - (z-1) + \dots + (-1)^p(z-1)^p + (-1)^{p+1}\frac{(z-1)^{p+1}}{z}, \quad \forall z \neq 0, \forall p \in \mathbb{N}^*. \quad (2.71)$$

Concrètement, en appliquant ce développement à $z = \frac{\hat{f}(x)}{Ef(x)}$ et $p = 1$, on arrive à l'écriture suivante :

$$\hat{r}_{NW}(x) = \frac{\hat{g}(x)}{Ef(x)} \left[1 - \frac{(\hat{f}(x) - Ef(x))}{Ef(x)} \right] + \frac{(\hat{f}(x) - Ef(x))^2}{(Ef(x))^2} \hat{r}_{NW}(x), \quad (2.72)$$

d'où l'on tire après passage à l'espérance :

$$E\hat{r}_{NW}(x) = \frac{E\hat{g}(x)}{Ef(x)} - \frac{A_1}{(Ef(x))^2} + \frac{A_2}{(Ef(x))^2}, \quad (2.73)$$

où

$$A_1 = E\hat{g}(x) (\hat{f}(x) - Ef(x)),$$

et

$$A_2 = E \left[\left(\hat{f}(x) - E\hat{f}(x) \right)^2 \hat{r}_{NW}(x) \right].$$

Pour calculer le terme A_1 , on utilise l'indépendance puis l'équidistribution des couples (X_i, Y_i) pour arriver à :

$$\begin{aligned} A_1 &= cov \left(\hat{g}(x), \hat{f}(x) \right) = \frac{1}{(nh)^2} \sum_{i=1}^n cov \left(Y_i K \left(\frac{x - X_i}{h} \right), K \left(\frac{x - X_i}{h} \right) \right) \\ &= \frac{1}{nh} E \left[Y \frac{1}{h} K^2 \left(\frac{x - X}{h} \right) \right] - \frac{1}{n} E \left[Y \frac{1}{h} K \left(\frac{x - X}{h} \right) \right] E \left[\frac{1}{h} K \left(\frac{x - X}{h} \right) \right]. \end{aligned}$$

Il suffit maintenant d'écrire chacune de ces espérances et de calculer les 3 intégrales qui leur correspondent en faisant un changement de variable du type $z = (x - u)/h$, exactement comme lors du calcul de (2.24), pour arriver à

$$\begin{aligned} A_1 &= \frac{1}{nh} \left(g(x) \int K^2(t) dt + o(1) \right) - \frac{1}{n} (g(x) + o(1)) (f(x) + o(1)) \\ &= \frac{1}{nh} g(x) \int K^2(t) dt + o\left(\frac{1}{nh}\right). \end{aligned} \quad (2.74)$$

Concernant maintenant le terme A_2 , puisque Y est bornée et puisque K est positif, on peut borner sûrement $\hat{r}_{NW}(x)$ par une constante C , et arriver ainsi à :

$$|A_2| \leq C var(\hat{f}(x)).$$

Et cette dernière variance se calcule exactement comme précédemment (Cf. Exercice 2.6), pour aboutir à

$$A_2 = O\left(\frac{1}{nh}\right). \quad (2.75)$$

En combinant les résultats (2.73)-(2.75) avec les formules (2.19) et (2.20) (dont on s'aperçoit en reprenant leurs preuves qu'elles restent vraies sous les hypothèses qui sont les nôtres maintenant), on arrive à

$$\begin{aligned} E\hat{r}_{NW}(x) &= \frac{g(x) + (h^2/2) \int t^2 K(t) dt g^{(k)}(x)}{f(x) + (h^2/2) \int t^2 K(t) dt f^{(k)}(x)} + o(h^2) + O\left(\frac{1}{nh}\right) \\ &= \frac{r(x) + (h^2/2) \int t^2 K(t) dt (g^{(k)} f - f^{(k)} g)(x) - (h^4/4) (\int t^2 K(t) dt)^2 f^{(k)}(x) g^{(k)}(x)}{f^2(x) - [(h^2/2) \int t^2 K(t) dt f^{(k)}(x)]^2} \\ &+ o(h^2) + O\left(\frac{1}{nh}\right), \end{aligned} \quad (2.76)$$

d'où l'on tire directement le résultat recherché (2.69).

- Preuve de (2.70). La preuve repose sur la même idée, c'est à dire une application du développement (2.71). On trouve, en utilisant (2.71) avec $z = \frac{\hat{f}(x)}{E\hat{f}(x)}$ et $p = 3$, que

$$\begin{aligned} \text{var}(\hat{r}_{NW}(x)) &= \frac{\text{var}(\hat{g}(x))}{(E\hat{f}(x))^2} - 4 \frac{E\hat{g}(x)\text{cov}(\hat{g}(x), \hat{f}(x))}{(E\hat{f}(x))^3} \\ &+ 3\text{var}(\hat{f}(x)) \frac{(E\hat{g}(x))^2}{(E\hat{f}(x))^4} + o\left(\frac{1}{nh}\right). \end{aligned} \quad (2.77)$$

La preuve de ce résultat est omise, car longue fastidieuse et relativement aisée. Elle est donnée par exemple dans Collomb (1976), Vieu (1994) ou Sarda et Vieu (2000). Pour calculer la variance de l'estimateur $\hat{g}(x)$, on utilise l'indépendance et l'équidistribution des couples (X_i, Y_i) pour arriver à

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

Il suffit maintenant d'écrire chacune de ces espérances et de calculer les 3 intégrales qui leur correspondent en faisant un changement de variable du type $z = (x - u)/h$, exactement comme lors du calcul de (2.24), pour arriver à

$$\begin{aligned} \text{var}(\hat{g}(x)) &= \frac{1}{nh} \left(f(x)\phi(x) \int K^2(t)dt + o(1) \right) - \frac{1}{n} (g(x) + o(1))^2 \\ &= \frac{1}{nh} f(x)\phi(x) \int K^2(t)dt + o\left(\frac{1}{nh}\right). \end{aligned} \quad (2.78)$$

La variance de l'estimateur de densité découle directement de ce résultat appliqué à $Y = 1$ (voir aussi Exercice 2.6) et on a

$$\text{var}(\hat{f}(x)) = \frac{1}{nh} f(x) \int K^2(t)dt + o\left(\frac{1}{nh}\right). \quad (2.79)$$

Quant à la covariance, elle a été calculée précédemment lorsque nous avons établi le résultat (2.74) et nous avons

$$\text{cov}(\hat{g}(x), \hat{f}(x)) = \frac{1}{nh} g(x) \int K^2(t)dt + o\left(\frac{1}{nh}\right). \quad (2.80)$$

En combinant maintenant les résultats (2.77)-(2.80), tout en utilisant les résultats (2.19) et (2.20) (dont on s'aperçoit en reprenant leurs preuves qu'ils restent vrais sous les hypothèses qui sont les nôtres maintenant), pour traiter les termes de la forme $(E\hat{g}(x))^p$ où $(E\hat{f}(x))^p$ qui apparaissent dans (2.77), on arrive au résultat recherché. \square

Le prochain résultat est établi sous un modèle de régression plus général que celui du Théorème 2.4.1, puisque l'hypothèse d'existence de dérivées continues pour les fonctions que l'on estime (*i.e.* pour f et r) est remplacée par la simple condition de continuité. Comme toujours en pareil cas, le gain obtenu ainsi en terme de généralité du modèle statistique est à mettre en balance avec la perte des vitesses de convergence.

Théorème 2.4.2 Convergence en moyenne quadratique ponctuelle sous condition de continuité. *Considérons le modèle (2.5) avec $k = 0$ et supposons que les conditions (2.6), (2.10), (2.63) et (2.65) soient réalisées. On a*

$$E [\hat{r}_{NW}(x) - r(x)]^2 \rightarrow 0. \quad (2.81)$$

Preuve du Théorème 2.4.2 - Dans toute cette preuve C désigne une constante générique. On considère, sans aucune perte de généralité, que $\int_{\mathbb{R}} K(t)dt = 1$. Les formules (2.73), (2.74) et (2.75) restent valables sous nos hypothèses, de sorte que nous avons :

$$E\hat{r}_{NW}(x) = \frac{E\hat{g}(x)}{E\hat{f}(x)} + \frac{1}{(E\hat{f}(x))^2}O\left(\frac{1}{nh}\right). \quad (2.82)$$

Il suffit maintenant d'appliquer les résultats (2.46) et (2.47), pour arriver à

$$E\hat{r}_{NW}(x) = \frac{g(x) + o(1)}{f(x) + o(1)} + O\left(\frac{1}{nh}\right), \quad (2.83)$$

ce qui s'écrit aussi

$$E\hat{r}_{NW}(x) \rightarrow r(x). \quad (2.84)$$

Il reste à traiter la variance. Notons que la décomposition (2.77) n'est basée que sur le développement (2.71). Notons aussi que les preuves de (2.79) et (2.80) ne font pas intervenir la dérivabilité de r ou de f , mais seulement leur continuité. Par conséquent les résultats (2.77), (2.79) et (2.80) restent valables sous les hypothèses qui sont les nôtres maintenant. Le seul calcul qu'il est nécessaire de refaire ici est celui de la variance de $\hat{g}(x)$ puisque le résultat (2.78) n'est pas applicable en l'état (en effet nous

ne disposons plus d'hypothèse de continuité sur la fonction ϕ). Reprenons donc ce calcul. On a

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

Pour calculer la première de ces espérances on commence par borner Y , puis on calcule l'intégrale correspondante par un changement de variable du type $z = (x - u)/h$. La deuxième de ces espérances se calcule directement par ce même changement de variable. On arrive ainsi à :

$$\begin{aligned} \text{var}(\hat{g}(x)) &= \frac{1}{nh} O(1) - \frac{1}{n} (g(x) + o(1))^2 \\ &= O\left(\frac{1}{nh}\right). \end{aligned} \tag{2.85}$$

On déduit finalement de (2.77), (2.79), (2.80) et (2.85) que

$$\text{var}(\hat{r}_{NW}(x)) \rightarrow 0, \tag{2.86}$$

ce qui, avec (2.84), achève la preuve de (2.81). \square

2.4.2 Erreur quadratique moyenne intégrée

Nous allons maintenant donner des versions uniformes sur un compact des deux théorèmes précédents. Pour cela, nous allons nous intéresser aux erreurs quadratiques moyennes intégrées, souvent notées dans la littérature *EQMI* (ou bien *MISE*) et définies par :

$$EQMI(\hat{r}_{NW}) = E \left(\int [\hat{r}_{NW}(x) - r(x)]^2 w(x) dx \right). \tag{2.87}$$

La fonction w est une fonction de poids vérifiant :

$$w \text{ est positive, bornée et à support compact } S. \tag{2.88}$$

Cette fonction est fixée a priori, et sera souvent dans la pratique prise égale à une indicatrice sur un intervalle borné de \mathbb{R} , ou bien égale à un produit d'une telle indicatrice par la densité f de X . Nous aurons aussi besoin d'une version uniforme de la condition (2.64), à savoir que

$$\phi(u) = E(Y^2 | X = u) \text{ est continue autour de } S. \tag{2.89}$$

Théorème 2.4.3 Erreur quadratique moyenne intégrée sous condition de dérivabilité. *Considérons le modèle (2.28) avec $k = 2$ et supposons que les conditions (2.29), (2.63), (2.65), (2.88) et (2.89) sont réalisées. On a*

$$EQMI(\hat{r}_{NW}) = B^2 h^4 + \frac{V}{nh} + o\left(h^4 + \frac{1}{nh}\right), \quad (2.90)$$

où

$$B^2 = \int B^2(x)w(x)dx \text{ et } V = \int V(x)w(x)dx, \quad (2.91)$$

$B(x)$ et $V(x)$ étant définis par (2.67) et (2.68).

Preuve du Théorème 2.4.3 - Il suffit de reprendre ligne à ligne les calculs effectués pour prouver le Théorème 2.4.1, en constatant que toutes les approximations faites sont basées sur des propriétés de continuité de f , r , ϕ , f'' ou r'' . Comme toutes ces fonctions sont supposées continues sur S qui est compact, elles sont donc uniformément continues sur S , et par conséquent les termes en $o(h^k)$ et $o(1/(nh))$ intervenant dans les résultats (2.69) et (2.70) sont uniformes en $x \in S$. Il suffit donc d'intégrer les formules (2.69) et (2.70) pour obtenir (2.90). \square

Pour terminer, nous allons donner un résultat analogue sous le modèle plus général $k = 0$, mais cela se fera au détriment de l'obtention des vitesses de convergence.

Théorème 2.4.4 Erreur quadratique moyenne intégrée sous condition de continuité. *Considérons le modèle (2.28) avec $k = 0$ et supposons que les conditions (2.29), (2.63), (2.65) et (2.88) soient réalisées. On a*

$$EQMI(\hat{r}_{NW}) \rightarrow 0. \quad (2.92)$$

Preuve du Théorème 2.4.4 - Il suffit de reprendre ligne à ligne la preuve du Théorème (2.4.2) en s'assurant, en utilisant les mêmes arguments de continuité uniforme sur S que ceux évoqués pour prouver le Théorème 2.4.3, que les convergences dans (2.84) et (2.86) sont uniformes en $x \in S$. \square

2.5 Choix du paramètre de lissage

2.5.1 Optimisation des vitesses de convergence

Par construction, les estimateurs à noyau \hat{r}_{NW} dépendent de deux paramètres : le noyau K et la largeur de fenêtre h . Dans la pratique, on aura besoin de décider

quels choix effectuer pour ces deux paramètres. Au vu des résultats asymptotiques des deux paragraphes précédents, il est clair que le paramètre h , en contrôlant la régularité de la fonction estimée, jouera un rôle essentiel. Nous nous en tiendrons ici à quelques idées générales concernant l'influence de ce paramètre de lissage h et à la présentation rapide de quelques méthodes permettant de le choisir de manière automatique. Le lecteur désireux d'approfondir ces deux terrains (choix de h et de K) pourra se reporter au Paragraphe 2.6 qui permettra de remonter à la littérature essentielle dans le domaine.

Tous les résultats de convergence pour lesquels nous avons été en mesure de spécifier les vitesses de convergence, c'est à dire les Théorèmes 2.3.1, 2.3.2, 2.4.1 et 2.4.3, mettent en évidence le rôle du paramètre de lissage h . En regardant par exemple le résultat et la preuve du Théorème 2.4.3, on constate que le terme de biais est le terme en h^4 , tandis que le terme de variance est en $1/(nh)$. L'un est proportionnel à h tandis que l'autre est inversement proportionnel à h . Notons ici que l'existence d'un biais n'est pas liée à la méthode d'estimation par noyau, mais que c'est une chose inhérente au modèle non-paramétrique lui-même (nous renvoyons à Collomb (1976) ou Sarda et Vieu (2000) pour une propriété d'inexistence d'estimateur non biaisé de la régression dans des contextes non-paramétriques).

Ainsi, une grande valeur de h se traduit par un estimateur fortement biaisé et alors qu'une trop petite valeur de h entraîne un estimateur à forte variabilité. Comment faut-il donc choisir h ?

D'un point de vue théorique il est facile de répondre à cette question. Concentrons nous par exemple sur le Théorème 2.4.3, et plus particulièrement sur son résultat :

$$EQMI(\hat{r}_{NW}) = B^2 h^4 + \frac{V}{nh} + o(h^4 + \frac{1}{nh}). \quad (2.93)$$

Il est aisé de minimiser (en h) les termes dominants de ce développement asymptotique, puisque il s'agit d'une fonction convexe en h . Le minimum est atteint pour la valeur optimale h_{opt} :

$$h_{opt} = \left(\frac{V}{4nB^2} \right)^{\frac{1}{5}}. \quad (2.94)$$

Le même type de raisonnement peut se faire à partir du résultat du Théorème 2.4.3, et on arrive immédiatement aux deux corollaires suivants.

Corollaire 2.5.1 Convergence optimale en moyenne quadratique ponctuelle. *Considérons le modèle (2.5) avec $k = 2$ et supposons que les conditions*

(2.6), (2.10), (2.64) et (2.65) soient réalisées. Supposons en outre que la fenêtre h soit de la forme :

$$h = Cn^{-\frac{1}{5}}, 0 < C < \infty. \quad (2.95)$$

Alors on a

$$E[\hat{r}_{NW}(x) - r(x)]^2 = O(n^{-\frac{4}{5}}). \quad (2.96)$$

Corollaire 2.5.2 Erreur quadratique moyenne intégrée optimale. *Considérons le modèle (2.28) avec $k = 2$ et supposons que les conditions (2.29), (2.65), (2.88), (2.89) et (2.95) soient réalisées. On a*

$$EQMI(\hat{r}_{NW}) = O(n^{-\frac{4}{5}}). \quad (2.97)$$

Le même type d'argument peut être développé à partir des résultats de convergence presque complète donnés dans les Théorèmes 2.3.1 et 2.3.2. Il s'agit maintenant de minimiser en h des expressions du type

$$O(h^k) + O\left(\sqrt{\frac{\log n}{nh}}\right), \quad (2.98)$$

et naturellement la nouvelle hypothèse sur h pour atteindre asymptotiquement le minimum devient

$$h = C \left(\frac{n}{\log n}\right)^{-\frac{1}{2k+1}}, 0 < C < \infty. \quad (2.99)$$

On obtient alors directement les deux résultats suivants comme conséquences directes des Théorèmes 2.3.1 et 2.3.2.

Corollaire 2.5.3 Convergence presque complète ponctuelle optimale. *Considérons le modèle (2.5) avec $k > 0$ et supposons que les conditions (2.6), (2.8), (2.9), (2.10) et (2.99) soient réalisées. On a*

$$\hat{r}_{NW}(x) - r(x) = O\left(\left(\frac{n}{\log n}\right)^{-\frac{k}{2k+1}}\right), p.co. \quad (2.100)$$

Corollaire 2.5.4 Convergence presque complète uniforme optimale. *Considérons le modèle (2.28) avec $k > 0$ et supposons que les conditions (2.8), (2.9), (2.29), (2.30) et (2.99) soient réalisées. On a*

$$\sup_{x \in S} |\hat{r}_{NW}(x) - r(x)| = O\left(\left(\frac{n}{\log n}\right)^{-\frac{k}{2k+1}}\right), p.co. \quad (2.101)$$

Ces résultats sont particulièrement intéressants d'un point de vue théorique, puisque l'on sait d'après des résultats généraux de Stone (1981) et (1982) que la vitesse optimale de convergence pour un modèle (2.28) est :

$$n^{-\frac{k}{2k+1}}, \text{ en norme } L_p, p < \infty, \quad (2.102)$$

et

$$\left(\frac{n}{\log n}\right)^{-\frac{k}{2k+1}}, \text{ en norme } L_\infty. \quad (2.103)$$

On a donc l'évidence de ce que les estimateurs à noyau, tout en restant d'une conception initiale relativement simple, atteignent moyennant un bon choix du paramètre de lissage les vitesses de convergence optimales pour des modèles non-paramétriques généraux. Nous renvoyons aussi à Konakov et Piterbarg (1984), Hall (1989b) ou Golubev et Nussbaum (1990) pour des visions un peu différentes de ces questions d'optimalité de vitesses de convergence.

2.5.2 Sur le choix automatique de fenêtre

Toutefois, ces résultats ne sont pas directement utilisables dans la pratique. Il existe des méthodes automatiques de sélection de ce paramètre, mais les décrire ici en détail nous entraînerait trop loin des objectifs initiaux de cet ouvrage. Nous souhaitons simplement décrire une de ces méthodes de choix de fenêtre. Pour bien insister sur le rôle de h nous adopterons pour la mesure d'erreur une notation qui fait apparaître explicitement ce paramètre :

$$EQMI(\hat{r}_{NW}) = EQMI(h).$$

La méthode que nous allons présenter est une des plus populaires tant sur le plan pratique que du point de vue des résultats asymptotiques dont on dispose à son sujet, et elle s'inspire des idées de validation croisée pour sélection de modèles. Elle consiste à choisir comme largeur de fenêtre

$$h_{CV} = \operatorname{argmin}_{h \in H} CV(h), \quad (2.104)$$

où

$$CV(h) = \sum_{i=1}^n [Y_i - r_{NW}^{-i}(X_i)]^2 w(X_i). \quad (2.105)$$

Dans ces définitions, H est un ensemble de valeurs possibles pour h et r_{NW}^{-i} est l'estimateur à noyau construit à partir de l'échantillon privé de l'observation (X_i, Y_i) :

$$\hat{r}_{NW}^{-i}(x) = \frac{\sum_{j, j \neq i} Y_j K\left(\frac{x-X_j}{h}\right)}{\sum_{j, j \neq i} K\left(\frac{x-X_j}{h}\right)}. \quad (2.106)$$

Nous admettrons le résultat suivant, qui a été démontré par Härdle et Marron (1985), et qui donne l'optimalité asymptotique de cette largeur de fenêtre en termes d'erreurs quadratiques. Notons aussi le papier précurseur de Rice (1984) sur ce thème.

Théorème 2.5.1 Validation Croisée : optimalité asymptotique. *Sous les hypothèses du Théorème 2.4.3, et si H ne contient que des largeurs de fenêtres vérifiant (2.95), alors on la propriété suivante*

$$\frac{\inf_{h \in H} EQMI(h)}{EQMI(h_{CV})} \rightarrow 1, p.s. \quad (2.107)$$

2.6 Compléments bibliographiques

2.6.1 Compléments sur les estimateurs à noyau de la régression

Comme nous l'exposons dans notre chapitre introductif, il était impossible de prétendre à une quelconque exhaustivité dans la présentation des résultats en estimation non-paramétrique d'une fonction de régression réelle. Les résultats que nous avons choisis d'énoncer dans les chapitres précédents avaient pour but essentiel d'insister sur le modèle non-paramétrique lui-même plutôt que sur la technique d'estimation. Nous souhaitons maintenant donner au lecteur quelques références clés afin de lui permettre de se forger une vue plus générale sur ce domaine.

Tout d'abord, pour ce qui concerne les estimateurs de type noyau que nous avons étudiés ici, il existe une littérature abondante à laquelle Gérard Collomb apporta une contribution déterminante comme en témoigne la compilation de ses travaux dans Collomb (1983) (voir aussi Révész, 1977, pour d'autres résultats précurseurs en ce domaine). Nous nous en sommes tenus dans cet ouvrage à des résultats de convergence en norme L^2 ou L^∞ . Nous renvoyons à Ioannides (1992) pour d'autres types de résultats en norme L^2 , et à Wand (1990) pour des résultats et des références récents à propos de résultats en norme L^1 . Nous renvoyons aussi à Mammen (1991a) et (1991b) pour des résultats analogues à ceux présentés ici mais sous des modèles de régularité un peu différents pour la fonction de régression à estimer.

De manière générale, la littérature récente sur les estimateurs à noyau concerne tant l'établissement de résultats théoriques plus fins que ceux que nous avons exposés ici que la diversité d'applications concrètes qu'ils peuvent susciter. Pour les aspects théoriques le lecteur trouvera de bons compléments et un large éventail de

références dans Bosq et Lecoutre (1987) et dans Schimek (2000), les manuscrits de Härdle (1990) et de Bowman et Azzalini (1997) étant plus utiles concernant les aspects appliqués des techniques de noyau, tandis que le livre de Härdle (1991) est consacré aux problèmes d'implémentation de ces techniques d'estimation. Citons finalement les travaux de Collomb (1981) et (1985), Vieu (1994b) et le Chapitre 1 de Sarda (2000) pour leur aspect purement bibliographique.

Pour ce qui concerne les questions de choix de paramètre de lissage qui ont été rapidement abordées ici, le lecteur pourra se reporter aux revues bibliographiques de Marron (1988) (qui bien que concernant les questions d'estimation de densité peut s'avérer intéressante en estimation de régression), de Vieu (1993) ou Herrmann (2000). Notons qu'à ce sujet, les principales alternatives aux techniques de validation croisée présentées au Paragraphe 2.5.2 qui précède sont celles basées sur les techniques de rééchantillonnage (*Cf. eg.* Cao-Abad, 1991) ou celles basées sur les techniques d'injection (*Cf. eg.* Herrmann, 1997). Pour ce qui concerne les techniques de rééchantillonnage, signalons les travaux de Léger et Altman (1990) et Mammen (1992) pour une présentation de ces méthodes dans un contexte dépassant largement celui du simple choix de paramètre de lissage. Citons finalement la technique moins connue basée sur les quantiles de régression et qui a été introduite par Kozek et Schuster (1990). Loader (1999) propose une étude comparative entre plusieurs de ces méthodes de choix de paramètre de lissage. Concernant les travaux en matière de choix de noyau que nous n'avons pas pu aborder dans cet ouvrage (voir cependant l'Exercice 2.8), nous renvoyons à l'article initial de Gasser et Müller (1979) qui permet de bien situer les problèmes, à celui de Berlinet (1993) qui propose une méthode automatique de choix de noyau ainsi qu'à celui de Vieu (1999) pour des résultats et références récents.

Il faut aussi noter que les estimateurs (2.4) peuvent parfois être étudiés et/ou modifiés lorsque l'objectif n'est pas d'estimer r mais plutôt certains aspects de r . Pour chacun de ces problèmes nous nous limitons à renvoyer aux références les plus récentes. Tout d'abord mentionnons la possibilité d'estimer les dérivées de la fonction r au moyen de techniques par noyau comme cela est exposé par exemple dans Gasser et Müller (1979), Georgiev (1984) ou Mack et Müller (1989a) et (1989b) (voir aussi Vieu (1999) pour un éventail de références récentes à ce sujet). Nous renvoyons à Boullaran *et al.* (1995) pour la présentation d'une modification des estimateurs (2.4) au problème de l'estimation des dérivées de r , et pour leur utilisation à l'estimation d'un point particulier de la courbe de r (extremum, point d'inflexion, ...). Nous renvoyons à Gasser et Müller (1979) pour une modification des estimateurs à noyau (à l'aide de pondération K non nécessairement symétriques) qui s'avère intéressante pour estimer les valeurs de $r(x)$ quand x est proche des bornes du support de X .

Cette idée de noyau dissymétrique a récemment été reprise par Wu et Chu (1993) pour estimer d'éventuelles discontinuités dans la courbe de r , permettant ainsi une utilisation des estimateurs de type (2.4) à des modèles plus généraux que ceux qui sont modélisés ici par (2.1), même avec $k = 0$. Notons à ce propos que, bien que portant essentiellement sur les questions d'estimation de densité, les travaux de Liebscher (1990) (liés aux idées de quasi-régularité introduites par Woodroffe, 1970) sont une contribution intéressante à la connaissance du comportement d'estimateurs non paramétriques en situation discontinue. Le travail récent de Couallier (1999) fait le point de la littérature et présente des résultats nouveaux en matière d'estimation non-paramétrique de cassures.

Pour conclure ce paragraphe, mentionnons que les estimateurs (2.4) peuvent parfois être utilisés en présence de régressions simultanées, comme par exemple dans le cadre de données longitudinales (*Cf. eg.* Hart et Wherly (1986) et (1993), Boularan *et al.*, 1994 et Kneip, 1995).

2.6.2 Utilisation des estimateurs à noyau au choix de sous-modèle

Il nous semble utile de mentionner à ce stade que les estimateurs non-paramétriques, et en particulier les estimateurs à noyau que nous avons étudiés ici, peuvent servir d'outil afin de valider certaines formes paramétriques pour la fonction de régression r , ou plus généralement certains sous modèles (paramétriques ou non) du modèle non-paramétrique général.

Les premiers travaux en ce sens eurent essentiellement pour objet de chercher à utiliser les estimateurs non paramétriques pour construire, sous un modèle paramétrique (linéaire le plus souvent) de nouveaux estimateurs paramétriques. Des travaux précurseurs en ce domaine sont ceux de Faraldo Roca et Gonzalez Manteiga (1987), Cristobal *et al.* (1987) et González Manteiga (1988), tandis que les références les plus récentes incluent celles de Kozek (1990) et (1991), Firth *et al.* (1991), González Manteiga et Cao Abad (1993) et Gonzalez Manteiga et Vilar Fernández (1995). Un problème voisin, lui aussi relativement souvent étudié ces dernières années, est celui de la validation d'un modèle paramétrique au moyen d'estimateurs non-paramétriques. Dans ce domaine nous nous limitons à mentionner quelques références clés comme les travaux de Hidalgo (1992), Härdle et Mammen (1993), Lee (1994), Kuchibhata et Hart (1996), Zheng (1996) et Fan et Li (1996), qui devraient permettre au lecteur de se faire une idée de la variété de situations dans lesquelles ce type de problème peut se poser.

2.6.3 Autres problèmes fonctionnels traitables par les techniques de noyau

Nous allons ouvrir maintenant une parenthèse assez rapide pour signaler que ces techniques de noyau s'adaptent très facilement à beaucoup de problème d'estimation fonctionnelle autres que l'estimation de régression. Nous avons vu tout au long des chapitres précédents, que l'on pouvait estimer la densité marginale de X par l'estimateur

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad (2.108)$$

dont quelques propriétés asymptotiques seront vues au travers des Exercices 2.1, 2.2, 2.5 et 2.6. Concernant ce problème, la littérature est très abondante, et nous nous contenterons de signaler les monographies de Devroye (1987), Izenman (1991), Scott (1992), Silverman (1986), Wand et Jones (1995) et Louani (2000).

Ces estimateurs de densité peuvent amener naturellement des estimateurs de la fonction de répartition F de la variable X définis par

$$\hat{F}(x) = \int_{-\infty}^x \hat{f}(t) dt. \quad (2.109)$$

Nous renvoyons à Shirata et Chu (1992) pour des propriétés de ces estimateurs, ainsi qu'à Lejeune et Sarda (1992) pour un estimateur voisin de (2.109).

En combinant les deux estimateurs (2.108) et (2.109), on peut construire directement un estimateur de la fonction de hasard $\lambda = f/(1 - F)$ de la variable X :

$$\hat{\lambda}(x) = \frac{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^n \int_x^{+\infty} K\left(\frac{t - X_i}{h}\right) dt}. \quad (2.110)$$

Un des premiers articles sur ces estimateurs fût celui de Rice et Rosenblat (1976). Nous renvoyons Youndjé *et al.* (1996) pour des propriétés récentes de ces estimateurs, et à Hassani *et al.* (1986) pour une large revue bibliographique sur ce thème.

Mentionnons aussi que ces techniques de type noyau sont aussi utilisables pour estimer la densité conditionnelle $f_{Y|X}$ de Y sachant X . A ce titre l'estimateur

$$\hat{f}_{Y|X=x}(y) = \frac{\frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) K\left(\frac{y - Y_i}{h}\right)}{\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)}, \quad (2.111)$$

est étudié par Youndjé (1993).

Pour terminer, mentionnons que les techniques de noyau peuvent aussi être utilisées en estimation de densité spectrale (*Cf.* Rachdi, 1998a), en estimation d'intensité de processus ponctuel (*Cf. eg.* Ramlau-Hansen, 1983, Diggle, 1985, Diggle et Marron, 1988, Leśkow, 1989, Dia, 1990, Brooks et Marron, 1990, Grégoire, 1993 ou Moncaup *et al.*, 1995), en estimation de quantiles conditionnels (Lejeune et Sarda, 1988, Samanta, 1989, Berlinet *et al.*, 1998) ou bien dans des modèles de diffusion (*Cf. eg.* Spokoiny, 2000) pour estimer le coefficient dérive ou celui de diffusion.

2.6.4 Alternatives aux méthodes de noyau en estimation de régression

Comme nous l'avons établi au Paragraphe 2.5, les estimateurs à noyau de type Nadaraya-Watson tels que définis par (2.4) sont optimaux en termes de vitesses de convergence pour des modèles de type (2.1), au sens de Stone (*Cf.* (2.102) et (2.103)). Cela signifie bien sûr que ces estimateurs sont bien adaptés à des modèles de type (2.1), mais cela n'exclue pas le fait que d'autres estimateurs peuvent aussi vérifier de telles propriétés d'optimalité et par là même être des concurrents potentiels pour ces estimateurs de Nadaraya-Watson. Nous allons nous limiter aux trois méthodes les plus communément reconnues (spline, polynômes locaux, ondelettes), et pour chacune d'entre elle à quelques références essentielles. Nous renvoyant à Collomb (1981) et (1985) ou pour une vision plus complète des choses. Concernant les relations liant ces diverses techniques d'estimation non-paramétrique nous renvoyons à l'article de Jennen-Steinmetz et Gasser (1988) et aux références qui y sont mentionnées.

Mentionnons tout d'abord que d'autres versions d'estimateurs à noyau existent dans la littérature, certaines étant adaptées au cas où la variable explicative X est certaine. Nous renvoyons à Chu et Marron (1991b) ou Jones *et al.* (1994) pour une présentation de ces versions d'estimateurs à noyau de régression, en mentionnant qu'une des versions les plus connues est l'estimateur par polynômes locaux initialement étudié dans le long article de Lejeune (1985) et qui n'a été curieusement rendu populaire que beaucoup plus tard par l'article de Fan (1993) et auquel Fan et Gijbels (1996) ont consacré un ouvrage général (voir aussi Fan et Gijbels (2000), Doksum *et al.* (2000) et Wu (2000) pour des références plus récentes). Les techniques de lissage par spline sont aussi des méthodes populaires alternatives aux techniques de noyau en estimation non-paramétrique de régression. Le récent travail de Eubank (2000) propose à la fois une bonne introduction à ces méthodes de lissage et un éventail de la bibliographie la plus récente sur le thème. Comme pour les estimateurs à noyau, les méthodes spline font intervenir un paramètre de lissage de manière déterminante (voir à ce sujet le récent travail général de van der Linde, 2000). Enfin, les méthodes d'ondelettes peuvent aussi être des alternatives aux techniques de noyau (*Cf.* An-

toniadis *et al.* (1994), Antoniadis (1997) et Nason et Silverman (2000) pour des descriptifs généraux de ces techniques). Finalement, mentionnons les estimateurs de type δ -suite qui constituent une classe générale d'estimateurs non-paramétriques (classe qui inclue entre autres les estimateurs à noyau) et dont les études les plus récentes sont celles de Marron et Härdle (1986) et Vidal (1999). Des versions robustes d'estimateurs non-paramétriques, basées sur les techniques de M -estimation, ont été étudiées par Boente et Fraiman (1989a), Hall et Jones (1990) et Fan *et al.* (1994).

2.7 Exercices

Sauf mention explicite du contraire, les définitions et notations utilisées dans les énoncés de ces exercices sont identiques à celles utilisées dans tout le chapitre.

Exercice 2.1 *Cet exercice a pour but d'établir des propriétés de convergence presque complète ponctuelle pour l'estimateur*

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

de la densité f de X au point x .

1. *Dans cette question on se place sous les hypothèses du Théorème 2.3.1.*

a. *En s'inspirant de ce qui a été fait pour prouver (2.14), montrer que l'estimateur de la densité vérifie*

$$E\hat{f}(x) - \int_{\mathbb{R}} K(t) dt f(x) = (-1)^k h^k \int z^k K(z) dz \frac{f^{(k)}(x)}{k!} + o(h^k). \quad (2.112)$$

b. *En s'inspirant de la preuve de (2.26) montrer qu'il existe une constante finie C telle que pour ϵ_0 suffisamment petit on ait :*

$$P \left[\left| E\hat{f}(x) - \hat{f}(x) \right| > \epsilon_0 \sqrt{\frac{\log n}{nh}} \right] \leq 2n^{-C\epsilon_0^2}. \quad (2.113)$$

En déduire une propriété de convergence presque complète vers 0, avec vitesse, de $E\hat{f}(x) - \hat{f}(x)$.

2. *En utilisant la question précédente, construire un ensemble d'hypothèses permettant de montrer que l'estimateur à noyau de la densité $\hat{f}(x)$ converge presque*

sûrement vers f en $\left(\frac{n}{\log n}\right)^{-\frac{k}{2k+1}}$.

3. Dans cette question on se place sous les hypothèses du Théorème 2.3.3, mais en allégeant la condition (2.8) en la remplaçant par

$$K \text{ est borné, absolument intégrable et vérifie } \lim_{|t| \rightarrow \infty} |tK(t)| = 0 \quad (2.114)$$

Pour $\alpha > 0$, on pose

$$I = \int_{|y| \leq \alpha} h^{-1}K(y/h)(f(x-y) - f(x))dy,$$

$$II = f(x) \int_{|y| > \alpha} h^{-1}K(y/h)dy,$$

$$III = \int_{|y| > \alpha} h^{-1}K(y/h)f(x-y)dy.$$

a. Montrer, en s'inspirant des calculs effectués lors de la preuve du Théorème 2.3.3, que

$$\forall \alpha > 0, \forall \epsilon > 0, \exists n_0, \forall n > n_0, |I| \leq \epsilon.$$

b. Montrer, en utilisant la propriété d'intégrabilité de K , que

$$\forall \alpha > 0, \forall \epsilon > 0, \exists n_1, \forall n > n_1, |II| \leq \epsilon.$$

c. Montrer, en utilisant la propriété (2.114), que

$$\forall \alpha > 0, \forall \epsilon > 0, \exists n_2, \forall n > n_2, |III| \leq \epsilon.$$

d. En déduire que $\hat{f}(x)$ est un estimateur asymptotiquement non biaisé de $f(x)$.

e. Montrer que sous les hypothèses de cette question, le résultat du Théorème 2.3.3 reste vrai.

Exercice 2.2 Cet exercice a pour but d'établir des propriétés de convergence presque complète uniforme pour l'estimateur \hat{f} de la densité déjà étudié dans l'Exercice 2.1.

a. En reprenant point par point la démarche de la question 1 de l'Exercice 2.1, et en s'inspirant des calculs effectués lors de la preuve du Théorème 2.3.2, établir sous les conditions du Théorème 2.3.2 une propriété de convergence uniforme sur un compact (en précisant les vitesses de convergence) pour l'estimateur à noyau \hat{f} de la

densité f .

b. En déduire un ensemble d'hypothèses permettant de montrer que les vitesses de convergence évoquées à la question a) peuvent atteindre $\left(\frac{n}{\log n}\right)^{-\frac{k}{2k+1}}$.

c. Montrer que le résultat du Théorème 2.3.4 reste vrai en remplaçant (2.8) par (2.114). (Indication : s'inspirer de la démarche suivie au cours de la question 3 de l'Exercice 2.1).

Exercice 2.3 a. Soit $\Delta_1, \dots, \Delta_n, \dots$ une suite de v.a.r. qui converge presque complètement vers une v.a.r. Δ , en ce sens que pour tout $\epsilon > 0$ on a

$$\sum_{i=1}^{\infty} P[|\Delta_i - \Delta| > \epsilon] < \infty.$$

Montrer qu'il y a aussi convergence presque sûre de $\Delta_1, \dots, \Delta_n, \dots$ vers Δ .

b. Soit U_n une suite (certaine) de réels strictement positifs. Montrer que si $\Delta_n = O(U_n)$ presque complètement, en ce sens qu'il existe $\epsilon > 0$ vérifiant

$$\sum_{i=1}^{\infty} P[|\Delta_i| > \epsilon U_n] < \infty,$$

alors on a aussi que $\Delta_n = O(U_n)$ presque sûrement.

Exercice 2.4 Soit K un noyau positif ou nul. Montrer que dès que $k > 2$, l'estimateur \hat{r}_{NW} associé à ce noyau ne peut pas atteindre les vitesses de convergence optimale en $\left(\frac{n}{\log n}\right)^{-\frac{k}{2k+1}}$.

Exercice 2.5 L'objectif de l'exercice est de démontrer les Théorèmes 2.3.5 et 2.3.6.

1. On suppose dans cette question que les hypothèses du Théorème 2.61 sont remplies. On justifiera tout d'abord le fait qu'on peut toujours supposer que le noyau K est d'intégrale 1.

a. Montrer que le biais de l'estimateur de densité se met sous la forme :

$$E\hat{f}(x) - f(x) = O(h^\beta).$$

b. Donner le même type de résultat pour $E\hat{g}(x)$.

c. Justifier (aucun calcul n'est nécessaire dans cette question) que l'on a

$$E\hat{f}(x) - \hat{f}(x) = O\left(\sqrt{\frac{\log n}{nh}}\right), \text{ p.co. et } E\hat{g}(x) - \hat{g}(x) = O\left(\sqrt{\frac{\log n}{nh}}\right), \text{ p.co.}$$

d. Conclure.

2. Démontrer le Théorème 2.3.6 en reprenant la démarche de la question précédente.

Exercice 2.6 a. En se plaçant sous les conditions du Théorème 2.4.1 mais avec $k > 0$ quelconque, montrer que la variance de l'estimateur de densité s'écrit :

$$\text{var}(\hat{f}(x)) = \frac{1}{nh} f(x) \int K^2(t) dt + o\left(\frac{1}{nh}\right). \quad (2.115)$$

b. En déduire une expression asymptotique de l'erreur quadratique de cet estimateur $E[\hat{f}(x) - f(x)]^2$.

Exercice 2.7 Enoncer et démontrer un résultat du type de celui du Théorème 2.4.3, mais pour l'estimateur de la densité \hat{f} et sous le modèle général :

f est k fois continûment dérivable autour d'un compact S , $k > 0$.

Exercice 2.8 1. Montrer qu'il n'existe qu'un seul noyau d'ordre 2 qui soit continu sur \mathbb{R} , qui s'écrive sous la forme

$$K(t) = \mathbf{1}_{[-1,+1]}(t) P_2(t),$$

P_2 étant un polynôme de degré ≤ 2 , et qui soit unitaire en ce sens que

$$\int_{\mathbb{R}} K(t) dt = 1.$$

Ce noyau sera noté K_2 .

2. Montrer qu'il n'existe qu'un seul noyau d'ordre 2 unitaire qui soit continu sur \mathbb{R} et qui s'écrive sous la forme

$$K(t) = \mathbf{I}_{[0,+1]}(t)P_1(t) + \mathbf{I}_{[-1,0]}(t)P_1(-t),$$

P_1 étant un polynôme de degré ≤ 1 . Ce noyau sera noté K_1 .

3. En s'inspirant du Corollaire 2.5.2, et en particulier en précisant la constante intervenant dans le résultat (2.97), expliquer pourquoi sous un modèle de type (2.5), la minimisation de la quantité

$$\left[\int t^2 K(t) dt \right] \left[\int K^2(t) dt \right]^2$$

peut être envisagée comme solution raisonnable au problème du choix du noyau K .

4. Montrer qu'au sens de ce critère K_2 est meilleur que K_1 .

NB. Le noyau K_2 est appelé noyau d'Epanechnikov, et ses propriétés d'optimalité ont été établies dans des cadres plus généraux que celui de cet exercice (Cf. eg. Gasser et Müller, 1979).

Chapitre 3

Prévision de Séries Temporelles

3.1 Introduction

3.1.1 Présentation du problème

Une des applications des méthodes présentées au Chapitre 2 réside en la possibilité d'adaptation quasi-immédiate à l'étude et à la prévision de séries temporelles. Considérons une série temporelle $\{(Z_i), i = 1, \dots, Z_\infty\}$, où chaque Z_i est une variable aléatoire réelle. Le problème de la prévision de valeurs futures Z_{n+s} ($s \in \mathbb{N}^*$) de ce processus, peut s'aborder au travers de l'estimation de la fonction d'autorégression :

$$R_k(t_1, \dots, t_k) = E(Z_{i+s} | Z_i = t_1, \dots, Z_{i-k+1} = t_k), \quad k \in \mathbb{N}^*, \quad t_j \in \mathbb{R}. \quad (3.1)$$

Dans ce Chapitre, nous nous en tiendrons à l'estimation de la fonction d'autorégression d'ordre 1, que nous noterons plus simplement

$$R(t) = E(Z_{i+1} | Z_i = t), \quad t \in \mathbb{R}, \quad (3.2)$$

tandis que le cadre général sera évoqué au Chapitre 4. L'hypothèse sous-jacente à tout ce chapitre est bien entendu l'existence d'une telle fonction R , indépendante de i , hypothèse satisfaite en particulier dès que le processus est stationnaire.

Une façon d'aborder les choses consiste à remarquer qu'en posant

$$Y_i = Z_{i+1} \text{ et } X_i = Z_i, \quad (3.3)$$

on se trouve en fait en face d'un problème d'estimation de la fonction de régression d'une variable réelle Y sur une variable réelle X , puisque l'on a de manière évidente que

$$R(t) = E(Y_i | X_i = t), \quad t \in \mathbb{R}.$$

Il est donc naturel de chercher à utiliser les méthodes de régression étudiées au Chapitre 2. Dans cette démarche, la seule difficulté consiste donc à adapter les techniques précédentes au cadre de l'estimation d'une fonction de régression à partir d'échantillon non indépendant. Nous supposons que tous les couples (X_i, Y_i) définis par (3.3) sont identiquement distribués, chacun ayant même loi qu'un couple (X, Y) .

3.1.2 Le cadre non-paramétrique

Pour reprendre les notations du Chapitre 2, on notera f la densité marginale de Z , densité supposée exister. Le cadre sous lequel nous envisageons ce problème d'estimation est de nature non-paramétrique, en ce sens que la Définition 1.2.1 est vérifiée pour le couple (X, Y) défini ci-dessus. Plus précisément notre modèle ne fera que des hypothèses de régularité sur les fonctions R et f . Lorsque nous nous intéressons à l'estimation de R en un point x fixé de \mathbb{R} tel que

$$f(x) > 0, \quad (3.4)$$

notre hypothèse non-paramétrique consistera à supposer que pour $k \in \mathbb{N}$ on a que

$$R \text{ et } f \text{ sont } k \text{ fois continûment dérivables autour de } x. \quad (3.5)$$

Lorsque nous nous intéresserons à une estimation globale de R sur un compact S tel qu'il existe $\theta > 0$ pour lequel

$$\inf_{t \in S} f(t) > \theta > 0, \quad (3.6)$$

notre hypothèse non-paramétrique consistera à supposer que pour $k \in \mathbb{N}$ on a que

$$R \text{ et } f \text{ sont } k \text{ fois continûment dérivables autour de } S. \quad (3.7)$$

3.1.3 Le prédicteur à noyau

Le fait que l'estimation de la fonction d'autorégression R puisse être vu comme un problème de régression particulier (*Cf.* (3.3)), et le fait que les estimateurs de Nadaraya-Watson définis en (2.4) combinent leurs bonnes propriétés mathématiques avec leurs facilités d'utilisation (*Cf.* Chapitre 2), nous amènent naturellement à proposer comme estimateur de la fonction R le prédicteur à noyau :

$$\hat{R}(x) = \frac{\sum_{i=1}^{n-s} Z_{i+s} K\left(\frac{x-Z_i}{h}\right)}{\sum_{i=1}^{n-s} K\left(\frac{x-Z_i}{h}\right)}, \quad \forall x \in \mathbb{R}. \quad (3.8)$$

L'introduction de ces prédicteurs remonte à Collomb (1984), et dans cette définition les paramètres h et K sont définis comme dans le cadre usuel de régression, c'est à dire que K est une fonction de pondération (non nécessairement positive) et que $h = h(n)$ est un réel strictement positif.

3.1.4 Présentation du Chapitre 3

Ce chapitre est organisé comme suit. Tout d'abord nous nous éloignerons un peu du problème initial de séries temporelles, en ce sens que nous nous consacrerons lors du Paragraphe 3.2 au problème général de l'estimation d'une fonction de régression d'une variable réelle Y sur une variable réelle X

$$r(x) = E(Y|X = x), \quad x \in \mathbb{R}, \quad (3.9)$$

que nous aborderons, contrairement au Chapitre 2, à partir d'un échantillon de couples $\{(X_i, Y_i), i = 1, \dots, n\}$ non nécessairement indépendants. En particulier nous définirons dans ce Paragraphe 3.2, une structure de dépendance sous laquelle nous travaillerons ultérieurement, et nous rappellerons quelques outils probabilistes disponibles dans ce cadre et qui joueront un rôle essentiel par la suite.

Le Paragraphe 3.3 établira des extensions au cadre dépendant des résultats de convergence du Chapitre 2. Nous donnerons des propriétés de convergence presque complète et de convergence en moyenne quadratique, ponctuelles sous un modèle de type (3.4), ou bien uniformes sous un modèle de type (3.6). L'obtention des vitesses de convergence sera liée à l'existence de dérivées continues (*i.e.* à une hypothèse $k > 0$ dans les modèles (3.5) ou (3.7)). Nous verrons en particulier que dans ce cadre dépendant relativement général, nos estimateurs peuvent atteindre les mêmes vitesses de convergence que dans un cadre i.i.d., vitesses dont l'optimalité sous des modèles de type (3.5) ou (3.7) a déjà été discutée au Paragraphe 2.5 du Chapitre 2.

Par le biais de (3.3), nous reviendrons ensuite lors du Paragraphe 3.4 au problème initial de prévision de séries temporelles, et nous verrons comment les résultats du Paragraphe 3.3 amènent directement des propriétés asymptotiques intéressantes pour l'estimateur \hat{R} de R .

Là aussi il était impossible de prétendre à une quelconque exhaustivité de notre propos, tant la littérature dans ce domaine est abondante. C'est la raison pour laquelle nous consacrons le Paragraphe 3.5 à une présentation rapide de la bibliographie récente dans ce domaine. Nous terminerons ce chapitre avec le Paragraphe 3.6 qui contient une série d'exercices qui ont pour but de se familiariser avec la

manipulation des outils de dépendance en estimation fonctionnelle. Notons que certains de ces exercices consistent à démontrer des extensions des résultats que nous présentons aux Paragraphes 3.3 et 3.4.

3.2 Le modèle de dépendance

3.2.1 Le mélange fort

La manière la plus habituelle qui est utilisée pour modéliser la dépendance dans des problèmes de prévision sous hypothèse non-paramétrique est de supposer que le processus vérifie une condition de mélange. La littérature fait état de nombreux types de conditions de mélange. Ces conditions de mélange ont été très souvent étudiées ces dernières années, tant d'un point de vue probabiliste que pour leurs possibilités d'application en prévision non-paramétrique. D'ailleurs, l'évolution des résultats concernant le prédicteur à noyau (3.8) a suivi celle des outils probabilistes disponibles en matière de variables aléatoires mélangeantes. Notre objectif ici n'est pas de proposer une vision exhaustive de toutes ces conditions de mélange et de tous ces outils. A ce titre nous recommandons la lecture du livre de Doukhan (1994) pour une discussion détaillée des diverses conditions de mélange existantes, ainsi que celui de Rio (1999) pour une présentation des outils probabilistes les plus récents.

Pour ce qui nous concerne ici, nous nous en tiendrons à la notion de mélange fort (ou α mélange) définie comme suit ; définition que nous allons donner dans un cadre général (*i.e.* pour des variables non nécessairement réelles).

Définition 3.2.1 Soit $\{\Delta_i, i \in \mathbb{Z}\}$ une famille de variables aléatoires à valeurs dans un même espace probabilisable (E, \mathcal{E}) . Pour tout couple (i, j) dans $\mathbb{Z} \cup \{-\infty, +\infty\}$, on note σ_i^j la tribu engendrée par $\{\Delta_k, i < k < j\}$. On appelle coefficients de mélange fort, les réels

$$\alpha(n) = \sup_{\{k \in \mathbb{Z}, A \in \sigma_{-\infty}^k, B \in \sigma_{n+k}^{+\infty}\}} |P(A \cap B) - P(A)P(B)|. \quad (3.10)$$

Définition 3.2.2 On dit qu'une famille $\{\Delta_i, i \in \mathbb{Z}\}$ de variables aléatoires à valeurs dans un même espace probabilisable (E, \mathcal{E}) est fortement mélangeante, ou α -mélangeante, si l'on a

$$\lim_{n \rightarrow \infty} \alpha(n) = 0. \quad (3.11)$$

Cette condition de mélange présente deux avantages. Tout d'abord, il s'agit de la moins restrictive parmi les différentes conditions de mélange existant dans la

littérature classique (*Cf. eg* Doukhan, 1994, Chapitre 1.1). Ensuite, les connaissances probabilistes sur ce type de dépendance sont suffisamment poussées (*Cf.* Rio, 1999) pour nous permettre de mener à bien l'étude de prévision non paramétrique qui est notre souci majeur ici.

La littérature fait en général état de deux types particuliers de mélange, selon la nature de la convergence vers 0 des coefficients $\alpha(n)$.

Définition 3.2.3 *On dit qu'une famille $\{\Delta_i, i \in \mathbb{Z}\}$ de variables aléatoires à valeurs dans un même espace probabilisable (E, \mathcal{E}) est algébriquement α -mélangeante, s'il existe deux constantes $c \in \mathbb{R}^{*+}$ et $a \in \mathbb{R}^{*+}$ telles que les coefficients de mélange vérifient*

$$\alpha(n) \leq cn^{-a} \quad (3.12)$$

Définition 3.2.4 *On dit qu'une famille $\{\Delta_i, i \in \mathbb{Z}\}$ de variables aléatoires à valeurs dans un même espace probabilisable (E, \mathcal{E}) est géométriquement α -mélangeante, s'il existe deux constantes $s \in \mathbb{R}^{*+}$ et $t \in]0, 1[$ telles que les coefficients de mélange vérifient*

$$\alpha(n) \leq st^n. \quad (3.13)$$

De manière évidente, la première de ces deux notions de mélange est plus générale que la seconde, puisque pour tout t et tout a on a $t^n = o(n^{-a})$. La plupart de nos résultats seront donnés sous la condition (3.2.3), plutôt que sous la condition générale (3.11), cela uniquement à fin de ne pas compliquer l'énoncé de nos théorèmes (et surtout de nos hypothèses) car cela comporterait le risque évident de masquer les aspects essentiels que nous souhaitons mettre en évidence. L'extension de nos résultats au cadre général (3.11) sera discutée, en lien avec la bibliographie existante, au Paragraphe 3.5.

3.2.2 Une inégalité exponentielle

L'outil que nous allons utiliser de manière déterminante dans les problèmes de convergence presque complète est l'inégalité exponentielle ci-dessous que l'on peut rencontrer dans Rio (1999, p. 90) sous l'appellation inégalité de Fuk-Nagaev. Cette inégalité est en fait une extension au cadre de variables fortement mélangeantes de l'inégalité de Bernstein que l'on peut trouver sous diverses formes dans Hoeffding (1963) (voir aussi le Lemme 2.3.1 pour une version simplifiée). En fait, le développement des résultats en estimation non paramétrique de régression sous dépendance s'est fait en parallèle avec celui de telles inégalités. Le premier résultat

concernant le prédicteur (3.8) est celui de Collomb (1984) qui dans son travail avait établi une première inégalité exponentielle pour processus mélangeant (voir aussi Bosq, 1975 pour une première version de cette inégalité) mais dans un cadre plus restrictif (*i.e.* le mélange uniforme ou ϕ -mélange) que le notre. Ensuite, dans le cadre α -mélangeant qui est le notre ici, le premier résultat disponible concernant le prédicteur (3.8) fût donné dans Györfi *et al.* (1989) grâce à l'utilisation d'une inégalité exponentielle de Carbon (1983), mais les limites de l'inégalité de carbon ne permettaient pas d'obtenir des vitesses de convergence optimale. Depuis, la puissance de telles inégalités a été sérieusement améliorée (*Cf.* Bosq, 1993), jusqu'à l'inégalité de Rio (1999). Le lecteur trouvera dans Rio (1999) mais aussi dans Bosq (1996) plusieurs versions de cette inégalité, dont certaines sous des énoncés plus généraux que ce qui va suivre. Cependant nous en tiendrons ici à cette version simplifiée qui est largement suffisante dans notre contexte, et que l'on trouve sous cette forme dans Rio (1999, p. 87, formule (6.19b)).

Lemme 3.2.1 Inégalité de type Fuk-Nagaev. *Soit $\{\Delta_i, i \in \mathbb{N}\}$ une famille de variables aléatoires à valeurs dans \mathbb{R} qui vérifient la condition de mélange fort (3.11) avec des coefficients à décroissance algébrique tels que définis en (3.12). On pose*

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

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

$$P \left[\left| \sum_{k=1}^n \Delta_k \right| > 4\epsilon \right] \leq 4 \left(1 + \frac{\epsilon^2}{rs_n^2} \right)^{-\frac{r}{2}} + 2ncr^{-1} \left(\frac{2r}{\epsilon} \right)^{a+1}. \quad (3.14)$$

3.2.3 Une inégalité de covariance

L'utilisation de l'inégalité exponentielle précédente va nécessiter le calcul d'expression de type s_n^2 , et donc nécessiter l'utilisation d'inégalités de covariance pour variables mélangeantes. La littérature abonde en inégalités de ce type, et nous renvoyons à Bosq (1996, Chapitre 1) et Rio (1999, Chapitre 1) pour une présentation de bon nombre d'entre elles. Pour ce qui nous concerne, nous fonderons nos preuves en nous tenant à la plus récente et la plus performante d'entre elles, celle de Rio (1993), dont le lecteur pourra aussi trouver une preuve dans Bosq (1996, Théorème 1.1) ou dans Rio (1999, Théorème 1.1). Concrètement, nous utiliserons la forme simplifiée suivante de l'inégalité de Rio. On trouvera cette forme simplifiée dans Bosq (1996, p. 19, formule (1.11)).

Lemme 3.2.2 Inégalité de covariance. Soit $\{\Delta_i, i \in \mathbb{N}\}$ une famille de variables aléatoires à valeurs dans \mathbb{R} qui vérifient la condition de mélange fort (3.11), et telle que $\|\Delta_i\|_\infty < \infty, \forall i$. On a pour tout $i \neq j$:

$$|\text{cov}(\Delta_i, \Delta_j)| \leq 4\|\Delta_i\|_\infty\|\Delta_j\|_\infty\alpha^{|i-j|}. \quad (3.15)$$

3.3 Regression sous mélange fort

3.3.1 Introduction

Dans ce Paragraphe 3.3 nous allons oublier temporairement le problème de série temporelle qui nous préoccupe pour nous placer dans le cadre général de l'estimation d'une fonction de régression sous dépendance. Ainsi, nous reprendrons le formalisme et les définitions qui étaient celles du Chapitre 2, à savoir que X et Y sont des variables aléatoires réelles pour lesquelles on cherche à estimer la fonction de régression r :

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

à partir d'un échantillon $\{(X_i, Y_i), i = 1, \dots, n\}$. La principale différence viendra du fait que nous ferons l'hypothèse de mélangeance suivante sur les couples (X_i, Y_i) :

$$\text{La suite } \{(X_i, Y_i), i = 1, \dots, n\} \text{ est } \alpha\text{-mélangeante}, \quad (3.16)$$

et nous supposons que pour tout $i \neq j$,

$$(X_i, X_j) \text{ admet une densité notée } f_{ij}. \quad (3.17)$$

L'estimateur que nous étudierons est l'estimateur usuel de Nadaraya-Watson déjà étudié au Chapitre 2

$$\hat{r}_{NW}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)}, \quad \forall x. \quad (3.18)$$

Ainsi, les résultats que nous allons présenter peuvent être vus comme des extensions au cadre mélangeant de ceux du Chapitre 2. Nous établissons maintenant des propriétés de convergence presque complète et en moyenne quadratique, en donnant chaque fois sous deux modèles différents (modèle de continuité et modèles de dérivabilité) un résultat ponctuel et un résultat uniforme.

3.3.2 Quelques résultats préliminaires

Pour étendre des résultats de convergence du cadre i.i.d. à un cadre de dépendance, la difficulté se traduit le plus souvent par la nécessité d'avoir à inclure des termes de type covariance dans les calculs habituels de variance. Vu la forme des estimateurs que nous allons étudier ici (Cf. (3.18)), nous allons avoir à appliquer des inégalités du type de celles présentées ci-dessus à des variables aléatoires de la forme :

$$\Delta_i = Y_i^l K\left(\frac{x - X_i}{h}\right) - EY_i^l K\left(\frac{x - X_i}{h}\right), \quad l = 0 \text{ ou } l = 1. \quad (3.19)$$

C'est la raison pour laquelle nous avons décidé, en guise de préliminaire, de donner un résultat général concernant la quantité

$$s_n^2 = \sum_{i=1}^n \sum_{j \neq i} cov(\Delta_i, \Delta_j). \quad (3.20)$$

Les hypothèses que nous nécessiterons sont voisines de celles utilisées au Chapitre 2 dans le cadre i.i.d. Nous introduirons les conditions suivantes :

$$K \text{ est intégrable, borné et à support compact,} \quad (3.21)$$

$$|Y| < M < \infty, \quad (3.22)$$

et nous aurons parfois besoin de la notion de noyau d'ordre k dont nous rappelons qu'elle s'écrit :

$$\int t^j K(t) dt = 0, \quad \forall j = 1, \dots, k-1 \text{ et } 0 < \left| \int t^k K(t) dt \right| < \infty. \quad (3.23)$$

Bien que nous utilisions des hypothèses et des arguments différents, la démonstration de la proposition suivante s'inspire des calculs de Bosq (1996, p. 43).

Proposition 3.3.1 *Sous les conditions (3.4), (3.5), (3.16), (3.17), (3.21) et (3.22), on a*

$$s_n^2 = o(nh) + O(n^2 \alpha((h \log n)^{-1})). \quad (3.24)$$

Preuve de la Proposition 3.3.1. En utilisant tout d'abord que Y est bornée, puis en utilisant une technique d'intégration par changement de variable, et en utilisant finalement que K est borné, on a :

$$\begin{aligned} |E\Delta_i, \Delta_j| &\leq C \int \int \left| K\left(\frac{x-u}{h}\right) K\left(\frac{x-v}{h}\right) (f_{ij}(u, v) - f(u)f(v)) \right| dudv \\ &\leq Ch^2 \int \int |K(z)K(t) (f_{ij}(x-hz, x-hz) - f(x-hz)f(x-hz))| dzdt. \end{aligned}$$

On arrive finalement à

$$\text{cov}(\Delta_i, \Delta_j) = E\Delta_i\Delta_j = O(h^2). \quad (3.25)$$

D'un autre côté, on peut aussi majorer cette covariance directement à partir de l'inégalité du Lemme 3.2.2, pour arriver à

$$|\text{cov}(\Delta_i, \Delta_j)| \leq C\alpha(|i - j|). \quad (3.26)$$

L'idée de la preuve consiste à introduire une suite u_n entière, et d'utiliser tantôt la borne (3.25) quand i est proche de j , tantôt la borne (3.26) quand i et j sont éloignés. On arrive ainsi à

$$\begin{aligned} s_n^2 &\leq C \left[\sum_{\{|i-j| \leq u_n\}} \sum h^2 + \sum_{\{|i-j| > u_n\}} \sum \alpha(|i-j|) \right] \\ &= O(h^2 n u_n + n^2 \alpha(u_n)). \end{aligned} \quad (3.27)$$

Il suffit alors de choisir $u_n = 1/(h \log n)$ pour obtenir le résultat (3.24). \square

Proposition 3.3.2 *Si S est un compact tel que (3.6) et (3.7) soient vérifiées, et si les conditions de la Proposition 3.3.1 sont vérifiées, alors on a*

$$\sup_{x \in S} s_n^2 = o(nh) + O(n^2 \alpha((h \log n)^{-1})). \quad (3.28)$$

Preuve de la Proposition 3.3.2. Il suffit de constater que les bornes obtenues en (3.25) et (3.26) sont indépendantes de x dans S . \square

Pour simplifier nos exposés, nous nous limiterons au cadre algébrique. Ainsi, de manière évidente on a le corollaire suivant.

Corollaire 3.3.1 *Supposons que les conditions de la Proposition 3.3.1 soient satisfaites et que les coefficients de mélange satisfont la condition algébrique (3.12). Supposons en outre que les coefficients de mélange sont liés à la largeur de fenêtre par la relation*

$$\exists \epsilon > 0, h^{a-1} = O(n^{-1-\epsilon}). \quad (3.29)$$

Alors on a

$$s_n^2 = o(nh). \quad (3.30)$$

Preuve du Corollaire 3.3.1. En reprenant le résultat (3.24), on a sous l'hypothèse (3.12) :

$$s_n^2 = o(nh) + O(n^2(h \log n)^a), \quad (3.31)$$

d'où l'on tire le résultat voulu. \square

De la même manière on obtient directement à partir de la Proposition 3.3.2 le corollaire suivant.

Corollaire 3.3.2 *Si S est un compact tel que (3.6) et (3.7) soient vérifiées, et si les conditions du Corollaire 3.3.1 sont vérifiées, alors on a*

$$\sup_{x \in S} s_n^2 = o(nh) + O(n^2(h \log n)^a). \quad (3.32)$$

Pour voir de manière encore plus directe la nature de notre hypothèse sur les coefficients de mélange, nous allons donner un nouveau corollaire de ce résultat lorsque la largeur de fenêtre est de la forme

$$h = C(n^{-\frac{1}{k+1}})(\log n)^\alpha, \quad \alpha \in \mathbb{R}, \quad (3.33)$$

forme dont l'optimalité a été vue dans le cadre i.i.d. lors du Paragraphe 2.5 du Chapitre 2, et qui sera confirmée dans le cadre dépendant dans ce chapitre (Cf. Paragraphe 3.5).

Corollaire 3.3.3 *Sous les conditions de la Proposition 3.3.1, pour les largeurs de fenêtre satisfaisant (3.33) et pour des coefficients de mélange qui satisfont la condition algébrique (3.12) avec*

$$a > 2k + 2, \quad (3.34)$$

on a

$$s_n^2 = o(nh). \quad (3.35)$$

Preuve du Corollaire 3.3.3. Il est immédiat que (3.29) découle de (3.33) et (3.34). \square

Nous terminons avec une version uniforme de ce dernier résultat qui découle directement du Corollaire 3.3.2.

Corollaire 3.3.4 *Si S est un compact tel que (3.6) et (3.7) soient vérifiées, et si les conditions du Corollaire 3.3.3 sont vérifiées, alors on a*

$$\sup_{x \in S} s_n^2 = o(nh). \quad (3.36)$$

3.3.3 Convergence presque complète ponctuelle

Nous allons tout d'abord donner des propositions générales qui concernent l'estimateur

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

de f , et l'estimateur

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

de g =rf. A partir de ces propositions générales nous obtiendrons quasi directement les propriétés de convergence presque complète qui nous intéressent.

Proposition 3.3.3 *Supposons que les conditions de la Proposition 3.3.1 soient vérifiées. Supposons que la condition de décroissance algébrique (3.2.3) soit satisfaite pour une valeur de a vérifiant avec la fenêtre h les hypothèses suivantes*

$$\exists \theta > 0, \exists c_1 > 0, \exists c_2 > 0, c_2 n^{\left(\frac{3-a}{a+1} + \theta\right)} \leq h \leq c_1 n^{\left(\frac{1}{1-a} - \theta\right)}. \quad (3.37)$$

Alors il existe $\nu > 0$ et $\epsilon > 0$ tel que l'on ait :

$$P \left[|E\hat{g}(x) - \hat{g}(x)| > \epsilon \sqrt{\frac{\log n}{nh}} \right] = O(n^{-1-\nu}), \quad (3.38)$$

et

$$P \left[|E\hat{f}(x) - \hat{f}(x)| > \epsilon \sqrt{\frac{\log n}{nh}} \right] = O(n^{-1-\nu}). \quad (3.39)$$

Preuve de la Proposition 3.3.3. Les deux résultats se démontrent de la même manière, et nous ferons les deux preuves simultanément en posant

$$\hat{\psi}(x) = \hat{f}(x) \text{ ou } \hat{g}(x).$$

Notons tout d'abord que la deuxième partie de la condition (3.37) implique que (3.29) est réalisée, et donc que (3.30) est vérifiée. Une application directe du Lemme 3.2.1 aux variables

$$\Delta_i = Y_i^l K\left(\frac{x - X_i}{h}\right) - EY_i^l K\left(\frac{x - X_i}{h}\right), \quad l = 0 \text{ ou } l = 1$$

amène alors en utilisant (3.30), que pour $\epsilon > 0$ et pour $r > 1$ on a

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

Ainsi on arrive à

$$P \left[\left| E\hat{\psi}(x) - \hat{\psi}(x) \right| > \epsilon \sqrt{\frac{\log n}{nh}} \right] \leq 4 \left(1 + \frac{\epsilon^2 \log n}{16r} \right)^{-\frac{r}{2}} \\ + 2ncr^{-1} \left(\frac{8r}{\epsilon} \right)^{a+1} (nh \log n)^{\frac{-(a+1)}{2}}.$$

On choisit alors r de telle sorte que $\log n = o(r)$, et on arrive à

$$P \left[\left| E\hat{\psi}(x) - \hat{\psi}(x) \right| > \epsilon \sqrt{\frac{\log n}{nh}} \right] \leq Ce^{-\frac{\epsilon^2 \log n}{32}} \\ + C\epsilon^{-(a+1)} n^{1-\frac{(a+1)}{2}} r^a h^{-\frac{a+1}{2}}.$$

On peut toujours choisir r sous la forme $r = n^b$, $b > 0$, et on arrive alors à

$$P \left[\left| E\hat{\psi}(x) - \hat{\psi}(x) \right| > \epsilon \sqrt{\frac{\log n}{nh}} \right] \leq Cn^{-\frac{\epsilon^2}{32}} \\ + C\epsilon^{-(a+1)} n^{1+ab-\frac{(a+1)}{2}} h^{-\frac{a+1}{2}}. \quad (3.40)$$

Grâce à la première partie de la condition (3.37), on peut trouver un b suffisamment petit tel qu'il existe $\nu > 0$ pour lequel

$$P \left[\left| E\hat{\psi}(x) - \hat{\psi}(x) \right| > \epsilon \sqrt{\frac{\log n}{nh}} \right] \leq Cn^{-\frac{\epsilon^2}{32}} + Cn^{-1-\nu}, \quad (3.41)$$

d'où pour ϵ suffisamment grand on a

$$P \left[\left| E\hat{\psi}(x) - \hat{\psi}(x) \right| > \epsilon \sqrt{\frac{\log n}{nh}} \right] < Cn^{-1-\nu}. \square \quad (3.42)$$

Avant de poursuivre il est utile de donner quelques précisions sur la condition (3.37) qui joue un rôle clé dans la démonstration précédente, et que nous avons volontairement gardé sous sa forme la plus générale afin de pouvoir en tirer le maximum dans ce qui va suivre. La remarque suivante, dont la preuve est immédiate, explicite deux cas particuliers pour lesquels cette condition est vérifiée.

Remarque. i) Pour que la condition (3.37) soit vérifiée, il faut nécessairement que

$$a > \frac{(5 + \sqrt{17})}{2}. \quad (3.43)$$

ii) Dans le cas particulier où $k > 0$ et où la largeur de fenêtre est de la forme habituelle :

$$h = C \left(\frac{\log n}{n} \right)^{\frac{1}{2k+1}}, \quad (3.44)$$

la condition (3.37) est vérifiée dès que

$$a > \max\{2k + 2; 3 + \frac{2}{k}\}. \quad (3.45)$$

Grâce à cette remarque nous allons pouvoir donner une version plus maniable de la Proposition 3.3.3.

Corollaire 3.3.5 *Supposons que les conditions de la Proposition 3.3.1 soient vérifiées et que $k > 0$. Supposons que la condition de décroissance algébrique (3.12) soit satisfaite pour une valeur de a vérifiant (3.45). Supposons enfin que la largeur de fenêtre soit celle définie en (3.44). Alors, il existe $\nu > 0$ tel que pour ϵ suffisamment grand, (3.38) et (3.39) soient réalisés.*

Nous sommes maintenant en mesure d'énoncer les résultats essentiels de ce Paragraphe 3.3.3, à savoir la convergence presque complète de l'estimateur \hat{r}_{NW} . Nous allons commencer par un résultat sous un modèle de dérivabilité dans lequel nous précisons les vitesses de convergence. Nous donnerons ensuite un résultat plus général en terme de modèle, puisqu'il s'agira d'un modèle de continuité, mais sans précision sur les vitesses de convergence. Pour des raisons de simplicité d'écriture, nous ne donnons nos résultats que pour des processus mélangeants à décroissance algébrique.

Théorème 3.3.1 **Vitesse de convergence presque complète ponctuelle sous condition de dérivabilité.** *Considérons le modèle de dérivabilité défini par (3.4) et (3.5) avec $k > 0$ et supposons que les conditions (3.12), (3.16), (3.17), (3.21), (3.22), (3.23), (3.44) et (3.45) soient réalisées. On a*

$$\hat{r}_{NW}(x) - r(x) = O \left(\left(\frac{n}{\log n} \right)^{-\frac{k}{2k+1}} \right), \quad p.co. \quad (3.46)$$

Preuve du Théorème 3.3.1. Il suffit de suivre le même cheminement que pour la démonstration du Théorème 2.3.1 qui est l'analogue de celui-ci dans le cadre i.i.d., en utilisant le Corollaire 3.3.5 qui précède pour traiter les termes affectés par la dépendance. Plus précisément on voit que l'établissement des résultats (2.13), (2.14) et (2.15) ne font pas intervenir la condition d'indépendance sur les couples (X_i, Y_i) . Ces résultats restent donc valables sous nos hypothèses. Quant aux résultats (2.16) et (2.17), ils découlent immédiatement des résultats (3.38) et (3.39) que nous venons d'établir lors du Corollaire 3.3.5. Il ne reste plus qu'à prouver (2.18), et cela se fait exactement comme dans le cadre indépendant à partir de (2.15) et (3.39). Finalement, on obtient à partir de (2.13)-(2.18) que

$$\hat{r}_{NW}(x) - r(x) = O(h^k) + O\left(\sqrt{\frac{\log n}{nh}}\right), \text{ p.co.}, \quad (3.47)$$

et il suffit alors d'utiliser l'hypothèse (3.44) sur h pour conclure à (3.46).□

Théorème 3.3.2 Convergence presque complète ponctuelle sous condition de continuité. *Considérons le modèle de continuité défini par (3.4) et (3.5) avec $k = 0$ et supposons que les conditions (3.12), (3.16), (3.17), (3.21), (3.22) et (3.37) soient réalisées. On a*

$$|\hat{r}_{NW}(x) - r(x)| \rightarrow 0, \text{ p.co.} \quad (3.48)$$

Preuve du Théorème 3.3.2. Ici aussi il suffit de suivre le même cheminement que pour la démonstration du Théorème 2.3.3 qui est l'analogue de celui-ci dans le cadre i.i.d., mais en utilisant la Proposition 3.3.3 qui précède pour traiter les termes affectés par la dépendance. Plus précisément on voit que les résultats (2.13), (2.46) et (2.47) restent valables sous nos hypothèses. Quant aux résultats (2.48) et (2.49), ils découlent immédiatement des résultats (3.38) et (3.39) que nous venons d'établir lors de la Proposition 3.3.5. Il ne reste plus qu'à constater que (2.37) est une conséquence immédiate de (2.47) et (3.39). Finalement, on obtient le résultat recherché (3.48) à partir de (2.13) et de (2.46)-(2.50).□

Remarque. *Comme explicité en (3.43), la condition (3.37) implique nécessairement que $a > (5 + \sqrt{17})/2$.*

3.3.4 Convergence presque complète uniforme

L'obtention de résultats analogues uniformes sur un compact S est subordonnée à une hypothèse supplémentaire sur le noyau K :

$$\exists \beta > 0, \exists C < \infty, \forall x \in S, \forall y \in S, |K(x) - K(y)| \leq C|x - y|^\beta. \quad (3.49)$$

Nous aurons besoin de renforcer la condition (3.37) de la façon suivante :

$$\exists \theta > 0, \exists c_1 > 0, \exists c_2 > 0, c_2 n^{\left(\frac{(3-a)\beta}{\beta(a+1)+2\beta+1} + \theta\right)} \leq h \leq c_1 n^{\left(\frac{1}{1-a} - \theta\right)}. \quad (3.50)$$

Remarque. Pour fixer les idées considérons le cas particulier où $\beta = 1$, et un modèle de dérivabilité (i.e. un modèle avec $k > 0$). La condition (3.50) est satisfaite dès que la largeur de fenêtre est de la forme habituelle :

$$h = C \left(\frac{\log n}{n} \right)^{\frac{1}{2k+1}}, \quad (3.51)$$

et quand le taux de mélange vérifie

$$a > \max\left\{2k + 2; 4 + \frac{4}{k}\right\}. \quad (3.52)$$

Théorème 3.3.3 **Vitesse de convergence presque complète uniforme sous condition de dérivabilité.** *Considérons le modèle de dérivabilité défini par (3.6) et (3.7) avec $k > 0$ et supposons que les conditions (3.12), (3.16), (3.17), (3.21), (3.22), (3.23), (3.50), (3.49) et (3.43) soient réalisées. On a*

$$\sup_{x \in S} |\hat{r}_{NW}(x) - r(x)| = O \left(\left(\frac{n}{\log n} \right)^{-\frac{k}{2k+1}} \right), \text{ p.co.} \quad (3.53)$$

Preuve du Théorème 3.3.3. En reprenant la preuve du Théorème 2.3.2 on s'aperçoit que les résultats (2.32), (2.33), (2.34) et (2.37) restent vrais sous nos hypothèses puisque leurs preuves ne font pas intervenir l'indépendance des couples (X_i, Y_i) . Il ne reste donc qu'à établir (2.35) et (2.36). La preuve sera donc terminée dès que nous aurons montré que

$$\sup_{x \in S} |E\hat{\psi}(x) - \hat{\psi}(x)| = O \left(\sqrt{\frac{\log n}{nh}} \right), \text{ p.co., pour } \hat{\psi} = \hat{f} \text{ ou } \hat{g}. \quad (3.54)$$

Soit $\epsilon > 0$. On utilise la même démarche que lors de la preuve du Théorème 2.3.2, et on recouvre S par un nombre fini d'intervalles

$$S \subset \cup_{k=1}^{k=\tau n} B_k \text{ où } B_k =]t_k - l_n; t_k + l_n],$$

et on écrit que

$$\begin{aligned}
& P \left[\sup_{x \in S} \left| E\hat{\phi}(x) - \hat{\phi}(x) \right| > \epsilon \sqrt{\frac{\log n}{nh}} \right] \\
& \leq P \left[\max_{k=1, \dots, \tau_n} \left| E\hat{\phi}(t_k) - \hat{\phi}(t_k) \right| > \frac{\epsilon}{2} \sqrt{\frac{\log n}{nh}} \right] \\
& + P \left[\sup_{x \in S} \left| E\hat{\phi}(x) - \hat{\phi}(x) - E\hat{\phi}(t(x)) + \hat{\phi}(t(x)) \right| > \frac{\epsilon}{2} \sqrt{\frac{\log n}{nh}} \right]. \quad (3.55)
\end{aligned}$$

On utilise maintenant (2.41) pour traiter le second terme de la partie droite de (3.55) et (3.40) pour traiter le premier, en remarquant la constante C intervenant dans (3.41) est indépendante de $x \in S$ (Cf. Proposition 3.3.2). On arrive alors, en reprenant les notations de la preuve de la Proposition 3.3.3, pour tout $b > 0$ à :

$$\begin{aligned}
P \left[\sup_{x \in S} \left| E\hat{\phi}(x) - \hat{\phi}(x) \right| > \epsilon \sqrt{\frac{\log n}{nh}} \right] & \leq \frac{C}{l_n} \left[n^{-\frac{\epsilon^2}{32}} + n^{1+ab-\frac{(a+1)}{2}} h^{-\frac{a+1}{2}} \right] \\
& + P \left[C \left(\frac{l_n^\beta}{h^{1+\beta}} \right) > \frac{\epsilon}{2} \sqrt{\frac{\log n}{nh}} \right]. \quad (3.56)
\end{aligned}$$

En prenant l_n sous la forme

$$l_n^\beta = h^{(\frac{1}{2}+\beta)} n^{-\frac{1}{2}}, \quad (3.57)$$

on arrive à

$$P \left[\sup_{x \in S} \left| E\hat{\phi}(x) - \hat{\phi}(x) \right| > \epsilon \sqrt{\frac{\log n}{nh}} \right] \leq \frac{C}{l_n} \left[n^{-\frac{\epsilon^2}{32}} + n^{1+ab-\frac{(a+1)}{2}} h^{-\frac{a+1}{2}} \right]. \quad (3.58)$$

En utilisant maintenant (3.57) et la première partie de (3.50), on voit que l'on peut toujours choisir b assez petit pour qu'il existe $\nu > 0$ pour lequel on ait

$$l_n^{-1} n^{1+ab-\frac{(a+1)}{2}} h^{-\frac{a+1}{2}} = O(n^{-1-\nu}).$$

En remarquant que la condition (3.57) permet d'écrire que

$$\exists \zeta > 0, l_n^{-1} = O(n^{\zeta}),$$

on arrive à

$$P \left[\sup_{x \in S} \left| E\hat{\phi}(x) - \hat{\phi}(x) \right| > \epsilon \sqrt{\frac{\log n}{nh}} \right] \leq C \left(n^{\zeta-\frac{\epsilon^2}{32}} + n^{-1-\nu} \right).$$

On peut toujours trouver ϵ tel que le terme de droite de cette inégalité soit celui d'une série convergente, et ainsi la preuve de (3.54) est achevée. \square

Le prochain résultat est établi sous un modèle de régression plus général que celui du Théorème 3.3.3, puisque l'hypothèse d'existence de dérivées continues est remplacée par la condition de continuité. Encore une fois le gain obtenu ainsi en terme de modèle statistique est à mettre en balance avec la perte des vitesses de convergence.

Théorème 3.3.4 Convergence uniforme presque complète sous condition de continuité. *Considérons le modèle de continuité défini par (3.6) et (3.7) avec $k = 0$ et supposons que les conditions (3.12), (3.16), (3.17), (3.21), (3.22), (3.49) et (3.50) soient réalisées. Alors on a*

$$\sup_{x \in S} |\hat{r}_{NW}(x) - r(x)| \rightarrow 0, \text{ p.co.} \quad (3.59)$$

Preuve du Théorème 3.3.4 - Cette preuve reprend point par point celle du Théorème 2.3.4. Il suffit de constater que (2.32), (2.54), (2.55) et (2.58) restent valables sous structure de dépendance (car leur preuve ne faisait pas intervenir l'indépendance des variables (X_i, Y_i)). Il ne reste plus qu'à établir (2.56) et (2.57), ce qui se fait en constatant que la preuve de (3.54) ci-dessus reste valable sous le modèle $k = 0$ qui est le notre dans ce théorème. \square

3.3.5 Convergence en moyenne quadratique

Nous allons énoncer maintenant des résultats de convergence en moyenne quadratique, sous des modèles de dérivabilité ou des modèles de continuité, et dans chaque cas nous donnerons une version ponctuelle et une version intégrée sur un compact. Nous reprendrons les notations du Paragraphe 2.4 du Chapitre 2. Les démonstrations en seront très rapides puisque nous avons, lors du Paragraphe 3.3.2, établi tous les résultats intermédiaires qui nous étaient nécessaires. En fait nous ne détaillerons que la première de ces preuves, les trois autres suivant clairement la même démarche et utilisant les mêmes arguments.

Théorème 3.3.5 Convergence en moyenne quadratique ponctuelle sous condition de dérivabilité. *Supposons que les conditions du Théorème 2.4.1 soient vérifiées. Supposons que la structure de dépendance soit celle définie par les conditions (3.12), (3.16), (3.17) avec*

$$a > 6 \text{ et } h = Cn^{-\frac{1}{5}}.$$

Alors on a

$$E[\hat{r}_{NW}(x) - r(x)]^2 = O\left(n^{-\frac{4}{5}}\right). \quad (3.60)$$

Preuve du Théorème 3.3.5. Il suffit de reprendre ligne à ligne la preuve du Théorème 2.4.1, en remarquant que tous les termes de covariance qui vont apparaître lors des calculs (c'est à dire lors de l'obtention des formules (2.74), (2.75), (2.78), (2.79) et (2.80)) sont de la forme (3.20) et sont par là même tous des $o(\frac{1}{nh})$ d'après le Corollaire 3.3.3. On arrive alors à

$$E[\hat{r}_{NW}(x) - r(x)]^2 = B^2(x)h^4 + V(x)\frac{1}{nh} + o(h^4 + \frac{1}{nh}), \quad (3.61)$$

où

$$B(x) = \frac{\int t^2 K(t) dt g^{(k)}(x) - r(x) f^{(k)}(x)}{2 f(x)}, \quad (3.62)$$

et

$$V(x) = \int K^2(t) dt \frac{\phi(x) - r^2(x)}{f(x)}. \quad (3.63)$$

Il suffit d'utiliser la condition sur h pour conclure. \square

La démonstration du résultat qui suit est naturellement omise car elle est similaire à celle qui vient de précéder.

Théorème 3.3.6 Erreur quadratique moyenne intégrée sous condition de dérivabilité. *Supposons que les conditions du Théorème 2.4.3 soient vérifiées. Supposons que la structure de dépendance soit celle définie par les conditions (3.12), (3.16), (3.17) avec*

$$a > 6 \text{ et } h = Cn^{-\frac{1}{5}}.$$

Alors on a

$$EQMI(\hat{r}_{NW}) = O\left(n^{-\frac{4}{5}}\right). \quad (3.64)$$

Théorème 3.3.7 Convergence en moyenne quadratique ponctuelle sous condition de continuité. *Supposons que les conditions du Théorème 2.4.2 soient vérifiées. Supposons que la structure de dépendance soit celle définie par les conditions (3.12), (3.16), (3.17) avec*

$$h^{(a-1)} = O(n^{-(1+\theta)}).$$

Alors on a

$$E[\hat{r}_{NW}(x) - r(x)]^2 \rightarrow 0. \quad (3.65)$$

Preuve du Théorème 3.3.7. Il suffit de reprendre ligne à ligne la preuve du Théorème 2.4.2, en invoquant les mêmes arguments que pour prouver 3.3.5 mais en les fondant sur le résultat du Corollaire 3.3.1 plutôt que sur ceux du Corollaire 3.3.3. \square

Le résultat suivant se démontre exactement de la même manière à partir du résultat du Corollaire 3.3.2.

Théorème 3.3.8 Erreur quadratique moyenne intégrée sous condition de continuité. *Supposons que les conditions du Théorème 2.4.4 soient satisfaites. Supposons que la structure de dépendance soit celle définie par les conditions (3.2.3), (3.16), (3.17) avec*

$$h^{(a-1)} = O(n^{-(1+\theta)}).$$

Alors on a

$$EQMI(\hat{r}_{NW}) \rightarrow 0. \quad (3.66)$$

3.4 Application à la prédiction

Nous allons maintenant revenir à notre problème de départ, à savoir la prévision de valeur future d'un processus Z à valeurs réelles. Nous allons donner des résultats asymptotiques concernant le prédicteur à noyau défini par (3.8) en tant qu'estimateur non-paramétrique de la fonction d'autorégression (3.2). Ce paragraphe sera relativement court, puisque nous avons déjà suffisamment balisé le terrain dans ceux qui précèdent. En fait les résultats concernant ce prédicteur sont des conséquences immédiates de ceux établis dans le Paragraphe 3.3 qui précède. Comme nous l'expliquons en paragraphe introductif, le lien entre estimation de régression et prévision du processus du Z se fera en posant

$$Y_i = Z_{i+s} \text{ et } X_i = Z_i. \quad (3.67)$$

Afin de faire dans la concision nous avons synthétisé toutes les propriétés asymptotique de l'estimateur (3.8) dans deux théorèmes. Le premier de ces théorèmes traite de résultats de convergence presque complète, et il se démontre directement à partir des théorèmes similaires établis précédemment pour l'estimateur de régression \hat{r}_{NW} .

Théorème 3.4.1 Propriétés presque complète du prédicteur à noyau.

i) Supposons que les conditions du Théorème 3.3.1 soient réunies pour les variables (X_i, Y_i) définies par (3.67). Alors on a

$$\hat{R}(x) - R(x) = O\left(\left(\frac{n}{\log n}\right)^{-\frac{k}{2k+1}}\right), \text{ p.co.} \quad (3.68)$$

ii) Supposons que les conditions du Théorème 3.3.2 soient réunies pour les variables (X_i, Y_i) définies par (3.67). Alors on a

$$\left|\hat{R}(x) - R(x)\right| \rightarrow 0, \text{ p.co.} \quad (3.69)$$

iii) Supposons que les conditions du Théorème 3.3.3 soient réunies pour les variables (X_i, Y_i) définies par (3.67). Alors on a

$$\sup_{x \in S} \left|\hat{R}(x) - R(x)\right| = O\left(\left(\frac{n}{\log n}\right)^{-\frac{k}{2k+1}}\right), \text{ p.co.} \quad (3.70)$$

iv) Supposons que les conditions du Théorème 3.3.4 soient réunies pour les variables (X_i, Y_i) définies par (3.67). Alors on a

$$\sup_{x \in S} |\hat{r}_{NW}(x) - r(x)| \rightarrow 0, \text{ p.co.} \quad (3.71)$$

De la même façon nous allons donner dans le théorème qui suit des propriétés de convergence en moyenne quadratique qui se démontrent directement à partir des résultats précédents sur l'estimateur \hat{r}_{NW} .

Théorème 3.4.2 Convergence en moyenne quadratique du prédicteur à noyau.

i) Supposons que les conditions du Théorème 3.3.5 soient réunies pour les variables (X_i, Y_i) définies par (3.67). Alors on a

$$E[\hat{R}(x) - R(x)]^2 = O\left(n^{-\frac{4}{5}}\right). \quad (3.72)$$

ii) Supposons que les conditions du Théorème 3.3.6 soient réunies pour les variables (X_i, Y_i) définies par (3.67). Alors on a

$$EQMI(\hat{R}) = O\left(n^{-\frac{4}{5}}\right). \quad (3.73)$$

iii) Supposons que les conditions du Théorème 3.3.7 soient réunies pour les variables (X_i, Y_i) définies par (3.67). Alors on a

$$E[\hat{R}(x) - R(x)]^2 \rightarrow 0. \quad (3.74)$$

iv) Supposons que les conditions du Théorème 3.3.8 soient réunies pour les variables (X_i, Y_i) définies par (3.67). Alors on a

$$EQMI(\hat{R}) \rightarrow 0. \quad (3.75)$$

3.5 Commentaires et compléments bibliographiques

3.5.1 Compléments sur les estimateurs à noyau

La conclusion essentielle à tirer des résultats des paragraphes précédents est que les techniques de prédiction par noyau, tout en étant d'une conception relativement simple, possèdent d'excellentes propriétés mathématiques. En effet, sous un cadre général de dépendance, les théorèmes précédents montrent que ces estimateurs atteignent des vitesses de convergence dont on sait qu'elles sont optimales dans un cadre d'indépendance (*Cf* Paragraphe 2.5), du moins dans des modèles de régularité du type (3.5) ou (3.7). Pour des raisons évidentes de clarté d'exposé et vu la multitude d'articles que l'on peut rencontrer et traitant de ce sujet, nous n'avons pas cherché à rendre les hypothèses de nos théorèmes les plus générales possibles. Concernant la nature même des résultats que nous donnons, le lecteur peut se faire une idée de l'évolution des recherches en consultant quelques articles relatifs à des périodes différentes. Commençons par citer les articles précurseurs de Robinson (1983) et Colomb (1984), et celui plus ancien de Rosenblatt (1971) (bien que dans un contexte différent d'estimation de densité). Ensuite, par exemple pour ce qui concerne les résultats de type L_2 , on pourra par exemple se limiter aux résultats initiaux de Vieu (1991) puis voir comment ceux-ci ont été améliorés par Kim et Cox (1995) ou Kim (1997) avant d'en venir à la version générale que nous exposons ici et qui est principalement issue sous cette forme de Bosq (1996). On pourra constater que cette évolution s'est faite en parallèle avec l'amélioration des inégalités de covariance (*i.e.* des inégalités du type de celle du Lemme 3.2.2). Pour ce qui concerne les résultats de type L_∞ , les premiers résultats en mélange uniforme font datent de Györfi *et al* (1989). Ces résultats ont subi plusieurs améliorations, comme par exemple dans Roussas (1990), Truong et Stone (1992), Tran (1993) ou Truong (1994), avant d'en venir à leur version la plus récente qui est présentée ici (voir aussi Bosq, 1986). Là aussi l'évolution de la qualité des résultats est liée à celle des outils probabilistes et en particulier à celle des inégalités exponentielles (*i.e.* des inégalités du type de celle du Lemme 3.2.1).

Il faut souligner que plusieurs de nos hypothèses peuvent être allégées. Nous renvoyons à Sarda et Vieu (1986), Diebolt et Laïb (1993), ou Bosq (1996, p.69) pour

l'exposé d'une technique de troncation qui permet de réduire l'hypothèse (3.22) en celle d'existence de moment (absolu ou exponentiel) sur la variable à prédire. Signalons aussi qu'il est possible d'alléger les conditions sur le taux algébrique de mélange a , comme dans Bosq (1987), mais en contrepartie on est amené à faire des hypothèses supplémentaires sur la loi des couples (X_i, X_j) (voir à cet effet l'Exercice 3.1). Signalons aussi que des résultats similaires aux nôtres peuvent être établis sous des conditions plus fortes (de type mélange géométrique) ; la perte en terme de généralité de l'hypothèse de dépendance est dans ce cas compensée par une rapidité des démonstrations (voir à cet effet l'Exercice 3.2). En fait nos théorèmes pourraient être prouvés sans préciser la forme du mélange (c'est à dire uniquement sous l'hypothèse générale (3.11)), mais cela nécessiterait des écritures lourdes du type de (3.24), afin de lier la largeur de fenêtre h et les coefficients $\alpha(n)$. Mentionnons aussi que la condition d'existence d'une densité commune aux variables Z_i peut être légèrement relâchée, comme l'explique Collomb (1984) dans un cadre de dépendance plus restrictif que le notre (mais nous pensons que les idées de Collomb sur ce point pourraient être utilisées aussi dans notre contexte de mélange fort). Pour terminer, mentionnons que l'on peut modéliser la dépendance avec des conditions moins restrictives que le α -mélange, avec des résultats moins performants en contrepartie sur les estimateurs. Citons à cet égard la notion de 2 - α -mélange introduite par Bosq (1996) et qui est légèrement plus générale que celle étudiée dans notre ouvrage, et sous lesquelles nos résultats doivent s'étendre aisément au prix d'un alourdissement sensible des notations et des calculs. Pour des modes de dépendance plus différents, citons les travaux de Yakowitz (1989) dans un cadre markovien non mélangeant, ceux de Delecroix (1996) et Laïb et Ould-Saïd (1996) dans un cadre ergodique, ceux de Hall et Hart (1990) pour les processus à longue mémoire, et enfin ceux de Bosq (1995), Maes (1999) et Lardjane (2000) pour des processus approximables et/ou des systèmes dynamiques. Toutefois il faut noter que le gain en terme de généralité du modèle de dépendance qui accompagne ces derniers travaux se solde naturellement par une perte sur la force des résultats concernant les estimateurs non-paramétriques.

Pour ce qui concerne les conditions de régularité que nous imposons sur les fonctions à estimer, r , f ou R , celles-ci peuvent évidemment être modifiées sans que la structure de dépendance ne pose le moindre problème, puisque ces diverses formes de régularité n'ont un effet que sur les termes non stochastiques de biais (voir à cet effet l'Exercice 3.3). Nous souhaitons aussi mentionner que des résultats analogues peuvent être obtenus dans un cadre de fonction r ou R discontinue (Cf. Couallier (2000)). Concernant les questions de choix optimal du paramètre de lissage, problème important mais que nous n'avons pas abordé ici, nous nous contenterons de renvoyer à quelques articles de base ainsi qu'aux articles bibliographiques et/ou

comparatifs de Chu et Marron (1991a), Hart (1996) et Quintela del Rio (1996). Il faut noter que dans le cadre de dépendance, l'éventail de techniques disponibles pour choisir le paramètre de lissage est moins large que dans le cadre usuel d'échantillon indépendant (*Cf* Paragraphe 2.6.1). Ces techniques se limitent essentiellement aux techniques de validation croisée (voir Härdle et Vieu, 1992, pour un résultat du type (2.5.1) sous mélange fort) ou à celles d'injection (voir Quintela del Rio, 1994), alors que les propriétés des méthodes de rééchantillonnage sont encore peu connues dans un cadre de dépendance (voir cependant Politis et Romano, 1993, 1994a et 1994b).

3.5.2 A propos des outils probabilistes

Le lecteur intéressé par une vision globale des outils probabilistes qui peuvent s'avérer utiles en estimation fonctionnelle et plus particulièrement en estimation par noyau pourra, outre les ouvrages généraux de Yoshihara (1992), (1993) et (1994), Doukhan (1994), Bosq (1996) et Rio (2000) déjà abondamment cités précédemment, se rapporter aux références suivantes. Par exemple, le comportement asymptotique de sommes de variables dépendantes est décrit, sous divers types de dépendance, dans Yoshihara (1978), Berkes et Philipp (1979), Yokoyama (1980), Philipp (1982), Peligrad (1985b), Roussas et Ioannides (1987), Lin (1993) et Kim (1993) et (1994). Des résultats relatifs au théorème de limite centrale sous dépendance peuvent par exemple être trouvés dans Basu (1985), Dehling *et al.* (1986), Glendinning (1988) et Peligrad (1985a), (1986), (1987) et Doukhan *et al.* (1994). Des résultats relatifs au comportement asymptotique de U -statistiques dépendantes se trouvent dans Yoshihara (1976), Lee (1990) et Kashimov (1993). Des résultats asymptotiques sur des tableaux triangulaires de variables dépendantes sont présentés dans Samur (1984). Des inégalités de covariance sont edonnées dans Rio (1993) et (1994).

3.5.3 Utilisation des régresseurs à noyau en choix de modèles

En s'inspirant de la démarche générale évoquée au Paragraphe 2.6.2 dans un contexte univarié et qui sera reprise dans un contexte multidimensionnel au Paragraphe 4.5.5, plusieurs auteurs ont cherché à utiliser les prédicteurs non-paramétriques étudiés ci-dessus afin de valider un modèle particulier (paramétrique ou non). Les références les plus récentes dans ce domaine sont celles de González Manteiga et Vilar Fernández (1987) et (1994), Vieu (1995), Hjellvik et Tjostheim (1995), Hidalgo (1999), Fan et Li (1999), Camlong (2000), Vilar Fernández et González Manteiga (2000) ou González Manteiga *et al.* (2000).

3.5.4 Utilisation des techniques de noyau dans d'autres cadres

Il faut aussi souligner que d'autres problèmes d'estimation fonctionnelle liées à des séries chronologiques peuvent être traités par des techniques de noyau similaires à celles que nous avons évoquées ici en régression. Pour chacun de ces problèmes nous nous limiterons à quelques références qui nous semblent essentielles en insistant à chaque fois sur celle qui, à notre connaissance, est la plus récente. Mentionnons tout d'abord que les idées précédemment évoquées dans un cadre de régression à partir de variable explicative aléatoire peuvent être adaptées au cas de variable explicative certaine (*Cf eg.* Roussas *et al.*, 1992). Concernant l'estimateur de densité (2.108) nous renvoyons à Bosq (1996) pour un ouvrage général ainsi qu'à Liebscher (1996) et Ango-Nzé et Rios (2000) pour les choses les plus récentes (voir aussi l'Exercice 3.4). Pour ce qui concerne l'estimateur de la fonction de hasard (2.110) nous renvoyons à Sarda et Vieu (1989) et Izenman et Tran (1990) pour des articles précurseurs, ainsi qu'à Estévez et Quintela del Rio (1999) pour des résultats et références récents. Pour ce qui concerne l'estimateur de la fonction de répartition (2.109) nous renvoyons à Cai et Roussas (1998 et 1999) et Sarda et Vieu (1989) pour des articles précurseurs, ainsi qu'à Estévez (2000) pour des résultats et références plus récents. Citons aussi les travaux de Youndjé (1993) en estimation de densité conditionnelle, ceux de Robinson (1994), Delgado et Robinson (1996), Rachdi (1998b) et (1998c) et Rachdi et Sabre (1998) en estimation de densité spectrale, ceux de Yakowitz (1979) et (1985) et de Doukhan (1983) relatifs à estimation de la transition de probabilité d'une chaîne markovienne et les travaux de Ould-Saïd (1993), Quintela del Rio et Vieu (1997) et Matzner-Lober (1998) concernant l'estimation de mode et/ou quantile conditionnel. Mentionnons aussi que ces techniques à noyau peuvent être utilisées en présence de plusieurs séries temporelles, comme par exemple dans le cadre de données longitudinales (*Cf eg.* Ferreira *et al.*, 1997, ou Nunez-Antón *et al.*, 1999). Enfin, mentionnons que les techniques à noyau peuvent aussi être utilisées dans des situations où les variables d'intérêt ne sont pas directement observées mais plutôt entâchées d'erreurs (Masry, 1993).

3.5.5 Alternatives aux techniques de noyau en prévision non-paramétrique

En guise de conclusion de ce Chapitre 3, nous souhaitons souligner que la prévision non-paramétrique est un domaine de la Statistique qui a été particulièrement étudié ces dernières années. Le lecteur pourra consulter la compilation de Franke *et al.* (1984) pour avoir une idée assez précise des choses à leur début. Ensuite, outre les ouvrages déjà cités de Györfi *et al.* (1989) et Bosq (1996), le lecteur pourra aussi consulter celui de Yoshihara (1994) ainsi que les récentes revues bibliographiques

de Tjostheim (1994) et de Härdle *et al.* (1997). Les aspects d'implémentation sont discutés par exemple dans Carbon et Delecroix (1993), Chen et Hafner (1995) et Matzner-Lober *et al.* (1998). Il faut mentionner aussi que d'autres techniques de lissage peuvent être utilisées en séries temporelles pour estimer R . C'est le cas par exemple des Splines (voir par exemple Burman (1991) ainsi que l'ouvrage général de Kohn, *et al.*, 2000), ou bien des estimateurs de type plus proches voisins (Yakowitz, 1987, Guerre, 2000), ou bien des ondelettes (voir par exemple les résultats récents de Masry 2000), ou bien des L -estimateurs (Boente et Fraiman, 1994) ou bien des M -estimateurs (Boente et Fraiman, 1989b, Truong, 1992), ou bien des polynômes locaux (Masry, 1991a et 1991b, Masry et Fan, 1997, Rios, 1997).

3.6 Exercices

Sauf mention explicite du contraire, les définitions et notations utilisées dans les énoncés de ces exercices sont identiques à celles utilisées dans tout le chapitre.

Exercice 3.1 *Supposons que la densité f_{ij} du couple (X_i, X_j) et la densité $f = f_i$ de X_i existent et posons $g_{ij} = f_{ij} - f_i f_j$. Supposons que*

$$|g_{ij}(u) - g_{ij}(v)| \leq C \|u - v\|, \forall (u, v) \in \mathbb{R}^2 \otimes \mathbb{R}^2,$$

et reprenons les notations et conditions du Corollaire 3.3.3.

1. *Montrer que*

$$\text{cov}(\Delta_i, \Delta_j) = O(h^4).$$

2. *En déduire que le résultat de ce Corollaire 3.3.3 reste vrai en allégeant l'hypothèse sur a . Préciser cette nouvelle hypothèse.*

Exercice 3.2 1. *Démontrer que sous les conditions de la Proposition 3.3.1 et pour un mélange géométrique (i.e. pour un mélange du type de (3.13)), on a*

$$s_n^2 = o(nh).$$

2. *Retrouver ainsi le résultat du Théorème 3.3.5 sous l'hypothèse de mélange géométrique.*

Exercice 3.3 On se place sous les conditions (et avec les mêmes notations) que dans les Théorèmes 2.3.5 et 2.3.6. On suppose que les couples (X_i, Y_i) sont mélangeants et qu'ils vérifient (3.12) avec

$$a > \max(2\beta + 2, 3 + 2/\beta).$$

On suppose l'existence d'une densité f_{ij} à chaque couple (X_i, X_j) , $i \neq j$, et on choisit une largeur de fenêtre de la forme

$$h = C \left(\frac{1}{\log n} \right)^{-\frac{1}{2\beta+1}}.$$

Montrer que les résultats des Théorèmes 2.3.5 et 2.3.6 restent valables. Justifier a posteriori cette condition sur h . (Indication : on pourra reprendre la démarche de l'Exercice 2.5.)

Exercice 3.4 Cet exercice a pour but d'établir des propriétés asymptotiques sous dépendance pour l'estimateur de la densité f de X défini par :

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

On se place sous les hypothèses du Théorème 2.4.1 mais avec k entier quelconque strictement positif. On considère la structure de dépendance générale définie par (3.11), sans aucune restriction sur la forme de la décroissance des coefficients de mélange, et on suppose l'existence d'une densité jointe f_{ij} à chaque couple (X_i, X_j) , $i \neq j$.

1. Montrer que

$$\text{var}(\hat{f}(x)) = \frac{1}{nh} \int_{\mathbb{R}} K^2(t) dt + o\left(\frac{1}{nh}\right) + o\left(\alpha\left(\frac{h}{\log n}\right)\right).$$

2. Donner un (ou plusieurs) ensemble d'hypothèses sur h et α permettant de montrer que l'estimateur $\hat{f}(x)$ atteint des vitesses optimales de convergence en moyenne quadratique, c'est à dire des vitesses du type :

$$E \left[\hat{f}(x) - f(x) \right]^2 = O(n^{-\frac{2k}{2k+1}}).$$

latex

Chapitre 4

Régression non-paramétrique vectorielle

4.1 Introduction

Nous nous plaçons dans ce chapitre dans le cadre de l'estimation de la fonction de régression

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

d'une variable réelle Y sur une variable X à valeurs dans \mathbb{R}^p , p étant un entier strictement positif. L'échantillon $\{(X_i, Y_i), i = 1, \dots, n\}$ est constitué de couples indépendants et ayant chacun même loi que (X, Y) . On notera f la densité marginale par rapport à la mesure de Lebesgues sur \mathbb{R}^p de la variable explicative X , (densité supposée exister). Dans tout ce chapitre, la norme que nous utiliserons sur \mathbb{R}^p est la norme euclidienne usuelle :

$$\|x\| = \sqrt{\sum_{j=1}^p (x_j^2)}.$$

Les modèles qui vont nous intéresser sont de type non-paramétrique. On a vu au cours du Chapitre 2 que l'estimateur à noyau défini par (2.4) possède de bonnes propriétés mathématiques dans le cas univarié. Ainsi il est naturel de le généraliser

de la manière suivante :

$$\hat{r}_{NW}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)}, \quad \forall x \in \mathbb{R}^p. \quad (4.1)$$

Dans cette définition K est une fonction de \mathbb{R}^p dans \mathbb{R} , $h = h_n$ est toujours un paramètre réel strictement positif, et on adopte la convention $0/0 = 0$.

Les démonstrations effectuées dans les chapitres précédents dans le cadre univarié ont fait apparaître l'intérêt de manipuler cet estimateur en considérant séparément les propriétés de convergence de son numérateur et celles de son dénominateur. C'est la raison pour laquelle, nous réécrivons l'estimateur précédent sous la forme :

$$\hat{r}_{NW}(x) = \frac{\hat{g}(x)}{\hat{f}(x)}, \quad (4.2)$$

avec

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

et

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

Comme dans le cas univarié, le dénominateur sera vu comme un estimateur de la densité f de X , tandis que le numérateur sera vu comme un estimateur de la fonction $g = rf$.

Ce chapitre est organisé de la manière suivante. Nous donnons tout d'abord des résultats de convergence presque complète (avec ou sans vitesse selon les modèles de régularité considérés) concernant l'estimateur (4.2), au Paragraphe 4.2.2 pour une estimation ponctuelle de r en un point fixé puis au Paragraphe 4.2.3 pour l'estimation uniforme de r sur un compact. Ensuite nous établirons des résultats analogues en terme de convergence en moyenne quadratique ponctuelle au Paragraphe 4.3.1 et en terme d'erreurs quadratiques moyennes intégrées au Paragraphe 4.3.2. Tous ces résultats sont des extensions au cadre multidimensionnel de ceux décrits au Chapitre 2. Ces résultats mettent en évidence deux phénomènes essentiels. Il s'agit tout d'abord le rôle du paramètre de lissage dont le bon choix va permettre aux estimateurs à noyau définis en (4.2) d'atteindre des vitesses de convergence optimales, puis du problème de la dimension p dont on verra qu'elle dégrade sensiblement ces vitesses de convergence. Le Paragraphe 4.4.1 discute du choix optimal de fenêtre et le Paragraphe 4.4.2 mettra en évidence ce fléau des grandes dimensions. Le Paragraphe

4.5 de conclusion décrira brièvement quelques pistes de solution à ce problème de dimension, pistes qui se traduisent naturellement avant tout par la construction de nouveaux modèles non-paramétriques. Le chapitre se terminera avec une série d'exercices qui permettront au lecteur de se familiariser avec toutes les notions vues précédemment et qui auront aussi pour objectif d'amener à la démonstration de certains résultats complémentaires qu'il était impossible de présenter de manière exhaustive dans le texte.

4.2 Convergence presque complète

4.2.1 Hypothèses générales et notations

Nous allons commencer par donner des résultats de convergence presque complète (*Cf.* Exercice 2.3 pour des détails sur cette notion) sous des modèles de régularité du type

$$r \text{ et } f \text{ sont } k \text{ fois continûment différentiables autour de } x, \quad (4.3)$$

x étant un point fixé de \mathbb{R}^p pour lequel on suppose que

$$f(x) > 0. \quad (4.4)$$

Les hypothèses sur la largeur de fenêtre seront les suivantes

$$\lim_{n \rightarrow \infty} h = 0 \text{ et } \lim_{n \rightarrow \infty} \frac{nh^p}{\log n} = \infty. \quad (4.5)$$

Pour ce qui concerne le noyau, nous supposerons que

$$K \text{ est borné, intégrable et à support compact.} \quad (4.6)$$

Pour préciser certaines vitesses de convergence, nous aurons besoin de la notion d'ordre d'un noyau multivarié. Cette notion, qui généralise celle de Gasser et Müller (1979) que nous avons rappelée en (2.9), fût introduite sous la forme de la Définition 4.2.1 ci-dessous dans Vieu (1991). Nous renvoyons à l'Exercice 4.1 pour un exemple de noyau vérifiant ce type de conditions. Nous introduisons les notations suivantes. On pose, pour tout p -uplet d'entiers positifs (i_1, \dots, i_p)

$$T_K(i_1, \dots, i_p) = \int_{\mathbb{R}^p} u_1^{i_1} \dots u_p^{i_p} K(u_1, \dots, u_p) du_1 \dots du_p, \quad (4.7)$$

et si k est fixé, $k \in \mathbb{N}^*$, on pose pour tout $j \in \{1, \dots, p\}$

$$T_K(j) = \int_{\mathbb{R}^p} u_j^k K(u_1, \dots, u_p) du_1 \dots du_p. \quad (4.8)$$

Définition 4.2.1 Nous dirons qu'une fonction K de \mathbb{R}^p dans \mathbb{R} est un noyau d'ordre k , $k \in \mathbb{N}^*$, lorsque :

$$\forall (i_1, \dots, i_p) \in \mathbb{N}^{*p}, \quad (\forall j, i_j < k) \Rightarrow (T_K(i_1, \dots, i_p) = 0), \quad (4.9)$$

$$\forall (i_1, \dots, i_p) \in \mathbb{N}^{*p}, \quad \forall j, T_K(j) \in \mathbb{R}^*. \quad (4.10)$$

Nous ferons donc parfois l'hypothèse que

$$K \text{ est un noyau d'ordre } k. \quad (4.11)$$

Finalement, comme dans le cas univarié, la condition suivante sera introduite afin d'alléger nos démonstrations :

$$|Y| < M < \infty. \quad (4.12)$$

4.2.2 Convergence presque complète ponctuelle

Notre premier résultat est une extension au cadre multidimensionnel du Théorème 2.3.1.

Théorème 4.2.1 Vitesse de convergence presque complète ponctuelle multivariée sous condition de dérivabilité. *Considérons le modèle (4.3) avec $k > 0$ et supposons que les conditions (4.4), (4.5), (4.6), (4.11) et (4.12) soient réalisées. On a*

$$\hat{r}_{NW}(x) - r(x) = O(h^k) + O\left(\sqrt{\frac{\log n}{nh^p}}\right), \text{ p.co.} \quad (4.13)$$

Preuve du Théorème 4.2.1 - La preuve est très voisine de celle du Théorème 2.3.1. Nous serons donc relativement brefs en nous contentant d'insister sur les parties de la démonstration pour lesquelles l'aspect multidimensionnel intervient de manière significative. Sans perte de généralité on considère que K est unitaire (*i.e.* K est d'intégrale 1). La décomposition (2.13) restant toujours valable, le résultat (4.13) sera prouvé dès que seront vérifiées les 5 propriétés suivantes :

$$E\hat{g}(x) - g(x) = O(h^k), \quad (4.14)$$

$$Ef(x) - f(x) = O(h^k), \quad (4.15)$$

$$E\hat{g}(x) - \hat{g}(x) = O\left(\sqrt{\frac{\log n}{nh^p}}\right), p.co., \quad (4.16)$$

$$E\hat{f}(x) - \hat{f}(x) = O\left(\sqrt{\frac{\log n}{nh^p}}\right), p.co., \quad (4.17)$$

et

$$\exists \delta > 0, \sum_{n=1}^{\infty} P[\hat{f}(x) \leq \delta] < \infty. \quad (4.18)$$

- Preuve de (4.14). En reprenant les calculs effectués pour prouver (2.14), on arrive en utilisant successivement l'équidistribution des couples (X_i, Y_i) , puis un conditionnement en X_i , puis en posant $z = (x - u)/h$ dans le calcul d'intégrale, puis en utilisant que K est unitaire, à l'expression :

$$E\hat{g}(x) - g(x) = \int (g(x - zh) - g(x)) K(z) dz. \quad (4.19)$$

Il suffit alors de développer la fonction g au voisinage de x , ce qui est possible au vu de la condition (4.3). Ceci s'écrit :

$$\begin{aligned} & g(x - zh) - g(x) \\ &= \\ & \sum_{j=1}^k \frac{(-h)^j}{j!} \sum_{i_1 + \dots + i_p = j} \left(z_1^{i_1}, \dots, z_p^{i_p} \left[\frac{\partial^j g}{\partial x_1^{i_1} \dots \partial x_p^{i_p}} \right] (x) + o(h^j) \right). \end{aligned} \quad (4.20)$$

Puisque K est à support compact, les $o(h^j)$ intervenant dans l'expression ci-dessus sont uniformes en $z = (z_1, \dots, z_p)$. Ainsi, on tire de (4.11), (4.19) et (4.20) que :

$$E\hat{g}(x) - g(x) = \frac{(-h)^k}{k!} \sum_{j=1}^p \left[\frac{\partial^k g}{\partial x_j^k} \right] (x) T_K(j) + o(h^k). \quad (4.21)$$

Ceci achève la preuve de (4.14).

- Preuve de (4.15). La démonstration suit les mêmes étapes et est donc par conséquent omise (*Cf.* Exercice 4.3). En fait on établit un résultat plus précis que celui recherché, à savoir que

$$E\hat{f}(x) - f(x) = \frac{(-h)^k}{k!} \sum_{j=1}^p \left[\frac{\partial^k f}{\partial x_j^k} \right] (x) T_K(j) + o(h^k). \quad (4.22)$$

- Preuve de (4.16). L'idée est d'appliquer le Lemme 2.3.1 aux variables

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

Pour cela il est nécessaire de trouver des majorants pour $|\Delta_i|$ et pour $E\Delta_i^2$. En reprenant les calculs ayant amené à (2.22) et (2.24), on arrive sans difficulté à

$$|\Delta_i| \leq \frac{C}{h^p}, \quad (4.23)$$

et à

$$E\Delta_i^2 \leq \frac{C}{h^p}. \quad (4.24)$$

Le Lemme 2.3.1 permet alors d'écrire que pour ϵ suffisamment petit on a

$$P[|E\hat{g}(x) - \hat{g}(x)| > \epsilon] \leq 2e^{-\frac{n\epsilon^2 h^p}{4C}}. \quad (4.25)$$

En prenant

$$\epsilon = \epsilon_0 \sqrt{\frac{\log n}{nh^p}},$$

on arrive pour tout ϵ_0 à

$$P \left[|E\hat{g}(x) - \hat{g}(x)| > \epsilon_0 \sqrt{\frac{\log n}{nh^p}} \right] \leq 2n^{-C\epsilon_0^2}. \quad (4.26)$$

On peut choisir $\epsilon_0 > 1/(\sqrt{C})$ et terminer ainsi la preuve de (4.16).

- Preuve de (4.17). La démonstration suit les mêmes étapes et est donc par conséquent omise (*Cf.* Exercice 4.3).

- Preuve de (4.18). La convergence presque complète de $\hat{f}(x)$ vers $f(x)$ découle de (4.15) et (4.17). Donc pour tout $\epsilon > 0$ on a :

$$\sum_{n=1}^{\infty} P \left[\left| \hat{f}(x) - f(x) \right| > \epsilon \right] < \infty,$$

et l'on peut alors terminer la preuve de (4.18) en prenant $\delta = \epsilon = f(x)/2$. \square

Lorsque l'on change les conditions de régularité sur les fonctions r et f on obtient des résultats analogues à celui du Théorème 4.2.1 qui précède mais avec des

modifications sur la vitesse de convergence vers 0 du biais. Ces modifications ayant été amplement décrites dans le cadre univarié, nous donnons les deux théorèmes suivants sans démonstration. Ces démonstrations sont renvoyées aux Exercices 4.4 et 4.5. Considérons le modèle défini par la condition suivante, où $\phi = r$ ou f :

$$\exists \beta > 0, \exists C < \infty, \exists \epsilon > 0, \forall y \in]x - \epsilon, x + \epsilon[, |\phi(x) - \phi(y)| \leq C \|x - y\|^\beta. \quad (4.27)$$

Théorème 4.2.2 Vitesse de convergence presque complète ponctuelle multivariée sous condition de Lipschitz. *Considérons le modèle défini par la condition (4.27), et supposons que les conditions (4.4), (4.5), (4.6) et (4.12) soient réalisées. On a*

$$|\hat{r}_{NW}(x) - r(x)| = O(h^\beta) + O\left(\sqrt{\frac{\log n}{nh^p}}\right), \text{ p.co.} \quad (4.28)$$

Théorème 4.2.3 Convergence presque complète ponctuelle multivariée sous condition de continuité. *Considérons le modèle (4.3) avec $k = 0$ et supposons que les conditions (4.4), (4.5), (4.6) et (4.12) soient réalisées. On a*

$$|\hat{r}_{NW}(x) - r(x)| \rightarrow 0, \text{ p.co.} \quad (4.29)$$

4.2.3 Convergence presque complète uniforme

Nous allons maintenant donner des versions uniformes sur un compact des résultats précédents. Nous considérerons des modèles non-paramétriques du type

$$r \text{ et } f \text{ sont } k \text{ fois continûment différentiables autour de } S, \quad (4.30)$$

et

$$\inf_{x \in S} f(x) > \theta. \quad (4.31)$$

Pour le reste, les conditions dont nous aurons besoin sont les mêmes que celles nécessitées pour l'obtention des résultats ponctuels ci-dessus, auxquelles s'ajoute la restriction suivante sur le noyau

$$\exists \beta > 0, \exists C < \infty, \forall x \in S, \forall y \in S, |K(x) - K(y)| \leq C \|x - y\|^\beta. \quad (4.32)$$

Comme nos démonstrations suivent la même démarche que dans le cas univarié, elles seront laissées à titre d'exercice.

Théorème 4.2.4 Vitesse de convergence presque complète uniforme multivariée sous condition de dérivabilité. *Considérons le modèle (4.30) avec $k > 0$ et supposons que les conditions (4.5), (4.6), (4.11), (4.12), (4.31) et (4.32) soient réalisées. On a*

$$\sup_{x \in S} |\hat{r}_{NW}(x) - r(x)| = O(h^k) + O\left(\sqrt{\frac{\log n}{nh^d}}\right), \text{ p.co.} \quad (4.33)$$

Preuve du Théorème 4.2.4. Voir Exercice 4.6. \square

Théorème 4.2.5 Convergence uniforme presque complète multivariée sous condition de continuité. *Considérons le modèle (4.30) avec $k = 0$ et supposons que les conditions (4.5), (4.6), (4.12), (4.31) et (4.32) soient réalisées. On a*

$$\sup_{x \in S} |\hat{r}_{NW}(x) - r(x)| \rightarrow 0, \text{ p.co.} \quad (4.34)$$

Preuve du Théorème 4.2.5. Voir Exercice 4.7. \square

Le dernier résultat est établi sous un modèle de régularité Lipschitzien, plus précisément sous la condition

$$\exists \beta > 0, \exists C < \infty, \forall x \in S, \forall y \in S, |\phi(x) - \phi(y)| \leq C \|x - y\|^\beta, \quad (4.35)$$

où ϕ désigne indifféremment f ou r .

Théorème 4.2.6 Vitesse de convergence presque complète uniforme multivariée sous condition de Lipschitz. *Considérons le modèle (4.35) et supposons que les conditions (4.5), (4.6), (4.12), (4.31) et (4.32) soient réalisées. On a*

$$\sup_{x \in S} |\hat{r}_{NW}(x) - r(x)| = O(h^\beta) + O\left(\sqrt{\frac{\log n}{nh^d}}\right). \text{ p.co.} \quad (4.36)$$

Preuve du Théorème 4.2.6. Voir Exercice 4.8. \square

4.3 Convergence en moyenne quadratique

4.3.1 Résultats ponctuels

Nous allons commencer par quelques résultats de convergence de type L_2 en un point fixé x de \mathbb{R}^p . Les quelques modifications mineures que nous sommes amenés à faire sur les hypothèses sont du même ordre que celles discutées dans le cas unidimensionnel au Paragraphe 2.4. Tout d'abord, la condition sur le paramètre de lissage peut être rendue légèrement moins restrictive en ce sens qu'on peut remplacer (4.5) par :

$$\lim_{n \rightarrow \infty} h = 0 \text{ et } \lim_{n \rightarrow \infty} nh^p = \infty. \quad (4.37)$$

Nous aurons aussi parfois besoin de rajouter la condition suivante sur la loi de (X, Y) :

$$\phi(u) = E(Y^2 | X = u) \text{ est continue au point } x. \quad (4.38)$$

Pour des raisons techniques qui seront explicitées ultérieurement, nous aurons besoin de nous restreindre à des noyaux positifs. Or (Cf. Exercice 4.2), cette condition est incompatible avec un noyau d'ordre supérieur à 2. Par conséquent, nous en resterons à un modèle de type (4.3) avec $k = 2$, et les conditions sur le noyau K deviennent alors :

$$K \text{ est borné, intégrable, positif, symétrique et à support compact.} \quad (4.39)$$

Le théorème qui suit est une version multidimensionnelle du Théorème 2.4.1. Comme les deux théorèmes se démontrent de manière analogue, nous ne donnerons que les étapes principales. La preuve détaillée fera l'objet de l'Exercice 4.9.

Théorème 4.3.1 Convergence en moyenne quadratique ponctuelle multivariée sous condition de dérivabilité. *Considérons le modèle (4.3) avec $k = 2$ et supposons que les conditions (4.4), (4.12), (4.37), (4.38) et (4.39) soient réalisées. On a*

$$E[\hat{r}_{NW}(x) - r(x)]^2 = B^2(x)h^4 + V(x)\frac{1}{nh^p} + o(h^4 + \frac{1}{nh^p}), \quad (4.40)$$

où

$$B(x) = \frac{1}{2f(x)} \sum_{j=1}^p T_K(j) \left(\left[\frac{\partial^k g}{\partial x_j^k} \right] (x) - \left[r(x) \frac{\partial^k f}{\partial x_j^k} \right] (x) \right) \quad (4.41)$$

et

$$V(x) = \int_{\mathbb{R}^p} K^2(t) dt \frac{(\phi(x) - r^2(x))}{f(x)}. \quad (4.42)$$

Preuve du Théorème 4.3.1 - On considère comme toujours que K est unitaire (*i.e.* K est d'intégrale 1). Le résultat énoncé va découler des 2 résultats suivants :

$$E\hat{r}_{NW}(x) - r(x) = B(x)h^2 + o(h^2), \quad (4.43)$$

et

$$Var\hat{r}_{NW}(x) = V(x) \frac{1}{nh^p} + o\left(\frac{1}{nh^p}\right). \quad (4.44)$$

- Preuve de (4.43). Concernant ce terme de biais, notons que le résultat (2.73) reste valable dans notre cadre multidimensionnel. En reprenant les preuves de (2.74) et (2.75), et en les adaptant à notre cadre multivarié, on arrive à

$$A_1 = \frac{1}{nh^p} g(x) \int_{\mathbb{R}^p} K^2(t) dt + o\left(\frac{1}{nh^p}\right), \quad (4.45)$$

et

$$A_2 = O\left(\frac{1}{nh^p}\right). \quad (4.46)$$

Finalement, (4.43) découle directement de (2.73), (4.21), (4.22), (4.45) et (4.46).

- Preuve de (4.44). La première chose consiste à établir, en suivant les mêmes arguments que pour démontrer (2.77), que l'on a

$$\begin{aligned} var(\hat{r}_{NW}(x)) &= \frac{var(\hat{g}(x))}{(E\hat{f}(x))^2} - 4 \frac{E\hat{g}(x)cov(\hat{g}(x), \hat{f}(x))}{(E\hat{f}(x))^3} \\ &+ 3var(\hat{f}(x)) \frac{(E\hat{g}(x))^2}{(E\hat{f}(x))^4} + o\left(\frac{1}{nh^p}\right). \end{aligned} \quad (4.47)$$

Ensuite, il suffit de reprendre les preuves de (4.16) et (4.17) et d'utiliser la condition (4.38) qui va permettre de préciser les constantes. Ainsi on arrive à :

$$var(\hat{g}(x)) = \frac{1}{nh^p} f(x)\phi(x) \int_{\mathbb{R}^p} K^2(t) dt + o\left(\frac{1}{nh^p}\right), \quad (4.48)$$

et

$$var(\hat{f}(x)) = \frac{1}{nh^p} f(x) \int_{\mathbb{R}^p} K^2(t) dt + o\left(\frac{1}{nh^p}\right). \quad (4.49)$$

Quant à la covariance, elle se calcule de manière analogue et nous avons

$$cov(\hat{g}(x), \hat{f}(x)) = \frac{1}{nh^p} g(x) \int_{\mathbb{R}^p} K^2(t) dt + o\left(\frac{1}{nh^p}\right). \quad (4.50)$$

La combinaison des résultats (4.48)-(4.50) et de ceux obtenus pour le biais de $\hat{f}(x)$ (Cf. (4.14)) et pour le biais de $\hat{g}(x)$ (Cf. (4.15)), permet de traiter tous les termes intervenant dans l'expression (4.47) et d'aboutir au résultat recherché. \square

Le prochain résultat est établi sous un modèle de régression plus général que celui du Théorème 4.3.1 (*i.e.* pour $k = 0$). Le gain obtenu ainsi en terme de généralité du modèle statistique est à mettre en balance avec la perte des vitesses de convergence. La preuve de ce résultat étant très voisine de celle établie dans le cas univarié (Cf. Théorème 2.4.2), elle est laissée en exercice (Cf. Exercice 4.10).

Théorème 4.3.2 Convergence en moyenne quadratique ponctuelle multivariée sous condition de continuité. *Considérons le modèle (4.3) avec $k = 0$ et supposons que les conditions (4.4), (4.12), (4.37) et (4.39) soient réalisées. On a*

$$E[\hat{r}_{NW}(x) - r(x)]^2 \rightarrow 0. \quad (4.51)$$

4.3.2 Résultats en moyenne quadratique intégrée

Nous allons maintenant donner des versions uniformes sur un compact des deux théorèmes précédents. Pour cela, nous allons nous intéresser aux erreurs quadratiques moyennes intégrées, notées *EQMI* et définies par

$$EQMI(\hat{r}_{NW}) = E \int_{\mathbb{R}^p} (\hat{r}_{NW}(x) - r(x))^2 w(x) dx, \quad (4.52)$$

avec une fonction de poids w vérifiant :

$$w \text{ est positive, bornée et à support compact } S. \quad (4.53)$$

Nous aurons besoin d'une version uniforme de la condition (4.38), à savoir que

$$\phi(u) = E(Y^2|X = u) \text{ est continue autour de } S. \quad (4.54)$$

Théorème 4.3.3 Erreur quadratique moyenne intégrée multivariée sous condition de dérivabilité. *Considérons le modèle (4.30) avec $k = 2$ et supposons que les conditions (4.31), (4.37), (4.39), (4.53) et (4.54) sont réalisées. On a*

$$EQMI(\hat{r}_{NW}) = B^2 h^4 + \frac{V}{nh^p} + o\left(h^4 + \frac{1}{nh^p}\right), \quad (4.55)$$

où

$$B^2 = \int_S B^2(x)w(x)dx \text{ et } V = \int_S V(x)w(x)dx, \quad (4.56)$$

$B(x)$ et $V(x)$ étant définis par (4.41) et (4.42).

Preuve du Théorème 4.3.3 - Il suffit de reprendre les calculs effectués pour prouver le Théorème 4.3.1, en constatant que toutes les approximations faites sont basées sur des propriétés de continuité de f , r , ϕ , f'' ou r'' . Comme toutes ces fonctions sont supposées continues sur S qui est compact, elles sont donc uniformément continues sur S , et par conséquent on peut vérifier que toutes ces approximations sont uniformes sur S . Ainsi le résultat (4.55) est obtenu par simple intégration du résultat (4.40). \square

Pour terminer, nous allons donner un résultat analogue sous le modèle plus général $k = 0$, mais cela se fera au détriment de l'obtention des vitesses de convergence. Très voisine de la preuve précédente, sa démonstration est laissée en exercice (Cf. Exercice 4.10).

Théorème 4.3.4 Erreur quadratique moyenne intégrée multivariée sous condition de continuité. *Considérons le modèle (4.30) avec $k = 0$ et supposons que les conditions (4.31), (4.37), (4.39) et (4.53) soient réalisées. On a*

$$EQMI(\hat{r}_{NW}) \rightarrow 0. \quad (4.57)$$

4.4 Le fléau des grandes dimensions

4.4.1 Vitesses optimales de convergence multivariée

La première des choses à remarquer, est que l'optimisation (en h) des vitesses de convergence établies dans le paragraphe précédent se fait de manière immédiate en procédant comme ce fût fait au Paragraphe 2.5 dans le cadre univarié. En procédant comme pour l'obtention des Corollaires 2.5.1, 2.5.2, 2.5.3 et 2.5.4 on démontre facilement les quatres résultats suivants.

Corollaire 4.4.1 Convergence optimale en moyenne quadratique ponctuelle multivariée. *Supposons que les conditions du Théorème 4.3.1 soient vérifiées. Supposons en outre que la fenêtre h soit de la forme :*

$$h = Cn^{-\frac{1}{4+p}}, \quad 0 < C < \infty. \quad (4.58)$$

Alors on a

$$E[\hat{r}_{NW}(x) - r(x)]^2 = O(n^{-\frac{4}{4+p}}). \quad (4.59)$$

Corollaire 4.4.2 Erreur quadratique moyenne intégrée optimale multivariée. *Supposons que les conditions du Théorème 4.3.3 soient vérifiées ainsi que l'hypothèse (4.58). On a*

$$EQMI(\hat{r}_{NW}) = O(n^{-\frac{4}{4+p}}). \quad (4.60)$$

Corollaire 4.4.3 Convergence presque complète ponctuelle optimale multivariée. *Supposons que les conditions du Théorème 4.2.1 soient vérifiées. Supposons en outre que la fenêtre h soit de la forme :*

$$h = C \left(\frac{n}{\log n} \right)^{-\frac{1}{4+p}}, \quad 0 < C < \infty. \quad (4.61)$$

Alors on a

$$\hat{r}_{NW}(x) - r(x) = O \left(\left(\frac{n}{\log n} \right)^{-\frac{k}{2k+p}} \right), \quad p.co. \quad (4.62)$$

Corollaire 4.4.4 Convergence presque complète uniforme optimale multivariée. *Supposons que les conditions du Théorème 4.2.4 soient vérifiées de même que (4.61). Alors on a*

$$\sup_{x \in S} |\hat{r}_{NW}(x) - r(x)| = O \left(\left(\frac{n}{\log n} \right)^{-\frac{k}{2k+p}} \right), \quad p.co. \quad (4.63)$$

Comme dans le cadre univarié, ces résultats sont particulièrement intéressants d'un point de vue théorique, puisque l'on sait d'après des résultats généraux de Stone (1981) et (1982) que la vitesse optimale de convergence pour un modèle (4.30) est :

$$n^{-\frac{k}{2k+p}}, \quad \text{en norme } L_p, \quad p < \infty, \quad (4.64)$$

et

$$\left(\frac{n}{\log n} \right)^{-\frac{k}{2k+p}}, \quad \text{en norme } L_\infty. \quad (4.65)$$

On a donc l'évidence de ce que les estimateurs à noyau, tout en restant d'une conception initiale relativement simple, atteignent moyennant un bon choix du paramètre de lissage les vitesses de convergence optimales pour des modèles non-paramétriques généraux.

4.4.2 Le fléau de la dimension

Les vitesses de convergence obtenues dans le paragraphe précédent sont optimales pour des modèles de type (4.30), mais elles n'en restent pas moins relativement mauvaises dès que p grandit. Ce problème, connu dans la littérature comme le fléau de la dimension, est lié à la rareté des données dans un espace à grande dimension. Nous renvoyons à Friedman et Stuetzle (1981) et à Huber (1985) pour une discussion détaillée de ce problème.

Le caractère local de l'estimateur \hat{r}_{NW} défini en (4.1) accentue encore l'importance de ce problème de rareté des données dans notre contexte non-paramétrique. En effet, cette rareté des données impose dans la pratique un choix de largeur de fenêtre h d'autant plus grand que p est grand afin de stabiliser la variance de l'estimateur, mais en contrepartie cela conduit à une détérioration du biais. Il est communément admis que dès que p dépasse 2 ou 3, à moins d'avoir à disposition des tailles d'échantillon gigantesques, les modèles non-paramétriques purs du type de ceux étudiés précédemment et définis par des conditions du type (4.3) ou (4.30) ne seraient être pleinement satisfaisants. Le petit exemple traité dans l'Exercice 4.11 permet de bien voir comment cette question de rareté des données prend très vite des proportions inquiétantes même pour des valeurs de p *a priori* petites.

Il est fondamental de noter ici que ce problème n'est absolument pas lié à la technique d'estimation par noyaux que nous utilisons ici, puisque les résultats exposés dans le Paragraphe 4.4.1 qui précède montrent l'optimalité de ces estimateurs. Il s'agit donc d'un problème lié uniquement à la nature des modèles purement non-paramétriques du type (4.3) ou (4.30). Nous renvoyons aux ouvrages récents de Gu (2000) et Nychka (2000) pour une description du comportement des estimateurs de type Spline face à ce problème de dimension, et à celui de Wu (2000) pour ce qui concerne les estimateurs de type polynômes locaux. Le travail récent de Härdle et Müller (2000) fournira de plus amples détails concernant le comportement des estimateurs à noyau multidimensionnels. Notons aussi que ce problème se pose dans d'autres contextes d'estimation fonctionnelle avec la même force qu'en estimation de régression (*Cf.* Scott, 1995 et 2000, pour des problèmes similaires en estimation de densité multivariée).

A partir de ces quelques références récentes que nous venons de mentionner le lecteur doit pouvoir remonter à la grande majorité des travaux traitant de ce problème et qu'il était, vu l'abondance de la littérature sur ce thème, impossible de présenter ici de manière exhaustive.

4.5 Discussion et compléments bibliographiques

4.5.1 Objectifs du paragraphe

L'objet de ce dernier paragraphe est de présenter succinctement la bibliographie relative à quelques possibilités de solution au problème de la dimension évoqué dans le Paragraphe 4.4.2 qui précède. Comme nous l'évoquons ci-dessus, ce problème est avant tout lié à la question de l'estimation d'une fonction multidimensionnelle. En ce sens, les possibilités de solution ne sont à rechercher que dans la construction de modèles de régression moins généraux que les modèles non-paramétriques purs de type (4.3) ou (4.30), sans tomber bien entendu dans l'excès des modèles paramétriques. Il est clair qu'une fois ce problème de modèle résolu, viendra la question de la construction de nouveaux estimateurs adaptés à ces nouveaux modèles.

4.5.2 Le modèle additif

Nous allons accorder une place privilégiée au modèle additif à la fois à cause de sa popularité et de la facilité d'interprétation qu'il offre. Ce modèle, pour lequel les premiers travaux remontent à Stone (1985), est un modèle non-paramétrique caractérisé par l'hypothèse suivante sur la fonction de régression r :

$$r(x^1, \dots, x^p) = \mu + \sum_{j=1}^{j=p} r^j(x^j), \quad (4.66)$$

où μ est un paramètre réel, et où pour raison d'identifiabilité on suppose que

$$\forall j, Er^j(X^j) = 0. \quad (4.67)$$

Il est clair que ce modèle, en contre partie de la perte de généralité qu'il présente par rapport à un modèle nonparamétrique pur de type (4.3) ou (4.30), offre l'avantage de remplacer le problème de l'estimation d'une fonction multidimensionnelle en celui de l'estimation de plusieurs fonctions à valeurs réelles. Sous ce modèle plusieurs estimateurs ont été développés. Ces estimateurs sont essentiellement des adaptations des estimateurs non-paramétriques classiques à ce nouveau cadre additif. Nous renvoyons à Stone et Ku (1986), Stone (1994) ou Gu (1992) pour des estimateurs de type Spline, à Buja *et al.* (1989) pour des estimateurs linéaires, à Opsomer et Ruppert (1997) pour des estimateurs linéaires locaux, à Härdle *et al.* (2000) pour des techniques d'estimation basées sur l'utilisation d'ondelettes et à Rodriguez *et al.* (2000) pour des techniques de vraisemblance locale directionnelle.

Récemment, les techniques à noyau qui sont au centre de notre propos dans cet ouvrage, ont été convenablement modifiées pour pouvoir avoir de bonnes propriétés mathématiques sous les modèles additifs. La construction des estimateurs à noyau adaptés à la situation additive se fait au travers d'une technique d'intégration marginale. Cette technique, initiée par Tjostheim et Auestad (1994), Linton et Nielsen (1995) et Linton et Härdle (1996) a été abondamment étudiée ces dernières années et nous renvoyons aux travaux récents de Sperlich *et al.* (1999), Härdle et Müller (2000) et de Camlong (2000) ainsi qu'aux références qu'ils mentionnent. Cette technique sera présentée dans un cadre général lors de l'Exercice 4.12.

En parcourant les références précédemment citées le lecteur pourra constater que la plupart de ces estimateurs peuvent, sous réserves de conditions de régularité sur les composantes fonctionnelles r^j , converger avec des vitesses de convergence indépendante de p . Ceci confirme l'intérêt du modèle additif (4.66) dans les problèmes multivariés puisqu'il existe des estimateurs qui, sous ce modèle, sont insensibles au fléau de la dimension.

4.5.3 Extensions du modèle additif

Il faut noter que plusieurs extensions de ce modèle additif ont été proposées dans la littérature. Citons par exemple le Modèle Additif Généralisé, dont le lecteur trouvera une présentation détaillée dans Hastie et Tibshirani (1986) et (1990), et qui est défini à partir d'une fonction de lien connue L par :

$$r(x^1, \dots, x^p) = L \left(\mu + \sum_{j=1}^{j=p} r^j(x^j) \right), \quad (4.68)$$

où μ et r^j vérifient toujours la condition d'identifiabilité (4.67). Le lecteur trouvera des résultats récents concernant les techniques d'estimation (essentiellement de type Spline) actuellement connues sous ce type de modèles dans Burman (1990) ou Stone (1994).

Ce modèle peut aussi être étendu afin de prendre en compte d'éventuelles interactions entre les variables. On parle alors de Modèle Additif Interactif (Buja *et al.*, 1989). Ce modèle est défini, toujours à partir d'une fonction de lien connue L , en posant

$$r(x^1, \dots, x^p) = L \left(\mu + \sum_{s \in \mathcal{S}} r^s(\bar{x}^s) \right), \quad (4.69)$$

où \mathcal{S} est l'ensemble de toutes les parties de $\{1, \dots, p\}$ et où \overline{X}^s est la sous-famille correspondante parmi les p variables explicatives X^1, \dots, X^p . Chaque fonction r^s est définie sur $\mathbb{R}^{Card(s)}$ et est supposée vérifier la condition d'identifiabilité

$$\forall j, Er^s(\overline{X}^s) = 0. \quad (4.70)$$

Les principaux travaux traitant de ce genre de modèle sont ceux de Hastie et Tibshirani (1987), Andrews et Whang (1990), Chen (1991), Stone (1994) et Linton et Härdle (1996).

4.5.4 Autres modèles à réduction de dimension

Une autre façon de généraliser le modèle (4.68) est de travailler avec une fonction de lien inconnue. Suivant l'approche de Breiman et Friedman (1985), ce type de modèle est appelé Modèle à Transformations Optimales et il s'écrit sous la forme :

$$E(T(Y)|X = (x^1, \dots, x^p)) = \sum_{j=1}^{j=p} r^j(x^j), \quad (4.71)$$

où T est une fonction de \mathbb{R} dans \mathbb{R} inconnue et où les composantes additives r^j sont supposées vérifier (4.67). Les techniques d'estimation sous ce modèle sont essentiellement des techniques itératives (*Cf. eg* Buja *et al.* 1989). Les résultats les plus récents dans ce cadre sont donnés par Burman (1990b) et (1991b) et par Horowitz (1998). Il est clair qu'un tel modèle peut aussi se généraliser pour faire intervenir des composantes interactives du type de celles définies en (4.69) et (4.70). On peut aussi envisager, comme dans Rodriguez *et al.* (2000) des modèles où certaines des composantes r^j sont de nature paramétrique.

Une autre façon de construire des modèles insensibles aux effets de dimension, consiste à supposer que l'influence de la variable p -dimensionnelle X sur la variable Y ne se fait qu'au travers d'un sous espace de dimension $q \leq p$. On parle alors de modèle à projections révélatrices, plus connu sous le nom de Projection Pursuit Regression (PPR) model. Ce modèle s'écrit sous la forme :

$$r(x^1, \dots, x^p) = \mu + \sum_{k=1}^{k=q} r^k({}^t\alpha_k \mathbf{X}), \quad (4.72)$$

où μ est un paramètre réel, où les α_k sont des vecteurs inconnus de \mathbb{R}^p , où les composantes additives r^k sont supposées vérifier les conditions d'identifiabilité suivantes :

$$\forall j, Er^k({}^t\alpha_k \mathbf{X}) = 0, \quad (4.73)$$

et où \mathbf{x} et \mathbf{X} désignent les vecteurs colonne (x^1, \dots, x^p) et (X^1, \dots, X^p) . Diacanis et Shashahani (1984) et H. Chen (1991) discutent précisément des conditions nécessaires à l'identifiabilité de ce modèle, tandis que des techniques d'estimation sont proposées par Friedman (1984), H. Chen (1991), Donoho et Jonstone (1989), Hall (1989) et Hall et Li (1993). Un cas particulier de ce modèle, très utilisé par les économètres pour ses facilités d'interprétation, est le cas du modèle à effet simple (Single Index Model) qui s'écrit, avec les notations précédentes, comme :

$$r(x^1, \dots, x^p) = \mu + r({}^t\alpha\mathbf{x}), \quad (4.74)$$

où

$$Er({}^t\alpha\mathbf{X}) = 0. \quad (4.75)$$

Nous renvoyons à Härdle *et al.* (1993), Bonneau et Delecroix (1992) ou Härdle *et al.* (1997) pour une étude de ce modèle. Notons que ce modèle se généralise directement de la manière suivante :

$$r(x^1, \dots, x^p) = \mu + r({}^t\alpha_1\mathbf{x}, \dots, {}^t\alpha_q\mathbf{x}), \quad (4.76)$$

où

$$\forall k, Er({}^t\alpha_k\mathbf{X}) = 0. \quad (4.77)$$

On parle alors de Modèle à Indices Multiples, pour lesquels une méthode d'estimation populaire est la méthode de régression inverse par tranches introduite par Li(1991) qui a été abondamment étudiée ces derniers temps et dont les résultats les plus récents ont été donnés par Hsing et Carroll (1992), Schott (1994), Ferré (1996), (1997) et (1998), Aragon et Saracco (1997), Saracco (1999) et Saracco et Gannoun A. (2000).

4.5.5 Compléments bibliographiques

Pour clore ce paragraphe, nous renvoyons à plusieurs ouvrages généraux qui permettront au lecteur de remonter à la plupart de la littérature existante, tant sur ce modèle additif, que sur ses diverses extensions, que sur d'autres modèles à objectif analogue. Citons la revue bibliographique sur les modèles à réduction de dimension de Pelegrina *et al.* (1994) et la compilation de Schimek (2000) qui permettront d'avoir une vision d'ensemble des choses.

Concernant les modèles additifs, les recherches ont été nombreuses ces dernières années et nous souhaitons ici nous en tenir à quelques références récentes qui doivent permettre au lecteur intéressé de remonter ainsi à la plupart de la littérature. Le travail récent de Mammen *et al.* (1999) est axé sur les techniques itératives d'estimation tandis que celui de Härdle, Huet *et al.* (2000) est axé sur les techniques

de rééchantillonnage, que celui de Fan *et al.* (1998) est axé sur les techniques d'estimation directe et que celui de Severance-Lossin et Sperlich (1999) est axé sur l'estimation des dérivées des composantes additives. Une question importante est celle de la validité du modèle additif, et plusieurs travaux ont abordé cette question sous forme de tests statistiques, dont les plus récents sont ceux de Gozalo et Linton (1999), Dette et von Lieres (2000) et Camlong (2000). Pour terminer, mentionnons les ouvrages généraux de Hastie et Tibshirani (1990), de Sperlich (2000), de Schimek et Turlach (2000) et le Chapitre 2 de Sarda (2000)

Pour terminer, il faut mentionner un champ important de recherches actuelles qui concerne l'utilisation d'estimateurs non paramétriques généraux (du type de l'estimateur \hat{r}_{NW} étudié dans ce chapitre) pour valider un modèle semi-paramétrique du type de ceux étudiés dans le Paragraphe 4.5 précédent. La démarche générale qui guide la plupart de ces études est similaire à celle évoquée au Paragraphe 2.6.2 dans un contexte univarié. Nous allons nous limiter à quelques références clés dans le domaine qui sont celles de Fan et Li (1996), Kuchibhata et Hart (1996) et Stute (1997). Dans le même ordre d'idées, certains auteurs se sont intéressés à des problèmes de validation de modèles paramétriques contre des alternatives de type de celles des modèles étudiés dans le Paragraphe 4.5 précédent (*Cf. eg.* Horowitz et Härdle (1994)). D'autres auteurs se sont intéressés à la question du choix de variables explicatives, toujours dans une optique de réduction de dimension liée au fléau de la dimension évoqué ci-dessus. C'est le cas en particulier de Zhang (1991), Vieu (1994b) et Lavergne et Vuong (1996). Un autre champ de recherche d'actualité relativement voisin de ceux que l'on vient d'évoquer, est celui où certaines contraintes paramétriques sont introduites sur certaines des variables explicatives. Dans ce cadre nous nous limiterons à citer citons les travaux récents de Aneiros (2001) concernant des modèles partiellement linéaires et ceux de Orbe (2000) concernant des modèles paramétriques à coefficients fonctionnels, travaux à partir desquels le lecteur devrait pouvoir remonter à la plupart de la littérature dans ces domaines.

4.6 Exercices

Sauf mention explicite du contraire, les définitions et notations utilisées dans les énoncés de ces exercices sont identiques à celles utilisées dans tout le chapitre.

Exercice 4.1 a. *Montrer que tout noyau multivarié s'écrivant comme produit de noyaux univariés d'ordre $k > 0$ au sens de (2.9) est un noyau d'ordre k au sens de la Définition 4.2.1.*

b. En déduire que

$$K(t_1, \dots, t_p) = \left(\frac{3}{4}\right)^p (1 - t_1^2) \times \dots \times (1 - t_p^2),$$

est un noyau p dimensionnel d'ordre 2.

Exercice 4.2 Démontrer que tout noyau p dimensionnel d'ordre $k > 2$ prend nécessairement des valeurs négatives.

Exercice 4.3 Cet exercice a pour but d'établir des propriétés de convergence presque complète ponctuelle pour l'estimateur

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

de la densité f de X au point x .

1. Dans cette question on se place sous les hypothèses du Théorème 4.2.1.

a. En s'inspirant de ce qui a été fait pour prouver (4.14), montrer que l'estimateur de la densité vérifie (4.22).

b. Montrer qu'il existe une constante finie C telle que pour tout $\epsilon_0 > 0$ on ait :

$$P \left[\left| E\hat{f}(x) - \hat{f}(x) \right| > \epsilon_0 \sqrt{\frac{\log n}{nh^p}} \right] \leq 2n^{-C\epsilon_0^2}. \quad (4.78)$$

En déduire une propriété de convergence presque complète vers 0, avec vitesse, de $E\hat{f}(x) - \hat{f}(x)$.

2. En utilisant les deux questions précédentes, construire un ensemble d'hypothèses permettant de montrer que l'estimateur à noyau de la densité $\hat{f}(x)$ converge presque sûrement vers f en $\left(\frac{n}{\log n}\right)^{-\frac{k}{2k+p}}$.

Exercice 4.4 En s'inspirant de la preuve des Théorèmes 2.3.3 et 4.2.1, démontrer le Théorème 4.2.3.

Exercice 4.5 En reprenant la démarche de l'Exercice 2.5, démontrer le résultat du Théorème 4.2.2.

Exercice 4.6 On se place sous les conditions du Théorème 4.2.4, et l'objet de cet exercice est de prouver ce théorème.

1. En s'inspirant des calculs effectués pour démontrer (4.14), établir que

$$\sup_{x \in S} |E\hat{g}(x) - g(x)| = O(h^k). \quad (4.79)$$

2. En utilisant (4.16) et en s'inspirant de la démonstration de (2.43), démontrer qu'il existe des constantes strictement positives A_1 , A_2 et A_3 telles que pour tout $\epsilon > 0$ on ait

$$P \left[\sup_{x \in S} |E\hat{g}(x) - \hat{g}(x)| > \epsilon \sqrt{\frac{\log n}{nh^p}} \right] \leq A_1 n^{-\frac{\epsilon^2}{A_2} + A_3}. \quad (4.80)$$

N.B. On pourra aussi s'inspirer de la preuve du Lemme 5.1.6 du Chapitre 5 pour traiter cette question.

3. Etablir des résultats analogues à (4.79) et (4.80) mais concernant l'estimateur de densité \hat{f} .

4. En déduire que le résultat du Théorème 4.2.4 est vrai.

5. Déduire un résultat analogue à celui du Théorème 4.2.4 mais pour l'estimateur de densité \hat{f} .

Exercice 4.7 On se place sous les conditions du Théorème 4.2.5, et l'objet de cet exercice est de prouver ce théorème. La démarche est analogue à celle suivie lors de l'Exercice 4.6.

1. Montrer que

$$\sup_{x \in S} |E\hat{g}(x) - g(x)| \rightarrow 0, \quad (4.81)$$

et

$$\sup_{x \in S} |E\hat{f}(x) - f(x)| \rightarrow 0. \quad (4.82)$$

2. Justifier du fait que les résultats de la question 2 de l'Exercice 4.6 sont applicables en l'état.

3. Dédurre des deux questions précédentes que le résultat du Théorème 4.2.5 est vrai.
4. Enoncer un résultat analogue à celui du Théorème 4.2.5 mais pour l'estimateur de densité \hat{f} .

Exercice 4.8 On se place sous les conditions du Théorème 4.2.6, et l'objet de cet exercice est de prouver ce théorème.

1. Montrer que

$$\sup_{x \in S} |E\hat{g}(x) - g(x)| = O(h^\beta). \quad (4.83)$$

2. Justifier du fait que les résultats de la question 2 de l'Exercice 4.6 sont applicables en l'état.
3. Conclure.
4. Enoncer un résultat analogue à celui du Théorème 4.2.6 mais pour l'estimateur de densité \hat{f} .

Exercice 4.9 Ecrire la démonstration complète du Théorème 4.3.1.

Exercice 4.10 On se place sous les hypothèses du Théorème 4.3.2.

1. Montrer que l'on a :

$$E\hat{r}_{NW}(x) \rightarrow r(x). \quad (4.84)$$

2. Montrer que l'on a :

$$\text{var}(r_{NW}(x)) = O\left(\frac{1}{nh^p}\right), \text{ p.co.} \quad (4.85)$$

3. En déduire le résultat (4.51).
4. Renforçons à présent les hypothèses en supposant vérifiées celles du Théorème 4.3.4. Montrer en procédant de manière analogue aux 3 questions précédentes que le résultat (4.57) est vrai.

Exercice 4.11 Soit X^1, \dots, X^p des variables aléatoires réelles supposées être indépendantes et suivant chacune une loi uniforme sur $[0, 1]$. On considère l'intervalle $I_h = [0, h]$, et on note $F_p(h)$ la proportion de X^j appartenant à I_h :

$$F_p(h) = \frac{1}{p} \sum_{j=1}^p \mathbf{1}_{\{X^j \in I_h\}}.$$

1. Exprimer $F_p(h)$ en fonction de p et h . Calculer la taille d'échantillon n nécessaire (en fonction de p) pour avoir en moyenne 10 données dans l'intervalle $I_{0.1}$. Commentaires.

2. Reprendre l'exercice en remplaçant la condition d'uniformité sur les variables X^j par la condition d'existence d'une densité pour (X^1, \dots, X^p) par rapport à la mesure de Lebesgues sur \mathbb{R}^p .

Exercice 4.12 Considérons le modèle additif défini par (4.66) et (4.67). On note f^j la densité marginale de chaque variable explicative réelle X^j , densité que l'on suppose exister.

1. Montrer que :

$$\mu = \int_{\mathbb{R}^p} r(x^1, \dots, x^p) f^1(x^1) \dots f^p(x^p) dx^1 \dots dx^p.$$

2. Montrer que pour tout j on a :

$$\mu + r^j(x^j) = \int_{\mathbb{R}^{p-1}} r(x^1, \dots, x^{j-1}, x^{j+1}, \dots, x^p) \prod_{i \neq j} (f^i(x^i) dx^i).$$

3. A partir de ce qui précède, proposer de manière empirique des estimateurs des composantes fonctionnelles additives $r^j(\cdot)$ du modèle (4.66).

Chapitre 5

Régression non-paramétrique pour variable aléatoire fonctionnelle

Ce chapitre est consacré à l'étude de deux modèles de régression non-paramétrique lorsque la variable explicative est à valeurs dans un espace vectoriel semi-normé alors que la variable réponse est scalaire. Un travail fondateur des techniques développées dans ce chapitre peut être trouvé dans Ferraty et Vieu (2000). Dans un premier temps, on s'intéresse à un échantillon indépendamment et identiquement distribué (i.i.d.) ; on parle alors de *cas indépendant*. Un estimateur non-paramétrique est introduit pour lequel nous établissons des propriétés asymptotiques. Dans un second temps, on propose une extension du cas indépendant au contexte des séries temporelles. On dispose alors d'un échantillon identiquement distribué mais pour lequel on ne suppose plus l'indépendance ; on parle génériquement du *cas dépendant*. Les résultats obtenus dans le cas indépendant sont alors généralisés afin de pouvoir traiter les séries chronologiques. Soulignons l'importance d'étudier le cas dépendant car il débouche, lui aussi, sur de nombreuses applications (prévision de la température du courant marin El Niño, prévention d'infarctus en fonction d'électro-cardiogramme, prévision de rendement de cultures, prévision de valeurs boursières,...).

D'un point de vue théorique, l'utilisation de v.a.f. introduit une importante difficulté supplémentaire puisqu'on ne peut plus se permettre de manipuler la fonction de densité aussi facilement que dans le cas réel (*Cf.* Chapitres 2 et 3) ou encore dans le cas vectoriel (*Cf.* Chapitres 4). On est donc amené à donner des écritures probabilistes qui nous conduisent à des hypothèses agissant directement sur la distribution de la v.a.f. plutôt que sur la densité comme dans le cas réel ou vectoriel. Par ailleurs, nous verrons au fil des développements asymptotiques que les techniques de démonstration utilisées s'inspirent en partie de celles employées dans les Chapitres précédents.

5.1 Échantillon i.i.d.

De même que dans les cas d'une variable explicative réelle ou vectorielle (*Cf.* Chapitres 2, 3 et 4), l'étude du modèle de régression pour variable aléatoire fonctionnelle nécessite une hypothèse sur la distribution de la v.a.f. Comme nous le verrons par la suite, cette dernière se traduit en termes de dimension fractale de la mesure de probabilité associée au régresseur. Le premier paragraphe introduit le modèle de régression que l'on souhaite approfondir et précise les hypothèses fondamentales que doit vérifier la v.a.f. Au second paragraphe, nous construisons un estimateur non-paramétrique en adaptant l'estimateur de Nadaraya-Watson à ce contexte fonctionnel. Les troisième et quatrième paragraphes sont consacrés aux propriétés asymptotiques de cet estimateur. Enfin, notons que la dimension fractale de la mesure de probabilité associée à la variable explicative fonctionnelle va jouer un rôle prépondérant puisqu'elle intervient directement dans les vitesses de convergence.

5.1.1 Présentation du modèle

On s'intéresse à un modèle de régression non-paramétrique de Y en X , où Y est une variable aléatoire réelle et X est une variable aléatoire à valeurs dans un espace vectoriel semi-normé $(\mathcal{H}, \|\cdot\|)$, X et Y étant définies sur un même espace probabilisé (Ω, \mathcal{A}, P) . Dans un tel cadre, nous cherchons à estimer l'opérateur fonctionnel Φ défini par

$$Y = \Phi(X) + \epsilon, \quad (5.1)$$

où ϵ est une variable réelle centrée et indépendante de X . De tels modèles recouvrent bien sûr les modèles classiques de régression quand $\mathcal{H} = \mathbb{R}$ ou \mathbb{R}^p , mais offrent surtout de nouvelles perspectives dans de nombreux domaines d'application lorsque la variable explicative X est de nature fonctionnelle ; on parle alors génériquement d'analyse de données fonctionnelles (*cf. eg.* Ramsay et Silverman, 1997).

Notre démarche repose sur une hypothèse dont l'originalité réside dans la possibilité de l'interpréter en termes de dimension fractale de la loi de probabilité de la v.a.f. X (*cf.* Exercice 5.1). Cette hypothèse fondamentale s'exprime de la façon suivante : supposons que pour un point fixé $x \in \mathcal{H}$, on ait :

$$\lim_{\alpha \rightarrow 0^+} \frac{P(X \in \mathcal{B}(x, \alpha))}{\alpha^{\delta(x)}} = c(x), \quad (5.2)$$

où $\delta(x)$ et $c(x)$ sont deux réels strictement positifs et où $\mathcal{B}(x, \alpha)$ désigne la boule de centre x et de rayon α pour la topologie associée à la semi-norme $\|\cdot\|$.

Nous verrons que cette hypothèse nous permet d'obtenir la convergence presque complète de notre estimateur (*cf.* Théorèmes 5.1.1 et 5.1.2), sous réserve que Φ soit suffisamment "régulier". En revanche, dès qu'on s'intéresse aux vitesses de convergence, nous devons renforcer l'hypothèse (5.2) par l'existence d'un réel $b(x)$ strictement positif tel que pour α au voisinage de 0 on ait :

$$P(X \in \mathcal{B}(x, \alpha)) = \alpha^{\delta(x)} c(x) + O(\alpha^{\delta(x)+b(x)}). \quad (5.3)$$

Les vitesses de convergence peuvent être ainsi précisées (*cf.* Théorème 5.1.3) pour des opérateurs Φ vérifiant une condition de régularité supplémentaire. Enfin, notons que les techniques de démonstration employées dans le cadre de v.a.f. s'inspirent largement de celles utilisées dans le cas réel (*Cf.* Chapitre 2). Par ailleurs, il est facile de faire le lien entre les hypothèses (5.2), (5.3) et celles usuellement employées dans le Chapitre 4) traitant du cas vectoriel (*cf.* partie 3) de l'Exercice 5.5).

5.1.2 Présentation de l'estimateur de Φ

On s'intéresse à l'estimation de l'opérateur $\Phi(x) = E(Y|X = x)$ à partir de n observations indépendantes $(X_i, Y_i)_{i=1, \dots, n}$ du couple (X, Y) . L'estimateur $\widehat{\Phi}_n(x)$ de $\Phi(x)$ est défini par :

$$\widehat{\Phi}_n(x) = \sum_{i=1}^{i=n} w_i(x) Y_i \quad (5.4)$$

avec

$$w_i(x) = \frac{K\left(\frac{\|X_i - x\|}{h}\right)}{\sum_{i=1}^{i=n} K\left(\frac{\|X_i - x\|}{h}\right)}, \quad (5.5)$$

h étant une suite de nombres positifs. Il s'agit d'une version de l'estimateur de Nadaraya-Watson adaptée au cas où la variable explicative est à valeurs dans un espace semi-normé. Son principe est le suivant : étant donné un élément x de \mathcal{H} , on détermine $\widehat{\Phi}_n(x)$ en moyennant les Y_i selon une pondération qui accorde plus de poids aux Y_i pour lesquels les X_i sont "proches" de x :

- la semi-norme $\|\cdot\|$ mesure la proximité entre les courbes,
- la largeur de fenêtre h contrôle le nombre de termes retenus dans le calcul de la moyenne pondérée,
- le noyau K est une fonction décroissante : plus X_i est proche de x c'est-à-dire plus la quantité $\|X_i - x\|$ est petite, plus le poids $w_i(x)$ est grand.

Notons que la principale différence avec le cas où le régresseur est un vecteur réside dans l'emploi de la semi-norme $\|\cdot\|$ pour mesurer la distance entre deux éléments de \mathcal{H} (hormis le cadre fonctionnel).

5.1.3 Etude de la convergence presque complète

Commençons par donner les hypothèses communes aux résultats asymptotiques auxquels nous allons nous intéresser :

- la largeur de fenêtre h est telle que :

$$\lim_{n \rightarrow \infty} h = 0 \text{ et } \lim_{n \rightarrow \infty} \frac{nh^{\delta(x)}}{\log n} = \infty, \quad (5.6)$$

- le noyau K est tel que

$$K \text{ Lipschitzien d'ordre 1 et support}(K) = [0, \xi] \text{ avec } \xi \in \mathbb{R}_*^+, \quad (5.7)$$

- la variable aléatoire réelle Y vérifie :

$$|Y| \leq M < \infty \text{ p.s.} \quad (5.8)$$

Théorème 5.1.1 Convergence presque complète ponctuelle. *Sous les hypothèses (5.2), (5.6)-(5.8) et si*

$$\Phi \text{ est continu en } x, \quad (5.9)$$

alors on a :

$$\lim_{n \rightarrow \infty} \widehat{\Phi}_n(x) = \Phi(x) \text{ p.co.} \quad (5.10)$$

Preuve. La démonstration de ce résultat est basé sur la décomposition suivante :

$$\widehat{\Phi}_n(x) - \Phi(x) = \frac{1}{\widehat{f}_n(x)} \{ \widehat{g}_n(x) - \Phi(x) C_\delta(x) \} - \frac{\Phi(x)}{\widehat{f}_n(x)} \{ \widehat{f}_n(x) - C_\delta(x) \} \quad (5.11)$$

en posant $\widehat{f}_n(x) = \frac{1}{nh^{\delta(x)}} \sum_{i=1}^n K\left(\frac{\|X_i - x\|}{h}\right)$, $\widehat{g}_n(x) = \frac{1}{nh^{\delta(x)}} \sum_{i=1}^n Y_i K\left(\frac{\|X_i - x\|}{h}\right)$ et où $C_\delta(x)$ est un réel strictement positif (dépendant de x ainsi que de la dimension fractale $\delta(x)$) qui sera précisé en cours de démonstration.

Lemme 5.1.1 *Sous l'hypothèse (5.7) et si*

- i) (5.2) est vérifiée, alors on a :

$$\lim_{n \rightarrow \infty} E \widehat{f}_n(x) = C_\delta(x). \quad (5.12)$$

- ii) (5.3) est vérifiée, alors on a :

$$E \widehat{f}_n(x) = C_\delta(x) + O(h^{b(x)}). \quad (5.13)$$

Preuve.

i) Comme K est Lipschitzien d'ordre 1, K est en particulier continu. on peut donc écrire :

$$K(t) = \lim_{J \rightarrow \infty} K_J(t) \text{ où } K_J(t) = \sum_{j=0}^{J-1} K(j\xi/J) \mathbf{1}_{[j\xi/J, (j+1)\xi/J]}.$$

On a

$$E\hat{f}_n(x) = \frac{1}{nh^{\delta(x)}} \sum_{i=1}^n E \left\{ \lim_{J \rightarrow \infty} K_J \left(\frac{\|X_i - x\|}{h} \right) \right\}$$

puis par équidistribution des X_i

$$E\hat{f}_n(x) = \frac{1}{h^{\delta(x)}} E \left\{ \lim_{J \rightarrow \infty} K_J \left(\frac{\|X - x\|}{h} \right) \right\}.$$

D'où (théorème de la convergence dominée)

$$E\hat{f}_n(x) = \frac{1}{h^{\delta(x)}} \lim_{J \rightarrow \infty} E \left\{ K_J \left(\frac{\|X - x\|}{h} \right) \right\},$$

ce qui amène à

$$E\hat{f}_n(x) = \lim_{J \rightarrow \infty} \sum_{j=0}^{J-1} K(j\xi/J) \left\{ \left(\frac{(j+1)\xi}{J} \right)^{\delta(x)} \frac{P(X \in \mathcal{B}(x, h(j+1)\xi/J))}{\left(\frac{(j+1)\xi h}{J} \right)^{\delta(x)}} - \left(\frac{j\xi}{J} \right)^{\delta(x)} \frac{P(X \in \mathcal{B}(x, hj\xi/J))}{\left(\frac{j\xi h}{J} \right)^{\delta(x)}} \right\}.$$

Posons

$$C_\delta(x) = \lim_{J \rightarrow \infty} c(x) \sum_{j=0}^{J-1} K(j\xi/J) \left\{ \left(\frac{(j+1)\xi}{J} \right)^{\delta(x)} - \left(\frac{j\xi}{J} \right)^{\delta(x)} \right\}.$$

On a :

$$\begin{aligned} E\hat{f}_n(x) - C_\delta(x) &= \\ \lim_{J \rightarrow \infty} \sum_{j=0}^{J-1} K(j\xi/J) &\left\{ \left(\frac{(j+1)\xi}{J} \right)^{\delta(x)} \left(\frac{P(X \in \mathcal{B}(x, h(j+1)\xi/J))}{\left(\frac{(j+1)\xi h}{J} \right)^{\delta(x)}} - c(x) \right) \right. \\ &\left. - \left(\frac{j\xi}{J} \right)^{\delta(x)} \left(\frac{P(X \in \mathcal{B}(x, hj\xi/J))}{\left(\frac{j\xi h}{J} \right)^{\delta(x)}} - c(x) \right) \right\}. \end{aligned} \quad (5.14)$$

D'où, grâce à la condition (5.2) :

$$\begin{aligned} Ef_n(x) - C_\delta(x) = \\ \lim_{J \rightarrow \infty} \sum_{j=0}^{J-1} \left\{ \left(K((j+1)\xi/J) \left(\frac{(j+1)\xi}{J} \right)^{\delta(x)} - K(j\xi/J) \left(\frac{j\xi}{J} \right)^{\delta(x)} \right) o(1) \right\} \\ - \lim_{J \rightarrow \infty} \sum_{j=0}^{J-1} \left\{ (K((j+1)\xi/J) - K(j\xi/J)) \left(\frac{(j+1)\xi}{J} \right)^{\delta(x)} o(1) \right\}. \end{aligned}$$

La 1ère limite tend vers 0 (les termes de la somme s'annulent 2 à 2). De plus, K Lipschitzien d'ordre 1 implique :

$$\begin{aligned} \left| \sum_{j=0}^{J-1} \left\{ (K((j+1)\xi/J) - K(j\xi/J)) \left(\frac{(j+1)\xi}{J} \right)^{\delta(x)} o(1) \right\} \right| \\ \leq \xi^{\delta(x)} \sum_{j=0}^{J-1} |(K((j+1)\xi/J) - K(j\xi/J))| o(1) \\ \leq \xi^{\delta(x)} \sum_{j=0}^{J-1} \frac{\xi}{J} o(1). \end{aligned}$$

Conclusion : pour tout $\epsilon > 0$, il existe un rang n_0 à partir duquel

$$|Ef_n(x) - C_\delta(x)| < \epsilon.$$

Il suffit maintenant de préciser $C_\delta(x)$: en posant $H(u) = K(u^{1/\delta(x)})$, on obtient :

$$C_\delta(x) = c(x) \int_0^{\xi^{\delta(x)}} H(u) du,$$

d'où l'on tire que

$$C_\delta(x) = c(x) \delta(x) \int_0^\xi K(v) v^{\delta(x)-1} dv.$$

ii) (5.14) et (5.3) permettent d'écrire :

$$\begin{aligned} \left| Ef_n(x) - C_\delta(x) \right| \leq \\ \lim_{J \rightarrow \infty} \left| \sum_{j=0}^{J-1} \left\{ \left(K((j+1)\xi/J) \left(\frac{(j+1)\xi}{J} \right)^{\delta(x)} - K(j\xi/J) \left(\frac{j\xi}{J} \right)^{\delta(x)} \right) O(h^{b(x)}) \right\} \right| \end{aligned}$$

$$+ \lim_{J \rightarrow \infty} \left| \sum_{j=0}^{J-1} \left\{ (K((j+1)\xi/J) - K(j\xi/J)) \left(\frac{(j+1)\xi}{J} \right)^{\delta(x)} O(h^{b(x)}) \right\} \right|.$$

Il suffit alors de reprendre les arguments développés précédemment pour conclure.

□

Lemme 5.1.2

$$\hat{f}_n(x) - E\hat{f}_n(x) = O\left(\sqrt{\frac{\log n}{nh^{\delta(x)}}}\right) \text{ p.co.} \quad (5.15)$$

Preuve. On a

$$\hat{f}_n(x) - E\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \Delta_{i,x} \text{ où } \Delta_{i,x} = \frac{1}{h^{\delta(x)}} \left\{ K\left(\frac{\|X_i - x\|}{h}\right) - EK\left(\frac{\|X_i - x\|}{h}\right) \right\}.$$

Il est clair que l'on a $|\Delta_{i,x}| \leq C_1/h^{\delta(x)}$. De plus, en reprenant les calculs faits précédemment, on a, d'une part,

$$\lim_{n \rightarrow \infty} \left\{ \frac{1}{h^{\delta(x)}} EK \left(\frac{\|X - x\|}{h} \right) \right\}^2 = \lim_{n \rightarrow \infty} (E\hat{f}_n(x))^2 = C_\delta^2(x),$$

et d'autre part, en remplaçant K par K^2 ,

$$\lim_{n \rightarrow \infty} \frac{1}{h^{\delta(x)}} EK^2 \left(\frac{\|X - x\|}{h} \right) = C'_\delta(x).$$

Comme

$$E\Delta_{i,x}^2 = \frac{1}{h^{\delta(x)}} \left\{ \frac{1}{h^{\delta(x)}} EK^2 \left(\frac{\|X - x\|}{h} \right) \right\} - \left\{ \frac{1}{h^{\delta(x)}} EK \left(\frac{\|X - x\|}{h} \right) \right\}^2,$$

il vient que

$$E\Delta_{i,x}^2 \leq C_2 \frac{1}{h^{\delta(x)}}.$$

On obtient alors en appliquant une version simplifiée d'une inégalité exponentielle de type Bernstein dans le cas indépendant (Cf. Hoeffding, 1963 ou le Lemme 2.3.1 du Chapitre 2)

$$P \left(|\hat{f}_n(x) - E\hat{f}_n(x)| > \eta \sqrt{\frac{\log n}{nh^{\delta(x)}}} \right) \leq C_3 \exp\left\{-n\eta^2 \frac{\log n}{nh^{\delta(x)}} \frac{1}{C_4}\right\},$$

ce qui équivaut à

$$P \left(|\hat{f}_n(x) - E\hat{f}_n(x)| > \eta \sqrt{\frac{\log n}{nh^{\delta(x)}}} \right) \leq Cn^{-C\eta^2}. \quad (5.16)$$

Ainsi, il existe η_0 tel que

$$\sum_{n=1}^{n=\infty} P[|\hat{f}_n(x) - E\hat{f}_n(x)| > \eta_0 \sqrt{\frac{\log n}{nh^{\delta(x)}}}] < \infty.$$

□

Lemme 5.1.3

$$\lim_{n \rightarrow \infty} E\hat{g}_n(x) = \Phi(x)C_\delta(x). \quad (5.17)$$

Preuve.

$$\begin{aligned} E\hat{g}_n(x) &= \frac{1}{h^{\delta(x)}} EYK\left(\frac{\|X-x\|}{h}\right), \\ &= \frac{1}{h^{\delta(x)}} E(\Phi(X)K\left(\frac{\|X-x\|}{h}\right)). \end{aligned}$$

Ceci se réécrit

$$E\hat{g}_n(x) = \Phi(x)E\hat{f}_n(x) + \frac{1}{h^{\delta(x)}} E(\Phi(X) - \Phi(x))K\left(\frac{\|X-x\|}{h}\right),$$

et implique que

$$|E\hat{g}_n(x) - \Phi(x)C_\delta(x)| \leq |\Phi(x)| \left| E\hat{f}_n(x) - C_\delta(x) \right| + \frac{1}{h^{\delta(x)}} E|\Phi(X) - \Phi(x)| \left| K\left(\frac{\|X-x\|}{h}\right) \right|.$$

Le premier terme de la partie droite de cette inégalité tend vers 0 à cause du Lemme 5.1.1 ; le second terme tend aussi vers 0 à cause de la continuité de Φ et en appliquant une nouvelle fois le Lemme 5.1.1 au noyau $|K|$.

□

Lemme 5.1.4

$$\hat{g}_n(x) - E\hat{g}_n(x) = O\left(\sqrt{\frac{\log n}{nh^{\delta(x)}}}\right) \quad p.co. \quad (5.18)$$

Preuve. Utilisons la décomposition suivante

$$\hat{g}_n(x) - E\hat{g}_n(x) = \frac{1}{n} \sum_{i=1}^n \Delta'_{i,x},$$

où

$$\Delta'_{i,x} = \frac{1}{h^{\delta(x)}} \left\{ Y_i K\left(\frac{\|X_i - x\|}{h}\right) - EY_i K\left(\frac{\|X_i - x\|}{h}\right) \right\}.$$

Il est clair, puisque Y est bornée, qu'en reprenant les arguments développés pour prouver (5.15), on a

$$|\Delta'_{i,x}| \leq C \frac{1}{h^{\delta(x)}} \text{ et } E\Delta'^2_{i,x} \leq C \frac{1}{h^{\delta(x)}}.$$

Par analogie avec la fin de la preuve de (5.15), il suffit d'utiliser à nouveau l'inégalité du Lemme 2.3.1 pour obtenir :

$$P \left(|\hat{g}_n(x) - E\hat{g}_n(x)| > \eta \sqrt{\frac{\log n}{nh^{\delta(x)}}} \right) \leq Cn^{-C\eta^2}. \quad (5.19)$$

Ainsi, il existe η_0 tel que

$$\sum_{n=1}^{n=\infty} P[|\hat{g}_n(x) - E\hat{g}_n(x)| > \eta_0 \sqrt{\frac{\log n}{nh^{\delta(x)}}}] < \infty. \quad \square$$

Finalement, la démonstration du Théorème 5.1.1 se déduit aisément de (5.11) ainsi que des quatre résultats intermédiaires précédents, à savoir (5.12), (5.15), (5.17) et (5.18).

□.

Intéressons-nous maintenant à la convergence presque complète uniforme sur un compact S de l'espace semi-normé $(\mathcal{H}, \|\cdot\|)$. Il est clair que l'obtention de résultats uniformes nécessite le renforcement des certaines hypothèses utilisées dans le paragraphe précédent. En particulier, on rend uniforme l'hypothèse (5.2) de la manière suivante :

$$\lim_{\alpha \rightarrow 0^+} \sup_{x \in S} \left\{ \frac{P(X \in \mathcal{B}(x, \alpha))}{\alpha^\delta} - c(x) \right\} = 0 \quad \text{et} \quad \inf_{x \in S} c(x) > 0, \quad (5.20)$$

où δ est un réel strictement positif ne dépendant pas de x . Notons que dans ce cas, cette hypothèse peut s'interpréter en termes de dimension de Hausdorff de la loi de la v.a.f. X (Bardet, 1997, p28). On peut alors énoncer le résultat suivant.

Théorème 5.1.2 Convergence presque complète uniforme. *Sous les hypothèses (5.6)-(5.8), (5.20),*

$$\Phi \text{ est uniformément continu sur } S, \quad (5.21)$$

alors on a :

$$\lim_{n \rightarrow \infty} \sup_{x \in S} \left| \widehat{\Phi}_n(x) - \Phi(x) \right| = 0 \quad p.co. \quad (5.22)$$

Preuve. La démonstration de ce résultat est en partie analogue à celle du Théorème 5.1.1. En effet, la décomposition (5.11) s'écrit maintenant de la façon suivante

$$\widehat{\Phi}_n(x) - \Phi(x) = \frac{1}{\widehat{f}_n(x)} \{ \widehat{g}_n(x) - \Phi(x)C_\delta \} - \frac{\Phi(x)}{\widehat{f}_n(x)} \{ \widehat{f}_n(x) - C_\delta \}, \quad (5.23)$$

où $C_\delta = c(x)\delta \int_0^\xi K(v)v^{\delta-1}dv$ est une constante strictement positive. En reprenant alors les calculs faits précédemment, on montre que :

$$\lim_{n \rightarrow \infty} E\widehat{f}_n(x) = C_\delta \text{ uniformément en } x, \quad (5.24)$$

ce qui implique que

$$\lim_{n \rightarrow \infty} \sup_{x \in S} |E\widehat{f}_n(x) - C_\delta| = 0. \quad (5.25)$$

Montrons maintenant le résultat suivant :

Lemme 5.1.5

$$\lim_{n \rightarrow \infty} \sup_{x \in S} |E\widehat{g}_n(x) - \Phi(x)C_\delta| = 0. \quad (5.26)$$

Preuve. En remplaçant K par $|K|$ dans la preuve de (5.24), on a :

$$\lim_{n \rightarrow \infty} \frac{1}{h^\delta} E \left| K \left(\frac{\|X - x\|}{h} \right) \right| = C_\delta''. \quad (5.27)$$

Reprenons alors la preuve du Lemme (5.1.3) ; on en déduit que :

$$|E\widehat{g}_n(x) - \Phi(x)C_\delta| \leq |\Phi(x)| \left| E\widehat{f}_n(x) - C_\delta \right| + \frac{1}{h^\delta} E |\Phi(X) - \Phi(x)| \left| K \left(\frac{\|X - x\|}{h} \right) \right|$$

et en utilisant (5.27) ainsi que l'uniforme continuité de Φ , pour tout ϵ strictement positif (ne dépendant pas de x), il existe n suffisamment grand tel que

$$|E\widehat{g}_n(x) - \Phi(x)C_\delta| \leq \epsilon.$$

□.

Remarque. (5.21) entraîne que Φ est uniformément borné sur S .

Dans ce qui suit, $\widehat{\psi}_n$ désignera indifféremment \widehat{f}_n ou \widehat{g}_n :

Lemme 5.1.6

$$\sup_{x \in S} \left| \widehat{\psi}_n(x) - E\widehat{\psi}_n(x) \right| = O\left(\sqrt{\frac{\log n}{nh^\delta}}\right) \text{ p.co.} \quad (5.28)$$

Preuve. S étant un compact de l'espace vectoriel topologique $(\mathcal{H}, \|\cdot\|)$, S est un fermé borné. Il existe donc un recouvrement fini de S tel que :

$$S \subset \bigcup_{k=1}^{\tau_n} \mathcal{B}(t_k, l_n),$$

avec $\tau_n l_n = C$ et posons

$$k(x) = \arg \min_{k=1, \dots, \tau_n} \|t_k - x\|.$$

On peut écrire :

$$\begin{aligned} P\left(\sup_{x \in S} \left| \widehat{\psi}_n(x) - E\widehat{\psi}_n(x) \right| > 2\eta\sqrt{\frac{\log n}{nh^\delta}}\right) &\leq \\ &\underbrace{P\left(\sup_{x \in S} \left| \widehat{\psi}_n(t_{k(x)}) - E\widehat{\psi}_n(t_{k(x)}) \right| > \eta\sqrt{\frac{\log n}{nh^\delta}}\right)}_A + \\ &\underbrace{P\left(\sup_{x \in S} \left| \widehat{\psi}_n(x) - \widehat{\psi}_n(t_{k(x)}) - E\widehat{\psi}_n(x) + E\widehat{\psi}_n(t_{k(x)}) \right| > \eta\sqrt{\frac{\log n}{nh^\delta}}\right)}_B \end{aligned}$$

- En ce qui concerne A :

$$\begin{aligned} A &\leq P\left(\max_{k=1, \dots, \tau_n} \left| \widehat{\psi}_n(t_k) - E\widehat{\psi}_n(t_k) \right| > \eta\sqrt{\frac{\log n}{nh^\delta}}\right) \\ &\leq \sum_{k=1}^{\tau_n} P\left(\left| \widehat{\psi}_n(t_k) - E\widehat{\psi}_n(t_k) \right| > \eta\sqrt{\frac{\log n}{nh^\delta}}\right) \\ &\leq \tau_n \max_{k=1, \dots, \tau_n} P\left(\left| \widehat{\psi}_n(t_k) - E\widehat{\psi}_n(t_k) \right| > \eta\sqrt{\frac{\log n}{nh^\delta}}\right). \end{aligned}$$

D'où, en utilisant (5.16), (5.19) et sachant que $\tau_n = C/l_n$, on a :

$$A \leq \frac{C}{l_n} n^{-C\eta^2}.$$

En prenant $l_n = n^{-\zeta}$ ($\zeta > 0$) on obtient :

$$A \leq Cn^{-C\eta^2 + \zeta}. \quad (5.29)$$

- En ce qui concerne B :

$$\begin{aligned}
\sup_{x \in S} \left| \widehat{\psi}_n(x) - \widehat{\psi}_n(t_{k(x)}) \right| &\leq \sup_{x \in S} \frac{C}{nh^\delta} \left| \sum_{i=1}^n \left\{ K \left(\frac{\|X_i - x\|}{h} \right) - K \left(\frac{\|X_i - t_{k(x)}\|}{h} \right) \right\} \right| \\
&\leq C \frac{1}{h^{\delta+1}} \sup_{x \in S} \|x - t_{k(x)}\| \\
&\leq C \frac{l_n}{h^{\delta+1}} \\
&\leq C \frac{n^{-\zeta}}{h^{\delta+1}}.
\end{aligned}$$

Par ailleurs,

$$\begin{aligned}
\sup_{x \in S} \left| E\widehat{\psi}_n(x) - E\widehat{\psi}_n(t_{k(x)}) \right| &\leq \frac{C}{h^\delta} \sup_{x \in S} E \left| K \left(\frac{\|X - x\|}{h} \right) - K \left(\frac{\|X - t_{k(x)}\|}{h} \right) \right| \\
&\leq C \frac{1}{h^{\delta+1}} \sup_{x \in S} \|x - t_{k(x)}\| \\
&\leq C \frac{l_n}{h^{\delta+1}} \\
&\leq C \frac{n^{-\zeta}}{h^{\delta+1}}.
\end{aligned}$$

D'où, en choisissant $\zeta = 1 + \delta^{-1}$, on a :

$$\frac{n^{-\zeta}}{h^{\delta+1}} = o \left(\sqrt{\frac{\log n}{nh^\delta}} \right),$$

ce qui implique, pour n suffisamment grand, que

$$P \left(\sup_{x \in S} \left| \widehat{\psi}_n(x) - \widehat{\psi}_n(t_{k(x)}) \right| + \sup_{x \in S} \left| E\widehat{\psi}_n(x) - E\widehat{\psi}_n(t_{k(x)}) \right| > \eta \sqrt{\frac{\log n}{nh^\delta}} \right) = 0.$$

Par conséquent, pour n suffisamment grand, on a $B = 0$.

Finalement :

$$P \left(\sup_{x \in S} \left| \widehat{\psi}_n(x) - E\widehat{\psi}_n(x) \right| > 2\eta \sqrt{\frac{\log n}{nh^\delta}} \right) \leq Cn^{-C\eta^2 + 1 + \delta^{-1}}.$$

Il est clair qu'on peut toujours trouver η tel que le terme de droite de cette inégalité soit le terme général d'une série convergente, ce qui termine la preuve du Lemme 5.1.6. \square

Enfin, il suffit d'écrire que

$$\sup_{x \in S} |\widehat{f}_n(x) - C_\delta| \leq \sup_{x \in S} |\widehat{f}_n(x) - E\widehat{f}_n(x)| + \sup_{x \in S} |E\widehat{f}_n(x) - C_\delta|$$

et

$$\sup_{x \in S} |\widehat{g}_n(x) - \Phi(x)C_\delta| \leq \sup_{x \in S} |\widehat{g}_n(x) - E\widehat{g}_n(x)| + \sup_{x \in S} |E\widehat{g}_n(x) - \Phi(x)C_\delta|,$$

d'utiliser (5.23), (5.25), (5.26), (5.28) pour en déduire (5.22), ce qui achève la démonstration du Théorème 5.1.2. \square

5.1.4 Vitesse de convergence presque complète

Théorème 5.1.3 *Vitesse de convergence presque complète ponctuelle. Sous les hypothèses (5.3), (5.6)-(5.8) et si*

$$|\Phi(u) - \Phi(v)| \leq C\|u - v\|^\beta, \quad (5.30)$$

$$h = h_0 \left(\frac{\log n}{n} \right)^{\frac{1}{2\gamma(x) + \delta(x)}} \quad \text{où } \gamma(x) = \min\{b(x), \beta\}, \quad (5.31)$$

alors

$$\Phi(x) - \widehat{\Phi}_n(x) = O \left(\left(\frac{\log n}{n} \right)^{\frac{\gamma(x)}{2\gamma(x) + \delta(x)}} \right) \quad p.co. \quad (5.32)$$

Preuve. Le résultat i) du Lemme 5.1.1 nous dit que

$$E\widehat{f}_n(x) = C_\delta(x) + O(h^{b(x)}). \quad (5.33)$$

Par ailleurs, au cours de la preuve du Lemme 5.1.3 on a vu que

$$|E\widehat{g}_n(x) - \Phi(x)C_\delta(x)| \leq |\Phi(x)| \left| E\widehat{f}_n(x) - C_\delta(x) \right| + \frac{1}{h^{\delta(x)}} E|\Phi(X) - \Phi(x)| \left| K\left(\frac{\|X - x\|}{h}\right) \right|.$$

Grâce à la condition (5.30) sur Φ et (5.33) on obtient

$$E\widehat{g}_n(x) = C_\delta(x)\Phi(x) + O(h^{b(x)}) + O(h^\beta). \quad (5.34)$$

En combinant ces deux résultats avec la décomposition

$$\widehat{\Phi}_n(x) - \Phi(x) = \frac{1}{\widehat{f}_n(x)} \{ \widehat{g}_n(x) - \Phi(x)C_\delta(x) \} - \frac{\Phi(x)}{\widehat{f}_n(x)} \{ \widehat{f}_n(x) - C_\delta(x) \},$$

puis en utilisant (5.33) et (5.34) ainsi que les Lemmes 5.1.2 et 5.1.4, on arrive à

$$\widehat{\Phi}_n(x) - \Phi(x) = O(h^\beta) + O(h^{b(x)}) + O\left(\sqrt{\frac{\log n}{nh^{\delta(x)}}}\right).$$

Il suffit de remplacer h par sa valeur pour obtenir (5.32). \square

Une version du Théorème 5.1.3 peut être obtenue moyennant quelques aménagements (*Cf.* Exercice 5.3).

5.2 Extension aux séries temporelles

Comme nous l'avons déjà précisé en début de ce chapitre, le cas dépendant joue un rôle important en statistique puisqu'il permet en particulier le traitement des séries chronologiques. D'un point de vue théorique, nous ferons appel à des outils déjà présentés au Chapitre 3, lesquels, combinés aux hypothèses sur la distribution de la v.a.f. nous permettrons l'obtention de résultats asymptotiques. Le premier paragraphe donne une description du problème que nous allons traiter. Le second paragraphe rappelle le modèle de dépendance que nous utilisons, lequel a déjà été abordé au Chapitre 3. Quant au troisième paragraphe, il nous conduit au cœur du sujet puisqu'il présente successivement le modèle de régression sous sa formulation la plus générale ainsi que l'estimateur. Quant aux derniers paragraphes, ils présentent les propriétés asymptotiques de l'estimateur proposé. Enfin, nous reprenons dans cette section les notations introduites dans la section 5.1 de ce Chapitre.

5.2.1 Position du problème

A titre d'exemple introductif, plaçons nous dans le cadre d'une variable aléatoire fonctionnelle Z à valeurs dans \mathcal{F} , espace de fonctions définies sur \mathbb{R} , et supposons que l'on observe $Z(t)$ pour t appartenant à l'intervalle $[0, T[$. Soit $0 = t_1 < t_2 < \dots < t_n < t_{n+1} = T$ une subdivision de $[0, T[$. À partir de Z , on peut construire n v.a.f X_1, \dots, X_n définies par :

$$X_i(t) = Z(t)\mathbf{1}_{[t_i; t_{i+1}[}(t), \quad \forall i = 1, \dots, n.$$

Il est clair que pour tout t appartenant à $[0, T[$, on a

$$Z(t) = \sum_{i=1}^n X_i(t).$$

Soit maintenant Y_1, \dots, Y_{n-1} , $n - 1$ v.a. réelles définies par :

$$Y_i = F(X_{i+1}), \quad \forall i = 1, \dots, n - 1,$$

où F est un opérateur **connu** défini sur \mathcal{F} et à valeurs dans \mathbb{R} . Ainsi, ayant à notre disposition $(X_i, Y_i)_{i=1, \dots, n-1}$, échantillon de taille $n - 1$ du couple (X, Y) , il est naturel de poser le problème de l'estimation de l'opérateur Φ défini par :

$$\Phi(x) = E(Y/X = x). \quad (5.35)$$

En particulier, peut-on estimer une valeur future de Z connaissant son passé? Le problème tel qu'il est posé par la relation (5.35) est étroitement lié avec le modèle

de régression posé dans le cas indépendant. C'est pourquoi, tant du point de vue de l'estimateur que du point de vue de l'étude théorique, on s'inspire largement de ce qui a été proposé dans le cas indépendant. Par ailleurs, pour fixer un peu plus les idées sur les applications qui peuvent découler de ce type de modèle, il suffit de préciser la forme que F peut éventuellement prendre :

- $Y_i = F(X_{i+1}) = X_{i+1}(t_{i+1} + \tau)$ avec $t_{i+1} + \tau \in [t_{i+1}; t_{i+2}[$, ce qui correspond à un objectif de prévision classique (*.i.e* quelle sera la valeur de Z à l'instant $T + \tau$?),
- $Y_i = F(X_{i+1}) = \sup_{t \in [t_{i+1}; t_{i+2}[} |X_{i+1}(t)|$, ce qui correspond à la prévision de maxima (crues, éruption volcaniques, valeurs boursières,...).

Au travers de ces diverses formes de F , on s'aperçoit de l'énorme potentialité de ce type de modélisation en termes d'applications. Notre objectif est donc de développer des méthodes capables d'appréhender de telles situations. Ainsi, nous aborderons dans les prochains paragraphes les aspects théoriques liés à une telle modélisation statistique.

5.2.2 Modèle de dépendance

Comme dans le Chapitre 3, nous modéliserons la dépendance par la notion de mélange fort appelé encore α -mélange. Afin d'éviter de trop nombreuses références au Chapitre 3, ceci dans le but de faciliter la lecture de ce document, nous rappelons brièvement les différentes notions liées au mélange fort. En revanche, pour plus de précisions, nous renvoyons le lecteur au Paragraphe 3.2 du chapitre 3.

Soit $\{\Delta_i, i \in \mathbb{Z}\}$ une famille de variables aléatoires à valeurs dans un même espace probabilisable $(\mathcal{H}, \mathcal{B}_{\mathcal{H}})$. Pour tout couple (i, j) dans $\mathbb{Z} \cup \{-\infty, +\infty\}$, on note σ_i^j la tribu engendrée par $\{\Delta_k, i < k < j\}$. On appelle coefficients de mélange fort, les réels $\alpha(n)$ définis par

$$\alpha(n) = \sup_{\{k \in \mathbb{Z}, A \in \sigma_{-\infty}^k, B \in \sigma_{n+k}^{+\infty}\}} |P(A \cap B) - P(A)P(B)|, \quad (5.36)$$

et on dit que la famille $\{\Delta_i, i \in \mathbb{Z}\}$ est fortement mélangeante, ou α -mélangeante, si l'on a

$$\lim_{n \rightarrow \infty} \alpha(n) = 0. \quad (5.37)$$

On définit alors deux types particuliers de mélange fort :

- s'il existe deux constantes $c \in \mathbb{R}^{*+}$ et $a \in \mathbb{R}^{*+}$ telles que les coefficients de mélange vérifient

$$\alpha(n) \leq cn^{-a} \quad (5.38)$$

on dit que la famille $\{\Delta_i, i \in \mathbb{Z}\}$ est algébriquement α -mélangeante,

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

$$\alpha(n) \leq st^n, \quad (5.39)$$

on dit que la famille $\{\Delta_i, i \in \mathbb{Z}\}$ est géométriquement α -mélangeante.

Nous rappelons aussi les outils fondamentaux pour l'étude théorique dans le cas d'une famille de variables aléatoires algébriquement α -mélangeante, à savoir l'inégalité de type Fuk-nagaev sous mélange algébrique et l'inégalité de covariances pour variables bornées :

- **Inégalité de type Fuk-Nagaev sous mélange algébrique.** Soit $\{\Delta_i, i \in \mathbb{N}\}$ une famille de variables aléatoires à valeurs dans \mathbb{R} qui vérifient la condition de mélange fort (5.37) avec des coefficients à décroissance algébrique tels que définis en (5.38). On pose

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

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

$$P \left[\left| \sum_{k=1}^n \Delta_k \right| > 4\epsilon \right] \leq 4 \left(1 + \frac{\epsilon^2}{rs_n^2} \right)^{-\frac{r}{2}} + 2ncr^{-1} \left(\frac{2r}{\epsilon} \right)^{a+1}. \quad (5.40)$$

- **Inégalité de covariance pour variables bornées.** Soit $\{\Delta_i, i \in \mathbb{N}\}$ une famille de variables aléatoires à valeurs dans \mathbb{R} qui vérifient la condition de mélange fort (5.37), et telle que $\|\Delta_i\|_\infty < \infty, \forall i$. On a pour tout $i \neq j$:

$$|\text{cov}(\Delta_i, \Delta_j)| \leq 4\|\Delta_i\|_\infty \|\Delta_j\|_\infty \alpha(|i-j|). \quad (5.41)$$

Enfin, pour plus de détails sur ces inégalités, et en particulier en ce qui concerne la bibliographie, nous conseillons au lecteur de revenir au § 3.2 du Chapitre 3.

5.2.3 Régression sous mélange fort ; présentation du modèle et définition de l'estimateur

Plutôt que de poser le modèle en termes de série temporelle, nous allons nous placer dans le cadre plus général du problème de régression d'une v.a.r. Y en une v.a.f. X , ceci sous condition de mélange fort. On s'intéresse donc au modèle :

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

où ϵ est une variable aléatoire réelle centrée et indépendante de X . Ceci revient à estimer $\Phi(x) = E(Y/X = x)$ à partir d'un échantillon $\{(X_i, Y_i), i = 1, \dots, n\}$

identiquement distribué. La principale différence avec ce qui a été fait au Paragraphe 5.1 vient de l'hypothèse de mélangeance s'appliquant aux couples (X_i, Y_i) :

$$\text{la suite } \{(X_i, Y_i), i = 1, \dots, n\} \text{ est } \alpha\text{-mélangeante.} \quad (5.42)$$

L'estimateur que nous étudierons est l'estimateur déjà présenté dans le Paragraphe 5.1 de ce chapitre, à savoir :

$$\widehat{\Phi}_n(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{\|X_i - x\|}{h}\right)}{\sum_{i=1}^n K\left(\frac{\|X_i - x\|}{h}\right)}, \quad \forall x \in \mathcal{H}. \quad (5.43)$$

Par ailleurs, pour pouvoir étendre au cas dépendant les résultats obtenus dans le cas indépendant, nous allons faire une hypothèse supplémentaire agissant sur la distribution conjointe du couple (X_i, X_j) pour $i \neq j$. En effet, nous verrons que la convergence presque complète ponctuelle de notre estimateur $\widehat{\Phi}_n(x)$ se déduit de l'existence d'un réel $\delta_0(x)$ strictement positif tel que, pour $i \neq j$, on ait

$$\lim_{\alpha \rightarrow 0^+} \frac{P((X_i, X_j) \in \mathcal{B}(x, \alpha) \times \mathcal{B}(x, \alpha))}{\alpha^{\delta(x) + \delta_0(x)}} = c(x). \quad (5.44)$$

Notons que lorsque (X_i, X_j) est un couple de variables aléatoires réelles, l'hypothèse (5.44) revient à supposer l'existence d'une densité pour la distribution conjointe du couple (X_i, X_j) , ce qui nous ramène à l'hypothèse (3.17) du Chapitre 3.

5.2.4 Étude asymptotique de s_n^2

Les résultats fournis par le Paragraphe 5.1 proviennent généralement d'une décomposition biais/variance. Sous l'hypothèse de dépendance, seuls les termes issus du calcul de la variance changent puisqu'on est contraint d'inclure des termes de covariance. C'est pourquoi, dans un premier temps, on va se focaliser sur la majoration (pour n suffisamment grand) de la quantité

$$s_n^2 = \sum_{i=1}^n \sum_{j=1}^n \text{cov}(\Delta_i, \Delta_j), \quad (5.45)$$

où les n variables aléatoires $\{\Delta_i\}_{i=1, \dots, n}$ sont définies par :

$$\Delta_i = Y_i^l K\left(\frac{\|x - X_i\|}{h}\right) - E\left(Y_i^l K\left(\frac{\|x - X_i\|}{h}\right)\right), \quad l = 0 \text{ ou } l = 1. \quad (5.46)$$

Remarque. Il faut d'abord s'assurer que les v.a.r. Δ_i ainsi définies forment bien une famille de v.a. fortement mélangeante. Dans ce but, rappelons que la v.a.f. X

est définie sur l'espace probabilisé (Ω, \mathcal{A}, P) et à valeurs dans un espace semi-normé \mathcal{H} . Il suffit alors de munir \mathcal{H} de $\mathcal{B}_{\|\cdot\|}$, la tribu engendrée par les ouverts de \mathcal{H} au sens de la topologie de la semi-norme $\|\cdot\|$ pour que la famille de v.a. $\{\Delta_i\}_i$ soit fortement mélangeante.

Proposition 5.2.1 *Sous les conditions (5.2), (5.6)-(5.8), (5.38) et (5.44), on a :*

$$s_n^2 = o(nh^{\delta(x)}) + O\left(n^2 (h^{\delta^*(x)} \log n)^a\right), \quad (5.47)$$

où $\delta^*(x) = \min\{\delta(x), \delta_0(x)\}$.

Preuve. En effet, K borné avec $\text{support}(K) = [0, \xi]$ implique que

$$K\left(\frac{\|x - X_i\|}{h}\right) \leq C \mathbf{1}_{\mathcal{B}(x, \xi h)}(X_i), \quad \forall i = 1, \dots, n.$$

Sachant que Y est borné, on en déduit que

$$|\text{cov}(\Delta_i, \Delta_j)| \leq CP((X_i, X_j) \in \mathcal{B}(x, \alpha) \times \mathcal{B}(x, \alpha)) + C'P(X_i \in \mathcal{B}(x, \alpha))P(X_j \in \mathcal{B}(x, \alpha)),$$

et en utilisant (5.2) et (5.44) on a :

$$|\text{cov}(\Delta_i, \Delta_j)| = O(h^{\delta(x) + \delta_0(x)}) + O(h^{2\delta(x)}) = O(h^{\delta(x) + \delta^*(x)}).$$

Par ailleurs, comme dans le Paragraphe 3.2, en remarquant d'après (5.41) que

$$|\text{cov}(\Delta_i, \Delta_j)| \leq C\alpha(|i - j|),$$

on peut écrire que :

$$s_n^2 \leq C \sum_{\{|i-j| \leq u_n\}} \sum_{\{|i-j| > u_n\}} h^{\delta(x) + \delta^*(x)} + C' \sum_{\{|i-j| > u_n\}} \alpha(|i - j|).$$

D'où

$$s_n^2 = O(h^{\delta(x) + \delta^*(x)} n u_n + n^2 \alpha(u_n)).$$

Il suffit maintenant de prendre $u_n = 1/(h^{\delta^*(x)} \log n)$ et d'utiliser la définition concernant le mélange algébrique pour obtenir (5.47). \square

Corollaire 5.2.1 *Supposons de plus que les coefficients de mélange sont liés à la largeur de fenêtre par la relation*

$$\exists \theta > 0, h^{a\delta^*(x) - \delta(x)} = O(n^{-1-\theta}). \quad (5.48)$$

Alors on a

$$s_n^2 = o(nh^{\delta(x)}). \quad (5.49)$$

Preuve. Il suffit de remarquer que

$$\frac{n^2 (h^{\delta^*(x)} \log n)^a}{nh^{\delta(x)}} = nh^{a\delta^*(x) - \delta(x)} (\log n)^a.$$

Puisque $h^{a\delta^*(x) - \delta(x)} = O(n^{-1-\theta})$, on a

$$\frac{n^2 (h^{\delta^*(x)} \log n)^a}{nh^{\delta(x)}} \leq C \frac{(\log n)^a}{n^\theta}. \quad \square$$

5.2.5 Convergence presque complète

Théorème 5.2.1 Convergence presque complète ponctuelle. *Sous les hypothèses (5.2), (5.6)-(5.9), (5.44) et en supposant que la condition de décroissance algébrique (5.38) soit satisfaite pour une valeur de a vérifiant*

$$\exists \theta > 0, \exists c_1 > 0, \exists c_2 > 0, c_2 n^{\left(\frac{3-a}{(a+1)\delta(x)} + \theta\right)} \leq h \leq c_1 n^{\left(\frac{1}{\delta(x) - a\delta^*(x)} - \theta\right)}. \quad (5.50)$$

alors :

$$\lim_{n \rightarrow \infty} \widehat{\Phi}_n(x) = \Phi(x) \quad p.co. \quad (5.51)$$

Preuve. La démonstration de ce résultat est calquée sur celle du Théorème 5.1.1. Notons que l'étude du biais, c'est-à-dire les Lemmes 5.1.1 et 5.1.3 restent inchangés. Autrement dit, l'hypothèse de dépendance n'affecte pas les calculs effectués dans la preuve de ces deux lemmes. En revanche, nous allons voir que la démonstration des lemmes 5.1.2 et 5.1.4 dans le cas de dépendance nécessite certains aménagements. Ainsi, pour montrer le Théorème 5.2.1, il suffit de prouver le résultat suivant :

Lemme 5.2.1 *Si les hypothèses (5.2), (5.6)-(5.8), (5.44) et (5.50) sont vérifiées, alors il existe $\nu > 0$ et $\eta > 0$ tel que l'on ait :*

$$P \left(\left| E\hat{f}(x) - \hat{f}(x) \right| > \eta \sqrt{\frac{\log n}{nh^{\delta(x)}}} \right) = O(n^{-1-\nu}), \quad (5.52)$$

et

$$P \left(\left| E\hat{g}(x) - \hat{g}(x) \right| > \eta \sqrt{\frac{\log n}{nh^{\delta(x)}}} \right) = O(n^{-1-\nu}). \quad (5.53)$$

Preuve. Les deux résultats se démontrent simultanément en posant

$$\hat{\psi}(x) = \hat{f}(x) \text{ ou } \hat{g}(x).$$

Notons tout d'abord que l'inégalité de droite de la condition (5.50) implique que (5.48) est réalisée, et donc que (5.49) est vérifiée. Une application directe de (5.40) aux variables

$$\Delta_i = Y_i^l K \left(\frac{\|X_i - x\|}{h^{\delta(x)}} \right) - EY_i^l K \left(\frac{\|X_i - x\|}{h^{\delta(x)}} \right), \quad l = 0 \text{ ou } l = 1$$

amène alors en utilisant (5.49), que pour $\epsilon > 0$ et pour $r > 1$ on a

$$\begin{aligned} P \left(\left| E\hat{\psi}(x) - \hat{\psi}(x) \right| > \epsilon \right) &= P \left(\left| \sum_{i=1}^n \Delta_i \right| > \epsilon n h^{\delta(x)} \right) \\ &\leq 4 \left(1 + \frac{\epsilon^2 n h^{\delta(x)}}{16r} \right)^{-\frac{r}{2}} + 2ncr^{-1} \left(\frac{8r}{nh^{\delta(x)}\epsilon} \right)^{a+1}. \end{aligned}$$

Ainsi on arrive à

$$\begin{aligned} P \left(\left| E\hat{\psi}(x) - \hat{\psi}(x) \right| > \eta \sqrt{\frac{\log n}{nh^{\delta(x)}}} \right) &\leq 4 \left(1 + \frac{\eta^2 \log n}{16r} \right)^{-\frac{r}{2}} + \\ &2ncr^{-1} \left(\frac{8r}{\eta} \right)^{a+1} (nh^{\delta(x)} \log n)^{-\frac{(a+1)}{2}} \end{aligned}$$

On choisit alors r de telle sorte que $\log n = o(r)$, et on arrive à

$$\begin{aligned} P \left(\left| E\hat{\psi}(x) - \hat{\psi}(x) \right| > \eta \sqrt{\frac{\log n}{nh^{\delta(x)}}} \right) &\leq C e^{-\frac{\eta^2 \log n}{32}} + \\ &C \eta^{-(a+1)} n^{1-\frac{(a+1)}{2}} r^a h^{-\frac{\delta(x)(a+1)}{2}}. \end{aligned}$$

On peut toujours choisir r sous la forme $r = C(\log n)^2$ et on arrive alors à

$$\begin{aligned} P \left(\left| E\hat{\psi}(x) - \hat{\psi}(x) \right| > \eta \sqrt{\frac{\log n}{nh^{\delta(x)}}} \right) &\leq C n^{-\frac{\eta^2}{32}} + \\ &C n^{\frac{1-a}{2}} h^{-\frac{\delta(x)(a+1)}{2}} (\log n)^{2a}. \quad (5.54) \end{aligned}$$

Grâce à l'inégalité de gauche de la condition (5.50), il existe $\nu > 0$ pour lequel

$$P \left(\left| E\hat{\psi}(x) - \hat{\psi}(x) \right| > \eta \sqrt{\frac{\log n}{nh^{\delta(x)}}} \right) \leq C n^{-\frac{\eta^2}{32}} + C n^{-1-\nu},$$

d'où pour η suffisamment grand on a

$$P \left(\left| E\hat{\psi}(x) - \hat{\psi}(x) \right| > \eta \sqrt{\frac{\log n}{nh^{\delta(x)}}} \right) < C n^{-1-\nu}. \quad \square$$

Remarque. Pour que la condition (5.50) soit vérifiée, il faut nécessairement que

$$\frac{3-a}{(a+1)\delta(x)} < \frac{1}{\delta(x) - a\delta^*(x)} < 0,$$

ce qui entraîne que le coefficient a de mélange algébrique doit nécessairement satisfaire :

$$a > \frac{3}{2} + \frac{\delta(x)}{\delta^*(x)} + \sqrt{\left(\frac{\delta(x)}{\delta^*(x)}\right)^2 + \frac{\delta(x)}{\delta^*(x)} + \frac{9}{4}}.$$

En particulier, pour $\delta(x) = \delta^*(x) = 1$, on retrouve bien la condition nécessaire (3.43) du Chapitre 3 que doit vérifier a dans le cas réel. Le prochain résultat que nous présentons est une version uniforme du Théorème 5.2.1. Bien entendu, comme précédemment, nous sommes amenés à renforcer certaines hypothèses. Dans ce but, on considère un compact S de \mathcal{H} .

Théorème 5.2.2 Convergence presque complète uniforme. *Si les hypothèses (5.6)-(5.8), (5.20), (5.21) sont satisfaites, si*

$$\lim_{\alpha \rightarrow 0^+} \sup_{x \in S} \left\{ \frac{P((X_i, X_j) \in \mathcal{B}(x, \alpha) \times \mathcal{B}(x, \alpha))}{\alpha^{\delta + \delta_0}} - c'(x) \right\} = 0, \quad (5.55)$$

où δ et δ_0 sont deux constantes strictement positives ne dépendant pas de x dans S , si la condition de décroissance algébrique (5.38) est vérifiée pour une valeur de a telle que

$$\exists \theta > 0, \exists c_1 > 0, \exists c_2 > 0, c_2 n^{\left(\frac{4-a}{\delta(a+1)+\delta+2} + \theta\right)} \leq h \leq c_1 n^{\left(\frac{1}{\delta-a\delta^*} - \theta\right)}, \quad (5.56)$$

avec $\delta^* = \min(\delta, \delta_0)$, alors on a :

$$\lim_{n \rightarrow \infty} \sup_{x \in S} \left| \widehat{\Phi}_n(x) - \Phi(x) \right| = 0 \text{ p.co.} \quad (5.57)$$

Remarque. Si $\delta = \delta_0 = 1$, la condition (5.56) s'écrit

$$\exists \theta > 0, \exists c_1 > 0, \exists c_2 > 0, c_2 n^{\left(\frac{4-a}{4+a} + \theta\right)} \leq h \leq c_1 n^{\left(\frac{1}{1-a} - \theta\right)}.$$

On retrouve bien ainsi la condition explicitée dans le cas réel pour un noyau Lipschitzien d'ordre 1.

Preuve. Notons qu'en remplaçant (5.44) par la condition (5.55) et en utilisant la partie droite de (5.56), le Corollaire 5.2.1 nous donne un résultat uniforme en x sur S , c'est-à-dire :

$$s_n^2 = o(nh^\delta), \quad \forall x \in S. \quad (5.58)$$

Par ailleurs, la trame de la démonstration de (5.57) est analogue à celle du Théorème 5.1.2. En effet, on peut toujours écrire que

$$\widehat{\Phi}_n(x) - \Phi(x) = \frac{1}{\widehat{f}_n(x)} \{ \widehat{g}_n(x) - \Phi(x)C_\delta \} - \frac{\Phi(x)}{\widehat{f}_n(x)} \{ \widehat{f}_n(x) - C_\delta \}, \quad (5.59)$$

où $C_\delta = c(x)\delta \int_0^\xi K(v)v^{\delta-1}dv$ est une constante strictement positive. En reprenant alors les calculs faits pour le Théorème 5.1.2 et en remarquant que toutes les approximations qui y étaient faites sont maintenant uniformes en x à cause de (5.20) on montre, d'une part, que

$$\lim_{n \rightarrow \infty} \sup_{x \in S} |E\widehat{f}_n(x) - C_\delta| = 0, \quad (5.60)$$

et d'autre part, que

$$\lim_{n \rightarrow \infty} \sup_{x \in S} |E\widehat{g}_n(x) - \Phi(x)C_\delta| = 0. \quad (5.61)$$

Dans ce qui suit, $\widehat{\psi}_n$ désigne indifféremment \widehat{f}_n ou \widehat{g}_n :

Lemme 5.2.2

$$\sup_{x \in S} |\widehat{\psi}_n(x) - E\widehat{\psi}_n(x)| = O\left(\sqrt{\frac{\log n}{nh^\delta}}\right) \text{ p.co.} \quad (5.62)$$

Preuve. S étant un compact de l'espace vectoriel topologique $(\mathcal{H}, \|\cdot\|)$, S est un fermé borné. Il existe donc un recouvrement fini de S tel que :

$$S \subset \bigcup_{k=1}^{\tau_n} \mathcal{B}(t_k, l_n),$$

avec $\tau_n l_n = C$ et posons

$$k(x) = \arg \min_{k=1, \dots, \tau_n} \|t_k - x\|.$$

On peut écrire :

$$\begin{aligned} P\left(\sup_{x \in S} |\widehat{\psi}_n(x) - E\widehat{\psi}_n(x)| > 2\eta\sqrt{\frac{\log n}{nh^\delta}}\right) &\leq \\ &\underbrace{P\left(\sup_{x \in S} |\widehat{\psi}_n(x) - \widehat{\psi}_n(t_{k(x)}) - E\widehat{\psi}_n(x) + E\widehat{\psi}_n(t_{k(x)})| > \eta\sqrt{\frac{\log n}{nh^\delta}}\right)}_A + \\ &\underbrace{P\left(\sup_{x \in S} |\widehat{\psi}_n(t_{k(x)}) - E\widehat{\psi}_n(t_{k(x)})| > \eta\sqrt{\frac{\log n}{nh^\delta}}\right)}_B. \end{aligned}$$

- En ce qui concerne A :

$$\begin{aligned} \sup_{x \in S} \left| \widehat{\psi}_n(x) - \widehat{\psi}_n(t_{k(x)}) \right| &\leq \sup_{x \in S} \frac{C}{nh^\delta} \left| \sum_{i=1}^n \left\{ K \left(\frac{\|X_i - x\|}{h} \right) - K \left(\frac{\|X_i - t_{k(x)}\|}{h} \right) \right\} \right| \\ &\leq C \frac{1}{h^{\delta+1}} \sup_{x \in S} \|x - t_{k(x)}\| \\ &\leq C \frac{l_n}{h^{\delta+1}} \end{aligned}$$

D'où, en choisissant $l_n = h^{\frac{\delta}{2}+1} n^{-\frac{1}{2}}$, on a :

$$\frac{l_n}{h^{\delta+1}} = o \left(\sqrt{\frac{\log n}{nh^\delta}} \right),$$

ce qui implique, pour n suffisamment grand, que

$$P \left(\sup_{x \in S} \left| \widehat{\psi}_n(x) - \widehat{\psi}_n(t_{k(x)}) \right| > \eta \sqrt{\frac{\log n}{nh^\delta}} \right) = 0.$$

Par conséquent, pour n suffisamment grand, on a $A = 0$.

- En ce qui concerne B :

$$\begin{aligned} B &\leq P \left(\max_{k=1, \dots, \tau_n} \left| \widehat{\psi}_n(t_{k(x)}) - E\widehat{\psi}_n(t_{k(x)}) \right| > \eta \sqrt{\frac{\log n}{nh^\delta}} \right) \\ &\leq \sum_{k=1}^{\tau_n} P \left(\left| \widehat{\psi}_n(t_{k(x)}) - E\widehat{\psi}_n(t_{k(x)}) \right| > \eta \sqrt{\frac{\log n}{nh^\delta}} \right) \\ &\leq \tau_n \max_{k=1, \dots, \tau_n} P \left(\left| \widehat{\psi}_n(t_{k(x)}) - E\widehat{\psi}_n(t_{k(x)}) \right| > \eta \sqrt{\frac{\log n}{nh^\delta}} \right). \end{aligned}$$

D'où, en utilisant (5.54) et sachant que $\tau_n = C/l_n$, il vient que :

$$B \leq \frac{1}{l_n} \left(Cn^{-\frac{\eta^2}{32}} + Cn^{\frac{1-a}{2}} h^{-\frac{\delta(a+1)}{2}} \right).$$

En prenant $l_n = h^{\frac{\delta}{2}+1} n^{-\frac{1}{2}}$ on obtient :

$$B \leq Cn^{\frac{1}{2}-\frac{\eta^2}{32}} h^{-\frac{\delta}{2}-1} + Cn^{\frac{1-a}{2}+\frac{1}{2}} h^{-\frac{\delta(a+1)}{2}-\frac{\delta}{2}-1}. \quad (5.63)$$

Il est clair qu'on peut choisir η suffisamment grand tel que le 1er terme de droite de l'inégalité (5.63) soit le terme général d'une série convergente. Par ailleurs, l'inégalité de gauche de la condition (5.56) nous permet de dire que le second terme de l'inégalité (5.63) est aussi le terme général d'une série convergente, ce qui termine la preuve du Lemme 5.2.2. \square

Enfin, il suffit d'écrire que

$$\sup_{x \in S} |\widehat{f}_n(x) - C_\delta| \leq \sup_{x \in S} |\widehat{f}_n(x) - E\widehat{f}_n(x)| + \sup_{x \in S} |E\widehat{f}_n(x) - C_\delta|$$

et

$$\sup_{x \in S} |\widehat{g}_n(x) - \Phi(x)C_\delta| \leq \sup_{x \in S} |\widehat{g}_n(x) - E\widehat{g}_n(x)| + \sup_{x \in S} |E\widehat{g}_n(x) - \Phi(x)C_\delta|,$$

d'utiliser (5.59), (5.60), (5.61), (5.62) pour en déduire (5.57), ce qui achève la démonstration du Théorème 5.2.2. \square

5.2.6 Vitesse de convergence presque complète

Théorème 5.2.3 Vitesse de convergence presque complète ponctuelle. *Sous les hypothèses (5.3), (5.6)-(5.8), (5.30) (i.e. $|\Phi(u) - \Phi(v)| \leq C\|u - v\|^\beta$), (5.31) (i.e. $h = h_0 \left(\frac{\log n}{n}\right)^{\frac{1}{2\gamma(x)+\delta(x)}}$ où $\gamma(x) = \min\{b(x), \beta\}$), (5.44) et (5.50), alors*

$$\Phi(x) - \widehat{\Phi}_n(x) = O\left(\left(\frac{\log n}{n}\right)^{\frac{\gamma(x)}{2\gamma(x)+\delta(x)}}\right) \quad p.co. \quad (5.64)$$

Remarque. Lorsque $h = h_0 \left(\frac{\log n}{n}\right)^{\frac{1}{2\gamma(x)+\delta(x)}}$, la condition (5.50) est vérifiée dès que

$$a > \max\left\{3 + 2\frac{\delta(x)}{\gamma(x)}, \frac{2}{\delta^*(x)}(\gamma(x) + \delta(x))\right\}.$$

Preuve. En reprenant la démonstration du Théorème 5.1.3, il vient, d'après (5.33), que

$$E\widehat{f}_n(x) = C_\delta(x) + O(h^{b(x)}),$$

et d'après (5.34), on a

$$E\widehat{g}_n(x) = C_\delta(x)\Phi(x) + O(h^{b(x)}) + O(h^\beta).$$

De plus, en utilisant le Lemme 5.2.1, on a

$$|E\widehat{f}_n(x) - \widehat{f}(x)| = O\left(\sqrt{\frac{\log n}{nh^{\delta(x)}}}\right) \quad p.co.,$$

et

$$|E\widehat{g}_n(x) - \widehat{g}(x)| = O\left(\sqrt{\frac{\log n}{nh^{\delta(x)}}}\right) \quad p.co.$$

Il suffit de combiner ces quatre derniers résultats avec la décomposition

$$\widehat{\Phi}_n(x) - \Phi(x) = \frac{1}{\widehat{f}_n(x)} \{ \widehat{g}_n(x) - \Phi(x) C_\delta(x) \} - \frac{\Phi(x)}{\widehat{f}_n(x)} \{ \widehat{f}_n(x) - C_\delta(x) \}$$

pour obtenir

$$\widehat{\Phi}_n(x) - \Phi(x) = O(h^\beta) + O(h^{b(x)}) + O\left(\sqrt{\frac{\log n}{nh^{\delta(x)}}}\right) \quad p.co.$$

Remplaçons maintenant h par $h_0 \left(\frac{\log n}{n}\right)^{\frac{1}{2\gamma(x)+\delta(x)}}$ dans cette dernière équation ; on a ainsi :

$$\Phi(x) - \widehat{\Phi}_n(x) = O\left(\left(\frac{\log n}{n}\right)^{\frac{\gamma(x)}{2\gamma(x)+\delta(x)}}\right) \quad p.co. \quad \square$$

5.3 Exercices

Exercice 5.1 *Un des points essentiels de ce qui précède réside dans l'interprétation de l'hypothèse fondamentale (5.2) en termes de dimension fractale. En effet, par définition, on appelle dimension (fractale) ponctuelle de la mesure de probabilité de X (Cf. eg. Pesin, 1993, page 533 ou Bardet, 1997, Page 34), le réel positif $\delta(x)$ tel que*

$$\lim_{\alpha \rightarrow 0^+} \frac{\log P(X \in \mathcal{B}(x, \alpha))}{\log \alpha} = \delta(x). \quad (5.65)$$

Montrer que l'hypothèse (5.2) implique que $\delta(x)$ est la dimension (fractale) ponctuelle de la mesure de probabilité de X .

Exercice 5.2 *On suppose que X est une v.a.f. à valeurs dans un espace de Hilbert H muni du produit scalaire $\langle \cdot, \cdot \rangle$. Montrer que les Théorèmes 5.1.1, 5.1.2 et 5.1.3 peuvent toujours s'appliquer lorsque $\widehat{\Phi}_n$ est défini par :*

$$\widehat{\Phi}_n(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{|\langle \phi, X_i - x \rangle|}{h}\right)}{\sum_{i=1}^n K\left(\frac{|\langle \phi, X_i - x \rangle|}{h}\right)}$$

où ϕ est un élément connu de H .

Exercice 5.3 *En vous inspirant de la démonstration du Théorème 5.1.3, en remplaçant (5.3) par*

$$\sup_{x \in S} \left| \frac{P(X \in \mathcal{B}(x, \alpha))}{\alpha^\delta} - c(x) \right| = O(\alpha^b) \quad \text{et} \quad \inf_{x \in S} c(x) > 0,$$

où b et δ sont deux réels strictement positifs ne dépendant pas de x et où $O(\alpha^b)$ ne dépend pas aussi de x ; montrer alors que

$$\sup_{x \in S} |\Phi(x) - \widehat{\Phi}_n(x)| = O\left(\left(\frac{\log n}{n}\right)^{\frac{\gamma}{2\gamma+\delta}}\right) \text{ p.co.}$$

Exercice 5.4 En vous inspirant de l'exercice précédent ainsi que des Théorèmes 5.2.1 et 5.2.2, donner une version uniforme du Théorème 5.2.3.

Exercice 5.5 Soit Y une variable aléatoire réelle (v.a.r) et X une variable aléatoire fonctionnelle à valeurs dans un espace vectoriel \mathcal{F} muni de la semi-norme $\|\cdot\|$. On s'intéresse au modèle de régression non-paramétrique

$$Y = R(X) + \varepsilon,$$

où R est un opérateur défini sur \mathcal{F} et à valeurs dans \mathbb{R} et ε une v.a.r. d'espérance nulle et indépendante de X . Soit $(X_i, Y_i)_{i=1, \dots, n}$ n v.a. identiquement et indépendamment distribuées de même loi que (X, Y) . Dans tout ce qui suit, x désigne un point fixé de \mathcal{F} . On définit alors un estimateur $\widehat{R}_n(x)$ de $R(x)$ par :

$$\widehat{R}_n(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{\|X_i - x\|}{h_n(x)}\right)}{\sum_{i=1}^n K\left(\frac{\|X_i - x\|}{h_n(x)}\right)},$$

où $h_n(x)$ est une suite de nombres positifs telle que :

$$\lim_{n \rightarrow +\infty} h_n(x) = 0 \text{ et } \lim_{n \rightarrow +\infty} \frac{nh_n(x)}{\log n} = +\infty,$$

et où K est un noyau borné tel que

$$\int_{\mathbb{R}} K(t) dt = 1, \quad \text{support}(K) = [0, a].$$

L'objectif de ce problème est d'étudier le comportement asymptotique de \widehat{R}_n sous l'hypothèse fondamentale suivante :

$$(H.1) \left\{ \begin{array}{l} \text{la v.a.r. } Z(x) = \|X - x\| \text{ admet une densité par rapport à la} \\ \text{mesure de Lebesgue sur } \mathbb{R}, \text{ notée } f_{Z(x)}, \text{ telle que : } f_{Z(x)}(0) \in \mathbb{R}_+^*. \end{array} \right.$$

On suppose de plus qu'il existe deux constantes C_1 et β strictement positives pour lesquelles $\forall (u, v) \in \mathbb{R}^2$, $|f_{Z(x)}(u) - f_{Z(x)}(v)| \leq C_1 |u - v|^\beta$, une constante C_2 strictement positive telle que, $\forall (x_1, x_2) \in \mathcal{F}^2$, $|R(x_1) - R(x_2)| \leq C_2 \|x_1 - x_2\|^\beta$ ainsi qu'une constante C_3 strictement positive telle que $|Y| < C_3 < +\infty$ p.s.

- 1) *Question préliminaire.* La largeur de fenêtre $h_n(x)$ dépend clairement de x ; $h_n(x)$ est alors appelée largeur de fenêtre locale. Expliquer rapidement pourquoi le fait de considérer une largeur de fenêtre locale peut s'avérer intéressant ?
- 2) *Vitesse de convergence ponctuelle.* Le but de cette première partie est de montrer que, dans ce cadre fonctionnel particulier, on peut obtenir les mêmes vitesses de convergence ponctuelle que dans le cas réel.

Posons :

$$\hat{f}_n(x) = \frac{1}{nh_n(x)} \sum_{i=1}^n K\left(\frac{\|X_i - x\|}{h_n(x)}\right).$$

et

$$\hat{g}_n(x) = \frac{1}{nh_n(x)} \sum_{i=1}^n Y_i K\left(\frac{\|X_i - x\|}{h_n(x)}\right).$$

2.1) Montrer qu'on peut écrire

$$E\hat{f}_n(x) - f_{Z(x)}(0) = \int_{\mathbb{R}} \frac{1}{h_n(x)} K\left(\frac{z}{h_n(x)}\right) (f_{Z(x)}(z) - f_{Z(x)}(0)) dz.$$

En déduire que

$$E\hat{f}_n(x) - f_{Z(x)}(0) = O(h_n^\beta(x)).$$

2.2) Montrer que

$$\hat{f}_n(x) - E\hat{f}_n(x) = O\left(\sqrt{\frac{\log n}{nh_n(x)}}\right) \text{ p.s.}$$

2.3) Montrer, d'une part, que

$$E\hat{g}_n(x) = \frac{1}{h_n(x)} E\left(R(X)K\left(\frac{\|X - x\|}{h_n(x)}\right)\right),$$

et d'autre part, que

$$E\hat{g}_n(x) - R(x)f_{Z(x)}(0) = O(h_n^\beta(x)).$$

2.4) Donner la vitesse de convergence presque sûre de $\hat{g}_n(x) - E\hat{g}_n(x)$.

2.5) En reprenant les questions 1.1), 1.2), 1.3) et 1.4), en déduire la vitesse de convergence presque sûre de $\hat{R}_n(x) - R(x)$.

2.6) Conclure.

3) Lien avec la dimension fractale ponctuelle. Cette partie s'intéresse à l'hypothèse (5.2) concernant la notion de dimension fractale ponctuelle. En effet, dont rappelons ici que le réel strictement positif $\delta(x)$ tel que

$$\lim_{\alpha \rightarrow 0^+} \frac{P(X \in \mathcal{B}(x, \alpha))}{\alpha^{\delta(x)}} = c(x) \in \mathbb{R}_+^*, \quad (5.66)$$

peut s'interpréter en terme de dimension fractale où $\mathcal{B}(x, \alpha)$ désigne la boule de centre x et de rayon α pour la topologie associée à la semi-norme $\|\cdot\|$.

3.1) Montrer que l'hypothèse fondamentale (H.1) concernant la distribution de $Z(x) = \|X - x\|$ implique que la dimension fractale ponctuelle de X est égale à 1.

3.2) Dans cette question, on suppose que (H.1) n'est pas nécessairement vérifiée. Considérons le cas particulier où $\mathcal{F} = \mathbb{R}^p$ et supposons que la distribution de X vérifie l'hypothèse suivante :

$$(H.2) \left\{ \begin{array}{l} \text{la mesure de probabilité de } X \text{ admet, par rapport à la mesure} \\ \text{de Lebesgue dans } \mathbb{R}^p, \text{ une densité } f_X \text{ continue telle que :} \\ f_X(x) > 0. \end{array} \right.$$

3.2.1) Posons $V_x(\alpha) = \int_{\mathcal{B}(x, \alpha)} dt$; montrer que

$$P(X \in \mathcal{B}(x, \alpha)) = V_x(\alpha)f_X(x) + o(V_x(\alpha)) \quad \text{quand } \alpha \rightarrow 0^+.$$

3.2.2) Montrer que $V_x(\alpha) = \alpha^p V_x(1)$; en déduire que $\delta(x) = p$.

3.2.3) Que vous inspirent les résultats obtenus aux questions 3.1) et 3.2.2) ?

3.3) On rappelle que pour tout $x = (x_1, \dots, x_p)$ appartenant à \mathbb{R}^p , $\|x\|_\infty = \max_{i \in \{1, \dots, p\}} |x_i|$ et plaçons-nous maintenant dans le cas particulier où $\mathcal{F} = \mathbb{R}^p$ et $\|\cdot\| = \|\cdot\|_\infty$.

3.3.1) Sous l'hypothèse (H.2), préciser $c(x)$.

3.3.2) Étudier la compatibilité entre les hypothèses (H.1) et (H.2) par un calcul direct, en supposant de plus que les p coordonnées du vecteur aléatoire X sont indépendantes.

3.3.3) Ceci confirme-t-il les conclusions explicitées à la question 3.2.3) ?

Chapitre 6

Un modèle de régression paramétrique pour variable fonctionnelle : le modèle linéaire fonctionnel

Ce chapitre concerne l'étude théorique du modèle (de régression) linéaire fonctionnel, directement issu des travaux de Cardot, Ferraty et Sarda (1999). Quelques approfondissements liés à des problèmes de lissage pourront être trouvés dans Cardot, Ferraty et Sarda (2000). Ce chapitre est conçu pour donner des éléments de théorie nécessaires aux développements asymptotiques liés au modèle linéaire fonctionnel. Par ailleurs, chronologiquement, le modèle linéaire fonctionnel a été introduit avant les modèles non-paramétriques pour variables aléatoires fonctionnelles (Ferraty et Vieu, 2000). En effet, l'usage intensif des modèles classiques de régression linéaire, tels que le modèle linéaire généralisé, ne sont parfois pas adaptés dans certaines études statistiques. C'est le cas lorsque les variables explicatives sont des points de discrétisation d'une même courbe. On trouve des exemples illustrant ce phénomène dans plusieurs domaines d'application : en chimie notamment où la variable à expliquer représente une caractéristique chimique d'un élément et les variables explicatives sont la discrétisation d'un même signal photométrique (cf. Frank et Friedman, 1993). Hastie, Buja et Tibshirani (1995) décrivent un autre exemple dans lequel une syllabe émise par une personne est digitalisée afin d'analyser la valeur du phonème : *ao* comme dans *water* ou *aa* comme dans *dark* (cf. également Hastie et Mallows, 1993, qui traitent un exemple en physique). D'autres applications, en particulier en météorologie, ont été proposées par Ramsay et Silverman (1997).

Dans un tel contexte, il arrive souvent que le nombre de variables explicatives

soit supérieur à la taille de l'échantillon et, par ailleurs d'après la nature même du problème, ces variables sont fortement corrélées entre elles. Le modèle de régression linéaire introduit doit donc tenir compte de ces contraintes ainsi que de la nature "fonctionnelle" des variables explicatives, d'où la terminologie de modèle linéaire fonctionnel.

Notons que les problèmes théoriques posés par le modèle linéaire fonctionnel se traduisent en termes d'estimation d'un opérateur linéaire. C'est pourquoi, le premier paragraphe rappelle quelques éléments de la théorie des opérateurs linéaires. Par ailleurs, l'aspect fonctionnel des données se formalise en considérant des variables aléatoires à valeurs dans un espace hilbertien, ce qui explique que le deuxième paragraphe donne des rappels et/ou compléments concernant des notions élémentaires autour des variables aléatoires hilbertiennes. Enfin, le troisième et dernier paragraphe rentre dans le vif du sujet puisqu'il est consacré aux résultats asymptotiques obtenus dans le cadre du modèle linéaire fonctionnel.

6.1 Éléments de la théorie des opérateurs linéaires

Ce paragraphe est consacré à des rappels et/ou compléments sur des notions élémentaires de la théorie des opérateurs linéaires. Nous nous en tiendrons à des notions basiques mais néanmoins nécessaires pour pouvoir appréhender les problèmes statistiques posés par le modèle de régression linéaire fonctionnelle. Ajoutons que ces rappels s'inspirent largement du cours de DEA de Jacques Dauxois. Toutefois, si le lecteur souhaite approfondir ses connaissances sur la théorie des opérateurs linéaires, il peut consulter Dunford et Schwartz (1963), Gohberg et Krejn (1971), Chatelin (1983), Kato (1976) et Nagy et Riesz (1952).

6.1.1 Les espaces de Hilbert

Définition 6.1.1 *On appelle espace de Banach tout espace vectoriel normé complet.*

Définition 6.1.2 *On appelle espace de Hilbert tout espace de Banach muni d'un produit scalaire.*

Soit H un espace de Hilbert (ou hilbertien); on note $\langle \cdot, \cdot \rangle_H$ le produit scalaire (p.s.) dans H et $\|\cdot\|_H$ la norme associée à ce produit scalaire.

Définition 6.1.3 *Une suite $(e_n)_{n \in \mathbb{N}^*}$ d'éléments de H est une base hilbertienne (ou dénombrable) orthonormale de H si :*

$$i) \forall (n, m) \in \mathbb{N}^* \times \mathbb{N}^*, \langle e_n, e_m \rangle_H = \delta_{nm},$$

$$ii) \forall x \in H, \exists (a_n)_{n \in \mathbb{N}^*} \in \mathbb{R}^{\mathbb{N}^*}, x = \sum_{i=1}^{\infty} a_n e_n.$$

Définition 6.1.4 H est dit **séparable** s'il existe une base dénombrable de H .

Exemple : soit une mesure σ -finie sur l'espace mesurable $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. On note $L^2(\lambda)$ le \mathbb{R} -espace vectoriel des fonctions réelles $\mathcal{B}_{\mathbb{R}}$ -mesurables de carré λ -intégrables. Alors, on montre que $L^2(\lambda)$, muni du p.s. $\langle f, g \rangle_H = \int_{\mathbb{R}} f(x)g(x)d\lambda(x)$, est un espace de Hilbert séparable.

6.1.2 Opérateurs linéaires

H, H_1 et H_2 désignent des espaces de Hilbert réels.

Définition 6.1.5 On appelle **opérateur linéaire** toute application linéaire de H_1 dans H_2 .

Définition 6.1.6 Soit $(x, y) \in H_1^* \times H_2^*$; le **produit tensoriel** de x par y , noté $x \otimes y$ est défini par :

$$x \otimes y : \begin{cases} H_1 & \rightarrow H_2 \\ u & \mapsto x \otimes y(u) = \langle u, x \rangle y \end{cases}$$

Il est clair que $x \otimes y$ est un opérateur linéaire.

Définition 6.1.7 T est un opérateur **borné** de H_1 dans H_2 si et seulement si il existe une constante M strictement positive telle que, pour tout x appartenant à H_1 , on a $\|T(x)\|_{H_2} \leq M\|x\|_{H_1}$.

Propriété 6.1.1 Tout opérateur linéaire est borné si et seulement si il est continu.

Propriété 6.1.2 Posons $\mathcal{L}(H_1, H_2)$ le \mathbb{R} -e.v. des opérateurs linéaires continus de H_1 dans H_2 et muni de la norme $\|\cdot\|_{\infty}$, appelée norme de la **convergence uniforme** et définie par :

$$\forall T \in \mathcal{L}(H_1, H_2), \|T\|_{\infty} = \sup_{\|x\|_{H_1}=1} \|T(x)\|_{H_2} = \sup_{x \in H_1^*} \frac{\|T(x)\|_{H_2}}{\|x\|_{H_1}}.$$

Alors $(\mathcal{L}(H_1, H_2), \|\cdot\|_{\infty})$ est un espace de Banach.

Remarque. Si Π_F est un projecteur de H dans F ($F \subset H$), alors $\|\Pi_F\|_{\infty} = 1$. En effet, on a, d'une part, $\|\Pi_F(x)\|_H \leq \|x\|_{\infty}$, et d'autre part, $x \in F \Rightarrow \Pi_F(x) = x$.

6.1.3 Convergences d'une suite d'opérateurs

Soit $(T_n)_{n \in \mathbb{N}^*}$ une suite d'opérateurs de $\mathcal{L}(H_1, H_2)$ où H_1 et H_2 sont deux espaces de Hilbert séparables. On définit alors les différents modes de convergences suivants :

- i) **Convergence faible** : $(T_n)_{n \in \mathbb{N}^*}$ converge faiblement vers T , noté $T_n \xrightarrow{w} T$ avec w pour *weak*, si et seulement si

$$\forall u \in H_1, \forall v \in H_2, \lim_{n \rightarrow \infty} \langle T_n(u), v \rangle_{H_2} = \langle T(u), v \rangle_{H_2} .$$

- ii) **Convergence forte** : $(T_n)_{n \in \mathbb{N}^*}$ converge forte vers T , noté $T_n \xrightarrow{s} T$ avec s pour *strong*, si et seulement si

$$\forall u \in H_1, \lim_{n \rightarrow \infty} T_n(u) = T(u) .$$

- iii) **Convergence uniforme** : $(T_n)_{n \in \mathbb{N}^*}$ converge uniformément vers T , noté $T_n \xrightarrow{u} T$, si et seulement si

$$\lim_{n \rightarrow \infty} \|T_n - T\|_{\infty} = 0 .$$

Remarque. $T_n \xrightarrow{u} T \Rightarrow T_n \xrightarrow{s} T \Rightarrow T_n \xrightarrow{w} T$ puisque
 $|\langle (T_n - T)(u), v \rangle_{H_2}| \leq \|v\|_{H_2} \|(T_n - T)(u)\|_{H_2} \leq \|u\|_{H_1} \|v\|_{H_2} \|T_n - T\|_{\infty} .$

6.1.4 Opérateurs linéaires adjoints

Définition 6.1.8 Dual topologique. On appelle *dual topologique* de H , le \mathbb{R} -e.v. H' contenant les formes linéaires continues sur H .

On utilise alors le crochet de dualité $\langle \cdot, \cdot \rangle_{H, H'}$ défini par :

$$\forall \phi \in H', \forall x \in H, \phi(x) = \langle x, \phi \rangle_{H, H'} .$$

et on a, par définition,

$$\forall \phi \in H', \|\phi\|_{H'} = \sup_{\|x\|_H=1} |\langle x, \phi \rangle_{H, H'}| = \sup_{x \in H/\{0\}} \frac{|\langle x, \phi \rangle_{H, H'}|}{\|x\|_H} ,$$

ce qui implique que $\forall x \in H, \forall \phi \in H', |\langle x, \phi \rangle_{H, H'}| \leq \|x\|_H \|\phi\|_{H'}$.

Pour tout x appartenant à H , on note M_x l'application définie par :

$$M_x : \begin{cases} H & \rightarrow \mathbb{R} \\ u & \mapsto M_x(u) = \langle u, x \rangle_H . \end{cases}$$

Il est clair que M_x est un élément de H' et donc que $M_x(u) = \langle u, M_x \rangle_{H, H'}$. par ailleurs, on a que $\|M_x\|_{H'} = \|x\|_H$.

Théorème 6.1.1 Théorème de Riesz. *L'application M , définie de H dans H' et qui à tout x associe M_x est une bijection, c'est-à-dire :*

$$\forall \phi \in H', \exists ! y \in H / \forall x \in H, \phi(x) = \langle y, x \rangle_H .$$

Remarque. Comme $\|M_x\|_{H'} = \|x\|_H$, M est un isomorphisme isométrique appelé isométrie canonique de H (sur H').

Définition 6.1.9 Opérateur adjoint. *Pour tout T de $\mathcal{L}(H_1, H_2)$, il existe un unique opérateur linéaire T^* de $\mathcal{L}(H_2, H_1)$, appelé adjoint de T , tel que :*

$$\forall x \in H_1, \forall y \in H_2, \langle T(x), y \rangle_{H_2} = \langle x, T^*(y) \rangle_{H_1} .$$

Exemple : $(x \otimes y)^* = y \otimes x$.

Propriété 6.1.3 *i) $(T \circ U)^* = U^* \circ T^*$, ii) $(T^*)^* = T$.*

Définition 6.1.10 $T \in \mathcal{L}(H)$ ($= \mathcal{L}(H, H)$) est dit **autoadjoint** si $T^* = T$; T est dit **autoadjoint positif** si, de plus, on a pour tout x appartenant à H , $\langle T(x), x \rangle_H \geq 0$.

Exemple : $T \circ T^*$ est autoadjoint positif.

Définition 6.1.11 Si $T \in \mathcal{L}(H)$ est autoadjoint positif, il existe un unique opérateur S de $\mathcal{L}(H)$, autoadjoint positif tel que

$$S \circ S = T .$$

Cet opérateur S , noté $T^{1/2}$, est appelé **racine carrée** de T .

6.1.5 Valeurs propres et valeurs singulières

Soit H un \mathbb{C} -espace hilbertien.

Définition 6.1.12 Soit $T \in \mathcal{L}(H)$. On note $sp(T)$ l'ensemble défini par :

$$sp(T) = \{ \lambda \in \mathbb{C} / \exists x \in H - \{0\}, Tx = \lambda x \} .$$

*i) Tout élément de $sp(T)$ est appelé **valeur propre** de T alors que $sp(T)$ est dit **spectre** de T .*

ii) Soit λ un élément de $sp(T)$; $E_\lambda = \ker(T - \lambda I)$ est appelé **sous-espace propre** associé à la valeur propre λ ; tout élément non nul de E_λ est appelé **vecteur propre** de T .

Définition 6.1.13 On appelle **valeur singulière** de T , toute valeur propre de l'opérateur $(T^* \circ T)^{1/2}$; si $S(T)$ est l'ensemble des valeurs singulières de T , on a par définition :

$$S(T) = sp\left((T^* \circ T)^{\frac{1}{2}}\right).$$

6.1.6 Opérateurs compacts

Définition 6.1.14 $T \in \mathcal{L}(H_1, H_2)$ est dit **compact** (ou *complètement continu*) si l'image par T de toute partie bornée de H_1 est relativement compacte (i.e. d'adhérence compacte) dans H_2 .

Définition 6.1.15 (équivalente pour des espaces hilbertiens) T est compact si et seulement si il transforme toute suite faiblement convergente dans H_1 en une suite fortement convergente dans H_2 .

Propriétés pour des opérateurs compacts, autoadjoints, positifs (ocap) :

Soit T (resp. S) un ocap de H :

- 1) $sp(T)$ est fini ou dénombrable et $sp(T) \subset \mathbb{R}^+$,
- 2) le seul point d'accumulation de $sp(T)$ est 0,
- 3) toute valeur propre λ non nulle est de multiplicité finie ($\dim E_\lambda < +\infty$),
- 4) $S(T) = sp(T)$,
- 5) E_λ est orthogonal à E_μ dès que $\lambda \neq \mu$,
- 6) Soit $(\lambda_1, \lambda_2, \dots)$ la suite strictement décroissante des valeurs propres de T (i.e. chaque valeur propre apparaît une seule fois); alors

$$\lambda_1 = \sup_{\|u\|_H=1} \langle Tu, u \rangle_H,$$

et

$$\lambda_k(T) = \sup_{\{u \in H / \|u\|_H=1, u \in (\otimes_{j=1}^{k-1} E_{\lambda_j})^\perp\}} \langle Tu, u \rangle_H,$$

- 7) T ocap $\Rightarrow \|T\|_\infty = \lambda_1(T)$,
- 8) Si S et T sont deux ocap alors $|\lambda_k(S) - \lambda_k(T)| \leq \|T - S\|_\infty (= \lambda_1(S - T))$,

9) Soit S et T deux opac et soit $v_k(S)$ (resp. $v_k(T)$) le vecteur propre S (resp. T) associé à la valeur propre $\lambda_k(S)$ (resp. $\lambda_k(T)$). Alors on a :

$$\|v_k(S) - v_k(T)\|_H \leq \frac{2\sqrt{2}}{\lambda_1(S) - \lambda_2(S)} \|S - T\|_\infty.$$

On appelle suite pleine décroissante des valeurs propres (non nulles) de l'opac T toute énumération $(\lambda_i(T))_{i \in I}$, où $I \subset \mathbb{N}^*$, chaque valeur propre figurant autant de fois que son ordre de multiplicité, la suite étant donnée par ordre décroissant.

Définition 6.1.16 Décomposition de Schmidt. Soit $(\lambda_i)_{i \in I}$ la suite pleine décroissante de valeurs propres (\neq) de l'opac T , $(e_i)_{i \in I}$ un système orthonormal de vecteurs propres (e_i est un vecteur propre unitaire associé à λ_i). Alors, on a :

$$T = \sum_{i \in I} \lambda_i e_i \otimes e_i \text{ dans } \mathcal{L}(H).$$

Conséquences :

- i) $T^{1/2} = \sum_{i \in I} \sqrt{\lambda_i} e_i \otimes e_i$,
- ii) si T^{-1} existe, alors $T^{-1} = \sum_{i \in I} (1/\lambda_i) e_i \otimes e_i$.

Proposition 6.1.1 Caractérisation d'un opérateur compact. $T \in \mathcal{L}(H)$ est un opérateur autoadjoint positif; T est compact si et seulement si il existe une suite pleine décroissante de valeurs propres $(\lambda_i)_{i \in I}$ tendant vers 0 avec i et une base $(e_i)_{i \in I}$ orthonormale telle que

$$T = \sum_{i \in I} \lambda_i e_i \otimes e_i.$$

Proposition 6.1.2 (Dunford et Schwartz, p. 1091). Soit $(T_n)_{n \in \mathbb{N}^*}$ une suite uniformément convergente vers T d'opérateurs compacts sur H et $(\lambda_i(T))_{i \in I}$ la suite pleine décroissante des valeurs propres non nulles de T . Alors, il existe des énumérations $(\lambda_i(T_n))_{i \in I}$ des valeurs propres non nulles des T_n (chaque valeur propre étant répétée suivant son ordre de multiplicité) telles que l'on ait, uniformément en i :

$$\lim_{n \rightarrow \infty} \sup_{i \in I} |\lambda_i(T_n) - \lambda_i(T)| = 0.$$

Soit $\Phi_p(T)$ l'application définie par :

$$\Phi_p(T) = \begin{cases} \left(\sum_{i \in I} |\lambda_i|^p \right)^{\frac{1}{p}} & \text{si } p \in [1, +\infty[, \\ \sup_{i \in I} |\lambda_i| & \text{si } p = +\infty, \end{cases}$$

et on note $\sigma_p(H) = \{T \in \mathcal{L}(H) ; \Phi_p(T) < +\infty\}$ et pour tout T appartenant à $\sigma_p(H)$, on pose

$$\|T\|_p = \Phi_p(T).$$

Propriété 6.1.4 Pour tout $(p, q) \in \mathbb{N}^2$ tels que $1 \leq p \leq q \leq +\infty$, on a :

$$\sigma_1(H) \subset \sigma_p(H) \subset \sigma_q(H) \subset \sigma_\infty(H),$$

ce qui revient à écrire de manière équivalente que :

$$\|T\|_1 \geq \|T\|_p \geq \|T\|_q \geq \|T\|_\infty.$$

Définition 6.1.17 Opérateur nucléaire. Tout élément de $\sigma_1(H)$ est appelé opérateur nucléaire sur H et la norme $\|\cdot\|_1$ est dite **norme trace** ($\|T\|_1 = \sum_{i \in I} |\lambda_i| = \text{tr}(T)$).

Proposition 6.1.3 Caractérisation d'un opérateur nucléaire. T est un opérateur nucléaire si et seulement si il existe une suite pleine de valeurs propres $(\lambda_i)_{i \in I}$ avec $\sum_{i \in I} |\lambda_i| < \infty$ et une base orthonormée $(e_i)_{i \in I}$ telles que

$$T = \sum_{i \in I} \lambda_i e_i \otimes e_i.$$

Proposition 6.1.4 Soit T un opérateur autoadjoint positif. T est un opérateur nucléaire si et seulement si il existe une base orthonormale $(e_i)_{i \in I}$ de H telle que $\sum_{i \in I} \langle T(e_i), e_i \rangle < +\infty$, cette somme ne dépendant pas de la base choisie.

Définition 6.1.18 Opérateur de Hilbert-Schmidt. Tout élément de $\sigma_2(H)$ est appelé opérateur de Hilbert-Schmidt et la norme $\|\cdot\|_2$ est dite **norme de Hilbert-Schmidt** ($\|T\|_2 = \sqrt{\sum_{i \in I} |\lambda_i|^2}$).

Définition 6.1.19 On appelle produit scalaire de Hilbert-Schmidt, noté $\langle \cdot, \cdot \rangle_2$, le p.s. défini par :

$$\langle \cdot, \cdot \rangle_2 : \begin{cases} \sigma_2(H) \times \sigma_2(H) & \rightarrow \mathbb{R} \\ (S, T) & \mapsto \langle S, T \rangle_2 = \sum_{i \in I} \langle S(e_i), T(e_i) \rangle_H = \text{tr}(S^* \circ T) = \text{tr}(T \circ S) \end{cases}$$

où $(e_i)_{i \in I}$ est une base orthonormée de H .

Remarque. Pour tout T appartenant à $\sigma_2(H)$, on a $\|T\|_2 = (\langle T, T \rangle_2)^{1/2}$.

Proposition 6.1.5 $(\sigma_2(H), \langle \cdot, \cdot \rangle_2)$ est un espace de Hilbert séparable

Proposition 6.1.6 Caractérisation des opérateurs de Hilbert-Schmidt. T est un opérateur de Hilbert-Schmidt si et seulement si pour toute base orthonormée $(e_i)_{i \in I}$ de H on a $\sum_{i \in I} \|T(e_i)\|_H^2 < \infty$.

6.2 Notions élémentaires autour des variables aléatoires hilbertiennes

L'objectif de ce paragraphe est de donner un certain nombre d'outils nécessaires à la manipulation des variables aléatoires hilbertiennes (v.a.h.). Rappelons que notre préoccupation principale reste le modèle de régression linéaire fonctionnelle. C'est pourquoi nous nous contenterons de donner uniquement les résultats qui nous paraissent les plus pertinents relativement au modèle de régression linéaire fonctionnelle. Pour ceux qui souhaitent approfondir leurs connaissances dans ce domaine, nous conseillons les ouvrages de Grenander (1963) et Parthasarathy (1967).

6.2.1 Notations et définition

Soit H un espace de Hilbert réel séparable (i.e. il existe une base dénombrable) muni du produit scalaire $\langle \cdot, \cdot \rangle$. On note $\|\cdot\|$ la norme associée à ce produit scalaire. \mathcal{B}_H désigne la tribu borélienne de H ; H' désigne le dual topologique (ensemble des formes linéaires continues) de H . Enfin, (Ω, \mathcal{A}, P) est un espace probabilisé.

Définition 6.2.1 Variable aléatoire hilbertienne. *Toute application mesurable de (Ω, \mathcal{A}) dans (H, \mathcal{B}_H) est dite variable aléatoire hilbertienne (v.a.h.).*

6.2.2 Intégration de v.a.h.

On pose

$$L_{\mathbb{R}}^1(\Omega, \mathcal{A}, P) = \left\{ Z : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}}) / \int_{\Omega} |Z(\omega)| dP(\omega) < \infty \right\},$$

et on rappelle que Z est P -intégrable si, par définition, Z appartient à $L_{\mathbb{R}}^1(\Omega, \mathcal{A}, P)$. Soit X une v.a.h. et y un élément de H ; si $\|X\|$ est P -intégrable, l'application qui, à tout élément ω de Ω associe le réel $\langle X(\omega), y \rangle$ est P -intégrable puisque $|\langle X(\omega), y \rangle| \leq \|X(\omega)\| \|y\|$. Par ailleurs, l'application

$$\phi \begin{cases} H & \rightarrow \mathbb{R} \\ y & \mapsto \phi(y) = \int_{\Omega} \langle X(\omega), y \rangle dP(\omega) \end{cases}$$

est une forme linéaire continue sur H ; d'après le théorème de Riesz, il existe un unique élément de H , noté EX , tel que

$$\phi(y) = \langle EX, y \rangle \quad \left(= \int_{\Omega} \langle X(\omega), y \rangle dP(\omega) \right)$$

Définition 6.2.2 **Espérance d'une v.a.h.** EX est appelée intégrale de X sur Ω ou encore espérance de X ; EX est aussi notée $\int_{\Omega} X(\omega)dP(\omega)$.

Proposition 6.2.1 Si X est une v.a.h. telle que $\|X\|$ est P -intégrable, alors

$$\|EX\| \leq E\|X\|.$$

Preuve. D'une part, on a :

$$\begin{aligned} \|EX\| &= \|\phi\|_{\infty}, \\ &= \sup_{y \neq 0} \frac{|\phi(y)|}{\|y\|}, \\ &= \sup_{y \neq 0} \frac{|\langle EX, y \rangle|}{\|y\|}, \end{aligned}$$

et d'autre part, on peut écrire :

$$\begin{aligned} |\langle EX, y \rangle| &= |E \langle X, y \rangle|, \\ &\leq E|\langle X, y \rangle|, \\ &\leq \|y\|E\|X\|, \end{aligned}$$

Comme

$$\|EX\| = \|\phi\|_{\infty} = \sup_{y \neq 0} \frac{|\langle EX, y \rangle|}{\|y\|},$$

il vient que $\|EX\| \leq E\|X\|$.

Définition 6.2.3 **V.a.h. intégrable.** X est une v.a.h. P -intégrable si et seulement si $\|X\| \in L^1_{\mathbb{R}}(\Omega, \mathcal{A}, P)$.

6.2.3 Outils asymptotiques pour v.a.h.

Loi des Grands nombres pour des v.a.h.

Soit $\{X_n\}_{n \in \mathbb{N}^*}$ une suite de v.a.h. définies sur (Ω, \mathcal{A}, P) et à valeurs dans (H, \mathcal{B}_H) , indépendantes et identiquement distribuées et posons $\bar{X}(\omega) = \frac{1}{n} \sum_{i=1}^n X_i(\omega)$.

i) **Convergence forte.** Si $E\|X_1\| < +\infty$, alors

$$P \left(\left\{ \omega \in \Omega / \|\bar{X}(\omega) - EX_1\| \xrightarrow{n \rightarrow \infty} 0 \right\} \right) = 1.$$

i) **Convergence en moyenne quadratique.** Si $E\|X\|^2 < \infty$, alors

$$E \left(\|\bar{X}(\omega) - EX_1\|^2 \right) \xrightarrow{n \rightarrow \infty} 0.$$

Théorème de la limite central pour v.a.h.

Soit (H, \mathcal{B}_H, Q) un espace probabilisé, H étant un espace de Hilbert réel séparable.

Définition 6.2.4 Fonction caractéristique. On appelle fonction caractéristique de Q , notée \widehat{Q} , la fonction de H dans \mathbb{C} définie par :

$$\forall x, x \in H, \widehat{Q}(x) = \int_H \exp\{i \langle x, y \rangle\} dQ(y).$$

Propriété 6.2.1 *i) \widehat{Q} est uniformément continue, ii) Si pour tout x élément de H , $\widehat{Q}_1(x) = \widehat{Q}_2(x)$, alors $Q_1 = Q_2$ (i.e. \widehat{Q} caractérise Q), iii) Si, pour tout f continue bornée, $\int_H f(x) dQ_n(x)$ converge vers $\int_H f(x) dQ(x)$ quand n tend vers $+\infty$, alors pour tout $x \in H$, $\widehat{Q}_n(x)$ converge vers $\widehat{Q}(x)$ quand n tend vers $+\infty$.*

Définition 6.2.5 Produit de convolution. On appelle produit de convolution de Q_1 et Q_2 (ou convolée de Q_1 et Q_2), noté $Q_1 * Q_2$, l'unique mesure de probabilité qui vérifie pour toute fonction f continue bornée :

$$\int_H f(x) dQ_1 * Q_2(x) = \int_H \int_H f(x+y) dQ_1(x) dQ_2(y).$$

Propriété 6.2.2 $\widehat{Q_1 * Q_2}(x) = \widehat{Q}_1(x) \widehat{Q}_2(x)$.

On dispose maintenant des notions nécessaires à la définition d'une variable aléatoire hilbertienne **gaussienne** :

Définition 6.2.6 La v.a.h. X est une v.a.h. gaussienne définie sur (H, \mathcal{B}_H) si et seulement si, sa loi de probabilité Q est telle que

$$\forall h \in H, \widehat{Q}(h) = \exp\{i \langle h, EX \rangle - \frac{1}{2} \langle \Gamma_X h, h \rangle\},$$

où $\Gamma_X = E((X - EX) \otimes X)$ est appelé **opérateur de covariance** de la v.a.h. X .

Notation et terminologie : on dit alors que X suit la loi $\mathcal{N}_H(EX, \Gamma_X)$, c-à-d la loi gaussienne d'espérance EX et d'opérateur de covariance Γ_X .

Remarque : X centrée (i.e. $EX = 0 \in H$) implique que $\Gamma_X = E(X \otimes X)$.

Théorème de la limite centrale. Soit $\{X_n\}_{n \in \mathbb{N}^*}$ une suite de v.a.h. définies sur (Ω, \mathcal{A}, P) et à valeurs dans (H, \mathcal{B}_H) , **indépendamment et identiquement distribuées**. Alors, la suite $\left\{ \frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - EX) \right\}_{n \in \mathbb{N}^*}$ converge en loi quand n tend vers l'infini vers la loi gaussienne $\mathcal{N}_H(0, \Gamma_X)$ (définie sur (H, \mathcal{B}_H)).

Inégalité exponentielle pour v.a.h.

Afin de compléter les résultats fondamentaux donnés précédemment, nous donnons ici un exemple d'inégalité exponentielle obtenue pour des suites v.a.h.. Pour plus de précisions, voir Yurinskii (1976).

Inégalité de type Bernstein pour des v.a.h. Soit X_1, X_2, \dots, X_n n v.a.h. indépendantes telles que :

- $\forall i = 1, \dots, n, \quad EX_i = 0,$
- $\forall i = 1, \dots, n, \exists (a_i, b) \in \mathbb{R}^+ \times \mathbb{R}^+ / E(\|X_i\|^m) \leq \frac{m!}{2} a_i^2 b^{m-2}, \forall m \geq 2.$

Alors, en posant $A_n = \sqrt{\sum_{i=1}^n a_i^2}$, on a :

$$P\left(\left\|\sum_{i=1}^n X_i\right\| \geq \varepsilon A_n\right) \leq 2 \exp\left\{-\frac{\varepsilon^2}{2\left(1 + c \frac{\varepsilon b}{A_n}\right)}\right\}.$$

6.3 Modèle linéaire fonctionnel

Ce paragraphe constitue l'élément central de ce chapitre puisqu'il propose une étude *empirique* (dans le sens où l'estimateur construit est dit empirique) du modèle linéaire fonctionnel. Après avoir introduit quelques notations et définitions, on présente dans un premier temps quelques propriétés élémentaires concernant les v.a.h.. Dans un second temps, on définit l'Analyse en Composantes Principales (ACP) Fonctionnelle, technique indispensable dès qu'on s'intéresse à une approximation (de rang donné) d'un opérateur linéaire. Enfin, une fois l'estimateur défini, on précise son comportement asymptotique. Notons que Soit Y une variable aléatoire réelle et $X = (X(t), t \in [0, 1])$ un processus réel à temps continu du second ordre définis sur le même espace probabilisé. On suppose que X et Y sont liés par la relation

$$Y = \int_0^1 \psi(t)X(t) dt + \epsilon, \quad (6.1)$$

où ψ est une fonction définie sur $[0, 1]$, de carré intégrable et ϵ une v.a.r. d'espérance 0, de variance σ^2 et indépendante de X .

Plusieurs auteurs ont proposé des estimateurs de la fonction ψ , reposant sur un échantillon $(X_i, Y_i), i = 1, \dots, n$. Frank et Friedman (1993) passent en revue et comparent différentes méthodes statistiques utilisées en chimie et qui s'appliquent

notamment dans le contexte ci-dessus : *Moindres Carrés pénalisés*, *Ridge Regression* et *Régression sur Composantes Principales*. Hastie et Mallows (1993) proposent un estimateur reposant sur les splines cubiques minimisant un critère de moindres carrés pénalisés. Marx et Eilers (1996) utilisent une base de B-splines et introduisent également une pénalisation dans les moindres carrés permettant ainsi une réduction de la dimension et une régularisation des coefficients de régression. Ils définissent ainsi des splines pénalisés ou P-splines. Nous renvoyons à ce dernier article pour une description détaillée de ces méthodes ainsi qu'une bibliographie.

Ce chapitre est consacré à une méthode d'estimation s'appliquant également dans un contexte plus général en ce sens qu'elle s'applique à des variables aléatoires hilbertiennes. Nous proposons alors, en nous inspirant des travaux de Bosq (1991) dans le cas des processus autorégressifs hilbertiens, un estimateur reposant sur l'analyse spectrale de l'opérateur de covariance empirique que l'on inverse ensuite dans l'espace vectoriel engendré par ses q_n vecteurs propres associés aux q_n plus grandes valeurs propres.

6.3.1 Modèle : définition et notations

Soit H un espace hilbertien séparable, $\langle \cdot, \cdot \rangle_H$ (resp. $\|\cdot\|_H$) désigne le produit scalaire (resp. la norme) dans H et posons

$$L_H^2(P) = \{Z : (\Omega, \mathcal{A}, P) \rightarrow (H, \mathcal{B}_H) / E(\|X\|_H^2) < \infty\}.$$

Dans tout ce qui suit, X désigne une v.a.h. appartenant à $L_H^2(P)$, Y une v.a.r. (i.e. Y est définie sur (Ω, \mathcal{A}, P) et à valeurs dans $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$), H' le dual topologique de H .

Définition 6.3.1 *On appelle modèle linéaire fonctionnel, le modèle défini par :*

$$Y = \Psi(X) + \varepsilon,$$

où Ψ est un élément de H' , ε est une v.a.r. indépendante de X telle que $E\varepsilon = 0$ et $Var\varepsilon = \sigma^2$.

La forme linéaire continue Ψ modélise une "relation linéaire" entre la v.a.h. X et la v.a.r. Y .

6.3.2 Résultats préliminaires

Rappel. $X(\omega) \otimes X(\omega) : \begin{cases} H & \rightarrow H \\ x & \mapsto X(\omega) \otimes X(\omega)(x) = \langle x, X(\omega) \rangle_H X(\omega). \end{cases}$

Proposition 6.3.1 $X \otimes X$ est une v.a.h. de (Ω, \mathcal{A}, P) dans (E, \mathcal{B}_E) P -intégrable où E est l'espace des opérateurs de Hilbert-Schmidt muni du produit scalaire de Hilbert-Schmidt (i.e. $E = (\sigma_2(H), \langle \cdot, \cdot \rangle_2)$).

Remarque. On a vu Chapitre 6.2 que $E = (\sigma_2(H), \langle \cdot, \cdot \rangle_2)$ est un espace de Hilbert séparable.

Preuve.

i) Montrons que $X(\omega) \otimes X(\omega)$ est un élément de E . En effet, soit $(e_i)_{i \in I}$ une base orthonormée de H ; on peut écrire :

$$\begin{aligned} \|X(\omega) \otimes X(\omega)\|_2^2 &= \sum_{i \in I} \langle X(\omega) \otimes X(\omega)(e_i), X(\omega) \otimes X(\omega)(e_i) \rangle_H \\ &= \sum_{i \in I} \langle X(\omega), e_i \rangle_H^2 \|X(\omega)\|_H^2 \\ &= \|X(\omega)\|_H^4. \end{aligned}$$

Or $X \in L_H^2(P)$ implique que $\|X(\omega)\|_H^2 < +\infty$ et donc

$$X(\omega) \otimes X(\omega) \in E.$$

ii) Montrons que $X(\omega) \otimes X(\omega)$ est une application mesurable de (Ω, \mathcal{A}, P) dans (E, \mathcal{B}_E) . Dans ce but, considérons les deux applications f et g définies par :

$$f : \begin{cases} L_H^2(P) & \rightarrow L_H^2(P) \times L_H^2(P) \\ X & \mapsto (X, X) \end{cases} \quad \text{et } g : \begin{cases} L_H^2(P) \times L_H^2(P) & \rightarrow L_H^2(P) \\ (X, Z) & \mapsto X \otimes Z. \end{cases}$$

Comme f et g sont continues, il vient que $g \circ f(X) = X \otimes X$ est une application mesurable de (Ω, \mathcal{A}, P) dans (E, \mathcal{B}_E) .

iii) Il reste à montrer que $X \otimes X$ est P -intégrable. Or, par définition, $X \otimes X$ est P -intégrable si et seulement si $E(\|X \otimes X\|_2) < +\infty$. Par ailleurs, d'après i) on a $\|X(\omega) \otimes X(\omega)\|_2 = \|X(\omega)\|_H^2$ et puisque $X \in L_H^2(P)$, il vient que $X \otimes X$ est P -intégrable.

Définition 6.3.2 Soit $\Gamma_X = E(X \otimes X)$; si $EX \neq 0$, Γ_X est appelé **opérateur du moment d'ordre 2** et si $EX = 0$, Γ_X est appelé **opérateur de covariance**.

Proposition 6.3.2 Γ_X est un opérateur autoadjoint, positif et nucléaire.

Preuve. Soit G_X l'opérateur linéaire défini par :

$$G_X : \begin{cases} L_{\mathbb{R}}^2(P) & \rightarrow H \\ f & \mapsto E(Xf) = \int_{\Omega} X(\omega)f(\omega)dP(\omega). \end{cases}$$

De plus, pour tout $g \in H$, on a :

$$\begin{aligned} \langle G_X f, g \rangle_H &= \langle E(Xf), g \rangle_H \\ &= \langle \int_{\Omega} X(\omega) f(\omega) dP(\omega), g \rangle_H \\ &= \int_{\Omega} \langle X(\omega), g \rangle_H f(\omega) dP(\omega) \\ &= E(\langle f, X, g \rangle_H) \\ &= \langle f, \langle X, g \rangle_H \rangle_{L^2_{\mathbb{R}}(P)}. \end{aligned}$$

Donc l'application qui à tout élément g de H associe l'élément $\langle X, g \rangle_H$ de $L^2_{\mathbb{R}}(P)$ est l'adjoint de G_X c'est-à-dire G_X^* . Par ailleurs, on a $\Gamma_X = G_X \circ G_X^*$ puisque

$$\begin{aligned} G_X \circ G_X^*(g) &= G_X(\langle X, g \rangle_H) \\ &= E(\langle g, X \rangle_H X) \\ &= (E(X \otimes X))(g). \end{aligned}$$

D'où :

- a) $\Gamma_X^* = \Gamma_X$,
- b) $\forall h \in H$, on a

$$\begin{aligned} \langle \Gamma_X h, h \rangle_H &= \langle G_X \circ G_X^* h, h \rangle_H, \\ &= \langle G_X^* h, G_X^* h \rangle_{L^2_{\mathbb{R}}(P)}, \\ &= \|G_X^* h\|_{L^2_{\mathbb{R}}(P)}^2, \end{aligned}$$

ce qui entraîne que $\forall h \in H$, $\langle \Gamma_X h, h \rangle_H \geq 0$.

c)

$$\begin{aligned} \sum_{i \in I} \langle \Gamma_X e_i, e_i \rangle_H &= \sum_{i \in I} \langle E(\langle e_i, X \rangle_H X), e_i \rangle_H, \\ &= \sum_{i \in I} E(\langle e_i, X \rangle_H^2), \\ &= E(\|X\|_H^2), \end{aligned}$$

et comme $X \in L^2_H(P)$, il vient que Γ_X est nucléaire.

Remarque. Γ_X opérateur nucléaire $\Rightarrow \Gamma_X$ opérateur de Hilbert-Schmidt $\Rightarrow \Gamma_X$ opérateur compact.

Définissons maintenant l'opérateur de covariance croisée. En effet, comme nous le verrons dans la suite, cet opérateur intervient directement dans l'écriture de l'estimateur que nous étudierons dans le modèle linéaire fonctionnel.

Définition 6.3.3 On appelle **opérateur du moment d'ordre 2 croisé** en X et Y , noté Δ_{XY} , l'opérateur défini par :

$$\Delta_{XY} : \begin{cases} H & \rightarrow \mathbb{R} \\ x & \mapsto \Delta_{XY}(x) = (E(X \otimes Y))(x). \end{cases}$$

Si $EX = 0$ ou $EY = 0$, Δ_{XY} est appelé **opérateur de covariance croisée**.

L'intérêt majeur de cet opérateur réside dans la proposition suivante :

Proposition 6.3.3 $\Delta_{XY} = \Psi\Gamma_X$.

Preuve. Pour tout x appartenant à H , on a :

$$\begin{aligned}\Delta_{XY} &= E(\langle x, X \rangle_H Y), \\ &= E(\langle x, X \rangle_H \Psi(X)) + E\langle x, X \rangle E\varepsilon, \\ &= \Psi E(\langle x, X \rangle_H X), \\ &= \Psi\Gamma_X(x).\end{aligned}$$

6.3.3 Analyse en Composantes Principale (ACP) Fonctionnelle

L'ACP fonctionnelle (Dauxois et Pousse, 1976, Dauxois, Pousse et Romain, 1982, Romain, 1979) va jouer un rôle important dans le modèle linéaire fonctionnel ; c'est elle qui va nous permettre d'obtenir une approximation linéaire de X (et donc de Γ_X). Cette approximation sera utilisée pour construire notre estimateur de Ψ .

Soit X une v.a.h. centrée appartenant à $L^2_H(P)$.

Définition 6.3.4 On appelle **ACP fonctionnelle** de X l'analyse spectrale de Γ_X qui permet d'écrire sa décomposition de Schmidt dans $\sigma_2(H)$

$$\Gamma_X = \sum_{k \in \mathbb{N}^*} \lambda_k v_k \otimes v_k$$

ainsi que la décomposition de X dans $L^2_H(P)$

$$X(\omega) = \sum_{k \in \mathbb{N}^*} \sqrt{\lambda_k} w_k(\omega) v_k$$

avec :

- $(\lambda_k)_{k \in \mathbb{N}^*}$ la suite pleine décroissante des valeurs propres non nulles de Γ_X appelées **valeurs principales**,
- $(v_k)_{k \in \mathbb{N}^*}$ la suite orthonormale dans H des vecteurs propres de Γ_X associés aux valeurs principales $(\lambda_k)_{k \in \mathbb{N}^*}$,
- $(w_k)_{k \in \mathbb{N}^*}$ la suite orthonormale dans $L^2_{\mathbb{R}}(P)$ des vecteurs propres de $W_X = G_X^* \circ G_X$ associés aux valeurs principales $(\lambda_k)_{k \in \mathbb{N}^*}$ appelés **composantes principales** et telles que

$$w_k = \frac{G_X^* v_k}{\|G_X^* v_k\|_{L^2_{\mathbb{R}}(P)}} = \frac{\langle X, v_k \rangle_H}{(E \langle X, v_k \rangle_H^2)^{1/2}}.$$

Remarques.

i) $sp(\Gamma_X) = sp(W_X)$. En effet,

$$\begin{aligned}
\lambda_k \in sp(\Gamma_x) &\Leftrightarrow \exists v_k \in H - \{0\} / \Gamma_X v_k = \lambda_k v_k, \\
&\Leftrightarrow G_X \circ G_X^* v_k = \lambda_k v_k, \\
&\Rightarrow G_X^* \circ G_X \circ G_X^* v_k = \lambda_k G_X^* v_k, \\
&\Rightarrow \lambda_k \in sp(W_x).
\end{aligned} \tag{*}$$

On montre de la même façon la réciproque.

ii) D'après (*), v_k vecteur propre de Γ_X implique que $G_X^* v_k$ vecteur propre de W_X , ce qui entraîne que

$$w_k = \frac{G_X^* v_k}{\|G_X^* v_k\|_{L^2_{\mathbb{R}}(P)}}$$

est vecteur propre orthonormé de W_X .

iii) w_k et $w_{k'}$ sont deux v.a.r. non corrélées dès que $k \neq k'$. En effet :

$$\begin{aligned}
E(w_k w_{k'}) &= \frac{E(\langle X, v_k \rangle_H \langle X, v_{k'} \rangle_H)}{(E \langle X, v_k \rangle_H^2)^{1/2} (E \langle X, v_{k'} \rangle_H^2)^{1/2}}, \\
&= \frac{E \langle X \otimes X(v_k), v_{k'} \rangle_H}{(E \langle X, v_k \rangle_H^2)^{1/2} (E \langle X, v_{k'} \rangle_H^2)^{1/2}}, \\
&= \frac{\langle \Gamma_X v_k, v_{k'} \rangle_H}{(E \langle X, v_k \rangle_H^2)^{1/2} (E \langle X, v_{k'} \rangle_H^2)^{1/2}}, \\
&= \frac{\lambda_k \langle v_k, v_{k'} \rangle_H}{(E \langle X, v_k \rangle_H^2)^{1/2} (E \langle X, v_{k'} \rangle_H^2)^{1/2}}, \\
&= 0.
\end{aligned}$$

Définition 6.3.5 Approximation linéaire d'ordre q .

i) On appelle approximation linéaire d'ordre q de Γ_X , noté Γ_X^q , l'opérateur défini par :

$$\Gamma_X^q = \sum_{k=1}^q \lambda_k v_k \otimes v_k.$$

ii) On appelle approximation linéaire d'ordre q de la v.a.h. X , noté X^q , la v.a.h. définie par :

$$X^q(\omega) = \sum_{k=1}^q \sqrt{\lambda_k} w_k(\omega) v_k.$$

Propriété 6.3.1 **Évaluation de l'approximation d'ordre q .**

$$\|\Gamma_X - \Gamma_X^q\|_2 = \left(\sum_{k \geq q} \lambda_k^2 \right)^{\frac{1}{2}} \quad \text{et} \quad E(\|X - X^q\|_H^2) = \sum_{k \geq q} \lambda_k.$$

Remarque. $\Gamma_X^q = E(X^q \otimes X^q) = \Gamma_{X^q}$.

Lien avec le modèle linéaire fonctionnel : dans le paragraphe suivant, on va introduire un estimateur de Ψ basé sur la relation $\Delta_{XY} = \Psi\Gamma_X$. En effet, si Γ_X^{-1} existe, on a

$$\Psi = \Delta_{XY}\Gamma_X^{-1}.$$

Or Γ_X n'est pas inversible dans $\mathcal{L}(H)$ (i.e. Γ_X^{-1} n'est pas borné). En effet, en utilisant la décomposition de Schmidt de Γ_X , il vient que $\Gamma_X^{-1} = \sum_{k \in \mathbb{N}^*} \lambda_k^{-1} v_k \otimes v_k$ avec $\lim_{k \rightarrow +\infty} \lambda_k = 0$. L'ACP fonctionnelle de X nous permet alors d'obtenir Γ_X^q , l'approximation d'ordre q de Γ_X , qui, elle, est inversible dans $\mathcal{L}(H)$ puisque $(\Gamma_X^q)^{-1} = \sum_{k=1}^q \lambda_k^{-1} v_k \otimes v_k$. On peut donc construire une approximation d'ordre q de Ψ de la manière suivante :

Définition 6.3.6 **Approximation d'ordre q de Ψ .** On appelle approximation d'ordre q de Ψ , notée Ψ_q , l'opérateur appartenant à H' définie par :

$$\Psi_q = \Delta_{X^q Y} (\Gamma_{X^q})^{-1}.$$

La proposition ci-dessous donne une autre expression de Ψ_q écrite à l'aide de projecteurs orthogonaux.

Proposition 6.3.4 Soit Π_q le projecteur orthogonal sur le s.e.v. H_q de H engendré par v_1, v_2, \dots, v_q ; on a alors :

$$\Psi_q = \Delta_{XY} \Pi_q (\Pi_q \Gamma_X \Pi_q)^{-1}. \quad (6.2)$$

Preuve. $X^q = \Pi_q X$ implique que, pour tout x appartenant à H , $\langle X^q, x \rangle_H = \langle \Pi_q X, x \rangle_H = \langle X, \Pi_q x \rangle_H$ puisque Π_q est autoadjoint. D'où :

$$\begin{aligned} \Delta_{X^q Y} &\stackrel{\text{def}}{=} E(\langle \cdot, \Pi_q X \rangle_H Y), \\ &= E(\langle \Pi_q(\cdot), X \rangle_H Y), \\ &= \Delta_{XY} \Pi_q. \end{aligned}$$

De plus, on a :

$$\begin{aligned}\Gamma_X^q &= \Gamma_{X^q}, \\ &\stackrel{\text{def}}{=} E(\langle \cdot, \Pi_q X \rangle_H \Pi_q X), \\ &= \Pi_q E(\langle \Pi_q(\cdot), X \rangle_H X), \\ &= \Pi_q \Gamma_X \Pi_q.\end{aligned}$$

6.3.4 Définition de l'estimateur de Ψ

Soit $(X_i, Y_i)_{i=1, \dots, n}$ un n -échantillon du couple (X, Y) (identiquement et **indépendamment** distribué). On s'intéresse alors aux versions empiriques de Γ_X et Δ_{XY} définies respectivement par :

$$\Gamma_n = \frac{1}{n} \sum_{i=1}^n X_i \otimes X_i \quad \text{et} \quad \Delta_n = \frac{1}{n} \sum_{i=1}^n X_i \otimes Y_i.$$

L'estimateur que nous allons introduire est inspiré de celui étudié par Bosq (1991) dans le cadre de processus autorégressifs hilbertiens. Par ailleurs, il est clair que $\dim(\mathcal{I}m \Gamma_n) \leq n$; on note $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_n \geq 0 = \hat{\lambda}_{n+1} = \dots$ les valeurs propres de Γ_n et $\hat{v}_1, \hat{v}_2, \dots$, les vecteurs propres orthonormés associés à ces valeurs propres. Soit $(q_n)_{n \in \mathbb{N}^*}$ une suite d'entiers positifs tels que $\lim_{n \rightarrow +\infty} q_n = +\infty$ avec $q_n \leq n$ et soit \hat{H}_{q_n} le sous espace vectoriel de H engendré par $\{\hat{v}_j ; j = 1, \dots, q_n\}$, $\hat{\Pi}_{q_n}$ étant la projection orthogonal sur \hat{H}_{q_n} :

$$\hat{\Pi}_{q_n} = \sum_{i=1}^{q_n} \hat{v}_i \otimes \hat{v}_i.$$

Définition 6.3.7 *En supposant que $\hat{\lambda}_{q_n} > 0$, l'estimateur $\hat{\Psi}_{q_n}$ de Ψ est défini par :*

$$\hat{\Psi}_{q_n} = \Delta_n \hat{\Pi}_{q_n} \left(\hat{\Pi}_{q_n} \Gamma_n \hat{\Pi}_{q_n} \right)^{-1}. \quad (6.3)$$

$\hat{\Psi}_{q_n}$ est tout simplement la version empirique de Ψ_q (voir (6.2)).

6.3.5 Étude asymptotique

Ce paragraphe s'intéresse aux propriétés asymptotiques de l'estimateur $\hat{\Psi}_{q_n}$ de Ψ . Autrement dit, nous allons préciser le comportement de $\hat{\Psi}_{q_n}$ par rapport à Ψ quand n tend vers l'infini.

Comportements asymptotiques de Γ_n

Proposition 6.3.5 $\{\Gamma_n\}_{n \in \mathbb{N}^*}$ est une suite de v.a. P -intégrables de (Ω, \mathcal{A}, P) dans (E, \mathcal{B}_E) qui converge presque sûrement vers Γ_X dans E .

Preuve. On sait que $\{X_i \otimes X_i\}_{i \in \mathbb{N}^*}$ est une suite de v.a. P -intégrables et indépendantes. Il suffit alors d'appliquer la loi forte des grands nombres pour obtenir la convergence presque sûre de $\frac{1}{n} \sum_{i=1}^n X_i \otimes X_i$ vers $E(X \otimes X)$ (dans E) quand n tend vers l'infini et le résultat énoncé s'en déduit immédiatement. \square

Remarque. Comme la convergence presque sûre se fait dans E (i.e. $\|\Gamma_n - \Gamma_X\|_2 \xrightarrow{n \rightarrow +\infty} 0$ p.s.) on en déduit la convergence uniforme presque sûre puisque $\|\Gamma_n - \Gamma_X\|_\infty \leq \|\Gamma_n - \Gamma_X\|_2$.

Proposition 6.3.6 Si $E\|X\|_H^4 < +\infty$, alors $E(\|\Gamma_n - \Gamma_X\|_2^2) \leq \frac{1}{n}E\|X\|_H^4$.

Preuve. $\|\Gamma_n - \Gamma_X\|_2^2 = \|\Gamma_n\|_2^2 + \|\Gamma_X\|_2^2 - 2 \sum_{i=1}^{+\infty} \langle \Gamma_n e_i, \Gamma_X e_i \rangle_H$ où $(e_i)_{i \in \mathbb{N}^*}$ est une b.o.n. de H , ce qui implique que

$$E(\|\Gamma_n - \Gamma_X\|_2^2) = E\|\Gamma_n\|_2^2 - \|\Gamma_X\|_2^2.$$

Par ailleurs,

$$\begin{aligned} \|\Gamma_n\|_2^2 &= \frac{1}{n^2} \sum_{j=1}^{+\infty} \sum_{k,k'=1}^n \langle X_k, e_j \rangle_H \langle X_{k'}, e_j \rangle_H \langle X_k, X_{k'} \rangle_H, \\ &= \frac{1}{n^2} \sum_{j=1}^{+\infty} \sum_{k=1}^n \langle X_k, e_j \rangle_H^2 \|X_k\|_H^2 + \underbrace{\frac{1}{n^2} \sum_{j=1}^{+\infty} \sum_{k \neq k'} \langle X_k \otimes X_k(e_j), X_{k'} \otimes X_{k'}(e_j) \rangle_H}_A \end{aligned}$$

Or,

$$\begin{aligned} E(A) &= \frac{n^2 - n}{n^2} \sum_{j=1}^{+\infty} \langle \Gamma_X e_j, \Gamma_X e_j \rangle_H \\ &= \|\Gamma_X\|_2^2 - \frac{1}{n} \|\Gamma_X\|_2^2. \end{aligned}$$

D'où :

$$\begin{aligned} E\|\Gamma_n - \Gamma_X\|_2^2 &= \frac{1}{n} E\|X\|_H^4 + \|\Gamma_X\|_2^2 - \frac{1}{n} \|\Gamma_X\|_H^2 - \|\Gamma_X\|_2^2, \\ &\leq \frac{1}{n} E\|X\|_H^4. \quad \square. \end{aligned}$$

Remarque. Comme précédemment, ce résultat reste valable pour la norme de la convergence uniforme.

Le prochain résultat que nous détaillons ci-dessous établit une borne supérieure entre les éléments spectraux de Γ_n et ceux de Γ_X . Il est essentiellement dû à Deville (1974) puis repris bien plus tard par Bosq (1991).

Proposition 6.3.7 Relation entre les éléments spectraux de Γ_n et Γ_X .

- i) $\forall j, j \in \mathbb{N}^*, |\hat{\lambda}_j - \lambda_j| \leq \|\Gamma_n - \Gamma_X\|_\infty,$
 ii) $\|\hat{v}_j - v'_j\| \leq a_j \|\Gamma_n - \Gamma_X\|_\infty$ avec

$$a_j = \begin{cases} \frac{2\sqrt{2}}{\lambda_1 - \lambda_2} & \text{si } j = 1, \\ \frac{2\sqrt{2}}{\min\{\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}\}} & \text{si } j > 1, \end{cases}$$

et

$$v'_j = \begin{cases} v_j & \text{si } \langle \hat{v}_j, v_j \rangle \geq 0, \\ -v_j & \text{si } \langle \hat{v}_j, v_j \rangle < 0. \end{cases}$$

Preuve.

- i) Voir Gohberg et Krejn (1971, p. 31).
 ii) Soit $j > 1$ et désignons par I l'application identité dans H : $\Gamma_X - \lambda_j I = \sum_{k \neq j} (\lambda_k - \lambda_j) v_k \otimes v_k$ et posons

$$(\Gamma_X - \lambda_j I)^+ = \sum_{l \neq j} \frac{1}{\lambda_l - \lambda_j} v_l \otimes v_l.$$

On a $(\Gamma_X - \lambda_j I)^+(\Gamma_X - \lambda_j I)(\hat{v}_j - v'_j) = \sum_{k \neq j} v_k \otimes v_k(\hat{v}_j)$, ce qui entraîne que

$$\begin{aligned} \|(\Gamma_X - \lambda_j I)^+(\Gamma_X - \lambda_j I)(\hat{v}_j - v'_j)\|_H^2 &= \sum_{k \neq j} \|v_k \otimes v_k(\hat{v}_j)\|_H^2, \\ &= \sum_{k \neq j} \langle v_k, \hat{v}_j \rangle_H^2, \\ &= 1 - \langle v'_j, \hat{v}_j \rangle_H^2. \end{aligned} \quad (6.4)$$

Par ailleurs,

$$(\Gamma_X - \lambda_j I)^+(\Gamma_X - \lambda_j I)(\hat{v}_j - v'_j) = (\Gamma_X - \lambda_j I)^+ \left\{ \underbrace{(\hat{\lambda}_j - \lambda_j)\hat{v}_j + (\Gamma_X - \Gamma_n)\hat{v}_j}_{u_j} \right\},$$

ce qui nous permet d'écrire que

$$\begin{aligned} \|(\Gamma_X - \lambda_j I)^+(\Gamma_X - \lambda_j I)(\hat{v}_j - v'_j)\|_H^2 &= \left\| \sum_{k \neq j} \frac{1}{\lambda_k - \lambda_j} \langle v_k, u_j \rangle_H v_k \right\|_H^2, \\ &\leq \frac{\sum_{k \neq j} \langle v_k, u_j \rangle_H^2}{\min^2(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})}, \\ &\leq \frac{\|u_j\|_H^2}{\min^2(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})}. \end{aligned} \quad (6.5)$$

Or $\|u_j\|_H \leq \|\Gamma_X - \Gamma_n\|_\infty + |\hat{\lambda}_j - \lambda_j|$, ce qui implique en utilisant le résultat i) de la proposition que

$$\|u_j\|_H^2 \leq 4\|\Gamma_X - \Gamma_n\|_\infty^2. \quad (6.6)$$

D'où (6.4), (6.5) et (6.6) entraînent que

$$1 - \langle v'_j, \hat{v}_j \rangle_H \leq \frac{4\|\Gamma_X - \Gamma_n\|_\infty^2}{\min^2(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})}.$$

Enfin, on peut écrire

$$\begin{aligned} \|\hat{v}_j - v'_j\|_H^2 &= \langle \hat{v}_j - v'_j, \hat{v}_j - v'_j \rangle_H, \\ &= 2(1 - \langle v'_j, \hat{v}_j \rangle), \\ &\leq 2(1 - \langle v_j, \hat{v}_j \rangle^2), \end{aligned}$$

puisque par définition de v'_j , on a $\langle v'_j, \hat{v}_j \rangle \in [0, 1]$.

Finalement, on a :

$$\|\hat{v}_j - v'_j\|_H^2 \leq \frac{8\|\Gamma_X - \Gamma_n\|_\infty^2}{\min^2(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})}. \quad \square$$

Convergence en probabilité de $\hat{\Psi}_{q_n}$

Théorème 6.3.1 *Si les hypothèses*

$$\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_{q_n} > 0 \text{ p.s.}, \quad (6.7)$$

$$\lambda_1 > \lambda_2 > \cdots > 0, \quad (6.8)$$

$$E\|X\|_H^4 < +\infty, \quad (6.9)$$

$$\lim_{n \rightarrow +\infty} n\lambda_{q_n}^4 = +\infty \text{ et } \lim_{n \rightarrow +\infty} \frac{n\lambda_{q_n}^2}{\left(\sum_{j=1}^{q_n} a_j\right)^2} = +\infty, \quad (6.10)$$

sont satisfaites, alors :

$$\|\widehat{\Psi}_{q_n} - \Psi\|_\infty \xrightarrow{n \rightarrow +\infty} 0 \text{ en proba.}$$

Preuve. Posons :

$$\Psi_{q_n} = \Delta_{XY} \Pi_{q_n} (\Pi_{q_n} \Gamma_X \Pi_{q_n})^{-1}.$$

Comme $\|\Psi - \widehat{\Psi}_{q_n}\|_\infty \leq \|\Psi - \Psi_{q_n}\|_\infty + \|\Psi_{q_n} - \widehat{\Psi}_{q_n}\|_\infty$, on va montrer les deux résultats suivants :

- i) $\|\Psi - \Psi_{q_n}\|_\infty \xrightarrow{n \rightarrow +\infty} 0$,
- ii) $\|\Psi_{q_n} - \widehat{\Psi}_{q_n}\|_\infty \xrightarrow{n \rightarrow +\infty} 0$ en probabilité.

i)

$$\begin{aligned} \|\Psi - \Psi_{q_n}\|_\infty^2 &= \sum_{j=1}^{+\infty} |(\Psi - \Psi_{q_n})(v_j)|^2, \\ &= \sum_{j > q_n} |\Psi(v_j)|^2, \\ &\leq \|\Psi\|_\infty^2. \end{aligned}$$

Comme Ψ appartient à H' , Ψ est aussi un opérateur borné et il vient que

$$\|\Psi - \Psi_{q_n}\|_\infty \xrightarrow{n \rightarrow +\infty} 0. \quad (6.11)$$

ii)

Lemme 6.3.1

$$\|\Psi_{q_n} - \widehat{\Psi}_{q_n}\|_\infty \leq \gamma_n \|\Gamma_X - \Gamma_n\|_\infty + \frac{1}{\widehat{\lambda}_{q_n}} \|\Delta_{XY} - \Delta_n\|_\infty$$

$$\text{où } \gamma_n = \|\Delta_{XY}\|_\infty \left\{ \frac{1}{\lambda_{q_n} \widehat{\lambda}_{q_n}} + 2 \left(\frac{1}{\lambda_{q_n}} + \frac{1}{\widehat{\lambda}_{q_n}} \right) \sum_{j=1}^{q_n} a_j \right\}.$$

Preuve. Pour les besoins de la démonstration, définissons l'opérateur intermédiaire $\widetilde{\Gamma}_{q_n}$:

$$\widetilde{\Gamma}_{q_n} = \sum_{j=1}^{q_n} \lambda_j \widehat{v}_j \otimes \widehat{v}_j.$$

Ecrivons ensuite :

$$\|\Psi_{q_n} - \widehat{\Psi}_{q_n}\|_\infty \leq \|\Delta_{XY} \Pi_{q_n}\|_\infty \left\| (\Pi_{q_n} \Gamma_X \Pi_{q_n})^{-1} - \widetilde{\Gamma}_{q_n}^{-1} \right\|_\infty + \left\| \Delta_{XY} \Pi_{q_n} \widetilde{\Gamma}_{q_n}^{-1} - \widehat{\Psi}_{q_n} \right\|_\infty \quad (6.12)$$

et étudions séparément chaque terme à droite de l'inégalité (6.12).

En utilisant $\|\Pi_{q_n}\|_\infty = 1$, on a $\|\Delta_{XY}\Pi_{q_n}\|_\infty \leq \|\Delta_{XY}\|_\infty$ et par conséquent, on obtient en utilisant la Proposition (6.3.7) :

$$\begin{aligned} \|\Delta_{XY}\Pi_{q_n}\|_\infty \left\| (\Pi_{q_n}\Gamma_X\Pi_{q_n})^{-1} - \tilde{\Gamma}_{q_n}^{-1} \right\|_\infty &\leq \|\Delta_{XY}\|_\infty \left\| \sum_{j=1}^{q_n} \frac{1}{\lambda_j} (v_j \otimes v_j - \hat{v}_j \otimes \hat{v}_j) \right\|_\infty \\ &\leq \frac{2\|\Delta_{XY}\|_\infty}{\lambda_{q_n}} \sum_{j=1}^{q_n} \|v_j - \hat{v}_j\|_H \\ &\leq \frac{2\|\Delta_{XY}\|_\infty}{\lambda_{q_n}} \|\Gamma_X - \Gamma_n\|_\infty \sum_{j=1}^{q_n} a_j. \end{aligned} \quad (6.13)$$

Décomposons ensuite le second terme :

$$\begin{aligned} \left\| \Delta_{XY}\Pi_{q_n}\tilde{\Gamma}_{q_n}^{-1} - \hat{\Psi}_{q_n} \right\|_\infty &\leq \|\Delta_{XY}\Pi_{q_n}\|_\infty \left\| \tilde{\Gamma}_{q_n}^{-1} - \left(\hat{\Pi}_{q_n}\Gamma_n\hat{\Pi}_{q_n} \right)^{-1} \right\|_\infty \\ &\quad + \left\| \Delta_n\hat{\Pi}_{q_n} - \Delta_{XY}\Pi_{q_n} \right\|_\infty \left\| \left(\hat{\Pi}_{q_n}\Gamma_n\hat{\Pi}_{q_n} \right)^{-1} \right\|_\infty. \end{aligned}$$

Or, en utilisant de nouveau le lemme (6.3.7) et en remarquant que les fonctions \hat{v}_j sont orthonormées puis que $\left\| \hat{\Pi}_{q_n} \right\|_\infty = 1$, on obtient :

$$\begin{aligned} \left\| \tilde{\Gamma}_{q_n}^{-1} - \left(\hat{\Pi}_{q_n}\Gamma_n\hat{\Pi}_{q_n} \right)^{-1} \right\|_\infty &= \sup_{\|x\|_H=1} \left\| \sum_{j=1}^{q_n} \left(\frac{1}{\lambda_j} - \frac{1}{\hat{\lambda}_j} \right) \hat{v}_j \otimes \hat{v}_j(x) \right\|_H \\ &\leq \sup_{\|x\|_H=1} \sum_{j=1}^{q_n} \frac{|\lambda_j - \hat{\lambda}_j|}{\lambda_j \hat{\lambda}_j} \|\hat{v}_j \otimes \hat{v}_j(x)\|_H \\ &\leq \frac{\|\Gamma_X - \Gamma_n\|_\infty}{\lambda_{q_n} \hat{\lambda}_{q_n}} \sup_{\|x\|_H=1} \sum_{j=1}^{q_n} \|\hat{v}_j \otimes \hat{v}_j(x)\|_H \\ &= \frac{\|\Gamma_X - \Gamma_n\|_\infty}{\lambda_{q_n} \hat{\lambda}_{q_n}} \left\| \hat{\Pi}_{q_n} \right\|_\infty \\ &= \frac{\|\Gamma_X - \Gamma_n\|_\infty}{\lambda_{q_n} \hat{\lambda}_{q_n}}. \end{aligned} \quad (6.14)$$

Par ailleurs, il est clair que

$$\left\| \left(\hat{\Pi}_{q_n}\Gamma_n\hat{\Pi}_{q_n} \right)^{-1} \right\|_\infty = \hat{\lambda}_{q_n}^{-1}. \quad (6.15)$$

Enfin, en utilisant les mêmes arguments que précédemment, on a :

$$\begin{aligned}
\left\| \Delta_n \widehat{\Pi}_{q_n} - \Delta_{XY} \Pi_{q_n} \right\|_{\infty} &\leq \|\Delta_{XY}\|_{\infty} \left\| \widehat{\Pi}_{q_n} - \Pi_{q_n} \right\|_{\infty} + \left\| \widehat{\Pi}_{q_n} \right\|_{\infty} \|\Delta_{XY} - \Delta_n\|_{\infty} \\
&\leq \|\Delta_{XY}\|_{\infty} \left\| \sum_{j=1}^{q_n} (v_j \otimes v_j - \widehat{v}_j \otimes \widehat{v}_j) \right\|_{\infty} + \|\Delta_{XY} - \Delta_n\|_{\infty} \\
&\leq \|\Delta_{XY}\|_{\infty} 2 \sum_{j=1}^{q_n} \|v_j - \widehat{v}_j\|_H + \|\Delta_{XY} - \Delta_n\|_{\infty} \\
&\leq 2 \|\Delta_{XY}\|_{\infty} \|\Gamma_X - \Gamma_n\|_{\infty} \sum_{j=1}^{q_n} a_j + \|\Delta_{XY} - \Delta_n\|_{\infty}. \quad (6.16)
\end{aligned}$$

Le résultat final est obtenu en introduisant les majorations (6.13), (6.14), (6.15) et (6.16) dans l'inégalité (6.3.7). \square

Montrons maintenant que $\left\| \Psi_{q_n} - \widehat{\Psi}_{q_n} \right\|_{\infty} \rightarrow 0$ en probabilité quand $n \rightarrow +\infty$. Dans ce but, commençons par donner le lemme établissant la convergence uniforme en moyenne quadratique des opérateurs de covariance empiriques.

Lemme 6.3.2 *Si X vérifie (6.9) alors :*

$$E \|\Delta_{XY} - \Delta_n\|_{\infty}^2 \leq \frac{\|\Psi\|_{\infty}^2 E \|X\|_H^4}{n} + \frac{\sigma^2}{n} E \|X\|_H^2. \quad (6.17)$$

Preuve. Puisque $E\Delta_n = \Delta_{XY}$, on a comme dans le cas réel :

$$E \|\Delta_{XY} - \Delta_n\|_{\infty}^2 = E \|\Delta_n\|_{\infty}^2 - \|\Delta_{XY}\|_{\infty}^2. \quad (6.18)$$

Dans un premier temps, utilisons l'indépendance entre les X_i pour décomposer :

$$\begin{aligned}
E \|\Delta_n\|_{\infty}^2 &= \frac{1}{n^2} \sum_{j \in \mathbb{N}} \sum_{i, i'=1}^n E (\langle X_i, e_j \rangle_H Y_i \langle X_{i'}, e_j \rangle_H Y_{i'}) \\
&= \frac{1}{n^2} \sum_{j \in \mathbb{N}} \sum_{i=1}^n E (\langle X_i, e_j \rangle_H Y_i)^2 \\
&\quad + \frac{1}{n^2} \sum_{j \in \mathbb{N}} \sum_{i \neq i'} E (\langle X_i, e_j \rangle_H Y_i \langle X_{i'}, e_j \rangle_H Y_{i'}) \\
&= \frac{1}{n} \sum_{j \in \mathbb{N}} E (\langle X, e_j \rangle_H Y)^2 + \|\Delta_{XY}\|_{\infty}^2 - \frac{1}{n} \|\Delta_{XY}\|_{\infty}^2 \quad (6.19)
\end{aligned}$$

Par ailleurs, en utilisant la Définition 6.3.1 du modèle linéaire fonctionnel et le fait que $\|X\|_H^2 = \sum_{j \in \mathbb{N}} \langle X, e_j \rangle_H^2$, on obtient alors

$$\begin{aligned} \frac{1}{n} \sum_{j \in \mathbb{N}} E(\langle X, e_j \rangle_H Y)^2 &= \frac{1}{n} \sum_{j \in \mathbb{N}} E(\langle X, e_j \rangle_H \{\Psi(X) + \varepsilon\})^2 \\ &= \frac{1}{n} \{E(\|X\|_H^2 \Psi(X)^2) + E(\|X\|_H^2 \varepsilon^2)\} \\ &\leq \frac{\|\Psi\|_\infty^2 E\|X\|_H^4 + \sigma^2 E\|X\|_H^2}{n} \end{aligned} \quad (6.20)$$

La majoration finale est ensuite déduite des inégalités (6.18), (6.19) et (6.20).

Nous pouvons à présent en déduire la preuve du Théorème en utilisant la majoration obtenue dans le Lemme 6.3.1. Cette borne est aléatoire et dépend de la valeur propre $\widehat{\lambda}_{q_n}$. Pour contourner ce problème, considérons l'événement suivant :

$$E_n = \left\{ \frac{\lambda_{q_n}}{2} < \widehat{\lambda}_{q_n} < \frac{3\lambda_{q_n}}{2} \right\}. \quad (6.21)$$

Sur E_n , on a $\frac{1}{\widehat{\lambda}_{q_n}} < \frac{2}{\lambda_{q_n}}$ et donc, d'après le Lemme 6.3.1 :

$$\left\| \Psi_{q_n} - \widehat{\Psi}_{q_n} \right\|_\infty \leq \delta_n \|\Delta_{XY}\|_\infty \|\Gamma_X - \Gamma_n\|_\infty + \frac{2}{\lambda_{q_n}} \|\Delta_{XY} - \Delta_n\|_\infty,$$

où $\delta_n = \frac{2}{\lambda_{q_n}^2} + \frac{6}{\lambda_{q_n}} \sum_{j=1}^{q_n} a_j$. On en déduit :

$$\begin{aligned} P\left(\left\| \Psi_{q_n} - \widehat{\Psi}_{q_n} \right\|_\infty > \eta, E_n\right) &\leq P\left(\|\Gamma_X - \Gamma_n\|_\infty > \frac{\eta}{2\delta_n \|\Delta_{XY}\|_\infty}\right) \\ &\quad + P\left(\|\Delta_{XY} - \Delta_n\|_\infty > \frac{\lambda_{q_n} \eta}{4}\right). \end{aligned} \quad (6.22)$$

Il vient alors :

$$P\left(\left\| \Psi_{q_n} - \widehat{\Psi}_{q_n} \right\|_\infty > \eta, E_n\right) \leq \frac{4\delta_n^2 \|\Delta_{XY}\|_\infty^2}{\eta^2} E\|\Gamma_X - \Gamma_n\|_\infty^2 + \frac{16}{\lambda_{q_n}^2 \eta^2} E\|\Delta_{XY} - \Delta_n\|_\infty^2. \quad (6.23)$$

Par ailleurs, on a

$$P(\overline{E}_n) = P\left(|\lambda_{q_n} - \widehat{\lambda}_{q_n}| > \frac{\lambda_{q_n}}{2}\right)$$

$$\begin{aligned}
&\leq P\left(\|\Gamma_X - \Gamma_n\|_\infty > \frac{\lambda_{q_n}}{2}\right) \\
&\leq \frac{4}{\lambda_{q_n}^2} E \|\Gamma_X - \Gamma_n\|_\infty^2.
\end{aligned} \tag{6.24}$$

On obtient donc d'après (6.23), (6.24), la Proposition 6.3.6 et le Lemme 6.3.2 :

$$\begin{aligned}
P\left(\|\Psi_{q_n} - \widehat{\Psi}_{q_n}\|_\infty > \eta\right) &\leq \frac{4\|\Delta_{XY}\|_\infty^2 E \|X\|_H^4 \delta_n^2}{\eta^2} + \\
&\frac{16}{\eta^2} (\|\Psi\|_\infty^2 E \|X\|_H^4 + \sigma^2 E \|X\|_H^2) \cdot \frac{1}{n\lambda_{q_n}^2} + \\
&4E \|X\|_H^4 \cdot \frac{1}{n\lambda_{q_n}^2}
\end{aligned} \tag{6.25}$$

Il suffit alors d'utiliser (6.11), (6.25), (6.9) et (6.10) pour terminer la preuve du Théorème. \square

6.4 Exercices

Exercice 6.1 $\forall(A, B) \in \sigma_2(H) \times \sigma_2(H)$ ($\sigma_2(H)$ étant l'ensemble des opérateurs de Hilbert-Schmidt), on appelle produit tensoriel dans $E = (\sigma_2(H), \langle, \rangle_2)$, l'opérateur noté \otimes_2 de $\sigma_2(H)$ dans lui-même défini par :

$$\forall C \in \sigma_2(H), \quad (A \otimes_2 B)(C) = \langle C, A \rangle_2 B.$$

En appliquant le Théorème de la limite centrale pour v.a.h., montrer que $\sqrt{n}(\Gamma_n - \Gamma_X)$ converge en loi vers une v.a.h. gaussienne $\mathcal{N}_{\sigma_2(H)}(0, K)$ avec $K = E((X \otimes X - \Gamma_X) \otimes_2 (X \otimes X - \Gamma_X))$.

Exercice 6.2 On reprend ici les notations introduites dans le cours. L'objectif de cet exercice, est de montrer la convergence presque sûre de $\widehat{\Psi}_{q_n}$ vers Ψ quand n tend vers $+\infty$. Dans ce but, en plus des hypothèses (6.7), (6.8) et (6.9) faites dans le cadre du Théorème 6.3.1, on suppose de plus :

$$\|X\|_H \geq c_1, \text{ p.s.}, \tag{6.26}$$

$$|\varepsilon| \geq c_2, \text{ p.s.}, \tag{6.27}$$

$$\lim_{n \rightarrow +\infty} \frac{n\lambda_{q_n}^4}{\log n} = +\infty \text{ et } \lim_{n \rightarrow +\infty} \frac{n\lambda_{q_n}^2}{\left(\sum_{j=1}^{q_n} a_j\right)^2 \log n} = +\infty. \tag{6.28}$$

1) On pose pour $i = 1, \dots, n$, $Z_i = X_i \otimes X_i - \Gamma_X$ (resp. $U_i = X_i \otimes \varepsilon_i$). En appliquant à la suite de v.a.h. $\{Z_i\}_{i=1, \dots, n}$ (resp. $\{U_i\}_{i=1, \dots, n}$) l'inégalité de type Bernstein pour des v.a.h., montrer qu'il existe deux constantes positives c_3 et c_4 telles que

$$P(\|\Gamma_n - \Gamma_X\|_\infty > \xi) \leq 2 \exp \left\{ -\frac{\xi^2 n}{2c_3(c_3 + c_4\xi)} \right\}$$

$$\left(\text{resp. } P(\|\Delta_n - \Delta_{XY}\|_\infty > \xi) \leq 2 \exp \left\{ -\frac{\xi^2 n}{2c_1c_2(c_1c_2 + c_4\xi)} \right\} \right).$$

2) Dédurre de ce qui précède que :

2.1) il existe une constante positive (indépendante de n), notée A_ζ , telle que

$$P \left(\|\Gamma_n - \Gamma\|_\infty > \frac{\zeta}{2\delta_n \|\Delta\|_\infty} \right) \leq 2 \exp \left\{ -A_\zeta \frac{n}{\delta_n^2} \right\}.$$

2.2) il existe une constante positive (indépendante de n), notée B_ζ , telle que

$$P \left(\|\Delta_n - \Delta\|_\infty > \frac{\lambda_{q_n} \zeta}{4} \right) \leq 2 \exp \left\{ -B_\zeta n \lambda_{q_n}^2 \right\}.$$

3) On rappelle que $E_n = \left\{ \frac{\lambda_{q_n}}{2} < \hat{\lambda}_{q_n} < \frac{3\lambda_{q_n}}{2} \right\}$. Montrer qu'il existe une constante positive C indépendante de n telle que :

$$P(\overline{E_n}) \leq 2 \exp \left\{ -Cn \lambda_{q_n}^2 \right\}.$$

4) Enfin, en utilisant ce qui précède, ainsi que l'inégalité (6.22), montrer que

$$P \left(\|\Psi_{q_n} - \hat{\Psi}_{q_n}\|_\infty > \zeta \right) \leq 2 \left\{ \exp \left\{ -A_\zeta \frac{n}{\delta_n^2} \right\} + \exp \left\{ -B_\zeta n \lambda_{q_n}^2 \right\} + \exp \left\{ -Cn \lambda_{q_n}^2 \right\} \right\}.$$

Conclure.