

N° d'ordre :

REPUBLIQUE ALGERIENNE DEMOCRATIQUE & POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR & DE LA RECHERCHE
SCIENTIFIQUE



UNIVERSITE DJILLALI LIABES
FACULTE DES SCIENCES
SIDI BEL-ABBES

Thèse de Doctorat

ES-Science

Présentée par KARA-ZAÏTRI Lydia

Spécialité : Mathématiques Appliquées
Option : Statistique

Intitulé

**Propriétés asymptotiques uniformes
d'estimateurs pour variables fonctionnelles
et applications**

Mémoire soutenu le 04 Septembre 2017

Devant le jury composé de :

Directeurs Pr. Ali Laksaci , Univ. Djillali Liabès, Sidi Bel Abbès, Algérie.
Pr. Philippe Vieu , Univ. Paul Sabatier, Toulouse, France.

Président Pr. Mohammed Attouch , Univ. Djillali Liabès, Sidi Bel Abbès, Algérie

Examineurs Pr. German Aneiros , Univ. de A Coruña, Espagne.
Pr. Abdelkader Gheriballah, Univ. Djillali Liabès, Sidi Bel Abbès, Algérie.
Pr. Mustapha Rachdi , Univ. Pierre Mendès, Grenoble Alpes, France.
Dr. Saâdia Rahmani , Univ. Djillali Liabès, Sidi Bel Abbès, Algérie.

*À ma mère, à mon époux, à mon fils.
Parce-qu'il est une autre forme de richesse que celle de l'or et l'argent...*

Remerciements

Mes premiers remerciements, et surtout les plus vifs, vont naturellement à mes directeurs de recherche qui, par leurs paroles, leurs écrits, leurs disponibilités, leurs conseils et leurs critiques, ont guidé mes réflexions et m'ont accompagnée pas à pas dans mon travail. Chers professeurs, Ali Laksaci, Mustapha Rachdi et Philippe Vieu, je vous sais gré de la confiance que vous avez placée en moi et de l'opportunité que vous m'avez offerte de faire partie de cette aventure. Me permettre de travailler sous vos directions respectives fut le plus grand honneur que l'on m'ait fait.

Je remercie le Professeur German Aneiros pour l'attention qu'il a manifestée à l'égard de cette thèse de recherche, en s'engageant à en être le rapporteur. Son expertise rajoute, sans nul doute, de la valeur à ce travail.

De même, je suis grandement reconnaissante à Messieurs les Professeurs Mohammed Attouch et Abdelkader Gheriballah de l'intérêt et du temps qu'ils ont bien voulu accorder à l'expertise de cette thèse, en acceptant d'en être les examinateurs. Soutenir ce travail devant vous, qui m'avez tant appris tout au long de mon parcours universitaire est, pour moi, un privilège certain.

Je tiens également à remercier le Docteur Saâdia Rahmani pour avoir eu l'amabilité d'accepter de rejoindre les membres du jury et de me faire l'honneur de contribuer à l'examen de ce mémoire.

En outre, je suis particulièrement reconnaissante à mon époux pour son soutien indéfectible au quotidien. Sa présence, sa compréhension et son implication ont été les piliers de mon travail. Merci de m'avoir aidée à progresser, à tes dépends parfois, dans cette phase éprouvante d'apprenti-chercheur.

Mes remerciements seraient incomplets, si je ne faisais pas mention de ma chère mère, ma plus fervente supportrice. À toi maman, ainsi qu'à mes beaux parents, à mes frères Nadir et Malik, et à mon bébé Jassim, à vous que j'ai quelque peu délaissés ces derniers mois de finalisation de thèse, je vous exprime au terme de ce parcours,

ma gratitude la plus profonde pour votre soutien moral et physique. Vos attentions et vos encouragements m'ont accompagnée tout au long de ces années de recherche.

À tous ceux qui ont, de près ou de loin, participé à l'élaboration de cette thèse, et que je ne pourrai nommer ici, je vous remercie de votre sollicitude aussi minime qu'elle ait pu être. Je citerai en particulier ma chère amie Meriem Benkalfat, pour sa relecture et ses suggestions avisées, et ma tante Malika Remaoun , pour avoir tant cru en moi et encouragée par tous les moyens.

À tous, merci !

Sommaire

<i>Résumé</i>	8
<i>Introduction générale</i>	11
1 <i>Contribution de la thèse</i>	12
1.1 <i>Problématique</i>	12
1.2 <i>Plan de la thèse</i>	15
1.3 <i>Liste des travaux</i>	16
2 <i>État de l'art</i>	17
2.1 <i>Contexte bibliographique</i>	17
2.2 <i>Présentation des modèles étudiés</i>	24
3 <i>Cadre de travail et résultats obtenus</i>	33
3.1 <i>Introduction</i>	33
3.2 <i>Hypothèses</i>	40
3.3 <i>Résultats de convergences obtenus</i>	44
3.4 <i>Choix automatique du paramètre de lissage</i>	51
I <i>Résultats asymptotiques</i>	57
4 <i>Convergence presque complète uniforme sur la fenêtre</i>	58
4.1 <i>Introduction</i>	59
4.2 <i>Cadre de travail</i>	61
4.3 <i>Opérateur de régression</i>	63
4.4 <i>Fonction de répartition conditionnelle</i>	67

4.5	<i>Densité de probabilité conditionnelle</i>	70
4.6	<i>Opérateur de hasard conditionnel</i>	71
4.7	<i>Application à la sélection du paramètre de lissage</i>	72
5	<i>Convergence presque complète uniforme sur le nombre de voisins</i>	77
5.1	<i>Introduction</i>	78
5.2	<i>Modèles et estimateurs</i>	80
5.3	<i>Résultats asymptotiques UNV</i>	82
5.4	<i>Application à la sélection du nombre de voisins</i>	84
II	<i>Études pratiques</i>	89
6	<i>Introduction</i>	90
7	<i>Étude simulée</i>	93
7.1	<i>Présentation de l'étude</i>	93
7.2	<i>Sélection automatique du nombre de voisins</i>	94
7.3	<i>Étude comparative</i>	97
8	<i>Application sur des données réelles</i>	99
8.1	<i>Description des données</i>	99
8.2	<i>Traitement statistique des données</i>	101
8.3	<i>Commentaires sur les résultats</i>	104
	<i>Conclusion et perspectives</i>	105
	<i>Bilan de la thèse</i>	106
	<i>Perspectives pour de futurs travaux</i>	108
	<i>Annexes</i>	112
A	<i>Outils mathématiques</i>	113
A.1	<i>Définitions mathématiques</i>	113
A.2	<i>Lemmes utilisés</i>	114

<i>B Démonstrations de nos résultats</i>	<i>116</i>
<i>B.1 Démonstrations relatives au chapitre 4</i>	116
<i>B.2 Démonstrations relatives au chapitre 5</i>	125

<i>Références bibliographiques</i>	<i>139</i>
------------------------------------	------------

Résumé

Nous nous proposons, dans ce travail de recherche, d'établir la convergence presque complète uniforme sur les paramètres de lissage de quelques estimateurs non paramétriques, en précisant pour chaque cas les vitesses de convergence. Ce qui nous permettra par la suite, de présenter des outils de sélection automatique du paramètre de lissage optimal par la validation croisée pour les estimateurs k NN.

Plus précisément, nous considérons ici quatre modèles non paramétriques, à savoir : la régression, la fonction de répartition conditionnelle, la densité conditionnelle et la fonction de hasard conditionnelle. Notre étude est effectuée sur un échantillon i.i.d d'une variable explicative fonctionnelle afin d'estimer une variable réponse réelle.

Nous nous intéresserons dans un premier temps aux estimateurs de type noyau, puis nous généraliserons les résultats obtenus aux cas d'estimateurs de type k NN. Enfin, nous développerons des exemples (simulés et réels) d'applications de ces méthodes.

Cette étude établit de nouveaux résultats d'uniformité sur le paramètre de lissage pour l'estimation non-paramétrique dans le cadre de variables fonctionnelles. À notre connaissance, notre contribution est le premier travail développant des règles générales de sélection automatique du paramètre de lissage d'estimateurs dans le cas fonctionnel pour les méthodes de type k NN, et a un impact direct dans le domaine de la statistique appliquée.

Mots clés — Analyse de données fonctionnelle ; Statistique non paramétrique fonctionnelle ; Estimateur à noyaux ; Estimateur k NN ; Estimateur data-driven ; Uniformité sur le paramètre de lissage ; Uniformité sur le nombre de voisins ; Petite boule de mesure de probabilité ; Régression fonctionnelle ; Distribution conditionnelle ; Densité conditionnelle ; Fonction de hasard conditionnelle.

Abstract

In this thesis, we propose to establish the uniform in bandwidth almost complete consistency of some nonparametric estimators, specifying in each case the rate of convergence. So that we will then present automatic optimal data-driven bandwidth selection tools by the cross-validation for the k NN estimators.

Specifically, we consider here four nonparametric models, namely : the regression, the conditional distribution, conditional density and conditional hazard function. Our study deals with an i.i.d sample of a functional regressor to estimate a real response variable.

We will focus first on kernel type estimators, and then we will generalize the results to the case of k NN type estimators. Finally, we will develop examples (simulated and real) of applications of these methods.

This work sets new uniform in bandwidth results for nonparametric estimation in the case of functional variables. As far as we know, our contribution is the first work developing general rules for automatic data-driven bandwidth selection in the functional case for the k NN type methods, and has a direct impact in the applied statistics field.

Key words— Functional data analysis ; Functional nonparametric statistics ; kernel estimator ; k NN estimator ; Data-driven estimator ; Uniformity in bandwidth ; Uniformity in nearest neighbors ; Small ball probability ; Functional regression ; Conditional distribution ; Conditional density ; Conditional hazard function.

ملخص

نقترح في هذه الأطروحة إنشاء التقارب شبه الكامل المنتظم على معامل السلسلة لبعض المقدرات غير البارامترية، وتحديد سرعة التقارب في كل حالة. هذا ما سيمكننا فيما بعد بتقديم أدوات الاختيار الأوتوماتيكي لمثل معامل سلاسة المقدرات ك ن ن بواسطة طريقة المصادقة المتقاطعة.

قد أجريت دراستنا على عينة مستقلة و متطابقة التوزيع من متغير تفسيري غير منتهى الأبعاد، لتقدير متغير جوايي حقيقي. وقد اعتبرنا أربع نماذج غير بارامترية، وهي : الإنحدار، دالة التوزيع الشرطي، دالة الكثافة الشرطية، و دالة الزهر الشرطية.

ركزنا عملنا في البداية على المقدرات من نوع النواة الكلاسيكي، ثم عممنا النتائج على المقدرات من نوع ك ن ن. أخيراً، عالجبنا أمثلة تطبيقية (مصطنعة و حقيقية) لهذه النماذج. تحدّد هذه الدراسة نتائج جديدة حول التناظم على معامل السلسلة للتقدير الغير البارامترى في إطار متغيرات غير منتهية الأبعاد.

على حد علمنا، مساهمتنا تمثل أول عمل ينشئ قواعد عامة للاختيار الأوتوماتيكي لمثل معامل سلاسة المقدرات للمناهج من نوع ك ن ن في حالة أبعاد غير منتهية ، و لها تأثير مباشر على مجال الإحصاء التطبيكي.

الكلمات الدالة- تحليل المعطيات غير المنتهية الأبعاد، إحصاء غير بارامترى، مقدرات من نوع النواة الكلاسيكية / ك ن ن، مقدرات مستندة إلى المعطيات، التناظم على معامل السلسلة / عدد الحيران، كرة الإحتمال الصغيرة، الإنحدار، دالة التوزيع الشرطي، دالة الكثافة الشرطية، دالة الزهر الشرطية.

Introduction générale

1.1 Problématique

Conséquence d'une société toujours plus informatisée, pratiquement plus aucun domaine ne néglige la multitude de données auxquelles le "*big data*" permet d'accéder. Cela explique l'émergence, de plus en plus manifeste, de l'étude de données de nature fonctionnelle. En effet, cette branche de la statistique n'est plus à présenter de nos jours, et nous ne comptons plus le nombre d'ouvrages qui s'y sont intéressés depuis des dizaines d'années¹.

Les travaux qui ont particulièrement retenu notre attention dans cette présente thèse, sont ceux traitants d'estimateurs à noyaux non paramétriques, lorsque la variable explicative est fonctionnelle et la variable réponse est réelle.

Dans la littérature mathématique, nous avons rencontré beaucoup de résultats asymptotiques théoriques à propos d'estimateurs à noyaux, offrant des vitesses de convergence très intéressantes. Mais, la qualité de chacun des estimateurs étudiés dépend fortement d'un paramètre de lissage qui intervient dans les deux termes des estimateurs : le biais et la variance.

Le choix de ce paramètre n'est pas chose aisée à faire, puisqu'il faut équilibrer entre

1. Nous renvoyons à [Chen et Zhang \[2014\]](#) pour une discussion générale à propos du big data, et à [Ramsay et Silverman \[2005\]](#) et [Ferraty et Vieu \[2006\]](#) pour l'estimation statistique avec des données fonctionnelles.

les deux termes biais-variance. En effet, plus le paramètre de lissage est petit et plus vite le terme biais tend vers zéro. Et inversement, plus le paramètre de lissage est grand et plus vite le terme variance tend vers zéro. Ainsi, un paramètre très petit donnera un estimateur asymptotiquement sans biais mais de variance conséquente, ce qui le rendra irrégulier. En revanche, un paramètre de lissage très grand aura pour résultat un estimateur régulier, mais avec un biais considérable.

Remarquons également que d'un point de vue pratique, en utilisant chacune des techniques de sélection de paramètre de lissage optimal qui ont été étudiées jusqu'à présent, nous rencontrons inévitablement le même problème pour chacun des estimateurs. À savoir, le fait de choisir un paramètre de lissage automatique nous renvoie systématiquement à une variable aléatoire.

De ce fait, nous nous sommes posés les questions suivantes :

- Si les résultats de convergence sont incontestables pour un paramètre de lissage déterministe, ne perdent-ils pas de leur qualité en présence d'un **paramètre de lissage aléatoire** ?
- Pourrait-on faire en sorte que les vitesses de convergence des estimateurs ne dépendent plus de ce paramètre de lissage sans pour autant perdre de leur efficacité ?

Nous répondrons à ces questions à travers cette thèse en étudiant la convergence uniforme sur le paramètre de lissage et non la convergence ponctuelle. Nous développerons par la suite une technique de choix systématique du paramètre de lissage.

En outre, le paramètre de lissage exprime concrètement la largeur de la fenêtre sur laquelle nous traitons les données. Il est évident qu'un paramètre de lissage optimal devrait dépendre desdites données et pas seulement de la taille de l'échantillon dont elles proviennent². On serait ainsi tentés d'étudier la convergence uniforme sur un paramètre de lissage qui serait orienté par les données. Ainsi, nous avons été amenés à nous poser une autre question : si nous obtenions des résultats de convergence uniforme sur un **paramètre de lissage local**, orienté donc par les données, les vitesses de convergence seraient-elles altérées ?

2. Une étude comparative sur ce point est disponible en consultant les articles de [Rachdi et Vieu \[2007\]](#) et de [Benhenni et collab. \[2007\]](#).

Un autre problème a retenu notre attention, à savoir le paramètre de lissage des estimateurs à noyaux étant défini comme un nombre réel positif, le fait d'en choisir un sur \mathbb{R}_+ n'est pas un exercice évident à réaliser, et encore moins lorsque ce paramètre dépend des données.

Dans le domaine de l'étude appliquée, la méthode de lissage k NN présente des avantages comparément à celle à noyaux. Elle consiste à définir pour chaque courbe x la largeur de la fenêtre $h(x)$ de telle sorte que la boule de centre x et de rayon $h(x)$ contienne exactement k données. De ce fait, elle s'adapte plus facilement à la structure locale des données, sans pour autant engendrer les complications que la méthode à noyau classique rencontre face à un paramètre de lissage adaptable localement. Le deuxième intérêt de la méthode de lissage k NN vient du fait que son paramètre de lissage soit un entier naturel qui prend ses valeurs dans un ensemble discret et non pas continu, ce qui économise un temps de calcul considérable en pratique.

C'est donc une méthode plus bénéfique que celle à noyaux, mais qui est pourtant peu étudiée dans le domaine fonctionnel. Cela s'explique peut-être par les difficultés théoriques que présente son étude. Difficultés dues à la nature aléatoire que prend la largeur de la fenêtre, lorsque le choix du nombre de voisins est prédéfini.

Sur ce, nous avons été amenés à nous poser ces deux dernières questions :

- Serait-il aussi intéressant d'avoir des résultats de convergence uniforme sur le nombre de voisins k pour des estimateurs de ce type ?
- Et si c'est le cas, ces résultats auraient-ils un impact sur la pratique ?

Au cours de notre travail, nous présenterons des estimateurs de type k NN pour quelques opérateurs conditionnels. Nous répondrons à la première question en effectuant une étude sur la convergence uniforme presque complète sur le nombre de voisins k de ces estimateurs. Nous nous concentrerons ensuite sur la dernière question en développant une application sur des données simulées puis sur une étude traitant de données réelles.

Ainsi, l'objectif de cette thèse est double. D'une part, nous cherchons à apporter de nouveaux résultats d'uniformité sur le paramètre de lissage à l'étude des convergences

d'estimateurs à noyaux et k NN de quelques opérateurs conditionnels. D'autre part, nous visons à établir de nouvelles méthodes pour rendre automatique le choix de ce paramètre de lissage qui intervient dans les estimations, directement applicables aux études de données réelles. D'un point de vue tant théorique que pratique, ce travail apporte de nouvelles perspectives sur les estimations non paramétriques fonctionnelles.

1.2 Plan de la thèse

Notre travail est scindé en trois parties :

Après une partie introductive (comprenant les trois premiers chapitres) dans laquelle nous donnons un bref historique sur les travaux réalisés concernant les estimations fonctionnelles, et où nous proposons une discussion sur les résultats obtenus, nous entamons notre travail par une partie théorique divisée en deux chapitres. Nous traitons, dans le 4^{ème} chapitre, des convergences uniformes en fenêtre h des quatre estimateurs à noyaux avec des données fonctionnelles, à savoir : régression, fonction de répartition conditionnelle, densité conditionnelle et fonction de hasard conditionnelle. Ces résultats purement théoriques nous permettent de développer, avec la technique de la validation croisée, des formules pour le choix automatique de la largeur optimale de la fenêtre h .

Dans le 5^{ème} chapitre, nous passons à l'étude des convergences uniformes des estimateurs de type k NN des quatre opérateurs suscités. Les résultats sont obtenus en appliquant directement les théorèmes établis au chapitre précédent. Nous proposons enfin des méthodes de sélection automatique du nombre optimal de voisins k .

Concernant la deuxième partie, nous l'avons consacrée aux études pratiques. Nous commençons au 7^{ème} chapitre par appliquer nos résultats sur des données simulées, afin de vérifier que le choix du paramètre de lissage par la méthode de la validation croisée est bien optimal. Nous poursuivons le travail au 8^{ème} chapitre par une application de nos résultats sur un cas réel d'étude sur données fonctionnelles, en appliquant chacun des opérateurs conditionnels étudiés. Une étude comparative entre

les méthodes de noyaux classiques et celles du k NN est proposée.

Enfin, nous achevons notre travail en dernière partie, par nos conclusions sur les résultats obtenus. Nous donnons également quelques perspectives de recherches que nous envisageons d'entreprendre, à court ou à plus long terme.

Les démonstrations détaillées des résultats théoriques sont données en annexe. Nous y énonçons également les théorèmes et définitions probabilistes auxquels nous avons eu recours tout au long de notre travail de recherche.

1.3 Liste des travaux

Publications

1. L. Kara-Zaitri, A. Laksaci, M. Rachdi, P. Vieu (2017). Uniform in bandwidth consistency for various kernel estimators involving functional data. *J. Nonparameter. Statist.*, **29**, 85–107.
2. L.Z. Kara, A. Laksaci, M. Rachdi, P. Vieu (2017). Data-driven k NN estimation in nonparametric functional data analysis. *J. of Multivariate Analysis*, **153**, 176–188.

Communications

L. Kara-Zaitri, A. Laksaci, M. Rachdi, P. Vieu (2017). Uniform in the smoothing parameter consistency results in functional regression. Poster à : *The 4th International Workshop on Functional and Operatorial Statistics. IWFOs 2017, A Coruña, SPAIN*

2.1 Contexte bibliographique

Tous les statisticiens qui manipulent des données s'accorderaient à dire que l'un des problèmes les plus importants qu'ils rencontrent dans leur domaine est l'étude de la relation qui puisse exister entre ces données. Cette étude statistique, a connu un tel développement depuis des dizaines d'années, qu'elle constitue aujourd'hui à elle seule une branche fondamentale de la statistique. En effet, qu'il s'agisse de problèmes de prédiction d'événements à venir, ou d'approximation de données manquantes, l'estimation statistique comble un réel besoin de compréhension et de prise de décision dans pratiquement tous les domaines de l'activité humaine : que ce soit pour la génétique, la santé, la météorologie, les sciences de l'ingénierie, les sciences humaines, l'industrie, pour toutes les branches de l'économie et pour bien d'autres secteurs.

De nos jours, nous ne comptons plus le nombre de travaux qui s'intéressent aux problèmes d'estimation statistique. Leurs études sont essentiellement classées en fonction de deux aspects : la nature des données dont ils disposent, et celle de la relation qui les relie.

Ainsi, l'étude est dite unidimensionnelle si chaque donnée dont nous disposons dans notre échantillon est sous forme d'un scalaire ; vectorielle (ou multidimensionnelle) si chaque donnée contient un nombre précis d'informations ; et fonctionnelle si elles contiennent un nombre infini d'informations, comme c'est le cas pour les courbes, les images ou les surfaces.

De même, l'étude est dite paramétrique si la relation à prédire est expliquée par un

nombre fini de paramètres réels, et non-paramétrique si nous ne pouvons écrire cette relation ni en fonction d'un nombre réduit de paramètres ni sous une forme explicite.

2.1.1 Analyse de données fini-dimensionnelles

Les problèmes d'inférence statistique ont probablement été traités depuis des centaines d'années. Mais les premiers travaux scientifiques publiés dans ce domaine remontent au début du 19ème siècle. Ces travaux s'intéressaient initialement aux questions d'estimation linéaire unidimensionnelle. En particulier, Legendre (1805) et Gauss (1809) ont introduit indépendamment la méthode d'estimation des moindres carrés en étudiant les positions des planètes et leurs trajectoires. Un peu moins d'un siècle après, en 1885, Galton a introduit la notion de régression et de corrélation lors d'une étude sur l'évolution de la taille des garçons en fonction de celle de leurs pères. Ce qui a amené Pearson à initier la notion de droite de régression en 1903.

Avec l'arrivée de l'outil informatique, les statisticiens avaient désormais davantage de possibilités de traiter les données et de les entrecroiser, et donc moins de précision sur la forme de la relation qui existe entre elles. Cela a donné le jour à un nouvel aspect de l'inférence, à savoir l'estimation non-paramétrique. Les recherches en ce sens ont été multiples, mais nous nous contenterons d'aborder les principaux travaux fondateurs.

En 1956, [Rosenblatt et collab.](#) ont proposé une discussion sur l'estimation non linéaire de la fonction de densité sur un échantillon de variables réelles indépendantes et identiquement distribuées. En 1961, [Tukey et collab.](#) ont introduit un nouvel estimateur non paramétrique pour la densité de probabilité, qui n'est autre que l'histogramme. Dans ces mêmes travaux, ils ont également proposé l'estimation de la régression par le régressogramme. Un an plus tard, [Parzen \[1962\]](#) a proposé une estimation à noyau uniforme pour la densité et a étudié sa convergence presque sûre. Il a utilisé ensuite ses résultats pour introduire un estimateur pour le mode. En 1966, [Cacoullos](#) a généralisé l'étude au cas multivarié.

C'est en 1964 que les modèles conditionnels d'estimation non paramétrique à noyaux auxquels nous nous intéressons au cours de cette thèse ont été initiés. Ces modèles introduits par les travaux de [Nadaraya \[1964\]](#) et ceux de [Watson \[1964\]](#) concernaient

l'estimation de la régression pour des données vectorielles. Nous renvoyons à l'article de Collomb [1981] pour une brève revue bibliographique des résultats initiaux relevant de ces estimateurs.

Signalons, à ceux désireux de s'approfondir dans le domaine, que les travaux de Stone renferment une riche collection de méthodes d'estimation non-paramétrique, où il traite de plusieurs propriétés asymptotiques. Ses publications demeurent, à elles seules, un bon répertoire de références bibliographiques. Nous pouvons citer, à titre d'exemple, les travaux qu'il a fait en 1977, 1982, 1984, où il traite de plusieurs estimateurs d'opérateurs conditionnels et discute les fonctions poids associées à chacun d'eux. Au cours de ces mêmes travaux, il a introduit la deuxième variante d'estimateurs auxquels nous nous intéressons dans notre thèse, à savoir les estimateurs de type k voisins les plus proches. Nous pouvons citer également ses recherches parues en 1985 qui traitent des modèles additifs. Ces modèles sont largement commentés dans Schimek et Turlach [1998].

L'analyse des modèles de dimensions finies est un domaine de la statistique toujours très étudié de nos jours. Mais n'étant pas le sujet de notre thèse, nous nous contenterons de ce rapide tour d'horizon qui nous permet d'introduire nos travaux, et renvoyons le lecteur à des recherches plus récentes les concernant. Toutefois, nous pouvons citer les livres de Hoel et collab. [1972], Wand et Jones [1994] et celui de Schimek [2012], qui regroupent et résument tous les outils et définitions nécessaires à l'étude des méthodes de lissage multivariées, et qui traitent en particulier des méthodes à noyaux que nous étudions dans cette présente thèse.

2.1.2 Analyse de données fonctionnelles

Au fil des années, les statisticiens ont fait face à des données définies sur des espaces dont la dimension était de plus en plus grande : des données dites fonctionnelles. Ils ont commencé tout naturellement à développer des méthodes de discrétisation de ces données pour les adapter aux méthodes vectorielles. Mais ils se sont vite aperçus que ces outils perdaient de leur efficacité face à des données aussi importantes. En effet, il s'avère que plus la dimension de la variable explicative augmente et moins les vitesses de convergence des estimateurs sont bonnes. Pour contourner ce paradoxe

connu sous le nom de " fléau de la dimension ", ils ont été amenés à introduire d'autres alternatives permettant de prendre en compte l'aspect continu des données et de tirer profit de toute la richesse de leurs observations.

À notre connaissance, les premiers travaux à avoir traité des données fonctionnelles remontent à 1958. En effet, dans ses travaux, [Rao \[1958\]](#) a étudié quelques méthodes statistiques pour la comparaison des courbes de croissance en envisageant l'analyse en composantes principales et l'analyse factorielle des données. Durant la même année, [Tucker \[1958\]](#) a lui aussi utilisé des techniques d'analyse factorielle pour la détermination des paramètres d'une relation fonctionnelle entre deux variables. En 1974, [Deville](#) a mis en évidence l'importance de l'étude de telles données et a proposé une vulgarisation des méthodes de l'analyse harmonique. Par la suite, [Ramsay \[1982\]](#) a proposé une extension des techniques d'analyse de données multi-variées au cas infini-dimensionnel.

Le premier modèle d'estimation de régression considéré dans le cas fonctionnel est un modèle linéaire, introduit et étudié par [Ramsay et Dalzell \[1991\]](#). Dans leur document, ils montrent comment la théorie des L-splines peut supporter des généralisations de la modélisation linéaire et analyse en composantes principales d'échantillons prélevés sur des fonctions aléatoires. En 1993, [Frank et Friedman](#) ont proposé une étude comparative des différentes méthodes paramétriques existantes pour l'estimation de la régression.

La théorie de l'analyse des données fonctionnelles a connu son plus grand essor en 1997 avec le livre de [Ramsay et Silverman](#), qui a été ensuite réédité en 2005. On y trouve un recueil important de méthodes statistiques "paramétriques" dans le cadre infini-dimensionnel. Dans le même esprit, le livre de [Bosq \[2000\]](#) a lui aussi fait état de différents aspects théoriques de l'étude des données fonctionnelles, en traitant des processus linéaires en temps continu. Plus récemment, l'analyse des modèles fonctionnels linéaires a été mise à l'honneur par plusieurs auteurs. Le lecteur désireux de s'initier à l'étude de ces modèles trouvera dans le livre de [Hsing et Eubank \[2015\]](#) toutes les définitions et explications fondamentales nécessaires à sa compréhension. Pour les lecteurs plus aguerris, il serait recommandé de se référer au livre de [Zhang \[2013\]](#) et à celui de [Horváth et Kokoszka \[2012\]](#) qui proposent des études approfondies des modèles paramétriques. Qu'il s'agisse de modèles simples ou conditionnels,

de cas d'observations indépendantes ou dépendantes, ou encore de variable réponses scalaire ou fonctionnelle, on y trouve un riche recueil de méthodes et de modèles différents à traiter. Dans la plupart des cas, des illustrations avec des données réelles sont abordées tout le long des deux ouvrages.

Pour des références entièrement dédiées aux études appliquées, les livres de [Ramsay et Silverman \[2002\]](#) et [Ramsay et collab. \[2009\]](#) sont particulièrement indiqués. Le premier se propose d'illustrer le fonctionnement des méthodes d'analyse fonctionnelle par des exemples pratiques variés, tandis que le second, plus technique, explique l'utilisation des deux langages mathématiques les plus populaires pour le traitement d'objets fonctionnels, à savoir les langages : R et Matlab. Pour plus de références, il serait recommandé de consulter les numéros spéciaux de [Manteiga et Vieu \[2007\]](#) et [Valderrama \[2007\]](#).

L'analyse paramétrique des données offre des résultats attrayants dans des espaces de dimension finie. Cependant, si les outils graphiques offrent un aperçu visuel clair de la relation entre des variables réelles, et qu'ils permettent donc de la modéliser par un nombre réduit de facteurs, ces outils ne sont hélas pas assez efficaces face à des variables de grandes dimensions. En effet, il est souvent difficile de modéliser d'une façon paramétrique la relation entre des données de type fonctionnel. C'est en cela que la modélisation non-paramétrique est préférable.

L'étude de modèles d'estimateurs "non paramétriques" sur des données fonctionnelles est plus récente que celle linéaire. Elle a été introduite en 1998 par [Gasser et collab.](#), lorsqu'ils ont proposé un estimateur non paramétrique du mode d'une variable fonctionnelle. L'idée adoptée pour établir des résultats asymptotiques était d'utiliser une approche fractale pour éviter de rencontrer le fameux fléau de la dimension. En 2000, [Ferraty et Vieu](#) ont utilisé la même approche pour généraliser l'estimateur de la régression introduit par Nadaraya et Watson au cas fonctionnel.

Nous renvoyons le lecteur au livre de [Ferraty et Vieu \[2006\]](#) pour une lecture plus approfondie et des références plus variées sur l'utilisation d'approches non paramétriques dans des problèmes d'estimation fonctionnelle statistique. Notons que depuis la parution de ce livre, des centaines de recherches ont été dirigées vers cette branche de la statistique qu'est "l'analyse non paramétrique des données fonctionnelles", et

ont proposé des études pour ces estimateurs sous différentes formes, tant sur le plan pratique que sur le plan théorique. Nous citerons les articles de [Geenens et collab. \[2011\]](#), [Cuevas \[2014\]](#), [Bongiorno et collab. \[2014\]](#) et [Goia et Vieu \[2016\]](#) pour une vaste sélection de références.

Vu l'étendu de la littérature disponible dans ce domaine, cette liste est très loin d'être exhaustive et de nombreuses réflexions et références intéressantes ont été laissées de côté par souci de concision et de lisibilité. Le but de cette thèse n'étant pas de donner un panorama complet de ce qui a été fait, nous avons délibérément choisi de ne donner que quelques exemples généraux de travaux que nous avons jugés pertinents pour notre recherche.

2.1.3 Choix du paramètre de lissage

D'un point de vue pratique, il est à noter que tous les estimateurs non paramétriques à noyaux dépendent fortement de deux paramètres, qui sont le noyau et la largeur de la fenêtre. Si la fonction noyau qui définit la forme du voisinage contenant les observations prises en compte dans l'étude n'a pas un véritable impacte sur la qualité de l'estimateur, il est évident que le paramètre de lissage qui contrôle la taille de ce voisinage joue un rôle déterminant. En effet, un choix judicieux de ce paramètre est primordial pour assurer à l'estimateur de bonnes performances. Mais le problème de sa sélection a fait l'objet de peu de travaux dans le cadre non paramétrique fonctionnel, et les recherches qui lui ont été consacrées ne ciblent, pour la grande majorité, que le modèle de régression.

Le problème de la sélection du paramètre de lissage a été abordé pour la première fois par [Rachdi et Vieu \[2007\]](#) en traitant de l'estimation de régression. L'idée était de proposer une version fonctionnelle des procédures de la validation croisée multivariée pour le choix automatique d'un paramètre de lissage global en présence d'un échantillon d'une distribution indépendante. Ces procédures sont connues pour être asymptotiquement optimales dans des contextes de régression. Durant la même année, l'article de [Benhenni et collab. \[2007\]](#) a également utilisé la généralisation de la méthode de validation croisée, et a montré que le fait d'utiliser la sélection d'un paramètre de lissage local plutôt que global peut considérablement améliorer la qualité de

l'estimateur non paramétrique de la régression. Quelques années plus tard, [Bouraine et collab. \[2010\]](#) ont généralisé ces résultats au cas d'index multi-fonctionnels. En 2013, [Laksaci et collab.](#) ont construit une règle de sélection du paramètre de lissage local entièrement axée sur les données par validation croisée et ont étudié ses propriétés asymptotiques lors de leur étude sur l'estimation de la densité conditionnelle. [Chagny et collab. \[2014\]](#) ont également proposé une règle de sélection du paramètre de lissage local orientée par les données mais avec d'autres méthodes, à savoir la méthode Goldenshluger et la méthode Lepski. Citons également les travaux de [Shang \[2014\]](#) qui proposent une approche Bayésienne pour estimer les paramètres de lissage optimaux dans un modèle de régression fonctionnelle non paramétrique qui admet des types mixtes de régresseurs avec des résidus de densité inconnue.

En outre, remarquons que l'une des préoccupations les plus importantes des problèmes de grande dimension est la prise en compte de la structure locale des éléments à traiter, mais la sélection d'un paramètre qui s'adapte localement à la forme des données renvoie systématiquement à une largeur de fenêtre aléatoire. Des recherches ont montré que l'utilisation de la méthode k NN est bien adaptée, particulièrement en présence de données aberrantes puisqu'elle prend en compte la rareté ou au contraire la densité des données dont elle dispose. Il semble que les travaux de [Burba et collab. \[2009\]](#) aient été les premiers à s'être intéressés à cette méthode dans le cas de données fonctionnelles et de réponse scalaire. Ils ont proposé un estimateur de type k NN pour l'opérateur de la régression non paramétrique, et en ont étudié les vitesses de convergence presque complète. Ils ont également proposé une étude comparative sur des données réelles entre la méthode du noyau classique et celle du type k NN. [Kudraszow et Vieu \[2013\]](#) ont établi la convergence uniforme presque complète de cet estimateur, et [Lian et collab. \[2011\]](#) a proposé une extension pour cet estimateur au cas de réponse également fonctionnelle. D'autres études ont traité de l'estimation k NN de la régression mais avec des méthodes différentes. Nous pouvons citer le travail de [Laloë \[2008\]](#) qui propose de traiter l'estimateur k NN de la régression en réduisant la dimension de l'espace fonctionnel, ou celui de [Biau et collab. \[2010\]](#) qui utilise la théorie du "*compact embedding*" pour étudier la vitesse de convergence de l'estimateur. Signalons que la littérature sur la question des méthodes k NN a été, elle aussi, jusqu'à présent peu développée et ne concerne principalement que l'opérateur de régression.

2.2 Présentation des modèles étudiés

Dans cette section, nous donnons un bref aperçu des travaux sur lesquels nous avons basé notre étude. Plus concrètement, nous présentons différentes méthodes à noyaux d'estimation d'une variable réelle Y à partir d'une autre variable X , toutes basées sur la distribution de Y conditionnée par X .

La présentation est divisée en quatre parties, chacune correspondante à un modèle d'opérateur conditionnel à estimer : régression, fonction de répartition conditionnelle, fonction de densité conditionnelle, et fonction de hasard conditionnelle.

Considérons dans tout ce qui suit une variable X à valeurs dans un espace \mathcal{F} (de dimension peut être infinie) muni d'une semi-métrique d , et une variable Y définie sur \mathbb{R} . Et prenons un échantillon de n observations $(X_i, Y_i)_{i=1;n}$ i.i.d et de même loi que le couple (X, Y) .

2.2.1 La régression

Pour étudier le comportement d'un phénomène à prédire (exprimé par la v.a Y) en fonction d'un autre phénomène explicatif (exprimé par la v.a X), l'outil le plus naturel et le plus souvent utilisé est la régression classique. Cet opérateur nous donne la valeur "*moyenne*" de Y en fonction de la valeur x que prend la variable X . Il est défini par :

$$m(x) = E[Y|X = x]$$

Ainsi, la relation entre Y et X est donnée par :

$$Y = m(X) + \epsilon$$

où le bruit ϵ exprime tous les facteurs explicatifs que nous ne pouvons pas observer, ou que nous ne voulons pas prendre en compte. Nous supposons dans tout ce qui va suivre que $\mathbb{E}(\epsilon/X) = 0$.

L'estimation la plus naturelle de la régression à un point $x \in \mathcal{F}$ consiste alors à calculer la moyenne pondérée des valeurs des Y_i observées en affectant à chacune

d'elles des poids de grandeurs différentes selon que les X_i qui leurs sont associés sont proches du point x ou pas. Ce que nous pouvons écrire mathématiquement par :

$$\hat{m}(x) = \frac{\sum_{i=1}^n Y_i W_i(x)}{\sum_{i=1}^n W_i(x)}$$

où le poids de Y_i est donné par la fonction W_i qui est à valeurs dans l'intervalle $[0; 1]$.

A - Contexte fini-dimensionnel

Nous nous plaçons dans un premier temps dans le cas où \mathcal{F} est un espace de \mathbb{R}^p (p étant un entier strictement positif) muni de sa norme euclidienne. Considérons le premier estimateur à noyau de la régression défini dans ce cadre et appelé "*régressogramme*". En partitionnant l'espace \mathcal{F} en plusieurs classes disjointes, et en notant C_x la classe de la partition qui contient x , le régressogramme considère des poids de la forme : $W_i(x) = \mathbb{1}_{C_x}(X_i)$. Cet estimateur présente comme inconvénient principal le fait que le choix de la finesse de la partition doit être effectué arbitrairement lors des calculs. Un inconvénient qui peut être évité en utilisant l'estimateur de "*la fenêtre mobile*" qui prend des poids de la forme $W_i(x) = \mathbb{1}_{|X_i - x| \leq h}$, où le paramètre h représente la largeur de la fenêtre et la fonction $\mathbb{1}$ renvoie à la fonction indicatrice. Mais cet estimateur présente également des inconvénients puisque, pour chaque largeur de fenêtre h choisie, il attribue un même poids $W_i(x) = 1$ à tous les Y_i correspondants (que les X_i de cette fenêtre soient très proches de x ou plus éloignés). Cette constance par morceaux le rend irrégulier.

C'est ainsi que l'estimateur de Nadaraya et Watson fut introduit en 1964 pour pallier les insuffisances de ces deux estimateurs primitifs. Les poids que ce nouvel estimateur attribue aux Y_i sont de la forme

$$W_i(x) = K \left(h^{-\frac{1}{d}}(x - X_i) \right),$$

où K est une fonction dite noyau, bornée, décroissante et à support dans $[0, 1]$, et le paramètre h est un réel positif à définir. Ainsi, plus la distance entre X_i et x est petite et plus le poids de Y_i est considérable, les Y_i correspondants au cas où $d(X_i, x) > h$ n'étant pas pris en compte.

En 1979, [Collomb](#) propose une version type k NN de l'estimateur à noyau de Nadaraya

et Watson en définissant les poids $W_i(x)$ comme suit :

$$W_i(x) = K(H_n^{-1}(x - X_i)) ,$$

où H_n est une variable aléatoire donnant la $k^{\text{ème}}$ plus grande valeur des $|x - X_i|$, $i = 1, n$. Ainsi, la largeur de la fenêtre n'est plus fixée mais fluctue selon les données, permettant de ce fait de ne considérer que les k données les plus similaires pour pondérer les Y_i . Ce côté aléatoire que prend la largeur de la fenêtre fait que ce nouvel estimateur prend davantage en considération les structures topologiques locales des données.

L'estimateur de Nadaraya et Watson a connu depuis son introduction des développements continus dans ce cadre multidimensionnel, et de nombreux auteurs s'y sont intéressés. Diverses recherches ont abordé des variantes et leurs propriétés asymptotiques sous différentes formes et avec différentes hypothèses, certaines plus restrictives que d'autres. Nous renvoyons, pour des références bibliographiques de ces travaux, des plus anciens aux plus récents, à [Collomb \[1981\]](#), [Sarda et Vieu \[2000\]](#) et [Schimek \[2013\]](#).

Un résultat de convergence a particulièrement retenu notre attention, celui de la convergence uniforme sur le paramètre de lissage h initié par [Einmahl et collab. \[2005\]](#). Notre travail s'inspire en grande partie des résultats qu'ils ont obtenus.

B - Contexte fonctionnel

Plaçons-nous à présent dans le cas où \mathcal{F} est un espace infini-dimensionnel. [Ferraty et Vieu \[2000\]](#) ont proposé une généralisation de l'estimateur à noyau introduit par Nadaraya et Watson à ce nouveau cadre fonctionnel, de la forme suivante :

$$\hat{m}(x) = \frac{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))Y_i}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))}$$

où K est une fonction noyau asymétrique à support dans $[0; 1]$, et $h = h_n$ une suite de réelles positive décroissante.

Ce qui les a conduits par la suite à définir un estimateur à noyau de type k NN en généralisant l'estimateur de Collomb au cadre fonctionnel. Ainsi ils définirent en [2009](#)

avec Burba l'estimateur suivant :

$$\widehat{m}_{kNN}(x) = \frac{\sum_{i=1}^n K(H_n^{-1}d(x, X_i))Y_i}{\sum_{i=1}^n K(H_n^{-1}d(x, X_i))}$$

où H_n est la variable aléatoire définie précédemment.

2.2.2 Fonction de répartition conditionnelle

La régression est certes l'outil le plus utilisé et le plus étudié en estimation, mais elle présente quelques inconvénients. Le premier obstacle qu'elle rencontre est celui des données aberrantes. En effet, la moyenne étant un opérateur sensible aux valeurs atypiques, l'estimation de la régression manque de robustesse dans certaines études, où par exemple elle est face à une distribution multimodale ou encore une distribution asymétrique. Son deuxième inconvénient, et peut-être le plus important, est que la régression se contente de résumer les données en une seule information, et ne peut décrire le comportement du processus au complet. Pour ce genre de besoins, la recherche de la fonction de répartition est plus adaptée que celle de la régression. En effet, la fonction de répartition, tout comme la fonction de densité de probabilité et la fonction caractéristique, a l'avantage de caractériser totalement la loi de la distribution de Y . L'estimation de la fonction de répartition est très utile par exemple pour estimer les chances qu'une observation appartienne à un intervalle donné, ou encore que la valeur à estimer ne dépasse pas un seuil fixé. Elle est aussi très pratique dans l'étude de l'estimation des quantiles.

A - Contexte fini-dimensionnel

La fonction de répartition de Y étant définie par :

$$F(y) = P(Y \leq y) ,$$

son estimateur le plus évident n'est autre que "la fonction de répartition empirique" définie par :

$$F_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Y_i \leq y\}} .$$

Mais même si cet estimateur rudimentaire est non biaisé et de variance minimale, et qu'il présente donc de bonnes propriétés de convergence uniforme, il a le défaut de ne pas prendre en compte les structures topologiques de l'espace sur lequel l'étude est effectuée.

Le deuxième estimateur logique qu'on pensa à introduire a été déduit de l'estimateur à noyau de la densité de probabilité introduit par Rosenblatt et Parzen et que nous présentons dans la section 2.2.3. Ainsi, l'estimateur à noyau de la fonction de répartition de Y est défini par :

$$\hat{F}(y) = \frac{1}{n} \sum_{i=1}^n H(h^{-1}(y - Y_i)) ,$$

où $H(x) = \int_{-\infty}^x K(y)dy$ et K est une fonction noyau.

La fonction de répartition de Y conditionnelle à X est donnée par :

$$F^x(y) = P(Y \leq y / X = x) .$$

Elle peut également être définie comme suit :

$$F^x(y) = \int_{-\infty}^y f^x(t)dt ,$$

où $f^x(y)$ est la densité de probabilité de Y sachant X définie par :

$$f^x(y) = \frac{f(x, y)}{f(x)} ,$$

de telle sorte que : $f(x, y)$ est la densité de probabilité conjointe du couple (X, Y) et $f(y)$ est la densité de probabilité de la variable X .

En 1969, **Roussas et collab.** a utilisé cette dernière définition pour introduire un estimateur de la fonction de répartition de Y conditionnelle à X par :

$$\hat{F}_R^x(y) = \int_{-\infty}^y \frac{q_n(x, x')}{p_n(x)} dx' ,$$

où $p_n(x)$ est l'estimateur à noyau traditionnel de la densité de la variable X donné par :

$$p_n(x) = \frac{1}{nh} \sum K(h^{-1/2}(x - X_i)) ,$$

et $q_n(x, x')$ est une adaptation à double noyaux de $p_n(x)$ pour estimer la densité conjointe du couple (X, Y) , définie comme suit :

$$q_n(x, x') = \frac{1}{nh} \sum K(h^{-1/2}(x - X_i)) K(h^{-1/2}(x' - X_i)) .$$

Dans le contexte fonctionnel, l'étude de la fonction de répartition conditionnelle n'est pas une extension directe de celle du contexte multivarié. Plus précisément, si le théorème de Jirina nous assure l'existence d'une distribution conditionnelle dans des espaces mesurables, cela n'est pas toujours le cas dans des espaces fonctionnels, où aucune mesure dominante n'est reconnue universellement. Nous nous devons donc de préciser que dans ce qui suit, il est implicitement supposé que la loi conditionnelle de Y sachant X est bien définie.

B - Contexte fonctionnel

Dans le cas où la variable X est dans un espace \mathcal{F} fonctionnel, l'estimateur de la fonction de répartition conditionnelle a été introduit par [Ferraty et collab. \[2006\]](#). Pour ce faire, ils se sont inspirés d'une adaptation de l'estimateur introduit par Roussas pour définir un estimateur comme suit :

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

où H est une fonction de répartition conditionnelle définie par :

$$H(\cdot) = \frac{\int_0^\cdot K(s)ds}{\int_0^{1/2} K(s)ds}$$

et $h_K = h_{K,n}$ (*resp.* $h_H = h_{H,n}$) est une suite de réels positifs.

Dans ce présent travail, nous apportons également une contribution à l'étude de cet estimateur.

2.2.3 Fonction de densité conditionnelle

Comme la densité de probabilité est obtenue en dérivant la fonction de répartition, et qu'inversement l'intégrale de la densité donne la fonction de répartition, il est vrai que la connaissance de l'une implique automatiquement celle de l'autre. Mais la fonction de densité a un net avantage sur le plan visuel, puisqu'elle donne un aperçu concret de quelques caractéristiques que l'autre fonction décrit d'une façon moins explicite, comme la symétrie, les pics ou les creux.

A - Contexte fini-dimensionnel

Dans le cas où \mathcal{F} est un intervalle de \mathbb{R} , [Tukey et collab. \[1961\]](#) furent les premiers à définir un estimateur non paramétrique pour la fonction de densité de probabilité, qui n'est autre que "l'Histogramme". Pour construire leur estimateur, ils ont découpé \mathcal{F} en classes C_j de largeurs l_j , $j = 1, \dots, k$. Ils ont alors proposé d'approcher la densité de probabilité f par une fonction en escalier définie sur la classe C_j qui contient le point y par :

$$\widehat{f}_T(y) = \sum_{j=1}^k \frac{p_j}{l_j} \mathbb{1}_{C_j}(y),$$

où p_j est la fréquence des points Y_i dans la classe C_j définie par :

$$p_j = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{C_j}(Y_i).$$

Cet estimateur présente le même inconvénient que celui du regressogramme. À savoir, le choix arbitraire de la largeur de la classe C_j fait que l'estimateur ne prend pas forcément en compte les points les plus proches de la valeur y où l'on effectue l'estimation. L'histogramme mobile est en ce sens plus approprié que son prédécesseur, puisqu'il est défini sur une classe C qui est choisie de manière à ce que y soit en son centre, *i.e.* $C = [y - h, y + h]$, où h est une largeur à fixer. L'histogramme mobile est donné par :

$$\widehat{f}_m(y) = \frac{1}{2hn} \sum_{i=1}^n \mathbb{1}_{|y - Y_i| \leq h}.$$

Ce dernier a encore été amélioré par [Parzen \[1962\]](#) qui a remplacé la fonction discrète $\frac{1}{2} \mathbb{1}_{[-1,1]}(h^{-1}(y - Y_i))$ par une fonction continue décroissante $K(h^{-1}(y - Y_i))$. C'est

alors qu'est né l'estimateur non paramétrique de la densité le plus répandu, qui est donné par :

$$\widehat{f}_p(y) = \frac{1}{nh} \sum_{i=1}^n K(h^{-1}(y - Y_i)) ,$$

où K est une fonction noyau.

Sa version multivariée (cas où \mathcal{F} est un espace de dimension p) a été initiée par [Cacoullos \[1966\]](#), et définie comme suit :

$$\widehat{f}_c(y) = \frac{1}{nh^p} \sum_{i=1}^n K(h^{-1}(y - Y_i)) .$$

Le modèle conditionnel auquel nous nous intéressons a été introduit pour la première fois par [Rosenblatt \[1969\]](#). En considérant le fait que le densité de Y conditionnée par X est définie par :

$$f^x(y) = \frac{f(x, y)}{f(x)}$$

Rosenblatt a utilisé l'estimateur de Parzen pour estimer $f(x)$, et sa version à deux dimensions pour estimer la densité conjointe $f(x, y)$ de la sorte :

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

et a ainsi défini l'estimateur de la densité conditionnelle comme suit :

$$\widehat{f}_R^x(y) = \frac{h_H^{-1} \sum_{i=1}^n K(h_K^{-1}(x - X_i)) K(h_H^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}(x - X_i))} .$$

B - Contexte fonctionnel

En s'inspirant des travaux initiaux de Rosenblatt, [Ferraty et collab. \[2006\]](#) ont initié l'estimateur de la densité de Y conditionnée par une variable fonctionnelle X comme suit :

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

2.2.4 Fonction de hasard conditionnelle

L'estimation de la fonction de hasard intervient dans l'analyse statistique des durées de vies. Son étude est très importante en théorie de fiabilité. Cette fonction, appelée aussi "taux de survie" sert à prévoir la durée écoulée avant l'apparition ou la disparition d'un phénomène, ou encore le temps nécessaire à un sujet pour passer d'un état à un autre. Elle est définie comme étant la probabilité instantanée qu'un phénomène sorte de son état à l'instant y . Autrement dit, si la variable aléatoire Y désigne le temps de vie du phénomène, la fonction de hasard est donnée par :

$$S(y) = \lim_{t \rightarrow 0} \frac{\mathbb{P}(y \leq Y < y + t / Y \geq y)}{t}.$$

En considérant que f est la densité de probabilité de la variable Y , et que F est sa fonction de répartition, la fonction de hasard peut être définie comme suit :

$$S(y) = \frac{f(y)}{1 - F(y)}.$$

En utilisant cette nouvelle définition, il nous suffit d'estimer les deux fonction f et F pour trouver l'estimateur de la fonction S . Nous pouvons également définir la fonction de hasard conditionnelle comme suit :

$$S^x(y) = \frac{f^x(y)}{1 - F^x(y)},$$

et utiliser les estimateurs des fonctions conditionnelles f^x et F^x pour estimer S^x .

Dans ce chapitre nous posons le cadre de travail, définissons les composantes topologiques de notre étude et mettons l'accent sur les choix des hypothèses que nous avons émises tout au long de nos travaux. Nous énonçons également les résultats principaux que nous avons obtenus.

3.1 Introduction

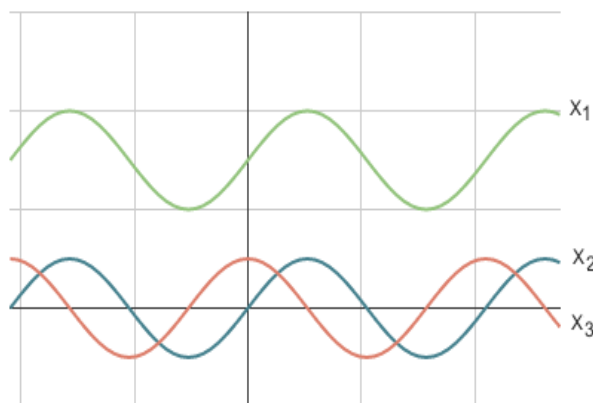
Cette thèse traite de l'estimation non paramétrique d'une variable aléatoire réelle Y connaissant le comportement d'un paramètre aléatoire fonctionnel X dont elle dépend. L'étude se base sur un échantillon de n observations $(X_i, Y_i)_{1 \leq i \leq n}$ indépendantes et de même distribution que le couple (X, Y) . Pour ce faire, prenons un espace fonctionnel \mathcal{F} sur lequel X est définie, et munissons-le d'une semi métrique d . Le choix de travailler sur un espace semi-normé plutôt que de se restreindre à un espace normé est motivé par deux aspects principaux, le premier est technique tandis que le second est plus théorique.

3.1.1 Semi-métriques et relation entre les données

Pour certaines études, le fait de mesurer les proximités entre les données avec une métrique peut être fort restrictif, puisque l'on risque de ne pas prendre en compte certaines caractéristiques fonctionnelles potentiellement primordiales à l'étude. En effet,

à la différence du cas vectoriel, les données fonctionnelles renferment des informations qui ne sont pas visibles directement, mais qui sont exprimées par leurs dérivées ou primitives, leurs intégrales, leurs périodicités ou leurs régularités, leurs variations, ou le déphasage entre elles.

Supposons, par exemple, que les variables fonctionnelles sont sous forme de courbes, et considérons les trois variables suivantes :

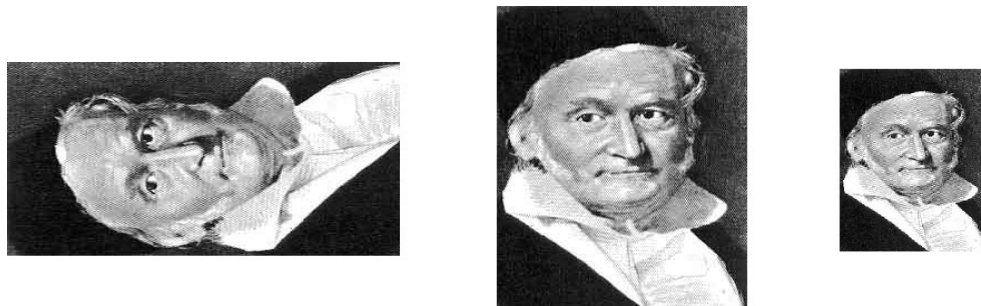


Si nous mesurons les distances entre ces courbes avec une norme quelconque, il est évident que nous obtiendrions des distances non nulles. Cela signifierait que les trois courbes sont différentes les unes des autres. Cependant, si l'étude ne prend en compte que la forme des courbes, nous aurions besoin d'une mesure de distance qui considérerait que X_1 et X_2 sont identiques et qu'elles sont toutes les deux différentes de X_3 . Dans ce cas précis, l'utilisation d'une norme ne serait pas adéquate, puisqu'il s'agirait d'avoir une distance nulle entre X_1 et X_2 alors que $X_1 \neq X_2$. Ainsi, l'utilisation d'une semi-norme qui traiterait des dérivées des fonctions serait plus appropriée. Si, au contraire, nous ne nous focalisons que sur la périodicité des trois fonctions, il serait préférable d'utiliser une semi-norme qui considérerait que les trois courbes sont équivalentes.

L'exemple que nous venons de voir traite de variables fonctionnelles sous forme de courbes, mais les mêmes idées pourraient s'appliquer à tout autre objet fonctionnel, comme des images par exemple. Les trois images suivantes¹ sont bel et bien identiques, et il faudrait donc mesurer les distances entre elles avec une semi-norme sur laquelle

1. Portrait de Carl Friedrich Gauss (1777 ; 1855)

les rotations et les changements d'échelles n'auraient aucune influence.



En résumé, des résultats sur des espaces fonctionnels semi-normés seront adaptés à des situations diverses. Cela répond tout aussi bien aux besoins d'études qui prennent en compte la moindre différence entre les données (sur des espaces de Banach ou de Hilbert ou autres espaces normés), qu'aux besoins d'études plus pointues qui ne se focalisent que sur certains aspects de divergence entre ces données.

3.1.2 Semi-métriques et mesure de concentration

Dans le même esprit que la sous-section précédente, nous allons discuter, d'un point de vue probabiliste cette fois-ci, de comment la qualité de l'estimation est liée à la mesure de proximité que nous choisissons d'utiliser. Par soucis de concision, nous ne traiterons que de l'estimation de l'opérateur de régression, mais les mêmes idées peuvent être adaptées à tout autre estimateur traité dans nos présents travaux.

Avant de commencer, faisons un bref retour au cadre vectoriel de très grande -mais finie- dimension, que l'on notera d . Nous savons que si, entre autres conditions, l'opérateur de régression est k fois différentiable (k étant un entier naturel), la vitesse de convergence de son estimateur est de l'ordre de

$$O(h^k) + O_{p.co.} \left(\sqrt{\frac{\log n}{nh^d}} \right)$$

où le paramètre h , qui quantifie la notion de proximité entre deux points, doit tendre autant que faire se peut, vers 0. Ce qui ne peut être réalisé que dans un espace dont la densité des points n'est pas négligeable.

Ainsi, la qualité de l'estimateur semble être étroitement liée à la "densité" des données

dans leur espace. Considérons donc la fonction de mesure de concentration suivante :

$$\forall r > 0, \phi_x(r) =: \mathbb{P}(X \in B(x, r))$$

appelée mesure d'une petite boule de probabilité. Bien évidemment, plus les variables sont éparpillées dans leur ensemble, moins la fonction ϕ est concentrée, et moins vite l'estimateur convergera. Pour avoir de bonnes propriétés asymptotiques nous avons donc besoin que ϕ soit la plus grande possible.

En 1982, [Stone](#) a montré que si le vecteur X est absolument continu par rapport à la mesure de Lebesgue, et si sa densité f est continue et strictement positive, alors :

$$\phi_x(r) \sim Ch^d \text{ quand } h \rightarrow 0$$

Il s'en suit que la vitesse de convergence de l'estimateur de la régression devient, avec un choix approprié de h , de l'ordre de :

$$O_{p.co.} \left\{ \left(\frac{\log n}{n} \right)^{\frac{k}{2k+d}} \right\}.$$

De plus, [Stone \[1982\]](#) a montré que cette vitesse de convergence ne pouvait être améliorée, et que par conséquent, la qualité de l'estimateur était inversement proportionnelle à la dimension de l'espace traité. Paradoxe connu sous le nom de fléau de la dimension.

Revenons à présent au cadre fonctionnel. [Ferraty et Vieu \[2006\]](#) ont montré que l'estimateur \hat{m} de la régression convergeait avec une vitesse de l'ordre de

$$O(h^\beta) + O_{p.co.} \left(\sqrt{\frac{\log n}{n\phi_x(h)}} \right)$$

où le paramètre β contrôle la régularité de l'opérateur de régression.

Dans leur article, [Ferraty et collab. \[2006\]](#) ont donné l'expression de la mesure d'une petite boule de probabilité pour plusieurs modèles différents de processus à temps continu. Ils ont également donné pour chaque cas la vitesse de convergence de l'estimateur basé sur le processus en question. Nous pouvons en citer les deux exemples suivants.

Exemple 1. Processus de diffusion

Considérons l'espace fonctionnel \mathcal{F} comme étant l'espace de Cameron-Martin² de l'espace $C([0, 1]; \mathbb{R}^p)$ ³. Prenons comme métrique d celle associée à la norme infinie :

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

et notons par P_w la mesure de probabilité de Wiener⁴ définie sur $C([0, 1]; \mathbb{R}^p)$. [Bogachev, 1998, p. 187] a montré que la mesure de Wiener d'une petite boule de probabilité centrée était de la forme

$$P_w (\|x\|_{\text{sup}} \leq h) \sim \frac{4}{\pi} e^{-\pi^2/8h^2}$$

Il s'en suit, d'après la définition de l'espace de Cameron-Martin \mathcal{F} , que :

$$P_w (\|x - x_0\|_{\text{sup}} \leq h) \sim C_{x_0} \frac{4}{\pi} e^{-\pi^2/8h^2} \quad , \quad \forall x_0 \in \mathcal{F} \quad (*)$$

À présent, notons par w_t le processus de Wiener, et considérons le processus de diffusion X_t^{diff} suivant :

$$X_t^{\text{diff}} = \int_0^t \Theta(s, X^{\text{diff}}) ds + w_t$$

où la fonction Θ est telle que l'équation précédente a une unique solution X^{diff} . Liptser et Shiryaev [1977] ont montré que si cette fonction vérifie l'hypothèse suivante :

$$\int_0^1 \Theta^2(s, X^{\text{diff}}) ds < \infty \quad ; \quad \text{p.s}$$

alors le processus X_t^{diff} a une mesure de probabilité absolument continue par rapport à la mesure P_w . Ainsi, sous l'hypothèse précédente, (*) implique que

$$P (\|X^{\text{diff}} - x_0\|_{\text{sup}} \leq h) \sim C_{x_0} e^{-\pi^2/8h^2}$$

2. Soit E un espace localement convexe avec une mesure μ . [Bogachev, 1998, p. 61] a défini l'espace de Cameron-Martin de E par $E^{CM} = \{x \in E / \mu(\cdot - x) \sim \mu(\cdot)\}$.

3. $C([0, 1]; \mathbb{R}^p)$ est l'ensemble des fonctions continues à valeurs dans \mathbb{R}^p définies sur $[0, 1]$.

4. Soient $(t_i)_{i=1,k} \in [0, 1]$ tels que $0 < t_1 < t_2 < \dots < t_k \leq 1$, et $(A_i)_{i=1,k}$ une famille de Borels de $[0, 1]$. En considérant la densité $f_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}$, la mesure de Wiener P_w est définie pour tout $X \in C([0, 1]; \mathbb{R}^p)$ par :

$$P_w (X(t_i) \in A_i, i = 1, k) = \int_{A_1} \dots \int_{A_k} f_{t_1}(x_1) f_{t_2-t_1}(x_2 - x_1) \dots f_{t_k-t_{k-1}}(x_k - x_{k-1}) dx_k \dots dx_1.$$

Ce qui donne finalement l'expression de la mesure de concentration d'un processus de diffusion sous la forme

$$\phi_x^{\text{diff}}(h) \sim C_x e^{-\pi^2/8h^2}$$

Il suffit alors de choisir un paramètre h tel que

$$h \sim C(\log n)^{-1/2}, \text{ p.co}$$

pour obtenir une vitesse de convergence de l'estimateur \hat{m} de l'ordre de

$$O_{p.co.}(\log n)^{-\beta/2}$$

Le lecteur intéressé par l'application de nos estimateurs à des processus de diffusion trouvera dans [Dabo-Niang \[2002\]](#) et [Banon \[1978\]](#) plus de détails concernant ce type de processus dans des contextes différents.

Exemple 2. Mouvement Brownien Fractionnel

Prenons maintenant l'espace $C([0, 1]; \mathbb{R})$, et \mathcal{F} l'espace de Cameron-Martin qui lui est associé. La métrique d utilisée ici est toujours celle associée à la norme infinie.

Considérons X^{MBF} le mouvement Brownien Fractionnel⁵ de paramètre α , $0 < \alpha < 2$.

En combinant les théorèmes 3.1 et 4.6 de [Li et Shao \[2001\]](#), nous obtenons :

$$C_{x_0} e^{h^{-2/\alpha}} \leq P(\|X^{\text{MBF}} - x_0\|_{\text{sup}} \leq h) \leq C'_{x_0} e^{h^{-2/\alpha}}, \forall x_0 \in \mathcal{F}$$

Ce qui exprime la mesure d'une petite boule de probabilité d'un processus de diffusion sous la forme

$$\phi_x^{\text{MBF}}(h) \sim C_x e^{h^{-2/\alpha}}$$

En choisissant ensuite le paramètre de lissage h tel que

$$h \sim C(\log n)^{-\alpha/2},$$

nous arrivons à une vitesse de convergence de l'ordre de

$$O_{p.co.}(\log n)^{-\alpha\beta/2}$$

5. Un mouvement Brownien Fractionnel β d'ordre α est un processus Gaussien centré tel que $\beta_0 = 0$ et $\forall t \neq s \in [0, 1], \mathbb{E}|\beta_t - \beta_s| = |t - s|^\alpha$.

Nous remarquons que pour les deux exemples cités, la mesure de concentration est de la forme

$$\phi_x(h) \sim C_x e^{Ch^{-2a}}, \quad a > 0.$$

Il s'avère que pour la grande majorité des processus stochastiques connus, la mesure d'une petite boule de probabilité est toujours sous forme exponentielle, et ce quelque soit la norme avec laquelle l'espace topologique est muni. Cette tendance exponentielle que prend ϕ fait qu'elle décroît rapidement autour du point zéro, ce qui mène, comme nous venons de le constater, à des vitesses de convergence plus lentes que celles atteintes dans le cadre multivarié fini.

Puisque la fonction ϕ est directement liée à l'effet de dispersion des données, et que cet effet est lui même contrôlé par la structure topologique de l'espace, il serait intéressant de changer de structure topologique au profit d'une autre structure où les données seraient plus concentrées. Dans cette optique, [Ferraty et Vieu \[2006\]](#), ont montré que le fait de travailler sur des structures topologiques liées à des semi-normes plutôt qu'à des normes, augmentait considérablement la concentration des données. Pour ce faire, ils ont construit une semi-métrique de projection pour laquelle le processus stochastique X est nécessairement de type fractal⁶.

En prenant pour espace fonctionnel \mathcal{F} un espace de Hilbert séparable, et en considérant $\langle \cdot, \cdot \rangle$ comme étant le produit scalaire sur une base orthonormée $\{e_i, i \in \mathbb{N}^*\}$, les auteurs ont construit une famille de semi-métrique d_k ($k \in \mathbb{N}^*$ fini) de projections sur des sous ensembles de dimension k telle que

$$d_k(X_1, X_2) = \sqrt{\sum_{i=1}^k \langle X_1 - X_2, e_i \rangle^2}.$$

Dans un tel contexte, ils ont montré que si X était une variable fonctionnelle de carré intégrable, et si la variable $X_k = (X^1, X^2, \dots, X^k)$ est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}^k , et de densité strictement positive et continue sur le point $x_k = (x^1, x^2, \dots, x^k)$, alors X est un processus de type fractal d'ordre k par rapport à la nouvelle semi-métrique d_k . La vitesse de convergence de

6. Un processus stochastique X est dit de type fractale d'ordre τ , s'il existe une constante C telle que la mesure de concentration associée ϕ_X est de la forme $\phi_X(h) \sim Ch^\tau$ lorsque $h \rightarrow 0$.

l'estimateur sur le nouvel espace est de l'ordre de :

$$O_{p.co.} \left\{ \left(\frac{\log n}{n} \right)^{\frac{2b}{2b+k}} \right\}.$$

Cette semi-métrique de projection d_k permet de réduire la dimension et de passer d'un espace fonctionnel de dimension infinie à un espace k -dimensionnel. Cette réduction de dimension mène à une mesure de concentration plus dense et qui décroît plus lentement autour du point zero. Il en résulte que la vitesse de convergence qu'atteint l'estimateur par ce type de semi-métriques de projection d'un espace de Hilbert est beaucoup plus attrayante que celle obtenue sur des espaces normés.

Le choix de la semi-métrique utilisée dépend donc de deux aspects principalement : la concentration des données et l'objectif du problème statistique étudié. Cela étant, le plus souvent en pratique, on choisira une semi-métrique dérivative définie comme suit :

$$\text{pour } x_1, x_2 \in \mathcal{F} \text{ et } q \in \mathbb{R};, d_q(x_1, x_2) = \int_{\mathbb{R}} \left(x_1^{(q)}(t) - x_2^{(q)}(t) \right)^2 dt$$

où $x^{(q)}$ désigne la dérivée d'ordre q de x . Il s'avère que cette famille de semi métriques, qui est de type projection, fait preuve d'une efficacité avérée face à des courbes qui ne sont pas nécessairement observées sur les mêmes points ou sur des grilles de mesure obligatoirement fines, mais elles présentent l'inconvénient de ne s'appliquer que sur des courbes lisses.

3.2 Hypothèses

Afin d'assurer de bonnes propriétés asymptotiques aux estimations effectuées, nous devons, bien évidemment et avant toute chose, imposer quelques conditions sur les modèles d'estimateurs utilisés. Cependant la qualité des estimations obtenues n'est pas liée qu'aux modèles choisis, mais dépend également d'autres aspects sur lesquels nous n'avons aucun contrôle. Ces aspects concernent principalement les composantes topologiques des espaces traités, et la relation qui existe entre les co-variables. Mais n'ayant aucune, ou très peu, d'informations sur leurs natures, nous nous contenterons de supposer qu'elles vérifient nos hypothèses.

La qualité de l'estimateur se résume en deux termes, biais et variance. Si nous voulons contrôler le terme probabiliste (variance) de l'estimateur, il est inutile d'appréhender la notion de densité de probabilité de la distribution de la variable fonctionnelle X , du moment que nous ne savons pas avec quelle mesure la déduire (aucune mesure de référence sur les espaces de dimension infinie n'étant universellement reconnue, comme la mesure de Lebesgue dans le cas fini-dimensionnel). Ainsi, pour maîtriser la concentration de la variable fonctionnelle X qui intervient dans le terme variance, nous utiliserons les considérations de mesure de "*petites boules de probabilités*" suivantes :

$$\text{pour tout } r > 0, \mathbb{P}(X \in B(x, r)) =: \phi_x(r) > 0 \quad (3.1)$$

où $B(x, r)$ est la boule fermée de centre x et de rayon r . Cette condition sur la fonction de concentration, largement commentée dans [Ferraty et collab. \[2006\]](#), [Ferraty et Vieu \[2006\]](#), [Geenens et collab. \[2011\]](#), est un outil fondamental dans ce cadre d'étude non paramétrique fonctionnelle. Cela nous assure que les données ne soient pas très éparées dans leur ensemble, ce qui nous garantit une certaine régularité sur l'estimateur obtenu.

Nous avons également besoin de contrôler le comportement local de cette fonction de concentration autour du point 0 pour arriver à borner le terme biais de l'estimateur. Il suffit pour cela de lui garantir une variation régulière autour de 0 de telle sorte que :

$$\text{pour tout } s \in (0, 1), \lim_{r \rightarrow 0} \frac{\phi_x(sr)}{\phi_x(r)} =: \tau_x(s) < \infty \quad (3.2)$$

Cette condition initiée par [Haan \[1971\]](#) n'est pas très restrictive. Le lecteur intéressé par son utilisation pourra trouver plus de détails la concernant dans les articles de [Ferraty et collab. \[2007\]](#) et [Mas \[2012\]](#).

La nature de l'opérateur à estimer a, elle aussi, un impact sur le biais de l'estimateur. En effet, plus il est régulier et moins son estimateur est biaisé. Le problème est que, les espaces fonctionnels considérés n'ont pas nécessairement de structure de Banach sur laquelle étudier une hypothétique différentiabilité, et nous ne pouvons donc pas nous contenter de supposer une classe D^k ou C^k pour l'opérateur⁷. Pour contourner

7. Une fonction est dite de classe :

- D^k si elle est k fois dérivables.
- C^k si elle est k fois dérivables telles que sa $k^{\text{ème}}$ dérivée est continue.

ce problème, nous devons donner une condition de régularité plus forte que la simple continuité. Ce qui est faisable en supposant les modèles de type Holderien définis pour β , β_1 et β_2 des réels positifs tels que :

$$H_{(\mathcal{F},\beta)} = \left\{ \begin{array}{l} f : \mathcal{F} \rightarrow \mathbb{R} / \exists C > 0, \forall x_1, x_2 \in \mathcal{F} : \\ |f(x_1) - f(x_2)| < C d^\beta(x_1, x_2) \end{array} \right\} \quad (\text{H1})$$

ou encore :

$$H_{(\mathcal{F} \times \mathbb{R}, \beta_1, \beta_2)} = \left\{ \begin{array}{l} f : \mathcal{F} \times \mathbb{R} \rightarrow \mathbb{R} / \exists C > 0, \forall x_1, x_2 \in \mathcal{F} \text{ et } \forall y_1, y_2 \in \mathbb{R} : \\ |f(x_1, y_1) - f(x_2, y_2)| < C \left(d^{\beta_1}(x_1, x_2) + |y_1 - y_2|^{\beta_2} \right) \end{array} \right\} \quad (\text{H2})$$

Autrement dit, cette hypothèse assure une variation lente de l'opérateur comparée à celle de X . Plus les paramètres β_i sont petits, plus les courbes de l'évolution de l'opérateur en fonction de X sont lisses, et moins l'estimateur obtenu sera biaisé.

Passons à présent au noyau K utilisé pour la construction de nos estimateurs. Comme nous l'avons vu dans le chapitre précédent, le poids associé à chaque Y_i doit être choisi de telle sorte que plus le point X_i correspondant est loin du point x et moins l'estimateur accorde de l'importance à Y_i . Il est donc naturel de considérer un noyau $K(\gamma^{-1}d(X_i, x))$ borné sur \mathbb{R}^+ qui décroît à mesure que la distance $d(X_i, x)$ augmente. Et pour faciliter les calculs nous le considérons à support dans $[0, \frac{1}{2}]$, mais nos résultats restent valables sur tout autre support compact de \mathbb{R}^+ .

Par ailleurs, nous avons besoin que la variable aléatoire $K(\gamma^{-1}d(X_i, x))$ ne soit pas centrée afin de pouvoir contrôler le terme biais de l'estimateur. Autrement dit il faut que l'espérance $E(K(\gamma^{-1}d(X_i, x)))$ soit strictement positive. En appliquant le théorème de Fubini, cette condition revient à supposer que K soit de classe C^1 et que :

$$K(1/2) - \int_0^{1/2} K'(s) \tau_x(s) ds > 0 \quad (3.3)$$

Toutes ces hypothèses énoncées jusqu'à présent, sont assez communes aux études de convergences ponctuelles d'estimateurs non paramétriques sur des données multivariées (de dimension finie ou infinie). Elles sont, pour la plupart, peu restrictives et facilement vérifiées en pratique.

Le principal changement que nous apportons dans nos calculs, est que pour obtenir l'uniformité de la convergence sur le paramètre de lissage, nous avons besoin

d'avoir des résultats de convergence sur le noyau aléatoire $K(\gamma^{-1}d(X, x))$ lorsque le paramètre tend vers 0. Là encore, puisque nous n'avons supposé l'existence d'aucune mesure de densité de probabilité, il nous faut un autre moyen pour imposer une telle convergence. Pour ce faire, considérons l'ensemble des fonctions $f_\gamma : \mathcal{F} \rightarrow \mathbb{R}^+$ suivant :

$$\mathcal{K} = \{f_\gamma(\cdot) = K(\gamma^{-1}d(x, \cdot)), \gamma > 0\}$$

Il nous suffit alors de faire en sorte que cette classe de fonctions soit de Donsker⁸ pour assurer la convergence de la fonction f_γ .

Une première méthode alternative pour obtenir une classe de Donsker serait d'utiliser les notions de probabilité extérieure et d'espérance extérieure sur cette classe de fonction. Ces notions ont été initiées par Hoffmann-Jørgensen [1974] et largement étudiées dans Van Der Vaart et Wellner [1996]. Mais nous avons préféré éviter les calculs contraignants que cette méthode engendrerait, et avons utilisé à la place une version du Théorème 2.5.2 de l'ouvrage [Van Der Vaart et Wellner, 1996, p.127] pour construire une classe de Donsker \mathcal{K} . Nous avons donc besoin d'imposer les deux conditions :

- \mathcal{K} vérifie la faible condition de mesurabilité suivantes :

$$\mathcal{K} \text{ est une classe mesurable ponctuellement} \quad (3.4)$$

- La taille de la classe \mathcal{K} doit être réduite de telle sorte que son entropie (Définition 5 p.114) soit uniformément $\sqrt{\log}$ -intégrable, c.à.d :

$$\sup_Q \int_0^1 \sqrt{1 + \log \mathcal{N}(\epsilon \|F\|_{Q,2}, \mathcal{K}, d_Q)} d\epsilon < \infty \quad (3.5)$$

où F est la fonction enveloppe (Définition 4 p.114) de \mathcal{K} , d_Q est la métrique de $L_2(Q)$ et $\|\cdot\|_{Q,2}$ est la norme de $L_2(Q)$.

Enfin, nous avons choisi de borner le paramètre de lissage entre deux suites de réelles tendant toutes les deux vers 0, de telle sorte que :

- Pour les estimateurs à noyau, on considère deux suites a_n et b_n telles que $a_n < b_n$ et :

$$\frac{\log n}{n \min(a_n, \phi_x(a_n))} \rightarrow 0 \quad (3.6)$$

8. Définition tirée du livre de [Van Der Vaart et Wellner, 1996, p.80] : Soit $(z_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires i.i.d. à valeurs dans un espace mesurable (\mathcal{E}, β) . Notons \mathcal{K} une classe de fonctions mesurables $f : \mathcal{E} \rightarrow \mathbb{R}$, et considérons le processus empirique \mathcal{G}_n indexé par \mathcal{K} . Nous disons que \mathcal{K} est une classe de Donsker s'il existe un Borel mesurable tendu \mathcal{G} de $l^\infty(\mathcal{K})$ (espace des suites des suites bornées f_γ de \mathcal{K}), tel que $\mathcal{G}_n \sim \mathcal{G}$ dans $l^\infty(\mathcal{K})$.

- Pour les estimateurs de type k NN, on considère deux suites $k_{1,n}$ et $k_{2,n}$ telles que $k_{1,n} < k_{2,n}$ et :

$$\frac{\log n}{\min\{n\phi_x^{-1}\left(\frac{k_{1,n}}{n}\right), k_{1,n}\}} \rightarrow 0 \quad (3.7)$$

3.3 Résultats de convergences obtenus

Dans cette section, nous énonçons les résultats de convergences uniformes d'estimateurs que nous avons traités le long de nos travaux. La présentation est divisée en quatre parties, chacune traitant d'un modèle conditionnel. Dans chaque cas, nous donnons les résultats de convergence pour les deux types d'estimateurs : noyau et k NN.

3.3.1 Régression

Pour borner le terme variance de l'estimateur de la régression, nous avons besoin de cette dernière hypothèse :

$$\text{Il existe } m \geq 2 \text{ et } C > 0, \text{ tels que : } \mathbb{E}[|Y|^m|X] < C_2 < \infty, \text{ p.s.} \quad (3.8)$$

A - Estimateur à noyau

Le théorème suivant concerne l'estimateur à noyau de la régression défini par :

$$\widehat{m}(x) = \frac{\sum_{i=1}^n K(h_K^{-1}d(x, X_i)) Y_i}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))}.$$

Théorème 3.1. — *Sous les hypothèses (3.1)-(3.6) et (3.8), et si l'opérateur m est de la famille (H1), nous avons :*

$$\sup_{a_n \leq h_K \leq b_n} |\widehat{m}(x) - m(x)| = O(b_n^\beta) + O_{p.co.} \left(\sqrt{\frac{\log n}{n\phi_x(a_n)}} \right).$$

Cet estimateur a largement été étudié dans plusieurs contextes différents. La convergence uniforme que nous avons abordée, et qui est énoncée dans la section 4.3 du chapitre 4, est une généralisation des résultats obtenus dans le cadre vectoriel par Einmahl et collab. [2005] au cadre fonctionnel de dimension non nécessairement finie.

B - Estimateur k NN

L'estimateur k NN de la régression est défini par :

$$\widehat{m}_k(x) = \frac{\sum_{i=1}^n K(H_{k,x}^{-1}d(x, X_i)) Y_i}{\sum_{i=1}^n K(H_{k,x}^{-1}d(x, X_i))},$$

où

$$H_{k,x} = \min \left\{ h \in \mathbb{R}^+ \text{ tels que } \sum_{i=1}^n \mathbb{1}_{B(x,h)}(X_i) = k \right\}.$$

Théorème 3.2. — *Sous les hypothèses (3.1)-(3.5), (3.7), (3.8), et si l'opérateur m est de la famille (H1), nous avons :*

$$\sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{m}_k(x) - m(x)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^\beta \right\} + O_{p.co.} \left(\sqrt{\frac{\log n}{k_{1,n}}} \right).$$

Les résultats de ce théorème, énoncés dans la section 5.3 du chapitre 5, sont une extension du théorème 3.2 de Burba et collab. [2009] au cas où la convergence est obtenue uniformément sur un nombre de voisins k_n compris entre deux suites réelles $k_{1,n}$ et $k_{2,n}$ tendant vers zero.

3.3.2 Fonction de répartition conditionnelle

A - Estimateur à noyau

L'estimateur de la fonction de répartition conditionnelle est donné par :

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

où H est une fonction de répartition conditionnelle qui vérifie :

$$H(\cdot) = \frac{\int_0^{\cdot} K(s)ds}{\int_0^{1/2} K(s)ds}.$$

Cet estimateur dépend de deux paramètres de lissage. L'uniformité doit donc être étudiée sur chacun d'eux. Ainsi les conditions (3.4) et (3.5) doivent être adaptées au noyau H également. Nous supposons alors que la classe :

$$\mathcal{K}_0 = \left\{ \cdot \mapsto \frac{1}{\int_0^1 K(s)ds} \int_0^{(\delta^{-1}(y-\cdot))} K(s)ds, \delta > 0 \right\} \text{ est mesurable ponctuellement} \quad (3.9)$$

et que

$$\sup_Q \int_0^1 \sqrt{1 + \log \mathcal{N}(\epsilon \|F_0\|_{Q,2}, \mathcal{K}_0, d_Q)} d\epsilon < \infty \quad (3.10)$$

où F_0 est la fonction enveloppe de \mathcal{K}_0 .

Aussi, nous prenons $h_H = h_{H,n}$ une suite de réels positifs qui appartient à un intervalle (c_n, d_n) avec $\lim_{n \rightarrow \infty} c_n = \lim_{n \rightarrow \infty} d_n = 0$.

Théorème 3.3. — *Sous les hypothèses (3.1)-(3.6), (3.9), (3.10), et si l'opérateur F^x est de la famille (H2), nous avons :*

$$\sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} |\widehat{F}^x(y) - F^x(y)| = O(b_n^{\beta_1}) + O(d_n^{\beta_2}) + O_{p.co} \left(\sqrt{\frac{\log n}{n\phi_x(a_n)}} \right).$$

L'étude de la convergence presque complète de cet estimateur a été réalisée par [Ferraty et collab. \[2006\]](#). Dans nos présents travaux, nous en avons étudié l'uniformité sur le paramètre de lissage h . Ce résultat de convergence est détaillé dans la section 4.4 et démontré dans l'annexe B.1.

B - Estimateur k NN

Le théorème suivant traite de l'estimateur de la fonction de répartition conditionnelle défini ci dessous :

$$\widehat{F}_k^x(y) = \frac{\sum_{i=1}^n K(H_{k,x}^{-1}d(x, X_i)) \mathbb{1}_{(Y_i \leq y)}}{\sum_{i=1}^n K(H_{k,x}^{-1}d(x, X_i))}.$$

Théorème 3.4. — *Sous les hypothèses (3.1)-(3.5) et (3.7), et si l'opérateur F^x est de la famille (H1), nous avons :*

$$\sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{F}_k^x(y) - F^x(y)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^\beta \right\} + O_{p.co.} \left(\sqrt{\frac{\log n}{k_{1,n}}} \right).$$

Ce théorème est traité dans la section 5.3. Remarquons que le fléau de la dimension infinie se manifeste ici dans le terme $\sqrt{\frac{\log n}{k_{1,n}}}$. En effet, la partie biais qui dépend de la régularité de F^x ne pose pas de problème, puisqu'elle diminue à mesure que la suite $k_{2,n}$ va vers zero. Mais la partie dispersion est inversement proportionnelle à la suite $k_{1,n}$ et augmente avec la diminution de cette dernière.

3.3.3 Fonction de densité conditionnelle

Pour obtenir nos résultats de convergence presque complète uniformément sur le paramètre de lissage de l'estimateur de la fonction de densité de probabilité, nous utilisons les mêmes hypothèses que pour l'estimateur de la régression à une différence près.

A - Estimateur à noyau

L'estimateur de la densité de probabilité conditionnelle suivant :

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

étant défini par deux noyaux, nous devons d'une part étudier l'uniformité sur les paramètres de lissages de chacun des deux noyaux, mais également poser une nouvelle conditions sur le paramètre h_H . Ainsi, nous remplaçons l'hypothèse (3.6) par la suivante :

L'intervalle (c_n, d_n) contenant h_H , et tel que $\lim_{n \rightarrow \infty} c_n = \lim_{n \rightarrow \infty} d_n = 0$, vérifie la condition :

$$\frac{\log n}{nc_n \min(a_n, \phi_x(a_n))} \rightarrow 0. \quad (3.11)$$

Théorème 3.5. — *Sous les hypothèses (3.1)-(3.5) et (3.11), et si l'opérateur f^x est de la famille (H2), nous avons :*

$$\sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} |\widehat{f}^x(y) - f^x(y)| = O(b_n^{\beta_1}) + O(d_n^{\beta_2}) + O_{p.co.} \left(\sqrt{\frac{\log n}{n c_n \phi_x(a_n)}} \right).$$

Tout comme dans les résultats initiaux de convergences presque complètes obtenus par Ferraty et collab. [2006], ce résultat de convergence uniforme sur h , introduit dans la section 4.5 du chapitre 4, est divisé en deux termes. Le terme déterministe dépend de la régularité de l'opérateur estimé (exprimé ici à travers les paramètres β_1 et β_2), tandis que le terme probabiliste dépend de la concentration des données qui apparaît dans la fonction de mesure d'une petite boule de probabilité ϕ_x . Les démonstrations de ce théorème sont données à l'annexe B.1.

B - Estimateur k NN

De même que l'estimateur classique de la fonction de densité de probabilité conditionnelle, l'estimateur k NN prend en compte, lui aussi, deux noyaux et deux paramètres de lissage. Il est défini par :

$$\widehat{f}_{k,\ell}^x(y) = L_{\ell,y}^{-1} \frac{\sum_{i=1}^n K(H_{k,x}^{-1}d(x, X_i)) K(L_{\ell,y}^{-1}|y - Y_i|)}{\sum_{i=1}^n K(H_{k,x}^{-1}d(x, X_i))},$$

où

$$L_{\ell,y} = \min \left\{ z \in \mathbb{R}^+ \text{ tels que } \sum_{i=1}^n \mathbb{1}_{(y-z, y+z)}(Y_i) = \ell \right\}.$$

De cette sorte, si nous considérons l'intervalle $(\ell_{1,n}, \ell_{2,n})$ contenant ℓ , tel que $\lim_{n \rightarrow \infty} \ell_{1,n} = \lim_{n \rightarrow \infty} \ell_{2,n} = 0$, nous devons poser l'hypothèse suivante :

$$\frac{n \ln n}{\ell_{1,n} \min \left\{ n \phi_x^{-1} \left(\frac{k_{1,n}}{n} \right), k_{1,n} \right\}} \rightarrow 0. \quad (3.12)$$

Théorème 3.6. — *Sous les hypothèses (3.1)-(3.5) et (3.12) et si l'opérateur f^x est de la famille (H2), nous avons :*

$$\sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{\ell_{1,n} \leq \ell \leq \ell_{2,n}} |\widehat{f}_{k,\ell}^x - f^x(y)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^{\beta_2} \right\} + O \left(\frac{\ell_{2,n}}{n} \right)^{\beta_3} + O_{p.co.} \left(\sqrt{\frac{n \log n}{\ell_{1,n} k_{1,n}}} \right).$$

Ce théorème est énoncé dans la section 5.3 et démontré à l'annexe B.2. Il est intéressant de remarquer qu'il n'y a pas de perte de vitesse de convergence comparée à celle de l'estimateur à noyau. En effet, il suffit de choisir une largeur de fenêtre h de telle sorte que le nombre de voisins $k = n \phi_x(h)$, pour retrouver l'estimateur précédent.

3.3.4 Fonction de hasard conditionnelle

La fonction de hasard conditionnelle étant construite à partir de la fonction de répartition conditionnelle et de la densité de probabilité conditionnelle, les hypothèses que nous posons pour obtenir la convergence presque complète uniforme sur le paramètre de lissage de son estimateur sont un mélange des hypothèses posées pour les estimateurs des deux opérateurs.

A - Estimateur à noyau

L'estimateur de la fonction de hasard conditionnelle n'est autre que :

$$\widehat{S}^x(y) = \frac{\widehat{f}^x(y)}{1 - \widehat{F}^x(y)}.$$

Théorème 3.7. — *Sous les hypothèses des théorèmes 3.3 et 3.5, nous avons :*

$$\sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} |\widehat{S}^x(y) - S^x(y)| = O(b_n^{\beta_1}) + O(d_n^{\beta_2}) + O_{p.co.} \left(\sqrt{\frac{\log n}{n c_n \phi_x(a_n)}} \right).$$

Il est claire en voyant ce résultat, dont le théorème est brièvement traité à la section 4.6, que la qualité de l'estimateur dépend fortement du comportement de la fonction de concentration ϕ_x lorsque la suite a_n tend vers zero.

B - Estimateur k NN

Pour estimer la fonction de hasard conditionnelle par la méthode k NN il suffit de considérer l'estimateur :

$$\widehat{S}_{k,\ell}^x(y) = \frac{\widehat{f}_k^x(y)}{1 - \widehat{F}_k^x(y)}.$$

Théorème 3.8. — *Sous les hypothèses des théorèmes 3.4 et 3.6, nous avons :*

$$\sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{\ell_{1,n} \leq \ell \leq \ell_{2,n}} |\widehat{S}_{k,\ell}^x(y) - S^x(y)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^{\beta_2} \right\} + O \left(\frac{\ell_{2,n}}{n} \right)^{\beta_3} + O_{p.co.} \left(\sqrt{\frac{n \log n}{\ell_{1,n} k_{1,n}}} \right).$$

Ce dernier théorème est donné à la section 5.3 et démontré à l'annexe B.2. Nous remarquons que la vitesse de convergence obtenue est identique à celle qu'atteint l'estimateur de densité conditionnelle qui intervient dans la construction de l'estimateur de la fonction de hasard.

3.4 Choix automatique du paramètre de lissage

3.4.1 Validation croisée pour les prédictions à noyaux

Les résultats de convergence uniforme sur le paramètre de lissage obtenus aux théorèmes 3.1, 3.3, 3.5 et 3.7 nous permettent de déduire, sans aucun calcul supplémentaire, la convergence des estimateurs lorsque le paramètre de lissage est aléatoire, en particulier s'il est sélectionné par validation croisée. Ce résultat est donné dans les corollaires suivants.

A - Opérateur de régression

Soit le paramètre de lissage optimal \tilde{h}_{CV} de l'estimateur de régression défini par validation croisée suivant :

$$\tilde{h}_{CV} = \arg \min_{h_K \in (a_n, b_n)} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2,$$

où \hat{Y}_i est le prédicteur de Y défini par :

$$\hat{Y}_i = \hat{m}^{-i}(X_i) = \frac{\sum_{\substack{j=1 \\ j \neq i}}^n K(h_K^{-1}d(X_i, X_j)) Y_j}{\sum_{\substack{j=1 \\ j \neq i}}^n K(h_K^{-1}d(X_i, X_j))}.$$

Corolaire 3.9. — *En considérant que \tilde{m}_{CV} est l'estimateur à noyau \hat{m} construit avec le paramètre de lissage \tilde{h}_{CV} , nous avons sous les hypothèses du théorème 3.1 :*

$$|\tilde{m}_{CV}(x) - m(x)| = O(b_n^\beta) + O_{p.co.} \left(\sqrt{\frac{\log n}{n\phi(a_n)}} \right).$$

B - Opérateurs conditionnels restants

Soient les paramètres de lissage $(\tilde{h}_{K,CV}, \tilde{h}_{H,CV})$ définis par :

$$(\tilde{h}_{K,CV}, \tilde{h}_{H,CV}) = \arg \min_{h_K \in (a_n, b_n), h_H \in (c_n, d_n)} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2,$$

où le prédicateur \hat{Y}_i est défini d'une des façons suivantes :

- *Médiane conditionnelle* : \hat{Y}_i est la solution de l'équation :

$$\hat{F}^{X_i, -i}(y) = \frac{\sum_{\substack{j=1 \\ j \neq i}}^n K(h_K^{-1}d(X_i, X_j)) H(h_H^{-1}(y - Y_j))}{\sum_{\substack{j=1 \\ j \neq i}}^n K(h_K^{-1}d(X_i, X_j))} = \frac{1}{2},$$

- *Mode conditionnel* : En considérant \mathcal{S} un compact de \mathbb{R} , \hat{Y}_i est la valeur de \mathcal{S} qui maximise la fonction :

$$\hat{f}^{X_i, -i}(y) = \frac{h_H^{-1} \sum_{\substack{j=1 \\ j \neq i}}^n K(h_K^{-1}d(X_i, X_j)) K(h_H^{-1}(y - Y_j))}{\sum_{\substack{j=1 \\ j \neq i}}^n K(h_K^{-1}d(X_i, X_j))},$$

Corolaire 3.10. — *En considérant que \tilde{F}^x , \tilde{f}^x et \tilde{S}^x sont respectivement les estimateurs \hat{F}^x , \hat{f}^x et \hat{S}^x construits avec les paramètres de lissage obtenus par validation croisée $\tilde{h}_{K,CV}$ et $\tilde{h}_{H,CV}$. Les trois résultats suivants sont alors vérifiés :*

- *Sous les conditions du théorème 3.3, nous avons :*

$$|\tilde{F}^x(y) - F^x(y)| = O(b_n^{\beta_1}) + O(d_n^{\beta_2}) + O_{p.co.} \left(\sqrt{\frac{\log n}{n \phi_x(a_n)}} \right).$$

- *Sous les conditions du théorème 3.5, nous avons :*

$$|\tilde{f}^x(y) - f^x(y)| = O(b_n^{\beta_1}) + O(d_n^{\beta_2}) + O_{p.co.} \left(\sqrt{\frac{\log n}{n c_n \phi_x(a_n)}} \right).$$

- *Sous les conditions du théorème 3.7, nous avons :*

$$|\tilde{S}^x(y) - S^x(y)| = O(b_n^{\beta_1}) + O(d_n^{\beta_2}) + O_{p.co.} \left(\sqrt{\frac{\log n}{n c_n \phi_x(a_n)}} \right).$$

3.4.2 Validation croisée pour les prédictions k NN

Les résultats de convergence uniforme sur le nombre de voisins obtenus aux théorèmes 3.2, 3.4, 3.6 et 3.8 nous permettent d'établir aisément la convergence des estimateurs lorsque le nombre de voisin dépend de tout l'échantillon. Ce point est abordé aux corolaires suivants.

A - Opérateur de régression

Soit le nombre de voisins optimal \tilde{k}_{CV} de l'estimateur de régression défini par validation croisée suivant :

$$\tilde{k}_{CV} = \arg \min_{k \in (k_{1,n}, k_{2,n})} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2,$$

où \widehat{Y}_i est le prédicateur de Y défini par :

$$\widehat{Y}_i = \widehat{m}_k^{-i}(X_i).$$

Corolaire 3.11. — *En considérant que $\widetilde{m}_{k,CV}$ est l'estimateur kNN \widehat{m} construit avec le paramètre de lissage \widetilde{k}_{CV} , nous avons sous les hypothèses du théorème 3.2 :*

$$|\widetilde{m}_{k,CV}(x) - m(x)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^\beta \right\} + O_{p.co.} \left(\sqrt{\frac{\log n}{k_{1,n}}} \right).$$

B - Opérateurs conditionnels restants

Considérons les nombres de voisins $(\widetilde{k}_{CV}, \widetilde{\ell}_{CV})$ définis par :

$$(\widetilde{k}_{CV}, \widetilde{\ell}_{CV}) = \arg \min_{k_{1,n} \leq k \leq k_{2,n}, \ell_{1,n} \leq \ell \leq \ell_{2,n}} \sum_{i=1}^n (Y_i - \widehat{Y}_i)^2$$

où le prédicateur \widehat{Y}_i est défini soit par la médiane conditionnelle :

$$\widehat{Y}_i = \widehat{F}^{(X_i^{-i})^{-1}}(1/2),$$

soit par le mode conditionnel :

$$\widehat{Y}_i = \arg \max \widehat{f}^{X_i^{-i}}(y).$$

Corolaire 3.12. — En considérant que \tilde{F}_k^x , $\tilde{f}_{k,\ell}^x$ et $\tilde{S}_{k,\ell}^x$ sont respectivement les estimateurs \hat{F}_k^x , $\hat{f}_{k,\ell}^x$ et $\hat{S}_{k,\ell}^x$ construits avec les paramètres de lissage obtenus par validation croisée $(\tilde{k}_{CV}, \tilde{\ell}_{CV})$. Les trois résultats suivants sont alors vérifiés :

- Sous les conditions du théorème 3.4, nous avons :

$$|\tilde{F}_k^x(y) - F^x(y)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^{\beta_1} \right\} + O_{p.co.} \left(\sqrt{\frac{\log n}{k_{1,n}}} \right).$$

- Sous les conditions du théorème 3.6, nous avons :

$$|\tilde{f}_{k,\ell}^x(y) - f^x(y)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^{\beta_2} \right\} + \left(\frac{\ell_{2,n}}{n} \right)^{\beta_3} + O_{p.co.} \left(\sqrt{\frac{n \log n}{\ell_{1,n} k_{1,n}}} \right).$$

- Sous les conditions du théorème 3.8, nous avons :

$$|\tilde{S}_{k,\ell}^x(y) - S^x(y)| = O(b_n^{\beta_1}) + O(d_n^{\beta_2}) + O_{p.co.} \left(\sqrt{\frac{\log n}{n c_n \phi_x(a_n)}} \right).$$

Ce résultat est traité à la section 5.4.4 .

3.4.3 Validation croisée pour les estimations k NN

Nos derniers résultats traitent de procédures de validation croisée que nous avons introduits pour minimiser l'erreur quadratique des estimateurs.

A - Fonction de répartition conditionnelle

Le nombre de voisins optimal est défini par :

$$\tilde{k}_{CV} = \arg \min_{k \in (k_{1,n}, k_{2,n})} \sum_{i=1}^n \{ \mathbb{1}_{(Y_i \leq Y_j)} - \hat{F}^{X_i^{-i}}(Y_j) \}^2,$$

où

$$\hat{F}^{X_i^{-i}}(y) = \frac{\sum_{j \neq i} K \{ H_{k,x}^{-1} d(X_i, X_j) \} \mathbb{1}_{\{Y_j \leq y\}}}{\sum_{j \neq i} K \{ H_{k,x}^{-1} d(X_i, X_j) \}}.$$

Corolaire 3.13. — *En considérant que $\tilde{F}_{k,CV}^x$ est l'estimateur kNN \hat{F}_k^x construit avec le nombre de voisins optimal \tilde{k}_{CV} , nous avons sous les hypothèses du théorème 3.4 :*

$$|\tilde{F}_{k,CV}^x(y) - F^x(y)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^{\beta_1} \right\} + O_{p.co.} \left(\sqrt{\frac{\log n}{k_{1,n}}} \right).$$

Ce résultat est traité à la section 5.4.2, et démontré à l'annexe B.2.

B - Densité de probabilité conditionnelle

Les nombres de voisins optimaux $(\tilde{k}_{CV}, \tilde{\ell}_{CV})$ de l'estimateur de la densité de probabilité conditionnelle $\hat{f}_{k,\ell}^x$ peuvent être définis par :

$$(\tilde{k}_{CV}, \tilde{\ell}_{CV}) = \arg \min_{k_{1,n} \leq k \leq k_{2,n}, \ell_{1,n} \leq \ell \leq \ell_{2,n}} \mathbf{CV}_{CD}(a, b).$$

Ici $\mathbf{CV}_{CD}(a, b)$ est la fonction de perte :

$$\mathbf{CV}_{CD}(a, b) = \frac{1}{n} \sum_{i=1}^n W_1(X_i) \int \left(\hat{f}_{(a,b)}^{X_i^{-i}} \right)^2(y) W_2(y) dy - \frac{2}{n} \sum_{i=1}^n \hat{f}_{(a,b)}^{X_i^{-i}}(Y_i) W_1(X_i) W_2(Y_i),$$

où W_1 et W_2 sont des fonctions de poids non négatives, et :

$$\hat{f}_{(a,b)}^{X_i^{-i}} = b^{-1} \frac{\sum_{j \neq i}^n K\{a^{-1}d(X_i, X_j)\} K(b^{-1}|y - Y_j|)}{\sum_{j \neq i}^n K\{a^{-1}d(X_i, X_j)\}}.$$

Corolaire 3.14. — *Si la fonction $\tilde{f}_{k,\ell}^x$ est l'estimateur \hat{f}_k^x construit avec les nombres de voisins optimaux $(\tilde{k}_{CV}, \tilde{\ell}_{CV})$, nous avons sous les hypothèses du théorème 3.6 :*

$$|\tilde{f}_{k,\ell}^x(y) - f^x(y)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^{\beta_2} \right\} + O \left(\frac{\ell_{2,n}}{n} \right)^{\beta_3} + O_{p.co.} \left(\sqrt{\frac{n \log n}{\ell_{1,n} k_{1,n}}} \right).$$

Ce corolaire est traité à la section 5.4.3, et démontré à l'annexe B.2.

Partie I

Résultats asymptotiques

Convergence presque complète uniforme sur la fenêtre

Chapitre 4

Les résultats de ce chapitre font l'objet d'une publication dans le journal "*Journal of Nonparametric Statistics*", doi : 10.1080/10485252.2016.1254780 (2017), sous le titre : "Uniform in bandwidth consistency for various kernel estimators involving functional data".

L'article a été co-produit par :

- Lydia Kara-Zaitri, *Université Djillali Liabès, Laboratoire de Mathématiques, Sidi Bel-Abbès, Algérie.*
- Ali Laksaci, *Université Djillali Liabès, Laboratoire de Mathématiques, Sidi Bel-Abbès, Algérie.*
- Mustapha Rachdi, *Université Grenoble Alpes, Laboratoire AGEIS EA 7407, Grenoble, France.*
- Philippe Vieu, *Université Paul Sabatier, Institut de Mathématiques, Toulouse, France.*

Convergence uniforme sur le paramètre de lissage de divers estimateurs à noyaux impliquant des données fonctionnelles

Ce chapitre traite de divers modèles non paramétriques comprenant la régression, la fonction de répartition conditionnelle, la densité conditionnelle et la fonction de hasard conditionnelle, quand les co-variables sont infinie-dimensionnelles. Sa principale contribution est de démontrer des résultats asymptotiques d'uniformité sur la largeur de la fenêtre pour les estimateurs à noyau de ces opérateurs fonctionnels.

4.1 Introduction

De nombreuses études ont été consacrées à l'analyse statistique des ensembles de données fonctionnelles, comme cela a été souligné récemment dans les monographies de Ramsay et Silverman [2005], Ferraty et Vieu [2006], Horváth et Kokoszka [2012], Zhang [2013] et Hsing et Eubank [2015], et dans plusieurs numéros spéciaux de diverses revues statistiques (voir Davidian et collab. [2004], Manteiga et Vieu [2007], Valderrama [2007], Goia et Vieu [2016] ou encore la présentation bibliographique de Cuevas [2014]).

L'analyse de la relation qui existe entre une variable aléatoire scalaire Y et une variable aléatoire fonctionnelle X est une question courante dans l'étude de l'analyse des données fonctionnelles, et les techniques non paramétriques sont maintenant largement utilisées à cette fin. La majeure partie de cette littérature utilise des techniques de lissage à noyaux pour estimer les opérateurs fonctionnels associés à la fonction de répartition conditionnelle de Y sachant X . Il s'avère que le comportement asymptotique de tout estimateur fonctionnel non paramétrique est fortement lié à un paramètre de lissage (i.e. la largeur de la fenêtre). Récemment, une vaste littérature a fourni des résultats de convergence uniforme sur la largeur de la fenêtre (ULF) dans des problèmes tels que l'estimation de la régression, de la densité conditionnelle et celle de la distribution conditionnelle dans le cadre multivarié habituel (i.e. lorsque X est de dimension finie). Pour plus de détails sur ce sujet, les articles suivants, pour

n'en citer que quelques uns, sont recommandés : Einmahl et collab. [2005], Varron [2012], Dony et collab. [2008], Chacón et Rodríguez-Casal [2010], Deheuvels et Ouadah [2013], Ouadah [2013a,b], Varron et Van Keilegom [2011], Mason et Swanepoel [2011], Bouzebda et collab. [2011] et Dony et Mason [2010].

À notre connaissance, il n'y a pas d'avancées en ce sens dans le cadre de dimension infinie, et le but principal de ce chapitre est de présenter quelques résultats fonctionnels de convergence ULF. L'une des principales conséquences des résultats de convergence ULF est de permettre directement de prendre en compte des paramètres de lissage aléatoires, et plus spécifiquement des paramètres de lissage orientés par les données. Par conséquent, les résultats ULF ont un impact important dans une grande variété d'applications. Bien que ce chapitre traite de résultats théoriques à propos des propriétés asymptotiques ULF d'estimateurs non paramétriques fonctionnels, nous présenterons également quelques conséquences directes sur le choix de la largeur de la fenêtre orientées par les données.

Le chapitre est organisé de la manière suivante : nous présentons notre cadre de travail général dans la section 4.2. Les propriétés asymptotiques de l'estimateur à noyau de la régression sont données dans la section 4.3. Dans la section 4.4, nous étudions l'estimation de la fonction de répartition conditionnelle. Puis, l'estimation de la densité conditionnelle est développée dans la section 4.5. Une application directe à la fonction de hasard conditionnel est discutée dans la section 4.6. Dans chaque cas, nous donnons la convergence presque complète (Définition1, p.113) uniformément sur la paramètre de lissage et spécifions les vitesses de convergence. Les propriétés asymptotiques des estimateurs sont obtenues dans des conditions générales liées à la fonction de concentration du régresseur fonctionnel, à la régularité de chaque opérateur cible et à l'entropie de l'espace fonctionnel. Dans la section 4.7, nous soulignons l'impact direct qu'ont ces résultats ULF sur la pratique concernant la sélection de largeurs de fenêtres orientées par les données. Les preuves techniques sont reportées à l'annexe B.1.

4.2 Cadre de travail

4.2.1 Généralités sur les modèles étudiés

Considérons un échantillon $(X_1, Y_1), \dots, (X_n, Y_n)$ de couples aléatoires indépendants et de même distribution que le couple (X, Y) qui prend ses valeurs dans un espace fonctionnel $\mathcal{F} \times \mathbb{R}$ muni d'une semi-métrique d^{**} . Nous visons à étudier la relation entre X et Y en estimant des opérateurs fonctionnels associés à la distribution conditionnelle de Y sachant X , tels que l'opérateur de régression :

$$m(x) = \mathbb{E}[Y|X = x], \quad (4.1)$$

la fonction de répartition conditionnelle :

$$F^x(\cdot) = \mathbb{P}(Y \leq \cdot | X = x), \quad (4.2)$$

la fonction de densité de probabilité conditionnelle :

$$f^x(\cdot) = (F^x)'(\cdot), \quad (4.3)$$

et la fonction de hasard conditionnelle :

$$S^x(\cdot) = \frac{f^x(\cdot)}{1 - F^x(\cdot)} \quad (4.4)$$

où g' indique la première dérivée de g et où $x \in \mathcal{F}$.

Nous souhaitons proposer des modèles très flexibles. D'une part, nous construisons des modèles généraux, où aucune famille spécifique de distributions (gaussienne par exemple) n'est supposée pour la loi conditionnelle de Y sachant X . D'autre part, nos modèles ne nécessitent pas de restriction paramétrique (linéarité par exemple) pour l'opérateur cible. À ce stade, il est intéressant de noter que l'énoncé des résultats de l'ULF, dans ce contexte de dimension infinie, n'est pas une extension simple de la littérature multivariée existante pour au moins deux raisons discutées ci-dessous.

4.2.2 Variables fonctionnelles et petites boules de probabilité

Tout d'abord, étant donné qu'aucune mesure de référence universelle n'est disponible (comme la mesure de Lebesgue pourrait être dans des espaces de dimension finie), il

n'y a aucune raison de supposer l'existence d'une fonction de densité pour la variable X . Une façon de surmonter cette difficulté est d'utiliser des considérations de petite boule de probabilité qui auront un impact direct sur la variance de nos estimateurs. Plus précisément, la concentration des données dans l'espace fonctionnel est contrôlée par une fonction de petite boule de probabilité définie, pour tout $r > 0$, par :

$$\mathbb{P}(X \in B(x, r)) =: \phi_x(r) > 0 \quad (4.5)$$

où $B(x, \varepsilon) = \{z \in \mathcal{F} \text{ tel que } d(z, x) \leq \varepsilon\}$.

À noter que le livre pionnier de [Ferraty et Vieu \[2006\]](#) commente largement les liens entre la théorie des petites boules de probabilité dans un cadre non paramétrique fonctionnel et la structure topologique sur l'espace fonctionnel \mathcal{F} . Une large littérature probabiliste donne une expression précise de la fonction de petite boule $\phi_x(\cdot)$ pour les processus stochastiques habituels (voir par exemple [Bogachev \[1998\]](#) ou [Li et Shao \[2001\]](#)).

Pour contrôler le terme biais des estimateurs non paramétriques, nous avons besoin d'avoir quelques informations sur la variabilité de la petite boule de probabilité, ce qui est obtenu par l'hypothèse suivante :

$$\text{Pour tout } s \in (0, 1), \lim_{r \rightarrow 0} \frac{\phi_x(sr)}{\phi_x(r)} = \tau_x(s) < \infty. \quad (4.6)$$

Cependant, cette condition n'est pas très restrictive. Elle a été utilisée dans des statistiques fonctionnelles par de nombreux auteurs (par exemple [Mas \[2012\]](#) et les références qui y sont citées). La fonction $\tau_x(\cdot)$ peut être explicitée pour de nombreux processus stochastiques classiques (voir [Ferraty et collab. \[2007\]](#), page 270).

4.2.3 Régularité des opérateurs fonctionnels

La seconde difficulté technique avec les problèmes infini-dimensionnels est le manque de structure de Banach dans les espaces fonctionnels. Une telle structure évite de recourir à des hypothèses de différentiabilité pour modéliser la régularité des opérateurs étudiés (comme c'est généralement le cas dans les statistiques non paramétriques de

dimension finie). Pour surmonter ce problème, nous utilisons des modèles de régularité de type Lipschitz qui ont un impact direct sur le biais des estimateurs. De telles conditions doivent évidemment être adaptées à chaque opérateur à estimer et c'est pourquoi elles ne sont pas présentées dans cette section (voir (4.8), (4.15) et (4.20) ci-après).

4.3 Opérateur de régression

Le but principal de cette section est d'étudier la convergence presque complète ULF de l'estimateur à noyau de l'opérateur de régression. L'estimateur à noyau fonctionnel $\widehat{m}(x)$ de l'opérateur de régression $m(x)$ est défini, pour tout $x \in \mathcal{F}$, par :

$$\widehat{m}(x) = \frac{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))Y_i}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))} \quad (4.7)$$

où $h_K = h_{K,n}$ est une suite de réels positifs qui appartient à un intervalle (a_n, b_n) avec $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = 0$.

Les propriétés asymptotiques de l'estimateur \widehat{m} ont été largement étudiées dans la littérature (voir [Ferraty et collab. \[2010\]](#), pour les avancées les plus récentes) mais seulement pour des paramètres de lissage fixes et déterministes.

4.3.1 Hypothèses sur le modèle et sur l'estimateur

Dans tout ce qui suit, considérons x un élément fonctionnel fixe dans \mathcal{F} . Afin d'établir la convergence presque complète ULF sur le point x , les hypothèses suivantes sont nécessaires :

- Il existe $\beta > 0$ et $C_1 > 0$, tels que :

$$\text{pour tout } x_1, x_2 \in N_x, |m(x_1) - m(x_2)| \leq C_1 d^\beta(x_1, x_2), \quad (4.8)$$

où N_x représente un voisinage fixe de x .

- Il existe $m \geq 2$ et $C_2 > 0$, tels que :

$$\mathbb{E}[|Y|^m | X] < C_2 < \infty, \text{ presque-sûrement.} \quad (4.9)$$

- Le noyau K est à support dans $(0, 1/2)$ et a une première dérivée continue sur $(0, 1/2)$. De plus, il existe deux constantes $0 < C_3 \leq C_4 < \infty$, telles que :

$$0 < C_3 \mathbb{1}_{(0,1/2)}(\cdot) \leq K(\cdot) \leq C_4 \mathbb{1}_{(0,1/2)}(\cdot) \quad (4.10)$$

et

$$K(1/2) - \int_0^{1/2} K'(s) \tau_x(s) ds > 0.$$

- La classe de fonctions :

$$\mathcal{K} = \{ \cdot \mapsto K(\gamma^{-1}d(x, \cdot)), \gamma > 0 \} \text{ est une classe mesurable ponctuellement} \quad (4.11)$$

telle que :

$$\sup_Q \int_0^1 \sqrt{1 + \log \mathcal{N}(\epsilon \|F\|_{Q,2}, \mathcal{K}, d_Q)} d\epsilon < \infty, \quad (4.12)$$

où le supremum est pris sur toutes les mesures de probabilité Q sur l'espace de probabilité $(\mathcal{X}, \mathcal{A})$ avec $Q(F^2) < \infty$ où F est la fonction enveloppe de l'ensemble \mathcal{K} . Ici, d_Q est la métrique $L_2(Q)$ et $\mathcal{N}(\epsilon, \mathcal{K}, d_Q)$ est le nombre minimal de boules ouvertes (selon la métrique $L_2(Q)$) de rayon ϵ qui sont nécessaires pour couvrir la classe de fonctions \mathcal{K} . Nous noterons par $\|\cdot\|_{Q,2}$ la norme $L_2(Q)$.

- La suite (a_n) vérifie :

$$\frac{\log n}{n \min(a_n, \phi_x(a_n))} \rightarrow 0. \quad (4.13)$$

4.3.2 Commentaires sur les hypothèses

Il est intéressant de remarquer que les conditions (4.8)-(4.13) traitent de diverses caractéristiques de l'espace infini-dimensionnel, comprenant la structure topologique sur \mathcal{F} et son entropie, la distribution de probabilité de X et le concept de mesurabilité pour la classe \mathcal{K} .

Les conditions sur le noyau

Les conditions sur le noyau ne sont pas très restrictives. La première partie de la condition (4.10) apparaît dans plusieurs études de noyaux fonctionnels et elle est facilement satisfaite pour une large classe de noyaux (voir H_4 dans [Ferraty et collab.](#)

[2007]). La seconde partie de cette condition (4.10), qui est ajoutée dans ce travail en tant qu'outil nécessaire à l'obtention de résultats ULF, est liée à la fonction τ_x et elle est aussi assez générale. Par exemple, quand τ_x est définie comme la masse de Dirac sur $1/2$, la seconde partie de (4.10) est vraie tant que $K'(s) \leq 0$ et $K(1/2) > 0$. D'autres exemples peuvent être dérivés de la Proposition 2 dans Ferraty et al. [Ferraty et collab. \[2007\]](#). Enfin, l'hypothèse (4.11) est une faible condition de mesurabilité qui permet d'établir des résultats d'uniformité. Notons qu'une façon alternative de traiter ce problème serait d'utiliser des notions de probabilité extérieure (voir [Van Der Vaart et Wellner \[1996\]](#), page 4), mais cela demanderait probablement des calculs plus compliqués.

La condition d'entropie

L'uniformité asymptotique est étroitement liée à la notion d'entropie et de compacité. La condition (4.12) est une condition sur l'uniformité de l'intégral de l'entropie utilisée pour caractériser la classe de fonctions Donsker (voir [Van Der Vaart et Wellner \[1996\]](#)) dont découlent des distributions limites uniformes. Notons que ce type d'hypothèses peut aussi être utile à l'estimation des moments de processus empiriques (voir [Van Der Vaart et Wellner \[2011\]](#)). Afin de souligner la vaste généralité de (4.12), notons qu'elle est moins restrictive que la condition usuelle de VC-class (voir [Einmahl et collab. \[2005\]](#)).

Les conditions sur les paramètres de lissage

La condition (4.13) exclut tous les paramètres de lissage très grands ou très petits pour lesquels la convergence ne peut pas être obtenue. Elle est vérifiée pour $a_n = O(\log n)^{-\nu_1}$ (pour un $\nu_1 > 0$ adéquat) tant que le processus X est de type exponentiel (c'est-à-dire lorsque la fonction de petite boule de probabilité décroît exponentiellement). Elle est également vérifiée pour $a_n = O(n/\log n)^{-\nu_2}$ (pour un $\nu_2 > 0$ adéquat) pour un processus fractal (c'est-à-dire lorsque la fonction de petite boule de probabilité décroît de façon polynomiale). Pour plus de détails se référer à [Ferraty et Vieu \[2006\]](#).

Autres conditions

Les conditions (4.8) et (4.9) sont classiques dans la régression non paramétrique. Notons simplement qu'une extension de nos résultats serait de considérer des modèles localement réguliers en posant un paramètre β dans (4.9) dépendant de x . Cette extension n'est pas traitée ici pour garder des notations plus simples.

4.3.3 Convergence ULF

Théorème 4.1. — *Sous les hypothèses (4.5)-(4.13), nous avons :*

$$\sup_{a_n \leq h_K \leq b_n} |\widehat{m}(x) - m(x)| = O(b_n^\beta) + O_{p.co.} \left(\sqrt{\frac{\log n}{n\phi_x(a_n)}} \right).$$

Preuve du théorème 4.1. : Commençons par écrire :

$$\widehat{m}(x) - m(x) = \widehat{B}(x) + \frac{\widehat{R}(x)}{\widehat{F}_D^x} + \frac{\widehat{Q}(x)}{\widehat{F}_D^x}$$

où

$$\widehat{Q}(x) := (\widehat{g}(x) - \mathbb{E}[\widehat{g}(x)]) - m(x)(\widehat{F}_D^x - \mathbb{E}[\widehat{F}_D^x]),$$

$$\widehat{B}(x) := \frac{\mathbb{E}[\widehat{g}(x)]}{\mathbb{E}[\widehat{F}_D^x]} - m(x) \quad \text{et} \quad \widehat{R}(x) := -\widehat{B}(x)(\widehat{F}_D^x - \mathbb{E}[\widehat{F}_D^x])$$

avec

$$\widehat{g}(x) = \frac{1}{n\phi_x(h_K)} \sum_{i=1}^n K(h_K^{-1}d(x, X_i)) Y_i \quad \text{et} \quad \widehat{F}_D^x = \frac{1}{n\phi_x(h_K)} \sum_{i=1}^n K(h_K^{-1}d(x, X_i)).$$

Nous concluons en combinant l'égalité ci-dessus avec les quatre résultats suivants pour lesquels les preuves sont données dans l'annexe B.1.1.

Lemme 4.2. — *Sous les hypothèses (4.5), (4.6) et (4.10)-(4.13), nous avons :*

$$\sup_{a_n \leq h_K \leq b_n} \left| \widehat{F}_D^x - \mathbb{E}[\widehat{F}_D^x] \right| = O_{p.co.} \left(\sqrt{\frac{\log n}{n\phi_x(a_n)}} \right).$$

Corolaire 4.3. — *Sous les hypothèses du Lemme 4.2, il existe un $C > 0$ tel que :*

$$\sum_{n=1}^{\infty} \mathbb{P} \left(\inf_{a_n \leq h_K \leq b_n} \widehat{F}_D^x < C \right) < \infty.$$

Lemme 4.4. — *Sous les hypothèses (4.5), (4.6), (4.8) et (4.10), nous avons :*

$$\sup_{a_n \leq h_K \leq b_n} \left| \widehat{B}(x) \right| = O(b_n^\beta).$$

Lemme 4.5. — *Sous les hypothèses du Théorème 4.1, nous avons :*

$$\sup_{a_n \leq h_K \leq b_n} |\widehat{g}(x) - \mathbb{E}[\widehat{g}(x)]| = O_{p.co.} \left(\sqrt{\frac{\log n}{n\phi_x(a_n)}} \right).$$

□

4.4 Fonction de répartition conditionnelle

Cette section traite de l'estimation de l'opérateur de la distribution conditionnelle $F^x(\cdot)$ définie dans (4.2). L'étude de cet opérateur est intéressante pour au moins deux raisons. Premièrement, il fournit une information complète sur la distribution conditionnelle de Y sachant X , alors que l'opérateur de régression n'est basé que sur son espérance. Ensuite, il peut être utilisé pour estimer le quantile conditionnel avec une application directe dans la prédiction de Y (voir Ferraty et Vieu [Ferraty et Vieu \[2006\]](#)). Considérons l'estimateur à double noyaux de $F^x(y)$ défini pour tout $y \in \mathbb{R}$, par :

$$\widehat{F}^x(y) = \frac{\sum_{i=1}^n K(h_K^{-1}d(x, X_i)) H(h_H^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))} \quad (4.14)$$

où H est une fonction de répartition conditionnelle définie par :

$$H(\cdot) = \frac{\int_0^\cdot K(s)ds}{\int_0^{1/2} K(s)ds}$$

et $h_K = h_{K,n}$ (*resp.* $h_H = h_{H,n}$) est une suite de réels positifs qui appartient à un intervalle (a_n, b_n) (*resp.* (c_n, d_n)), avec $\lim_{n \rightarrow \infty} c_n = \lim_{n \rightarrow \infty} d_n = 0$.

Bien que ce type d'estimateurs à noyaux ait été largement étudié dans la littérature ces dernières années, ces études concernent exclusivement des paramètres de lissage fixes. Dans cette section, nous établissons la convergence ULF avec les vitesses correspondantes (voir Théorème 4.6 ci-dessous) et soulignons les applications pour la construction automatique de paramètres de lissage orientés par les données dans la section 4.7.

L'obtention de l'uniformité asymptotique sur la largeur de la fenêtre de l'estimateur (4.14) nécessite les mêmes hypothèses générales que dans le cadre de la régression telles que discutées dans la section 4.3, seulement quelques-unes d'entre elles doivent être légèrement adaptées.

Rappelons que x est un élément fonctionnel fixe et N_x est un voisinage fixe de x . Le modèle non paramétrique de la distribution conditionnelle est construit en supposant l'existence de $\beta_1, \beta_2 > 0$ et $C_6 > 0$ tels que pour tout $(y_1, y_2) \in \mathbb{R}^2$ et tout $(x_1, x_2) \in N_x \times N_x$ nous avons :

$$|F^{x_1}(y_1) - F^{x_2}(y_2)| \leq C_6 (d(x_1, x_2)^{\beta_1} + |y_1 - y_2|^{\beta_2}). \quad (4.15)$$

Puisque l'estimateur de la fonction de répartition conditionnelle nécessite deux noyaux, la condition (4.11) doit, elle aussi, être modifiée. Nous considérons la classe de fonctions :

$$\mathcal{K}_0 = \left\{ \cdot \mapsto \frac{1}{\int_0^1 K(s) ds} \int_0^{(\delta^{-1}(y-\cdot))} K(s) ds, \text{ pour } \delta > 0 \right\}$$

et supposons que

$$\mathcal{K}_0 \text{ est une classe mesurable ponctuellement,} \quad (4.16)$$

telle que

$$\sup_Q \int_0^1 \sqrt{1 + \log \mathcal{N}(\epsilon \|F_0\|_{Q,2}, \mathcal{K}_0, d_Q)} d\epsilon < \infty \quad (4.17)$$

où F_0 est la fonction enveloppe de l'ensemble \mathcal{K}_0 .

Théorème 4.6. — *Sous les hypothèses (4.5), (4.6), (4.10)- (4.13) et (4.15)-(4.17) nous avons :*

$$\sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} |\widehat{F}^x(y) - F^x(y)| = O(b_n^{\beta_1}) + O(d_n^{\beta_2}) + O_{p.co} \left(\sqrt{\frac{\log n}{n\phi_x(a_n)}} \right).$$

Preuve du théorème 4.6. : Définissons :

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

et considérons la décomposition suivante :

$$\widehat{F}^x(y) - F^x(y) = \widetilde{B}(x, y) + \frac{\widetilde{R}(x, y)}{\widehat{F}_D^x} + \frac{\widetilde{Q}(x, y)}{\widehat{F}_D^x} \quad (4.18)$$

où

$$\widetilde{Q}(x, y) := (\widehat{F}_N^x(y) - \mathbb{E}[\widehat{F}_N^x(y)]) - F^x(y)(\widehat{F}_D^x - \mathbb{E}[\widehat{F}_D^x])$$

$$\widetilde{B}(x, y) := \frac{\mathbb{E}[\widehat{F}_N^x(y)]}{\mathbb{E}[\widehat{F}_D^x]} - F^x(y) \quad \text{et} \quad \widetilde{R}(x, y) := -\widetilde{B}(x, y)(\widehat{F}_D^x - \mathbb{E}[\widehat{F}_D^x]).$$

La preuve découle du lemme 4.2 et du corollaire 4.3 énoncés dans la section 4.3 combinés avec les deux lemmes suivants dont la démonstration est reportée à l'annexe B.1.2

Lemme 4.7. — *Sous les hypothèses (4.5), (4.6), (4.10) et (4.15) nous avons :*

$$\sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} \widetilde{B}(x, y) = O(b_n^{\beta_1}) + (d_n^{\beta_2}).$$

Lemme 4.8. — *Sous les hypothèses du Théorème 4.6, nous avons :*

$$\sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} |\widehat{F}_N^x(y) - \mathbb{E}[\widehat{F}_N^x(y)]| = O_{p.co} \left(\sqrt{\frac{\log n}{n\phi_x(a_n)}} \right).$$

□

4.5 Densité de probabilité conditionnelle

La fonction de densité de probabilité conditionnelle (4.3) est largement utilisée dans la statistique, que ce soit pour faire des prédictions (via le mode conditionnel) ou encore pour la construction d'intervalles de confiance ou même pour l'estimation des valeurs de probabilité extrêmes. Un estimateur non paramétrique usuel de l'opérateur (4.3) est donné pour tout $y \in \mathbb{R}$ par :

$$\hat{f}^x(y) = \frac{h_H^{-1} \sum_{i=1}^n K(h_K^{-1}d(x, X_i)) K(h_H^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))}. \quad (4.19)$$

Un tel estimateur a été largement étudié dans la récente littérature statistique fonctionnelle (voir Ferraty et collab. [2010] et les références qui y sont citées pour les récentes avancées dans ce domaine), mais seulement pour des paramètres de lissage fixes. Dans cette section, nous établissons des résultats de convergence ULF (voir le Théorème 4.9) qui seront utilisés pour la construction automatique de paramètres de lissage orientés par les données (voir Section 4.7).

Le modèle non paramétrique est construit sous les conditions de régularité suivantes. Supposons l'existence de $\beta_1, \beta_2 > 0$ et $C_7 > 0$ tels que pour tout $(y_1, y_2) \in \mathbb{R}^2$ et tout $(x_1, x_2) \in N_x \times N_x$:

$$|f^{x_1}(y_1) - f^{x_2}(y_2)| \leq C_7 (d(x_1, x_2)^{\beta_1} + |y_1 - y_2|^{\beta_2}), \text{ pour } C_7 > 0. \quad (4.20)$$

Les hypothèses techniques sont les mêmes que pour la distribution conditionnelle (voir Section 4.4), la seule différence est la restriction sur le paramètre de lissage suivante :

$$\frac{\log n}{nc_n \min(a_n, \phi_x(a_n))} \rightarrow 0. \quad (4.21)$$

Théorème 4.9. — *Sous les hypothèses (4.5), (4.6), (4.10)-(4.13), (4.20) et (4.21), nous avons :*

$$\sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} |\hat{f}^x(y) - f^x(y)| = O(b_n^{\beta_1}) + O(d_n^{\beta_2}) + O_{p.co} \left(\sqrt{\frac{\log n}{n c_n \phi_x(a_n)}} \right).$$

Preuve du théorème 4.9 . : Ce résultat est basé sur la décomposition :

$$\widehat{f}^x(y) - f^x(y) = \ddot{B}(x, y) + \frac{\ddot{R}(x, y)}{\widehat{F}_D^x} + \frac{\ddot{Q}(x, y)}{\widehat{F}_D^x}$$

où

$$\ddot{Q}(x, y) := (\widehat{f}_N^x(y) - \mathbb{E}[\widehat{f}_N^x(y)]) - f^x(y)(\widehat{F}_D^x - \mathbb{E}[\widehat{F}_D^x])$$

$$\ddot{B}(x, y) := \frac{\mathbb{E}[\widehat{f}_N^x(y)]}{\mathbb{E}[\widehat{F}_D^x]} - f^x(y) \quad \text{et} \quad \ddot{R}(x, y) := -\ddot{B}(x, y)(\widehat{F}_D^x - \mathbb{E}[\widehat{F}_D^x])$$

avec

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

Ensuite, la preuve du Théorème 4.9 requière les deux lemmes suivants.

Lemme 4.10. — *Sous les hypothèses (4.5), (4.6), (4.10), (4.13) et (4.20) nous avons :*

$$\sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} \ddot{B}(x, y) = O(b_n^{\beta_1}) + O(d_n^{\beta_2}), \quad (4.22)$$

Lemme 4.11. — *Sous les hypothèses du Théorème 4.9, nous avons :*

$$\sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} |\widehat{f}_N^x(y) - \mathbb{E}[\widehat{f}_N^x(y)]| = O_{p.co.} \left(\sqrt{\frac{\log n}{n c_n \phi_x(a_n)}} \right). \quad (4.23)$$

Les preuves de ces deux lemmes sont données dans l'annexe B.1.2. □

4.6 Opérateur de hasard conditionnel

La construction de l'opérateur de hasard conditionnel (voir (4.4)) est liée aux opérateurs de distribution et de densité conditionnelles. Dans cette section, nous montrons comment des résultats ULF peuvent être déduits de ceux des Sections 4.4 et 4.5. L'estimateur à noyaux $\widehat{S}^x(y)$ est défini pour tout $y \in \mathbb{R}$ tel que $F^x(y) < 1$, par :

$$\widehat{S}^x(y) = \frac{\widehat{f}^x(y)}{1 - \widehat{F}^x(y)}.$$

Il existe quelques résultats dans la littérature à propos de cet estimateur (voir [Ferraty et collab. \[2008\]](#) et [Quintela-Del-Río \[2008\]](#)) et le but de cette section est de les généraliser au cadre ULF. L'application des résultats ULF sur le choix automatique du paramètre de lissage orienté par les données est reportée à la Section 4.7.

Théorème 4.12. — *Sous les hypothèses des Théorèmes 4.6 et 4.9, nous avons :*

$$\sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} |\widehat{S}^x(y) - S^x(y)| = O(b_n^{\beta_1}) + O(d_n^{\beta_2}) + O_{p.co} \left(\sqrt{\frac{\log n}{n c_n \phi_x(a_n)}} \right).$$

Preuve du théorème 4.12. : La preuve de ce théorème est basée sur le fait que :

$$|\widehat{S}^x(y) - S^x(y)| \leq \frac{1}{|1 - \widehat{F}^x(y)|} \left[|\widehat{f}^x(y) - f^x(y)| + \frac{|f^x(y)|}{|1 - F^x(y)|} |\widehat{F}^x(y) - F^x(y)| \right].$$

Le Théorème 4.12 est obtenu directement à partir des Théorèmes 4.6 et 4.9 et du Corolaire 4.13 ci dessous qui est une conséquence du Théorème 4.6.

Corolaire 4.13. — *Sous les conditions du Théorème 4.12, il existe un $\delta > 0$ tel que :*

$$\sum_{n=1}^{\infty} \mathbb{P} \left\{ \sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} |1 - \widehat{F}^x(y)| < \delta \right\} < \infty.$$

□

4.7 Application à la sélection du paramètre de lissage

Dans les études pratiques, lors de l'utilisation d'estimateurs non paramétriques, on se retrouve confronté au problème du choix de la largeur de la fenêtre. Un tel choix devrait être fait en fonction des données dont on dispose, ce qui implique que les paramètres de lissage utilisés dans la pratique sont des variables aléatoires qui dépendent de tout l'échantillon statistique. La théorie asymptotique du choix du paramètre de lissage orienté par les données dans les statistiques multivariées a toujours été liée à une sorte d'uniformité sur la paramètre de lissage (voir par exemple le Lemme 1 dans le papier précurseur de [Härdle et Marron \[1985\]](#)) même si le terme ULF est apparu bien plus tard dans la littérature mathématique. Dans cette section nous allons discuter de la façon avec laquelle les résultats ULF présentés précédemment dans ce chapitre permettent un large éventail d'applications faciles au choix du paramètre

de lissage orienté par les données dans le cadre fonctionnel. Cela est décrit dans la Sous-section 4.7.1 pour le modèle de régression et ensuite plus brièvement dans la Sous-section 4.7.2 pour les autres problèmes d'estimation (à savoir les opérateurs conditionnels de : la fonction de répartition, la densité et la fonction de hasard).

4.7.1 Choix du paramètre de lissage dans la régression

Considérons le problème de régression défini dans la Section 4.3 et supposons que \tilde{h}_K est une variable aléatoire prenant ses valeurs dans l'intervalle (a_n, b_n) , où a_n et b_n sont tels que (4.13) soit satisfaite. Désignons maintenant par \tilde{m} le même estimateur à noyau que \hat{m} mais en utilisant le paramètre de lissage aléatoire \tilde{h}_K . Nous obtenons ainsi, comme conséquence directe du Théorème 4.1 :

$$|\tilde{m}(x) - m(x)| = O(b_n^\beta) + O_{p.co.} \left(\sqrt{\frac{\log n}{n\phi(a_n)}} \right). \quad (4.24)$$

Dans la pratique, on utilise des paramètres de lissage orientés par les données dépendants de l'échantillon statistique :

$$\tilde{h}_K = \tilde{h}_K(X_1, Y_1, \dots, X_n, Y_n).$$

Ainsi, les résultats de convergence peuvent être obtenus pour tout type de sélecteur automatique de paramètre de lissage orienté par les données, comme conséquence à (4.24). La validation croisée est l'une des approches les plus populaires qui consiste à minimiser le critère de prédiction des moindres carrés suivant :

$$\tilde{h}_{CV} = \arg \min_{h_K \in (a_n, b_n)} CV(h_K) \quad \text{où} \quad CV(h_K) = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2,$$

où \hat{Y}_i est ce que l'on appelle la prédiction "leave-one-out" de Y_i , ce qui peut être traduit par $\hat{Y}_i = \hat{m}^{-i}(X_i)$ avec :

$$\hat{m}^{-i}(x) = \frac{\sum_{j \neq i} K(h_K^{-1}d(x, X_j))Y_j}{\sum_{j \neq i} K(h_K^{-1}d(x, X_j))}. \quad (4.25)$$

Le résultat suivant est une conséquence directe de (4.24).

Corolaire 4.14. — *Sous les conditions du Théorème 4.1, si \tilde{m}_{CV} est l'estimateur \hat{m} construit avec le paramètre de lissage choisi par validation croisée \tilde{h}_{CV} , alors nous avons :*

$$|\tilde{m}_{CV}(x) - m(x)| = O(b_n^\beta) + O_{p.co.} \left(\sqrt{\frac{\log n}{n\phi(a_n)}} \right).$$

Ce résultat rend valide cette procédure courante qui consiste à utiliser des paramètres de lissage choisis par validation croisée permettant l'utilisation de régresseurs à noyaux fonctionnels entièrement automatiques. Cela complète la littérature standard dans le cadre multivarié (voir [Härdle et Marron \[1985\]](#)) ainsi que les premières avancées existantes dans le cadre fonctionnel (voir [Rachdi et Vieu \[2007\]](#), [Benhenni et collab. \[2007\]](#) et [Aneiros-Pérez et Vieu \[2011\]](#)).

Par souci de simplicité, nous avons traité le plus utilisé des sélecteurs de paramètres de lissage, qui est la procédure de validation croisée. Mais nous pouvons parvenir à des résultats similaires à ceux obtenus dans le Corollaire 4.14 avec tout autre sélecteur de paramètres de lissage. Cela peut inclure des paramètres de lissage basés sur des idées Bayésiennes (voir par exemple [Shang \[2014\]](#)) ou alors sur une approche minimax (voir [Chagny et collab. \[2014\]](#)).

4.7.2 Choix du paramètre de lissage dans d'autres problèmes

Les mêmes idées peuvent être développées pour traiter les paramètres de lissage aléatoires dans d'autres contextes que celui de la régression classique. Pour éviter les répétitions fastidieuses, les situations étudiées dans les sections 4.4, 4.5 et 4.6 seront présentées ensemble. Soient \tilde{h}_K et \tilde{h}_H des variables aléatoires prenant respectivement des valeurs dans les intervalles (a_n, b_n) et (c_n, d_n) . Notons par \tilde{F}^x , \tilde{f}^x et \tilde{S}^x les estimateurs obtenus en utilisant les largeurs de fenêtres aléatoires \tilde{h}_K et \tilde{h}_H dans les estimateurs \hat{F}^x , \hat{f}^x and \hat{S}^x . Ainsi, en tant qu'applications directes des Théorèmes 4.6, 4.9 et 4.12, on obtient les résultats suivants :

$$|\tilde{F}^x(y) - F^x(y)| = O(b_n^{\beta_1}) + O(d_n^{\beta_2}) + O_{p.co.} \left(\sqrt{\frac{\log n}{n\phi_x(a_n)}} \right),$$

$$|\tilde{f}^x(y) - f^x(y)| = O(b_n^{\beta_1}) + O(d_n^{\beta_2}) + O_{p.co.} \left(\sqrt{\frac{\log n}{n c_n \phi_x(a_n)}} \right),$$

et

$$|\tilde{S}^x(y) - S^x(y)| = O(b_n^{\beta_1}) + O(d_n^{\beta_2}) + O_{\text{p.co.}} \left(\sqrt{\frac{\log n}{n c_n \phi_x(a_n)}} \right).$$

Ces résultats généraux sur les largeurs de fenêtres aléatoires ont des applications directes dans des problèmes de sélection de paramètres de lissage orientés par les données, comme ce qui est présenté dans les quelques exemples suivants :

Exemple 1. En se basant sur l'estimateur de la fonction de répartition conditionnelle, une prédiction statistique peut être effectuée au moyen de la médiane conditionnelle. La prévision leave-one-out de Y_i est définie comme étant la solution $\hat{Y}_i^{(1)}$ de l'équation suivante :

$$\hat{F}^{X_i, -i}(y) = \frac{1}{2},$$

où $\hat{F}^{x, -i}$ est définie comme dans (4.25) en ne prenant pas en compte le couple (X_i, Y_i) . Les candidats naturels pour les choix des paramètres de lissage orientés par les données sont alors :

$$(\tilde{h}_{K,CV}^{(1)}, \tilde{h}_{H,CV}^{(1)}) = \arg \min_{h_K \in (a_n, b_n), h_H \in (c_n, d_n)} \sum_{i=1}^n (Y_i - \hat{Y}_i^{(1)})^2.$$

Exemple 2. À partir de l'estimateur de la densité conditionnelle, la prédiction peut être faite en utilisant le mode conditionnel. Plus précisément, Y_i peut être prédite comme étant la solution du problème de maximisation suivant :

$$\hat{Y}_i^{(2)} = \arg \max \hat{f}^{X_i, -i}(y),$$

où $\hat{f}^{x, -i}$ est défini par la technique leave-one-out comme dans (4.25). Ainsi, d'autres candidats naturels pour les choix des paramètres de lissage orientés par les données se profilent comme suit :

$$(\tilde{h}_{K,CV}^{(2)}, \tilde{h}_{H,CV}^{(2)}) = \arg \min_{h_K \in (a_n, b_n), h_H \in (c_n, d_n)} \sum_{i=1}^n (Y_i - \hat{Y}_i^{(2)})^2.$$

Les résultats suivants sont des conséquences directes des Théorèmes 4.6, 4.9 et 4.12.

Corolaire 4.15. — Soit $j = 1$ ou $j = 2$, et notons par $\tilde{F}^{(j),x}$, $\tilde{f}^{(j),x}$ et $\tilde{S}^{(j),x}$ respectivement les estimateurs \hat{F}^x , \hat{f}^x et \hat{S}^x construits avec les paramètres de lissage obtenus par validation croisée $\tilde{h}_{K,CV}^{(j)}$ et $\tilde{h}_{H,CV}^{(j)}$. Les trois résultats suivant sont alors vérifiés :

— Sous les conditions du Théorème 4.6, nous avons :

$$|\tilde{F}^{(j),x}(y) - F^x(y)| = O(b_n^{\beta_1}) + O(d_n^{\beta_2}) + O_{p.co.} \left(\sqrt{\frac{\log n}{n \phi_x(a_n)}} \right).$$

— Sous les conditions du Théorème 4.9, nous avons :

$$|\tilde{f}^{(j),x}(y) - f^x(y)| = O(b_n^{\beta_1}) + O(d_n^{\beta_2}) + O_{p.co.} \left(\sqrt{\frac{\log n}{n c_n \phi_x(a_n)}} \right).$$

— Sous les conditions du Théorème 4.12, nous avons :

$$|\tilde{S}^{(j),x}(y) - S^x(y)| = O(b_n^{\beta_1}) + O(d_n^{\beta_2}) + O_{p.co.} \left(\sqrt{\frac{\log n}{n c_n \phi_x(a_n)}} \right).$$

Ces derniers ont un intérêt majeur dans la pratique puisqu'ils rendent l'étude complètement automatique en utilisant les estimateurs \hat{F} , \hat{f} et \hat{S} . Alors que ce problème a déjà été légèrement traité pour la régression (voir la discussion à la fin de la section 4.7.1), la littérature est très faible concernant le problème de distribution conditionnelle (pour autant que nous le sachions le seul travail à s'y être intéressé est celui de [Chagny et collab. \[2014\]](#)), et notre travail semble être le premier en ce sens traitant de la densité conditionnelle et de l'opérateur de hasard conditionnel.

Convergence presque complète uniforme sur le nombre de voisins

Chapitre 5

Les résultats de ce chapitre font l'objet d'une publication dans le journal " *Journal of Multivariate Analysis*", **153**, 176–188 (2017), sous le titre : "Data-driven k NN estimation in nonparametric functional data analysis".

L'article a été co-produit par :

- Lydia Kara-Zaitri, *Université Djillali Liabès, Laboratoire de Mathématiques, Sidi Bel-Abbès, Algérie.*
- Ali Laksaci, *Université Djillali Liabès, Laboratoire de Mathématiques, Sidi Bel-Abbès, Algérie.*
- Mustapha Rachdi, *Université Grenoble Alpes, Laboratoire AGEIS EA 7407, Grenoble, France.*
- Philippe Vieu, *Université Paul Sabatier, Institut de Mathématiques, Toulouse, France.*

Estimation de type k NN orientée par les données dans l'analyse non paramétrique des données fonctionnelles

Dans ce chapitre, les estimateurs de type voisins les plus proches (k NN) sont introduits pour l'analyse non paramétrique d'échantillons statistiques impliquant des données fonctionnelles. La théorie asymptotique est traitée pour plusieurs opérateurs différents, dont la régression, la densité conditionnelle, la distribution conditionnelle et la fonction de hasard conditionnelle.

Le point principal ici est d'apporter des méthodes de sélection du nombre de voisins optimal afin de rendre les méthodes proposées entièrement automatiques. En utilisant les résultats du chapitre précédent, nous établissons des résultats théoriques de convergences uniformes sur le nombre de voisins (UNV) pour des estimateurs fonctionnels de type k NN.

5.1 Introduction

Grâce à leur flexibilité et à leur efficacité, les méthodes de lissage non paramétriques de type k voisins les plus proches (k NN) ont fait l'objet d'une grande attention dans la littérature statistique pour l'analyse de données multivariées. Les recherches dans ce sens ont été initiées par Cover [1968] et bon nombre de documents sont maintenant disponibles dans divers contextes d'estimation tels que la régression (Collomb [1981], Devroye et collab. [1994], Stone [1977]), la discrimination (Devroye et Wagner [1982], Li [1985]), l'estimation de la densité (Beirlant et collab. [2008], Devroye et Wagner [1977], Moore et Yackel [1977]) et l'analyse de regroupement (Tran et collab. [2006]). Le livre de Györfi et collab. [2006] fournit une étude approfondie des estimateurs k NN dans le cadre de dimension finie.

L'intérêt principal de l'approche k NN, comparée à celle des estimateurs à noyaux classiques, est le fait qu'elle comprenne un paramètre de lissage adaptable localement permettant de contrôler une certaine hétérogénéité locale des données. Comme la structure locale des données a plus d'influence à mesure que la dimension de l'espace

grandit (voir [Scott \[2015\]](#)), l'approche k NN est particulièrement bien adaptée aux problèmes multivariés. Dans les problèmes de dimension infinie le besoin de construire des estimateurs adaptables localement est encore plus conséquent ; voir [Benhenni et collab. \[2007\]](#) pour des études empiriques. Par conséquent, on s'attend à ce que les idées du k NN aboutissent à des méthodes statistiques intéressantes pour l'analyse des données fonctionnelles, et c'est le point que nous voulons aborder dans ce travail.

L'inférence statistique pour l'analyse fonctionnelle des données (AFD) a été fortement étudiée récemment ; voir [Bongiorno et collab. \[2014\]](#), [Horváth et Kokoszka \[2012\]](#), [Hsing et Eubank \[2015\]](#), [Ramsay et Silverman \[2005\]](#), [Zhang \[2013\]](#) pour une sélection d'ouvrages généraux sur le sujet. Les idées non paramétriques ont été popularisées par le livre de [Ferraty et Vieu \[2006\]](#) et prennent maintenant une grande place dans la littérature de l'AFD ; se référer les discussions des enquêtes récentes menées par [Cuevas \[2014\]](#) et [Goia et Vieu \[2016\]](#). Il existe notamment une liste étendue de contributions dans les statistiques fonctionnelles non paramétriques concernant les estimateurs à noyaux standards ; voir par exemple [Chagny et Roche \[2016\]](#), [Ezzahrioui et Ould-Said \[2008\]](#), [Ferraty et collab. \[2010\]](#), [Masry \[2005\]](#), [Rachdi et collab. \[2014\]](#) pour une sélection d'œuvres dans ce domaine. Cependant, l'étude des méthodes k NN est encore assez limitée et est principalement orientée vers l'estimation de la régression [Biau et collab. \[2010\]](#), [Burba et collab. \[2009\]](#), [Kudraszow et Vieu \[2013\]](#), [Laloë \[2008\]](#), [Lian et collab. \[2011\]](#), ou vers les discriminations de courbes [Cérou et Guyader \[2006\]](#). En outre, il convient de souligner que la faible documentation sur les méthodes k NN fonctionnelles concerne un nombre fixe de voisins alors que ce nombre dépend des données dans la pratique.

Le but principal de notre travail est donc d'énoncer un nombre de résultats asymptotiques traitant de plusieurs opérateurs cibles différents et permettant un choix automatique d'un nombre de voisins orienté par les données. Les principaux outils pour atteindre cet objectif sont des résultats uniformes sur le nombre de voisins (UNV). Ainsi, nous établissons en premier lieu la convergence presque complète (p.co) UNV. En fait, le caractère UNV permet de contrôler le comportement asymptotique des estimateurs même si le nombre de voisins est aléatoire, ce qui conduit à des applications directes à des nombres de voisins orientés par les données. Selon nos connaissances, cette contribution est la première à traiter des propriétés UNV pour les estimateurs k NN, car même dans le cas standard multivarié, ce type de littérature ne concerne que

les estimateurs à noyaux. Nous renvoyons à [Deheuvels et Ouadah \[2013\]](#) pour une liste de références sur la convergence uniforme sur la largeur de la fenêtre (ULF) des estimateurs à noyaux multivariés classiques, et à [Kara et collab. \[2017\]](#) pour l'extension fonctionnelle (Chapitre 4 ci dessus).

Ce chapitre est organisé comme suit : nous présentons nos modèles et leurs estimateurs dans la Section 5.2. Comme résultats préliminaires, nous donnons dans la Section 5.3 quelques propriétés asymptotiques pour une sélection d'estimateurs k NN aboutissant ainsi directement aux principaux résultats de ce chapitre (voir Section 5.4) sur le comportement asymptotique des estimateurs k NN en utilisant un nombre aléatoire de voisins.

5.2 Modèles et estimateurs

Soit $(X_1, Y_1), \dots, (X_n, Y_n)$ un échantillon de n couples aléatoires indépendants et de même distribution que le couple (X, Y) qui est un vecteur aléatoire à valeurs dans $\mathcal{F} \times \mathbb{R}$, où \mathcal{F} est un espace semi-métrique. Dans ce qui suit, d est une semi-métrique sur \mathcal{F} , x est un point fixe de \mathcal{F} , N_x est un voisinage fixe de x , et la boule fermée de centre x et de rayon α est notée :

$$B(x, \alpha) = \{y \in \mathcal{F} \text{ tels que } d(y, x) \leq \alpha\}.$$

Nous étudions des propriétés asymptotiques d'estimateurs de type k NN sur trois modèles conditionnels non paramétriques. Le premier étant le modèle de régression :

$$m(x) = \mathbb{E}(Y|x = x). \quad (5.1)$$

Ce modèle a été largement étudié dans le contexte fonctionnel ; voir [Ferraty et collab. \[2010\]](#) pour des avancées récentes. Dans la plus grande partie des travaux existants, la relation entre le modèle non paramétrique et l'espace de données est exprimée à travers la condition Lipschitzienne suivante. Supposons que pour certaines constantes $\beta > 0$ et $C_1 > 0$ nous avons :

$$\forall x_1, x_2 \in N_x \quad |m(x_1) - m(x_2)| \leq C_1 d^{\beta_1}(x_1, x_2). \quad (5.2)$$

L'estimateur de type k NN de la régression est donné par :

$$\widehat{m}_k(x) = \frac{\sum_{i=1}^n K\{H_{k,x}^{-1}d(x, X_i)\}Y_i}{\sum_{i=1}^n K\{H_{k,x}^{-1}d(x, X_i)\}}$$

où

$$H_{k,x} = \min \left\{ h \in \mathbb{R}^+ \text{ tels que } \sum_{i=1}^n \mathbb{1}_{B(x,h)}(X_i) = k \right\},$$

et $\mathbb{1}_A$ est la fonction indicatrice sur l'ensemble A .

Le second opérateur que nous allons étudier est la fonction de répartition conditionnelle qui est définie comme suit :

$$F^x(\cdot) = \mathbb{P}(Y \leq \cdot | X = x). \quad (5.3)$$

En fait, F^x peut être vue comme un cas particulier de la fonction de régression m , où la variable réponse Y est remplacée par la fonction indicatrice $\mathbb{1}_{\{Y \leq \cdot\}}$. Ainsi, l'estimateur k NN fonctionnel de F^x peut être défini comme :

$$\widehat{F}_k^x(y) = \frac{\sum_{i=1}^n K\{H_{k,x}^{-1}d(x, X_i)\} \mathbb{1}_{\{Y_i \leq y\}}}{\sum_{i=1}^n K\{H_{k,x}^{-1}d(x, X_i)\}} \quad (5.4)$$

et le modèle non paramétrique est caractérisé par l'hypothèse que pour un $\beta_2 > 0$ et un $C_2 > 0$:

$$\forall x_1, x_2 \in N_x \quad \forall y \in \mathbb{R} \quad |F^{x_1}(y) - F^{x_2}(y)| \leq C_2 d(x_1, x_2)^{\beta_2}. \quad (5.5)$$

De la même façon, nous considérons également la fonction de densité de probabilité conditionnelle définie, sur tout point où F^x est différentiable, par :

$$f^x(\cdot) = (F^x)'(\cdot) \quad (5.6)$$

et pour laquelle l'estimateur de type k NN est défini par :

$$\widehat{f}_{k,\ell}^x(y) = L_{\ell,y}^{-1} \frac{\sum_{i=1}^n K\{H_{k,x}^{-1}d(x, X_i)\} K(L_{\ell,y}^{-1}|y - Y_i|)}{\sum_{i=1}^n K\{H_{k,x}^{-1}d(x, X_i)\}}$$

où

$$L_{\ell,y} = \min \left\{ z \in \mathbb{R}^+ \text{ tels que } \sum_{i=1}^n \mathbb{1}_{(y-z, y+z)}(Y_i) = \ell \right\}$$

et où ℓ est une suite d'entiers naturels appartenant à l'intervalle $(\ell_{1,n}, \ell_{2,n})$.

Nous faisons remarquer que pour éviter de recourir à des notations additionnelles et

fastidieuses, nous utiliserons la même fonction noyau K pour pondérer la variable fonctionnelle X et la variable scalaire Y , mais les résultats obtenus dans la suite restent vrais pour deux noyaux différents.

L'aspect non paramétrique de ce modèle est caractérisé par l'hypothèse que pour tous $x_1, x_2 \in N_x$ et pour tous y_1, y_2 dans un voisinage fixe de y nous avons, pour $C_3, \beta_3, \beta_4 > 0$:

$$|f^{x_1}(y_1) - f^{x_2}(y_2)| \leq C_3 \{d(x_1, x_2)^{\beta_3} + |y_1 - y_2|^{\beta_4}\}. \quad (5.7)$$

Il est clair qu'en combinant ces deux derniers estimateurs, la méthode k NN peut être également utilisée pour l'estimation de l'opérateur de hasard défini comme suit :

$$S^x(\cdot) = \frac{f^x(\cdot)}{1 - F^x(\cdot)} \quad (5.8)$$

et l'estimateur correspondant est alors défini pour tout $y \in \mathbb{R}$ tel que $F^x(y) < 1$, par :

$$\widehat{S}_{k,\ell}^x(y) = \frac{\widehat{f}_k^x(y)}{1 - \widehat{F}_k^x(y)}.$$

5.3 Résultats asymptotiques UNV

Commençons d'abord par énoncer toutes les hypothèses nécessaires à l'obtention des résultats asymptotiques.

(H.1) Pour tout $r > 0$, $\mathbb{P}\{X \in B(x, r)\} =: \phi_x(r) > 0$ tel que, pour tout $s \in (0, 1)$:

$$\lim_{r \rightarrow 0} \frac{\phi_x(sr)}{\phi_x(r)} = \tau_x(s).$$

(H.2) La classe de fonctions :

$$\mathcal{K} = \{\cdot \mapsto K\{\gamma^{-1}d(x, \cdot)\}, \gamma > 0\}$$
 est une classe mesurable ponctuellement

telles que

$$\sup_Q \int_0^1 \sqrt{1 + \ln \mathcal{N}(\epsilon \|F\|_{Q,2}, \mathcal{K}, d_Q)} d\epsilon < \infty,$$

où le supremum est pris sur toutes les mesures de probabilité Q sur l'espace \mathcal{F} avec $Q(F^2) < \infty$ et où F est la fonction enveloppe de l'ensemble \mathcal{K} . Ici, d_Q est la

métrique $L_2(Q)$ et $\mathcal{N}(\epsilon, \mathcal{K}, d_Q)$ est le nombre minimal de boules ouvertes (avec la métrique $L_2(Q)$) de rayon ϵ nécessaire pour couvrir l'ensemble de fonctions \mathcal{K} . Nous noterons par $\|\cdot\|_{Q,2}$ la norme $L_2(Q)$.

(H.3) Le noyau K est à support dans $(0, 1/2)$ et sa première dérivée est continue sur $(0, 1/2)$ de telle sorte que :

$$0 < C_4 \mathbb{1}_{(0,1/2)}(\cdot) \leq K(\cdot) \leq C_5 \mathbb{1}_{(0,1/2)}(\cdot)$$

et

$$K(1/2) - \int_0^{1/2} K'(s) \tau_x(s) ds > 0,$$

(H.4) La suite des entiers naturels $(k_{1,n})$ est telle que :

$$\frac{\log n}{\min \left\{ n \phi_x^{-1} \left(\frac{k_{1,n}}{n} \right), k_{1,n} \right\}} \rightarrow 0.$$

Mentionnons que la plupart de nos hypothèses sont standards dans le contexte non paramétrique fonctionnel; voir [Ferraty et Vieu \[2006\]](#). La seule condition qui est spécifique à notre étude de l'uniformité sur le nombre de voisins est la condition (H.2). La première partie de (H.2) est une petite restriction sur la mesurabilité qui nous permet d'établir nos résultats uniformes. Notons qu'une façon alternative de traiter ce problème serait d'utiliser des notions de probabilités extérieures (voir [Van Der Vaart et Wellner \[1996\]](#), p. 4) mais cela nécessiterait des calculs nettement plus compliqués à faire. En outre, pour obtenir la convergence UNV, il convient de noter que l'uniformité asymptotique est étroitement liée aux notions d'entropie et de compacité. La seconde partie de (H.2) est une condition sur l'intégrale de l'entropie utilisée pour caractériser les fonctions de classe Donsker (voir [Van Der Vaart et Wellner \[1996\]](#)) et qui permet d'obtenir une distribution limite uniforme. Notons que ce type d'hypothèses pourrait aussi être utile pour l'estimation des moments de processus empiriques (voir [Van Der Vaart et Wellner \[1996\]](#)). La large généralité de l'hypothèse (H.2) est davantage soulignée par le fait qu'elle soit moins restrictive que la condition habituelle de VC-class (voir [Einmahl et collab. \[2005\]](#)).

Théorème 5.1. — *Sous les hypothèses (H.1)–(H.4) nous avons :*

•

$$\sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{F}_k^x(y) - F^x(y)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^{\beta_1} \right\} + O_{p.co.} \left(\sqrt{\frac{\log n}{k_{1,n}}} \right).$$

• Si $\mathbb{E}(|Y|^q|X) < C_5 < \infty$, presque sûrement pour $q \geq 2$ et $C_5 > 0$, nous avons alors :

$$\sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{m}_k(x) - m(x)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^{\beta} \right\} + O_{p.co.} \left(\sqrt{\frac{\log n}{k_{1,n}}} \right).$$

• De plus, si nous remplaçons l'hypothèse (H.4) par l'hypothèses (H4') suivante :

$$\frac{n \log n}{\ell_{1,n} \min \left\{ n \phi_x^{-1} \left(\frac{k_{1,n}}{n} \right), k_{1,n} \right\}} = o(1),$$

nous obtenons alors :

$$\sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{\ell_{1,n} \leq \ell \leq \ell_{2,n}} |\widehat{f}_{k,\ell}^x(y) - f^x(y)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^{\beta_2} \right\} + O \left(\frac{\ell_{2,n}}{n} \right)^{\beta_3} + O_{p.co.} \left(\sqrt{\frac{n \log n}{\ell_{1,n} k_{1,n}}} \right),$$

et

$$\sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{\ell_{1,n} \leq \ell \leq \ell_{2,n}} |\widehat{S}_{k,\ell}^x(y) - S^x(y)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^{\beta_2} \right\} + O \left(\frac{\ell_{2,n}}{n} \right)^{\beta_3} + O_{p.co.} \left(\sqrt{\frac{n \log n}{\ell_{1,n} k_{1,n}}} \right).$$

La preuve de ce théorème est donnée à l'annexe B.2.

5.4 Application à la sélection du nombre de voisins

Rappelons que l'intérêt principal de la méthode k NN (parmi les approches à noyaux usuelles) est de fournir un paramètre de lissage adaptable localement qui est facile à sélectionner en pratique, puisqu'il ne dépend que d'un seul paramètre. Toutefois, les avancées théoriques pour de telles procédures du choix d'un nombre de voisins

orienté par les données n'ont toujours pas été établies dans la littérature, et nous allons montrer dans ce qui suit l'impact des résultats théoriques de l'uniformité sur le nombre de voisins établis dans la Section 5.3 dans ce contexte.

Nous allons utiliser la technique la plus populaire de sélection basée sur les données, celle de la validation croisée. Nous concluons ensuite cette section par une discussion sur comment des résultats similaires restent valides pour tout autre outil de sélection automatique. Dans la suite, quand cela n'est pas spécifié, la minimisation sur le nombre de voisins k (resp. ℓ) est prise sur toutes les valeurs $k \in (k_{1,n}, k_{2,n})$ (resp. $\ell \in (\ell_{1,n}, \ell_{2,n})$).

5.4.1 Opérateur de régression

La procédure "leave-one-out" de validation croisée (Rachdi et Vieu [2007]) consiste à minimiser le critère de prédiction de l'erreur de moindre carré suivant :

$$k_{CV} = \arg \min_{k \in (k_{1,n}, k_{2,n})} CV(k),$$

où

$$CV(k) = \sum_{i=1}^n \{Y_i - \widehat{m}_k^{-i}(X_i)\}^2 \quad \text{et} \quad \widehat{m}_k^{-i}(x) = \frac{\sum_{j \neq i} K\{H_{k,x}^{-1}d(x, X_j)\}Y_j}{\sum_{j \neq i} K\{H_{k,x}^{-1}d(x, X_j)\}}. \quad (5.9)$$

La valeur k_{CV} sélectionnée est une fonction aléatoire qui dépend de tout l'échantillon statistique $(X_1, Y_1), \dots, (X_n, Y_n)$ ce qui rend l'étude directe de l'estimateur k NN correspondant plus difficile à effectuer. Les résultats UNV obtenus dans la Section 5.3 nous permettent d'obtenir facilement la vitesse de convergence de

$$\widetilde{m}_{k,CV} = \frac{\sum_{i=1}^n K\{H_{k_{CV}(X_1, Y_1, \dots, X_n, Y_n), x}^{-1}d(x, X_j)\}Y_j}{\sum_{i=1}^n K\{H_{k_{CV}(X_1, Y_1, \dots, X_n, Y_n), x}^{-1}d(x, X_j)\}}.$$

Ceci est énoncé dans le corolaire suivant.

Corolaire 5.2. — *Sous les conditions du Théorème 5.1, nous avons :*

$$|\widetilde{m}_{k,CV}(x) - m(x)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^\beta \right\} + O_{p.co.} \left(\sqrt{\frac{\log n}{k_{1,n}}} \right).$$

Ce résultat est une conséquence directe de la première assertion du Théorème 5.1.

5.4.2 Fonction de répartition conditionnelle

Dans le cadre multivarié, [De Gooijer et Gannoun \[2000\]](#) ont proposé quelques critères de validation croisée qui peuvent être adaptés à notre étude fonctionnelle de la façon suivante :

$$\text{CV}_{GG}(k) = \sum_{i=1}^n \{ \mathbb{1}_{\{Y_i \leq Y_j\}} - \widehat{F}_k^{X_i^{-i}}(Y_j) \}^2,$$

où

$$\widehat{F}_k^{X_i^{-i}}(y) = \frac{\sum_{j \neq i} K \{ H_{k,x}^{-1} d(X_i, X_j) \} \mathbb{1}_{\{Y_j \leq y\}}}{\sum_{j \neq i} K \{ H_{k,x}^{-1} d(X_i, X_j) \}}.$$

Le corolaire suivant garantit un bon comportement asymptotique à l'estimateur orienté par les données construit avec le nombre optimal de voisins qui minimise la loi CV_{GG} .

Corolaire 5.3. — *Sous les conditions du Théorème 5.1, si $\widetilde{F}_k^x(y)$ est l'estimateur de $F_k^x(y)$ construit avec la procédure de validation croisée CV_{GG} , nous avons alors :*

$$|\widetilde{F}_k^x(y) - F^x(y)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^{\beta_1} \right\} + O_{p.co.} \left(\sqrt{\frac{\log n}{k_{1,n}}} \right).$$

Ce résultat est une conséquence directe de la deuxième assertion du Théorème 5.1.

5.4.3 Densité de probabilité conditionnelle

Le problème de la sélection du paramètre de lissage dans l'estimation de la densité conditionnelle a suscité une certaine attention dans la littérature statistique multivariée. Les idées populaires de validation croisée (voir [Youndjé et collab. \[1993\]](#) et les références qui y sont citées) peuvent être adaptés au cadre fonctionnel de la façon suivante. L'objectif est de minimiser l'erreur suivante :

$$\text{err}_1(\widehat{f}_{(a,b)}, f) = \int \int \left\{ \widehat{f}_{(a,b)}^x(y) - f^x(y) \right\}^2 W_1(x) W_2(y) dP_X(x) dy,$$

$$\text{err}_2(\widehat{f}_{(a,b)}, f) = \frac{1}{n} \sum_{i=1}^n \left\{ \widehat{f}_{(a,b)}^{X_i}(Y_i) - f^{X_i}(Y_i) \right\}^2 \frac{W_1(X_i) W_2(Y_i)}{f^{X_i}(Y_i)}$$

ou

$$\text{err}_3(\widehat{f}, f) = \int \int \mathbb{E} \left\{ \widehat{f}_{(a,b)}^x(y) - f^x(y) \right\}^2 W_1(x) W_2(y) dP_X(x) dy,$$

où W_1 et W_2 sont deux fonctions de ponctuation non négatives. Ces erreurs théoriques sont impossibles à calculer en pratique, et le critère "leave-one-out" de la validation croisée peut être construit pour en faire une approximation dans une certaine façon entièrement basée sur les données :

$$\begin{aligned} \mathbf{CV}_{CD}(a, b) = & \frac{1}{n} \sum_{i=1}^n W_1(X_i) \int \left(\widehat{f}_{(a,b)}^{X_i^{-i}} \right)^2(y) W_2(y) dy \\ & - \frac{2}{n} \sum_{i=1}^n \widehat{f}_{(a,b)}^{X_i^{-i}}(Y_i) W_1(X_i) W_2(Y_i) \end{aligned} \quad (5.10)$$

où

$$\widehat{f}_{(a,b)}^{X_i^{-i}} = b^{-1} \frac{\sum_{j \neq i}^n K\{a^{-1}d(X_i, X_j)\} K(b^{-1}|y - Y_j|)}{\sum_{j \neq i}^n K\{a^{-1}d(X_i, X_j)\}}.$$

Le paramètre de lissage bidimensionnel $(k_{\mathbf{CV}_{CD}}, \ell_{\mathbf{CV}_{CD}})$ est sélectionné par la procédure suivante :

$$(k_{\mathbf{CV}_{CD}}, \ell_{\mathbf{CV}_{CD}}) = \arg \min_{k_{1,n} \leq k \leq k_{2,n}, \ell_{1,n} \leq \ell \leq \ell_{2,n}} \mathbf{CV}_{CD}(k, \ell).$$

Les résultats UNV, établis dans la Section 5.3, permettent d'obtenir la vitesse de convergence suivante pour l'opérateur fonctionnel de densité conditionnelle orienté par les données et obtenu par validation croisée.

Corolaire 5.4. — *Sous les conditions du Théorème 5.1, si $\widetilde{f}_{k,\ell}$ est l'estimateur de densité conditionnelle construit par les procédures de validation croisée \mathbf{CV}_{CD} , nous avons :*

$$|\widetilde{f}_{k,\ell}^x(y) - f^x(y)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^{\beta_2} \right\} + O \left(\frac{\ell_{2,n}}{n} \right)^{\beta_3} + O_{p.co.} \left(\sqrt{\frac{n \log n}{\ell_{1,n} k_{1,n}}} \right).$$

Ce résultat est une conséquence directe de la troisième assertion du Théorème 5.1.

5.4.4 Autres méthodes orientées par les données

Alors que les procédures de validation croisée décrites plus haut visent à approcher l'erreur quadratique de l'estimation, des méthodes alternatives pour le choix du paramètre de lissage pourraient être introduites dans le but d'optimiser le pouvoir prédictif de la méthode. Cela peut être fait en minimisant l'un des critères de prédiction

suivants :

$$(\tilde{k}_{\text{CV}}^{(1)}, \tilde{\ell}_{\text{CV}}^{(1)}) = \arg \min_{k_{1,n} \leq k \leq k_{2,n}, \ell_{1,n} \leq \ell \leq \ell_{2,n}} \sum_{i=1}^n (Y_i - \widehat{Y}_i^{(1)})^2, \quad (5.11)$$

ou

$$(\tilde{k}_{\text{CV}}^{(2)}, \tilde{\ell}_{\text{CV}}^{(2)}) = \arg \min_{k_{1,n} \leq k \leq k_{2,n}, \ell_{1,n} \leq \ell \leq \ell_{2,n}} \sum_{i=1}^n (Y_i - \widehat{Y}_i^{(2)})^2 \quad (5.12)$$

où la prédiction est effectuée au moyen soit de la médiane conditionnelle, c.à.d :

$$\widehat{Y}_i^{(1)} = \widehat{F}_k^{(X_i^{-i})^{-1}}(1/2)$$

soit du mode conditionnel, c.à.d :

$$\widehat{Y}_i^{(2)} = \arg \max_{\mathcal{S}} \widehat{f}_{k,\ell}^{X_i^{-i}}(y),$$

où \mathcal{S} est un compact de \mathbb{R} .

À notre connaissance, de tels sélecteurs n'ont reçu aucune attention mathématique par le passé (y compris dans les situations multivariées habituelles), probablement à cause des difficultés techniques. Cependant, la grande généralité des résultats UNV mentionnés ci-dessus nous permet, comme conséquence directe et facile du Théorème 5.1, d'établir les vitesses de convergence pour les estimateurs orientés par les données correspondants :

Corolaire 5.5. — Soit $j = 1$ ou $j = 2$, et notons par $\widetilde{F}_k^{(j),x}$, $\widetilde{f}_{k,\ell}^{(j),x}$ et $\widetilde{S}_{k,\ell}^{(j),x}$ respectivement les estimateurs \widehat{F}_k^x , $\widehat{f}_{k,\ell}^x$ et $\widehat{S}_{k,\ell}^x$ qui sont construits en utilisant le nombre optimal de voisins $\tilde{k}_{\text{CV}}^{(j)}$ et $\tilde{\ell}_{\text{CV}}^{(j)}$. Alors, les trois résultats suivants sont obtenus :

$$|\widetilde{F}_k^{(j),x}(y) - F^x(y)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^{\beta_1} \right\} + O_{p.co.} \left(\sqrt{\frac{\log n}{k_{1,n}}} \right),$$

$$|\widetilde{f}_{k,\ell}^{(j),x}(y) - f^x(y)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^{\beta_2} \right\} + \left(\frac{\ell_{2,n}}{n} \right)^{\beta_3} + O_{p.co.} \left(\sqrt{\frac{n \log n}{\ell_{1,n} k_{1,n}}} \right),$$

et

$$|\widetilde{S}_{k,\ell}^{(j),x}(y) - S^x(y)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^{\beta_2} \right\} + O \left(\frac{\ell_{2,n}}{n} \right)^{\beta_3} + O_{p.co.} \left(\sqrt{\frac{n \log n}{\ell_{1,n} k_{1,n}}} \right).$$

Partie II

Études pratiques

Si les recherches théoriques des statisticiens se focalisent de plus en plus sur le traitement des données fonctionnelles, c'est essentiellement parce que de plus en plus de domaines y ont recours. Aujourd'hui, grâce au développement des logiciels informatiques et des outils de stockage des données, le big data se retrouve au cœur de presque toutes les branches des connaissances humaines. Dans ce chapitre, nous mettons en relief la pluralité des domaines qui ont recours aux méthodes statistiques pour traiter des données fonctionnelles. Il convient de préciser que cette liste n'offre qu'une vision réduite des nombreux champs d'applications statistiques existants.

Domaine médical : La médecine est l'une des spécialités qui traitent le plus des données fonctionnelles. Dans l'étude de survie par exemple, les médecins enregistrent continuellement des données à propos de l'évolution de l'état de santé de leurs patients. Ces renseignements leur servent ensuite à prévoir le temps nécessaire à la guérison du patient, ou à l'évolution de la maladie dans un sens ou dans l'autre.

D'un autre côté, la plupart des appareils utilisés dans le domaine de la santé fournissent des données sous forme de courbes. L'électrocardiogramme en est un exemple représentatif. Cet appareil enregistre l'activité cardiaque dans le temps, et les informations précieuses qu'il fournit sont utilisées, entre autres, pour détecter des troubles du rythme cardiaque, et/ou pour prévenir des maladies cardio-vasculaires.

Traitement du signal : Un signal est souvent modélisé par un

processus stochastique, ce qui explique l'omniprésence de la théorie de l'estimation fonctionnelle dans divers systèmes de traitement du signal, comme le radar par exemple. Ce système est utilisé pour la détection de la position et de la vitesse d'objets comme les voitures, les avions ou même les précipitations. Pour déterminer la localisation de l'objet, le radar transmet des ondes électromagnétiques qui sont réfléchies par l'objet, provoquant un écho qui est reçu par l'antenne du radar après un temps t . L'écho pouvant être faussé par des parasites environnementaux, l'interprétation directe du délai d'aller-retour t ne donne pas l'emplacement exacte de l'objet étudié. Des techniques d'analyse statistique de ces échos sont nécessaires pour estimer le vrai temps t_0 que les ondes prendraient pour l'aller-retour s'il n'y avait pas de parasites. L'estimation de ce délai t_0 engendre celle de la position de l'objet.

Météorologie : L'amélioration des techniques d'observation de phénomènes atmosphériques permet aujourd'hui de collecter continuellement des données concernant des variables météorologiques (précipitations, vitesse du vent, formation des nuages...etc). L'analyse de ces données fonctionnelles, collectées sur des grilles de temps très rapprochées, est une pratique courante dans différents domaines.

En agriculture, par exemple, les prévisions des précipitations à venir servent à anticiper les quantités des récoltes et, par conséquent, les gains ou les pertes à venir. En ingénierie, les opérateurs des centrales éoliennes utilisent des courbes de puissance pour estimer ou prévoir la production d'électricité dans des conditions de vent données.

Autres champs d'application :

- En cryptologie, on s'intéresse à la biométrie des signatures manuscrites pour l'authentification des documents.
- En sciences environnementales l'on cherche à analyser les données de pollution atmosphériques et à étudier leur impacte sur l'écologie.
- En finance, les "traders" cherchent à anticiper le cours des actions en bourse permettant ainsi de préserver la croissance économique.

- En génétique, les biologistes analysent l'expression des gènes au cours du temps dans le but d'avoir une meilleure compréhension de la fonction des gènes et des interactions entre certains de leurs effets.
- D'autres exemples d'applications comme la consommation d'énergie, l'agro-alimentaire ou encore la criminologie sont cités et traités dans les livres de Ramsay et Silverman [2002], Ferraty et Vieu [2006], Ramsay et collab. [2009] et Zhang [2013].

Nous renvoyons aux trois sites Web suivants, qui permettent à tout praticien de mettre en œuvre facilement des techniques statistiques dans l'analyse des données fonctionnelles. Le lecteur y trouvera des sources de routines pour les logiciels R ou Matlab, et des explications sur leur utilisation.

<http://eio.usc.es/pub/gi1914/index.php/en/software->

Pour des techniques d'estimation linéaire :

<http://www.psych.mcgill.ca/misc/fda/>

Pour des estimations non-paramétriques :

<http://www.math.univ-toulouse.fr/staph/npfda/>

Dans cette deuxième partie de notre thèse nous nous proposons d'illustrer la faisabilité et le bon comportement de nos méthodes sur des échantillons finis. Nous commencerons au chapitre 7 par une étude sur des données simulées. Nous passerons ensuite au chapitre 8, où nous appliquerons nos résultats sur des données réelles.

L'objectif de cette section est double. Premièrement, nous allons montrer que les procédures automatiques k NN peuvent facilement être implémentées. Ensuite, nous allons les comparer avec les procédures à noyaux usuelles afin de souligner le fait que le caractère local de l'approche k NN permet de réduire efficacement les erreurs de la prédiction.

7.1 Présentation de l'étude

Considérons les modèles non paramétriques fonctionnels suivants :

$$\forall_{i \in \{1, \dots, n=300\}} \quad Y_i = m(X_i) + \varepsilon_i, \quad (7.1)$$

où les ε_i sont générés indépendamment suivant la distribution $\mathcal{N}(0, 0.05)$ et sont supposés être indépendants de X_i pour tout $i \in \{1, \dots, n\}$. Les variables explicatives fonctionnelles échantillonnées sont générés comme suit. Pour tout $i \in \{1, \dots, n\}$:

$$\forall t \in (0, \pi) \quad X_i(t) = \cos(2a_i t) + \sin(2t + b_i) + c_i t,$$

où $a_i \sim \mathcal{N}(0, 1)$, $b_i \sim \mathcal{N}(3, 1)$, et $c_i \sim \mathcal{U}(0, 1)$. Les courbes X_1, \dots, X_n sont alors discrétisées sur la même grille générée à partir de 100 mesures équidistantes sur l'intervalle $(0, \pi)$. Les courbes sont tracées dans la Figure 7.1.

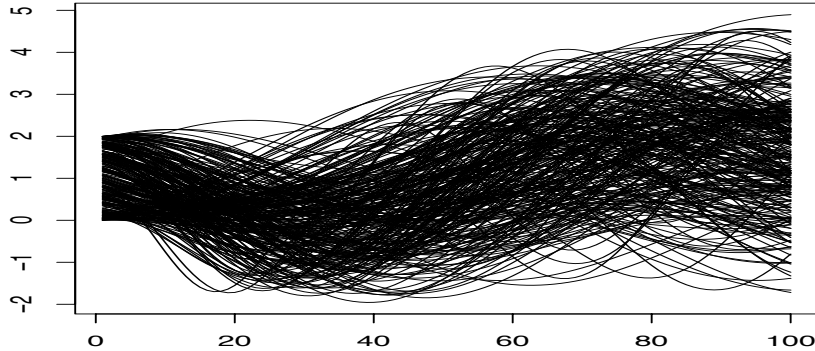


FIGURE 7.1 – Échantillon de 300 courbes

La réponse scalaire Y_i est définie dans (7.1) en utilisant l'opérateur de régression :

$$m(x) = 5 \exp \left[\frac{1}{\int_0^\pi \{1 + x^2(t)\} dt} \right].$$

Lors de la construction des divers estimateurs k NN nous avons utilisé la semi-métrique :

$$d(u, v) = \int_0^\pi \{u(t) - v(t)\}^2 dt$$

et la fonction noyau

$$K(t) = \frac{3}{4} (1 - t^2) \mathbb{1}_{(0,1)}.$$

Les routines utilisées pour calculer les estimateurs fonctionnels sont disponibles sur le site web www.lsp.ups-tlse.fr/staph/npfda.

7.2 Sélection automatique du nombre de voisins

Tout d'abord, nous allons décrire l'influence qu'a le nombre de voisins sur le comportement des estimateurs. Afin de couvrir une grande variété d'opérateurs (régression, distribution conditionnelle et densité conditionnelle), nous examinerons différents prédicateurs (espérance conditionnelle, médiane conditionnelle et quantile conditionnel). Ainsi, les techniques utilisées pour la sélection des voisins sont celles qui sont traitées

dans (5.9), (5.11) et (5.12). Nous avons calculé les erreurs de prédiction correspondantes :

$$\text{MSE}_{reg}(k) = \frac{1}{n} \sum_{i=1}^n (Y_i - \widehat{Y}_i^{(reg^k)})^2 \quad \text{où} \quad \widehat{Y}_i^{(reg^k)} = \frac{\sum_{j \neq i} K\{H_{k,x}^{-1}d(x, X_j)\} Y_j}{\sum_{j \neq i} K\{H_{k,x}^{-1}d(x, X_j)\}},$$

$$\text{MSE}_{median}(k) = \frac{1}{n} \sum_{i=1}^n (Y_i - \widehat{Y}_i^{(median^k)})^2 \quad \text{où} \quad \widehat{F}^{(X_i^{-i})} \left(\widehat{Y}_i^{(median^k)} \right) = 1/2,$$

et

$$\text{MSE}_{mode}(k, \ell) = \frac{1}{n} \sum_{i=1}^n (Y_i - \widehat{Y}_i^{(mode^{k,\ell})})^2 \quad \text{où} \quad \widehat{Y}_i^{(mode^{k,\ell})} = \arg \max \widehat{f}^{X_i^{-i}}(y).$$

Ces erreurs sont évaluées sur une suite de nombres de voisins définies par $\{5, 10, 15, \dots, 300\}$. Les estimateurs $\widehat{Y}_i^{(reg^k)}$, $\widehat{Y}_i^{(median^k)}$ et $\widehat{Y}_i^{(mode^{k,\ell})}$ sont obtenus par les R-routines appelées respectivement :

`funopare.kNN.gcv`, `funopare.mode.lcv` et `funopare.quantile.lcv`¹.

Les résultats sont résumés dans la Figure 7.2 pour le prédicateur de la régression et celui de la médiane conditionnelle. Concernant le mode conditionnel, puisque l'estimateur inclut deux paramètres ℓ et k , les résultats sont présentés dans la Figure 7.3 par deux graphes : la figure de gauche montre le comportement de l'erreur en fonction de ℓ (quand k est fixé à sa valeur optimale k_{opt}) tandis que la figure de droite montre le comportement de l'erreur en fonction de k (quand ℓ est fixée à sa valeur optimale ℓ_{opt}). Dans chacun des graphiques, la ligne horizontale montre la valeur minimale de l'erreur du prédicateur.

La large variabilité des erreurs MSE montre que les méthodes k NN sont très sensibles au choix du nombre de voisins, et montre ainsi dans chaque cas l'importance du choix de la valeur optimale qui est celle obtenue en minimisant le critère discuté ci-dessus.

1. Toutes ces routines sont disponibles sur www.lsp.ups-tlse.fr/staph/npfda

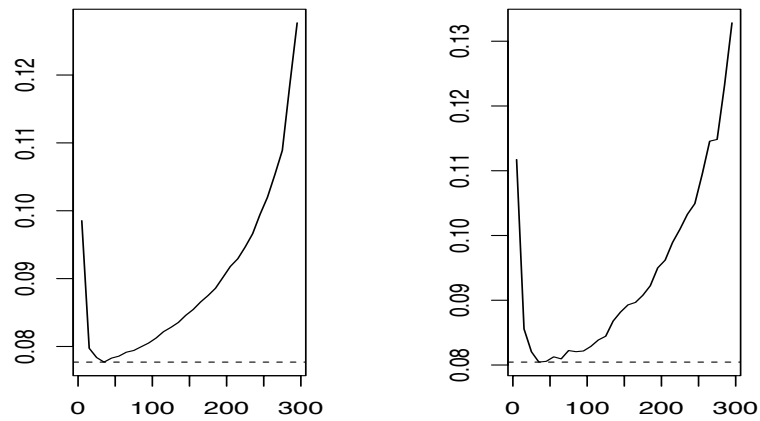
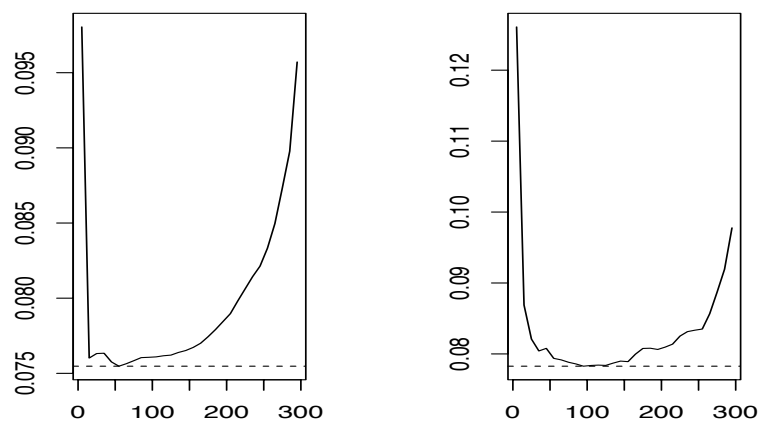


FIGURE 7.2 – Erreurs de prédiction en fonction du nombre de voisins.

À gauche, prédicteur de régression. À droite, prédicteur de la médiane conditionnelle

FIGURE 7.3 – Erreurs de prédiction en fonction des nombres de voisins pour le prédicteur du mode conditionnel. À gauche, k est fixé. À droite, ℓ est fixé

7.3 Étude comparative

Le but de cette section est de comparer l'approche k NN aux approches à noyaux classiques. À cette fin, nous avons calculé les erreurs optimales avec les approches k NN, notamment :

$$k\text{NNMSE}(reg) := \text{MSE}_{reg}(k.opt),$$

$$k\text{NNMSE}(median) := \text{MSE}_{median}(k.opt),$$

et

$$k\text{NNMSE}(mode) := \text{MSE}_{mode}(k.opt, \ell.opt),$$

et les avons comparées à celles obtenues avec un paramètre fixe global :

$$\text{KERMSE}(reg) := \frac{1}{n} \sum_{i=1}^n (Y_i - \tilde{Y}_i^{(reg^{a.opt})})^2 \quad \text{où} \quad \tilde{Y}_i^{(reg^a)} = \frac{\sum_{j \neq i} K\{a^{-1}d(x, X_j)\}Y_j}{\sum_{j \neq i}^n K\{a^{-1}d(x, X_j)\}},$$

$$\text{KERMSE}(median) := \frac{1}{n} \sum_{i=1}^n (Y_i - \tilde{Y}_i^{(median^{a.opt})})^2 \quad \text{où} \quad \tilde{Y}_i^{(median^a)} = \tilde{F}_a^{(X_i^{-i})^{-1}}(1/2),$$

et

$$\text{KERMSE}(mode) := \frac{1}{n} \sum_{i=1}^n (Y_i - \tilde{Y}_i^{(mode^{a.opt, b.opt})})^2 \quad \text{où} \quad \tilde{Y}_i^{(mode^{a,b})} = \arg \max \tilde{f}_{a,b}^{X_i^{-i}}(y),$$

avec

$$\tilde{F}_a^x(y) = \frac{\sum_{i=1}^n K\{a^{-1}d(x, X_i)\} \mathbf{1}(Y_i \leq y)}{\sum_{i=1}^n K\{a^{-1}d(x, X_i)\}},$$

et

$$\tilde{f}_{a,b}^x(y) = b^{-1} \frac{\sum_{i=1}^n K\{a^{-1}d(x, X_i)\} K(b^{-1}|y - Y_i|)}{\sum_{i=1}^n K\{a^{-1}d(x, X_i)\}}.$$

Il convient de souligner ici que dans les trois derniers cas, les paramètres de lissage optimaux $a.opt$ et/ou $b.opt$ sont choisis parmi une suite de quantiles du vecteur distances entre les variables fonctionnelles (resp. entre les variables réponses).

	Mode	Médiane	Régression
Méthode à noyau	0.081	0.11	0.11
Méthode k NN	0.046	0.087	0.086

TABLE 7.1 – Résultats MSE

Tous les résultats sont résumés dans le Tableau 7.1, où nous pouvons observer que la méthode k NN conduit à une réduction importante des erreurs de prédiction (d'au moins 25%) pour chaque prédicteur (qu'il soit basé sur la régression, la distribution conditionnelle ou sur la densité conditionnelle). Dans cet exemple, le prédicteur basé sur le mode conditionnel semble être plus efficace que les deux autres qui sont basés sur l'espérance et la médiane conditionnelles.

Application sur des données réelles

Dans ce chapitre, nous traitons un cas d'étude sur la qualité du sucre en utilisant la spectroscopie de fluorescence. Le but de ce travail est double. Nous cherchons d'un côté, à montrer que les techniques de choix du paramètre de lissage optimal initiées dans les sections 4.7 et 5.4 sont faciles d'utilisation et d'implémentation. D'un autre côté, nous voulons montrer que les prédicteurs définis à partir des paramètres de lissage optimaux mènent à des résultats intéressants. Nous nous proposons pour ce faire d'étudier les trois prédicteurs : moyenne, médiane et mode conditionnels, obtenus par les deux méthodes d'estimation : noyau et k NN. Nous proposons ensuite de comparer entre les résultats obtenus en calculant leurs erreurs quadratiques moyennes.

8.1 Description des données

En chimie organique et en biochimie, la spectroscopie de fluorescence permet de déduire des informations à propos de composées chimiques. L'étude consiste à émettre, grâce à un spectrofluoromètre, des lumières avec des longueurs d'ondes précises pour exciter les électrons des molécules, qui à leur tour réagissent en émettant un spectre de lumières de différentes longueurs d'ondes.

L'étude que nous analysons ici nous vient de l'industrie du sucre et a été traitée par Bro [1999]. Des échantillons de sucre ont été prélevés dans une usine de production en Scandinavie. Au total, 268 échantillons ont été dissous séparément dans de l'eau non tamponnée et mesurés au spectrofluoromètre. Pour chaque échantillon traité,

l'expérience consistait à envoyer successivement sept niveaux de lumières d'excitation différents, de longueurs d'ondes : 230, 240, 255, 290, 305, 325 et 340 nanomètres. Les intensités du spectre d'émission provoqué par chacune des longueurs d'ondes précédentes sont ensuite mesurées sur des plages comprises entre 275nm et 560nm avec un saut de 0.5 d'intervalle, ce qui donne 571 points de mesure pour chaque spectre émis.

La figure 8.1 suivante représente l'intensité de la lumière en fonction de la longueur d'onde émise pour chacune des sept longueurs d'ondes d'excitation. L'axe des abscisses est divisé en sept plages représentant les sept niveaux d'excitation dans un ordre croissant. Chacune des plages comprend les longueurs d'ondes de 270nm à 560nm.

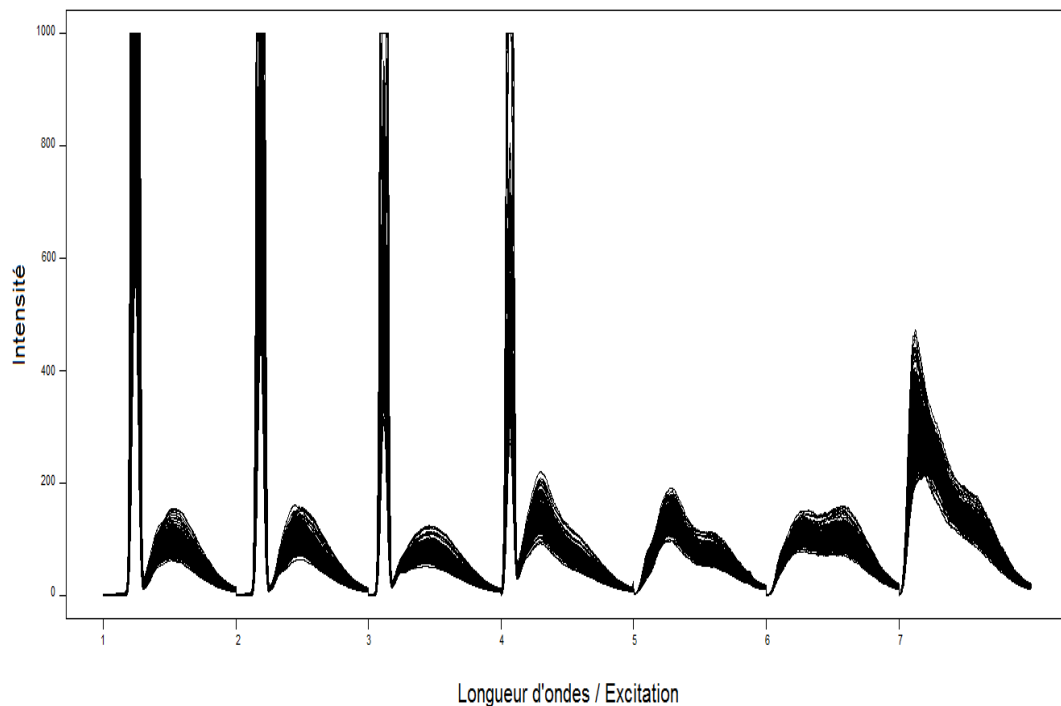


FIGURE 8.1 – Courbes de fluorescence des 268 échantillons $(X_i)_{i=1,268}$

La variable réponse Y à laquelle nous nous intéressons ici exprime la couleur que prend le sucre. Elle est prélevée au même rythme que les échantillons de sucre mesurés par fluorométrie, et déterminée comme étant l'absorption à 420 nm d'une solution filtrée à

la membrane de sucre ajustée au pH=7. La couleur est donnée en une unité dérivée de l'absorbance où 45 est la couleur maximale autorisée de sucre standard. Cette couleur est de loin si faible, qu'elle n'est pas importante pour le consommateur, mais elle est intéressante pour le contrôle des processus de fabrication et de la qualité du sucre. pour plus de détails, visiter le site web http://www.models.life.ku.dk/Sugar_Process. Les données y sont disponibles sous Matlab.

8.2 Traitement statistique des données

Afin d'évaluer l'efficacité de nos prédicteurs k NN, l'ensemble des données présentées plus haut est divisé en deux parties de façon totalement aléatoire. Une famille d'apprentissage contenant 218 échantillons $(X_i, Y_i)_{i=1,218}$ que nous utilisons pour construire nos estimateurs, et une famille de test qui contient les 50 échantillons restants $(X_i, Y_i)_{i=219,268}$. Nous supposons que les valeurs y_i de la famille de test sont inconnues et nous cherchons à les prédire.

Pour estimer les valeurs des $(Y_j)_{j=219,268}$ de l'échantillon test, nous utilisons les prédicteurs : moyenne, mode et médiane conditionnels des deux méthodes noyau et k NN.

Méthode à noyau : Nous considérons les estimateurs à noyaux optimaux :

$$\hat{m}_{\text{opt}}(x) = \frac{\sum_{i=1}^{218} K\{h_{\text{opt}.1}^{-1}d(x, X_i)\}Y_i}{\sum_{i=1}^{218} K\{h_{\text{opt}.1}^{-1}d(x, X_j)\}},$$

$$\hat{F}_{\text{opt}}^x(y) = \frac{\sum_{i=1}^{218} K\{h_{\text{opt}.2}^{-1}d(x, X_i)\}\mathbf{1}(Y_i \leq y)}{\sum_{i=1}^{218} K\{h_{\text{opt}.2}^{-1}d(x, X_i)\}},$$

et

$$\hat{f}_{\text{opt}}^x(y) = h_{\text{opt}.y}^{-1} \frac{\sum_{i=1}^{218} K\{h_{\text{opt}.x}^{-1}d(x, X_i)\}K(h_{\text{opt}.y}^{-1}|y - Y_i|)}{\sum_{i=1}^{218} K\{h_{\text{opt}.x}^{-1}d(x, X_i)\}},$$

où $h_{\text{opt}.1}$, $h_{\text{opt}.2}$ et $(h_{\text{opt}.x}, h_{\text{opt}.y})$ sont les largeurs de fenêtres optimales des estimateurs \hat{m} , \hat{F}^x et \hat{f}^x respectivement, obtenues par validation croisée et définies dans la section 4.7.

En notant x_j la valeur que prend la variable X_j , les prédicateurs des Y_j sont alors définis pour tout j entre 219 et 268 par :

$$\widehat{Y}_j^{\text{reg}} = \widehat{m}_{\text{opt}}(x_j),$$

$$\widehat{Y}_j^{\text{médiane}} \text{ est telle que : } \widehat{F}_{\text{opt}}^{x_j}(\widehat{Y}_j^{\text{médiane}}) = 1/2,$$

et

$$\widehat{Y}_j^{\text{mode}} = \arg \max_{y \in \mathcal{S}} \widehat{f}_{\text{opt}}^{x_j}(y),$$

où l'ensemble $\mathcal{S} = \{Y_i, i = 1, 218\}$.

Méthode k NN : De même, les estimateurs de type k NN optimaux étant :

$$\widehat{m}_{k_{\text{opt}}}(x) = \frac{\sum_{i=1}^{218} K\{H_{k_{\text{opt}.1}}^{-1} d(x, X_i)\} Y_i}{\sum_{i=1}^{218} K\{H_{k_{\text{opt}.1}}^{-1} d(x, X_i)\}},$$

$$\widehat{F}_{k_{\text{opt}}}^x(y) = \frac{\sum_{i=1}^{218} K\{H_{k_{\text{opt}.2}}^{-1} d(x, X_i)\} \mathbf{1}(Y_i \leq y)}{\sum_{i=1}^{218} K\{H_{k_{\text{opt}.2}}^{-1} d(x, X_i)\}},$$

et

$$\widehat{f}_{(k,\ell)_{\text{opt}}}^x(y) = L_{\text{opt}}^{-1} \frac{\sum_{i=1}^{218} K\{H_{k_{\text{opt}}}^{-1} d(x, X_i)\} K(L_{\text{opt}}^{-1} |y - Y_i|)}{\sum_{i=1}^{218} K\{H_{k_{\text{opt}}}^{-1} d(x, X_i)\}},$$

où $k_{\text{opt}.1}$, $k_{\text{opt}.2}$ et $(k_{\text{opt}}, \ell_{\text{opt}})$ sont les nombres de voisins optimaux des estimateurs \widehat{m}_k , \widehat{F}_k^x et $\widehat{f}_{k,\ell}^x$ respectivement, obtenues par validation croisée et définies dans la section 5.4. Les prédicateurs des Y_j sont alors définis par :

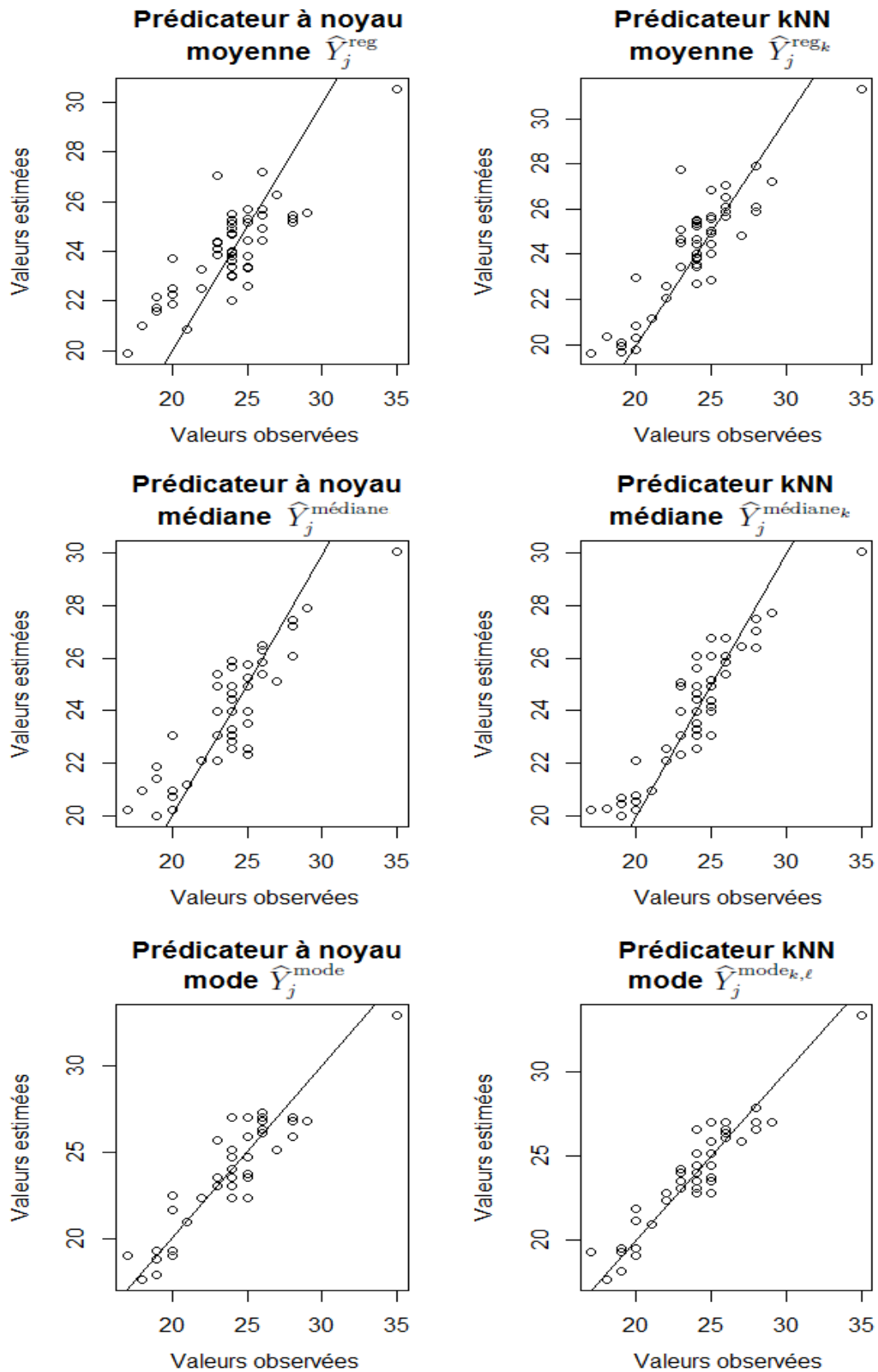
$$\widehat{Y}_j^{\text{reg}_k} = \widehat{m}_{k_{\text{opt}}}(x_j),$$

$$\widehat{Y}_j^{\text{médiane}_k} \text{ est telle que : } \widehat{F}_{k_{\text{opt}}}^{x_j}(\widehat{Y}_j^{\text{médiane}_k}) = 1/2,$$

et

$$\widehat{Y}_j^{\text{mode}_{k,\ell}} = \arg \max_{y \in \mathcal{S}} \widehat{f}_{(k,\ell)_{\text{opt}}}^{x_j}(y),$$

La figure 8.2 suivante donne les résultats des six prédictions effectuées.

FIGURE 8.2 – Prédications : méthode à noyau *vs* méthode *k*NN

8.3 Commentaires sur les résultats

Nous remarquons aux allures des nuages de points obtenus que la méthode k NN donne de meilleurs résultats que ceux obtenus par la méthode classique. Ceci est confirmé par le tableau 8.3 suivant qui regroupe les erreurs quadratiques moyennes

$$\text{MSE} = \frac{1}{50} \sum_{i=219}^{268} (Y_i - \tilde{Y}_i)^2$$

où \tilde{Y}_i est l'un des six prédicteurs traités.

	Mode	Médiane	Régression
Méthode à noyau	2.61	3.75	3.12
Méthode k NN	1.88	2.84	2.36

FIGURE 8.3 – Résultats MSE

Tout comme dans l'étude simulée du chapitre 7, nous remarquons que la méthode k NN est plus efficace que celle classique, puisque les erreurs de prédictions k NN sont largement inférieures (plus de 25% plus petites) à celles des prédictions à noyau. Nous concluons donc que nos procédures de choix automatique du nombre de voisins optimale traitées à la section 5.4 sont efficaces et faciles d'utilisation, et confirmons que le choix du paramètre de lissage détermine la qualité de l'estimateur dans le sens où un paramètre de lissage orienté par les données mène à des résultats plus attrayants que ceux obtenus par un paramètre de lissage global. Nous soulignons également le fait que dans ce genre d'études, où les données ne sont pas régulières, la prédiction par le mode ou par la médiane sont préférables à celle par la régression.

Conclusion et perspectives

Bilan de la thèse

Notre travail a commencé par fournir des résultats théoriques nouveaux sur différents estimateurs à noyaux classiques. Ces résultats prolongent les idées de l'uniformité sur la largeur de la fenêtre au contexte d'espaces statistiques fonctionnels (de dimensions finies ou infinies). En effet, nous avons réussi, au chapitre 4, à obtenir des propriétés de convergences uniformes sur la largeur de la fenêtre pour les estimateurs de chacun des quatre opérateurs conditionnels : régression, distribution conditionnelle, densité conditionnelle et hasard conditionnel, en présence de variables explicatives fonctionnelles. Dans chacun des cas les vitesses de convergence ont été mentionnées.

Ces résultats purement théoriques, bien que souvent utilisés en pratique, n'ont à notre connaissance jamais été établis dans le contexte d'espaces infini-dimensionnels. Plus encore, les seuls résultats d'uniformité sur le paramètre de lissage qui existent dans la littérature sont ceux traitant de l'estimateur de la régression sur des espaces de dimension finie. Ainsi, les propriétés que nous avons obtenues au chapitre 4 complètent les résultats asymptotiques existants traitant des convergence sur le paramètre de lissage. Quelques corollaires directs dans ce sens ont été énoncés dans la section 4.7 pour les estimateurs en utilisant des largeurs de fenêtres entièrement automatisées et orientées par les données.

Les résultats d'uniformité sur la largeur de la fenêtre ont un grand impact dans le domaine de l'appliqué. En effet, le bon comportement de toute technique de lissage non paramétrique est toujours lié, dans la pratique, à des paramètres de lissage orientés par les données. Le caractère local de ces paramètres fait qu'ils sont aléatoires, et complique donc considérablement les calculs asymptotiques. Nos résultats de convergence uniforme nous ont permis de traiter, au chapitre 5, d'estimateurs incluant des pa-

ramètres de lissage aléatoires tout en évitant les calculs fastidieux qu'ils engendrent d'habitude. Ainsi, nous avons obtenu des vitesses de convergence uniformes sur le nombre de voisins pour les estimateurs de type k NN de chacun des quatre opérateurs (voir Section 5.3). Cette étude nous a permis d'apporter une contribution à la connaissance des estimateurs k NN sur des régresseurs fonctionnels.

L'utilité principale de ce nouveau type de résultats asymptotiques est de permettre d'en dériver la théorie asymptotique pour les estimateurs k NN orientés par les données (voir Section 5.4). Une attention particulière est accordée, dans la Section 5.4, aux procédures de sélection basées sur des idées de validation croisée. Les résultats de la Section 5.4 sont, à notre connaissance, les premiers dans la littérature concernant des opérateurs autres que la régression. En outre, et sans doute plus important encore, ils sont les premiers dans l'étude k NN fonctionnelle.

Nous tenons à souligner qu'en dépit du caractère technique lié à l'uniformité asymptotique sur le nombre de voisins, les résultats mathématiques obtenus dans la Section 5.3 sont non seulement intéressants en eux-mêmes mais surtout pour leur impact directement applicable, puisque les conséquences dérivées dans la Section 5.4 rendent les approches k NN automatiquement utilisables en pratique. Ces faits sont appuyés par les simulations présentées dans le chapitre 7, suivies d'une étude pratique au chapitre 8.

Notre travail de recherche est présenté sous une forme plus ou moins générale permettant de traiter plusieurs problèmes d'estimation d'opérateurs différents (régression, densité, fonction de distribution cumulée, et les opérateurs de hasard), et pour une large gamme de sélecteurs de nombre de voisins orientés par les données. Il convient de souligner que, même si l'étude des données fonctionnelles était le but principal de ce travail, il a été traité de telle manière qu'il s'applique directement au cadre multivarié. Il suffit pour cela de prendre $\mathcal{F} = \mathbb{R}^p$. Dans une telle situation multivariée, tout comme les résultats d'uniformité sur le paramètre de lissage, les règles orientées par les données de la Section 5.4 sont des extensions de celles existantes dans la littérature.

Perspectives pour de futurs travaux

Les techniques développées dans cette thèse nous semblent très intéressantes, et nous espérons les utiliser dans d'autres contextes qui s'inscrivent dans la continuité de nos travaux. Nous présentons dans ce qui suit quelques questions ouvertes dans l'analyse des données fonctionnelles pour lesquelles les idées de l'uniformité sur le paramètre de lissage développées ici peuvent être intéressantes.

Modèle semi-paramétrique

Une première idée d'extension possible concerne des modèles liés à la réduction de dimension. Alors que notre travail se concentre sur des modèles fonctionnels entièrement non paramétriques, la littérature récente sur l'analyse des données fonctionnelles a mis en évidence l'intérêt des modèles semi-paramétriques du fait qu'ils soient intermédiaires entre les modèles flexibles (mais de très grande dimension) et ceux linéaires de faible dimension (mais trop restrictifs). Ce champ comprend par exemple des modèles fonctionnels à indice unique (voir [Goia et Vieu \[2015\]](#), pour les dernières avancées), des modèles de projection (voir [Chen et collab. \[2011\]](#)), ou des modèles partiellement linéaires (voir [Aneiros-Pérez et Vieu \[2006, 2008\]](#) ou [Lian \[2011\]](#)).

À notre connaissance, la grande majorité de la littérature proposée concernant ces modèles traite d'estimateurs à largeurs de fenêtres fixes et nous pensons que les idées d'uniformité sur la largeur de la fenêtre pourraient probablement être utilisées avec succès pour déduire des résultats asymptotiques sur des paramètres de lissage aléatoires, ouvrant la voie aux problèmes du choix de paramètres de lissage orientés par

les données et d'estimateurs de type k NN. En effet, l'étude des méthodes k NN orientée par les données dans la modélisation fonctionnelle semi-paramétrique est, à notre connaissance, un domaine totalement sous-développé.

Une autre façon de réduire la dimension dans la régression non paramétrique fonctionnelle consiste à réduire la courbe fonctionnelle à quelques points qui ont un impact sur la réponse. Le besoin d'un développement méthodologique dans ce sens a été mis en évidence dans divers travaux récents (voir par exemple la section 6 de l'étude bibliographique de Cuevas [2014]). Il n'y a pas beaucoup de travaux dans cette direction (voir cependant Aneiros et Vieu [2014, 2015], et les références qui y sont citées) et ceux existants sont tous basés sur des largeurs de fenêtres fixes. Les idées développées dans cette thèse pourraient également être intéressantes pour ces modèles épars.

Modèle local linéaire

Une deuxième extension que nous envisageons d'entreprendre à court terme, est de traiter des modélisations localement linéaires. Il s'avère que cette méthode d'estimation, introduite par Baïllo et Grané [2009] dans le contexte de données fonctionnelles, est asymptotiquement plus efficace que les méthodes à noyaux classiques. En effet, la modélisation locale linéaire prend en compte l'influence locale des observations et donne, de ce fait, un biais moins important que les méthodes à noyaux. Des études récentes ont utilisé cette méthode d'estimation dans plusieurs cadres différents, nous pouvons citer les nombreux travaux de Demongeot et collab. [2011, 2013, 2014 et 2016] qui analysent des estimateurs dans le cadre de distribution conditionnelle ou de variable réponse fonctionnelle, et les travaux de Wang et collab. [2015] qui traitent des données dépendantes.

À notre connaissance les techniques de pondération k NN n'ont jamais été traitées pour le modèle local linéaire, et nous pensons que les techniques d'uniformité sur la largeur de la fenêtre sont susceptibles de mener à des résultats dans ce sens.

Moments d'ordre

Dans la statistique inférentielle théorique, il est de coutume de traiter les termes "biais" et "variance" des estimateurs pour juger de leur efficacité sous certaines conditions.

Récemment, [Geenens \[2015\]](#) est allé chercher plus loin que ces deux moments usuels, et a réussi à donner l'expression exacte des moments de tout ordre et de tout type (simples, centrés, absolus) de l'estimateur à noyau classique de la régression en fonction de la largeur de la fenêtre et de la mesure de concentration. Ses résultats, obtenus sous des conditions moins restrictives que celles habituelles, lui ont permis de déduire la normalité asymptotique et les probabilités des grandes déviations de l'estimateur.

Nous pensons que l'extension de nos résultats en utilisant ces techniques pourrait aboutir à de bonnes propriétés asymptotiques. Et nous envisageons d'entamer nos travaux de recherches dans ce sens sous peu.

Modèle bi-fonctionnel

La littérature à propos des estimations non paramétriques de modèles conditionnels quand la variable réponse est fonctionnelle est très limitée. En effet le fait de chercher à prédire toute une trajectoire en se basant sur une donnée explicative (fini ou infini dimensionnelle) n'est pas une simple extension du contexte de la prédiction d'une variable réelle. Ce sujet a fait l'objet de peu d'avancées scientifiques, et ne traite principalement que de la régression classique. Ces modèles d'estimation, introduits par [Dabo-Niang et Rhomari \[2009\]](#), ont toutefois donné des résultats intéressants de convergence uniforme sur un compact de l'espace fonctionnel dans les travaux de [Ferraty et collab. \[2011\]](#). On pourrait penser à appliquer les mêmes méthodes utilisées dans nos travaux pour prolonger leurs résultats au cadre doublement uniforme et apporter ainsi une nouvelle contribution à l'analyse des modèles bi-fonctionnels.

En outre, si la sélection d'un paramètre de lissage orienté par les données est primordiale quand la variable explicative est de dimension infinie, le besoin de la sélection de tels paramètres est encore plus important lorsque la variable réponse est aussi de

nature fonctionnelle. Dans ce sens, on pourrait espérer que les mêmes méthodes de sélection d'un paramètre de lissage local développées ici puissent être étendues au cas bi-fonctionnel.

Données dépendantes

Enfin, un dernier type d'extension de nos idées concernerait des échantillons statistiques dépendants avec des applications possibles aux séries temporelles. La littérature traitant des estimateurs à noyaux fonctionnels dépendants a été assez développée (voir Masry [Masry \[2005\]](#) ou Ferraty et Vieu [Ferraty et Vieu \[2006\]](#)) mais toujours pour des largeurs de fenêtres fixes, et les idées utilisées dans ce document peuvent s'avérer intéressantes dans ce cadre également. Notons que cette extension devrait être plus difficile à obtenir que les précédentes, car il s'agirait de trouver un nouveau cadre de travail probabiliste, étant donné que celui utilisé ici est spécifique aux échantillons i.i.d.

En conclusion, nous soulignons le fait que ces idées de l'uniformité sur le paramètre de lissage devraient avoir un bel avenir dans l'analyse des données fonctionnelles, pour établir des résultats sur des paramètres de lissage orientés par les données. Nous espérons que notre travail pourrait contribuer à promouvoir de nouvelles avancées dans ce sens.

Annexes

A.1 Définitions mathématiques

1. *Convergence presque complète*

Soit $(z_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles.

- Nous disons que (z_n) converge presque-complètement (p.co.) vers zéro si pour tout $\epsilon > 0$:

$$\sum_{n=1}^{\infty} \mathbb{P}(|z_n| > \epsilon) < \infty .$$

- De plus, pour une suite déterministe de réels positifs $(u_n)_{n \in \mathbb{N}}$ qui tend vers zéro, s'il existe $\epsilon > 0$ tel que :

$$\sum_{n=1}^{\infty} \mathbb{P}(|z_n| > \epsilon u_n) < \infty ,$$

nous disons que la vitesse de convergence est d'ordre u_n , et nous notons

$$z_n = O_{a.co.}(u_n) .$$

Ce mode de convergence implique la convergence presque-sûre et celle en probabilité.

2. *Normes et semi-normes*

Soit E un espace quelconque.

- Une semi-norme définie sur E , est une application $\|\cdot\| : E \rightarrow \mathbb{R}^+$ telle que :

1. $\forall \lambda \in \mathbb{R} : \|\lambda x\| = \lambda \|x\|$;
2. $\forall x, y \in E : \|x + y\| \leq \|x\| + \|y\|$.

- Une norme définie sur E est une semi-norme qui vérifie la condition :

$$\|x\| = 0 \implies x = 0$$

- La (semi) métrique d associée à la (semi) norme $\|\cdot\|$ est l'application :

$$\begin{aligned} \|\cdot\| : E \times E &\longrightarrow \mathbb{R}^+ \\ (x, y) &\longmapsto d(x, y) = \|x - y\| \end{aligned}$$

3. Classe mesurable ponctuellement

Une classe de fonctions \mathcal{C} est dite mesurable ponctuellement s'il existe un sous-ensemble dénombrable \mathcal{C}_0 tel que, pour toute fonction $g \in \mathcal{C}$, il existe une suite de fonctions $(g_m)_{m \in \mathbb{N}}$ dans \mathcal{C}_0 telles que : $|g_m(z) - g(z)| = o(1)$.

4. Fonction enveloppe

Une fonction enveloppe G d'une classe de fonctions \mathcal{C} est une fonction mesurable telle que : $\sup_{g \in \mathcal{C}} |g(z)| \leq G(z)$, pour tout z .

5. Entropie

Soit (C, d) un ensemble d'un espace de fonctions réelles normé. Notons par $\mathcal{N}(\epsilon, C, d)$ le nombre minimal de boules ouvertes de rayon ϵ (par rapport à la mesure d) nécessaires pour couvrir l'ensemble C . Les centres des boules n'appartiennent pas forcément à C .

La quantité $\log(\mathcal{N}(\epsilon, C, d))$ est appelée la ϵ -entropie de Kolmogorov de l'ensemble C .

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

A.2 Lemmes utilisés

Lemme A.1. — (voir Théorème 2.14.1 dans [Van Der Vaart et Wellner, 1996, p.239]).

Soient Z_1, Z_2, \dots, Z_n des variables aléatoires indépendantes et identiquement distribuées sur \mathcal{F} , et considérons \mathcal{C} une classe mesurable ponctuellement de fonctions $g : \mathcal{F} \rightarrow \mathbb{R}$ avec une fonction enveloppe F , alors :

$$\|\alpha_n(g)\|_{\mathcal{C}} \leq CJ(1, \mathcal{C}) \|F\|_{p \vee 2}$$

où

$$\alpha_n(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (g(Z_i) - \mathbb{E}g(Z_i)) \text{ et } J(1, \mathcal{C}) = \sup_Q \int_0^1 \sqrt{1 + \log \mathcal{N}(\epsilon \|F\|_{Q,2}, \mathcal{C}, d_Q)} d\epsilon$$

avec $\|\cdot\|_p = \mathbb{E}^{1/p}[\cdot]^p$ et $(s \vee t)$ indiquant le maximum entre s et t .

Lemme A.2. — (voir Théorème 3.1 dans *Dony et collab. [2009]*).

Soit \mathcal{C} une classe mesurable ponctuellement de fonctions $g : \mathcal{F} \rightarrow \mathbb{R}$ vérifiant :

$$\mathbb{E}\|\alpha_n(g)\|_{\mathcal{C}} \leq C\|F\|_2$$

où F est une fonction enveloppe de \mathcal{C} . Alors, pour tout $A \in \mathcal{A}$, nous avons :

$$\mathbb{E}\|\alpha_n(g \cdot \mathbb{1}_A)\|_{\mathcal{C}} \leq 2C\|F \cdot \mathbb{1}_A\|_2$$

Lemme A.3. — (Inégalité de Bernstein voir Théorème 3.1 dans *Dony et collab. [2009]*)

Soient Z_1, Z_2, \dots, Z_n des variables aléatoires indépendantes et identiquement distribuées sur \mathcal{F} , et considérons \mathcal{C} une classe mesurable ponctuellement de fonctions $g : \mathcal{F} \rightarrow \mathbb{R}$ avec une fonction enveloppe F . Supposons que pour un $H > 0$:

$$\mathbb{E}[F^p(Z)] \leq \frac{p!}{2} \sigma^2 H^{p-2} \text{ où } \sigma^2 \geq \mathbb{E}[F^2(Z)].$$

Alors, pour $\beta_n = \mathbb{E}[\|\sqrt{n}\alpha_n(g)\|_{\mathcal{C}}]$, nous avons pour tout $t > 0$:

$$\mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\alpha_n(g)\|_{\mathcal{C}} \geq \beta_n + t \right\} \leq \exp \left(-\frac{t^2}{2n\sigma^2 + 2tH} \right)$$

Démonstrations de nos résultats

Dans toutes nos démonstrations, quand il n'y aura pas de confusion possible, nous noterons par C et C' toute constante générique positive. Nous noterons également pour $i = 1, \dots, n$:

$$K_i(x) = K(h_K^{-1}d(x, X_i)) \quad \text{et} \quad H_i(y) = H(h_H^{-1}(y - Y_i)).$$

B.1 Démonstrations relatives au chapitre 4

Nous donnons dans ce qui suit les preuves des lemmes et corolaires énoncés au chapitre 4. Nous commençons par donner à la sous-section B.1.1 les démonstrations détaillées des résultats de convergence sur le modèle de régression traités à la section 4.3. Afin d'éviter les répétitions fastidieuses, les résultats des sections 4.4, 4.5 et 4.6 seront démontrés de manière plus synthétique à la sous-section B.1.2.

B.1.1 Preuves de la Section 4.3

Preuve du lemme 4.2. : Nous devons montrer qu'il existe $\eta_0 > 0$ de telle sorte que :

$$\sum_n \mathbb{P} \left\{ \sup_{a_n \leq h_K \leq b_0} \sqrt{\frac{n\phi_x(a_n)}{\log n}} \left| \widehat{F}_D^x - \mathbb{E}[\widehat{F}_D^x] \right| \geq \eta_0 \right\} < \infty, \text{ pour un } b_0 > 0.$$

En suivant les idées de [Einmahl et collab. \[2005\]](#), notre démonstration se base en grande partie sur l'inégalité de Bernstein pour des processus empiriques établis dans le lemme A.3. En prenant :

$$h_{K,j} = 2^j a_n \text{ et } L(n) = \max\{j : h_{K,j} \leq 2b_0\},$$

Il est clair que :

$$b_0 \leq h_{K,L(n)}$$

Ainsi,

$$[a_n, b_0] \subset \bigcup_{1 \leq j \leq L(n)} [h_{K,j-1}, h_{K,j}]$$

d'où,

$$\sup_{a_n \leq h_K \leq b_0} \sqrt{\frac{n\phi_x(a_n)}{\log n}} \left| \widehat{F}_D^x - \mathbb{E}[\widehat{F}_D^x] \right| \leq \max_{1 \leq j \leq L(n)} \sup_{h_{K,j-1} \leq h_K \leq h_{K,j}} \sqrt{\frac{n\phi_x(h_K)}{\log n}} \left| \widehat{F}_D^x - \mathbb{E}[\widehat{F}_D^x] \right|.$$

Nous écrivons ensuite la différence :

$$\widehat{F}_D^x - \mathbb{E}[\widehat{F}_D^x] := \frac{1}{\sqrt{n}\phi_x(h_K)} \alpha_n(K),$$

où $\alpha_n(K) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (K_i - \mathbb{E}[K_i])$ correspond au processus empirique basé sur les variables X_1, X_2, \dots, X_n . Nous considérons ensuite la classe de fonctions suivante :

$$\mathcal{G}_{K,j} = \{z \mapsto K(\gamma^{-1}d(x, z)) \text{ où } h_{K,j-1} \leq \gamma \leq h_{K,j}\}.$$

Ainsi,

$$\begin{aligned} & \mathbb{P} \left\{ \sup_{a_n \leq h_K \leq b_0} \sqrt{\frac{n\phi_x(a_n)}{\log n}} \left| \widehat{F}_D^x - \mathbb{E}[\widehat{F}_D^x] \right| \geq \eta_0 \right\} \\ & \leq \sum_{j=1}^{L(n)} \mathbb{P} \left\{ \frac{1}{\sqrt{n\phi_x(h_{K,j}/2) \log n}} \|\sqrt{n}\alpha_n(K)\|_{\mathcal{G}_{K,j}} \geq \eta_0 \right\} \\ & \leq L(n) \max_{j=1, \dots, L(n)} \mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k}\alpha_k(K)\|_{\mathcal{G}_{K,j}} \geq \eta_0 \sqrt{n\phi_x(h_{K,j}/2) \log n} \right\}. \end{aligned}$$

Nous devons alors étudier :

$$\mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k}\alpha_k(K)\|_{\mathcal{G}_{K,j}} \geq \eta_0 \sqrt{n\phi_x(h_{K,j}/2) \log n} \right\}.$$

Pour cela, on applique le lemme A.3 de l'inégalité de Bernstein. On pose alors :

$$\beta_n = \mathbb{E} [\|\sqrt{n}\alpha_n(K)\|_{\mathcal{G}_{K,j}}] \quad \text{et} \quad \sigma^2 = \mathbb{E} [G_{K,j}^2(X)]$$

où $G_{K,j}$ est la fonction enveloppe de la classe $\mathcal{G}_{K,j}$. En utilisant (4.10) il s'en suit que :

$$G_{K,j}(z) \leq C \mathbb{1}_{B(x, h_{K,j}/2)}(z).$$

Par conséquent, pour tout $p \geq 1$, et sous la condition (4.5), nous avons :

$$\mathbb{E} [G_{K,j}^p(X)] \leq C^p \phi_x(h_{K,j}/2). \quad (\text{B.1})$$

En particulier :

$$\sigma^2 = O(\phi_x(h_{K,j}/2)).$$

Concernant le terme β_n , en combinant (4.11) et (4.12) avec le lemme A.1. Nous obtenons :

$$\mathbb{E}[\|\alpha_n(g)\|_{\mathcal{G}_{K,j}}] \leq CJ(1, \mathcal{K})\|F\|_2$$

où F est la fonction enveloppe de \mathcal{K} .

Les conditions du lemme A.2 étant vérifiées pour la classe $\mathcal{G}_{K,j}$, nous avons :

$$\mathbb{E}\|\alpha_n(g \cdot \mathbb{1}_{B(x, h_{K,j}/2)})\|_{\mathcal{G}_{K,j}} \leq 2CJ(1, \mathcal{K})\|F \cdot \mathbb{1}_{B(x, h_{K,j}/2)}\|_2.$$

Ce qui donne finalement :

$$\beta_n = \mathbb{E} [\|\sqrt{n}\alpha_n(K)\|_{\mathcal{G}_{K,j}}] \leq C\sqrt{n\phi_x(h_{K,j}/2)}.$$

Nous pouvons à présent appliquer l'inégalité de Bernstein avec

$$\beta_n = O\left(\sqrt{n\phi_x(h_{K,j}/2)}\right), \quad \sigma^2 = O(\phi_x(h_{K,j}/2)), \quad D = C \quad \text{et} \quad t = \eta_0/2\sqrt{n\phi_x(h_{K,j}/2)\log n},$$

et nous obtenons :

$$\begin{aligned} & \mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k}\alpha_k(K)\|_{\mathcal{G}_{K,j}} > \eta_0\sqrt{n\phi_x(h_{K,j}/2)\log n} \right\} \\ & \leq \mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k}\alpha_k(K)\|_{\mathcal{G}_{K,j}} > \beta_n + t \right\} \\ & \leq \exp \left\{ -\eta_0^2 \frac{\log n}{8 + C\sqrt{\frac{\log n}{n\phi_x(h_{K,j}/2)}}} \right\} \\ & \leq n^{-C'\eta_0^2}. \end{aligned}$$

La dernière inégalité est une conséquence de (4.13).

Et puisque $L(n) \leq 2 \log n$, nous avons :

$$L(n) \max_{j=1, \dots, L(n)} \mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k} \alpha_k(K)\|_{\mathcal{G}_{K,j}} > \eta_0 \sqrt{n \phi_x(h_{K,j}/2) \log n} \right\} \leq C(\log n) n^{-C' \eta_0^2}.$$

Il suffit maintenant de choisir η_0 telle que $C' \eta_0^2 > 1$, pour obtenir :

$$n \sup_{a_n \leq h_K \leq b_n} \left| \widehat{F}_D^x - \mathbb{E}[\widehat{F}_D^x] \right| = O_{p.co.} \left(\sqrt{\frac{\log n}{n \phi(a_n)}} \right).$$

□

Preuve du corolaire 4.3. : Par de simples arguments analytiques, nous considérons à partir de (4.5), (4.6) et (4.10)

$$\mathbb{E} \left[\frac{1}{n \phi_x(h_K)} \sum_{i=1}^n K(h_K^{-1} d(x, X_i)) \right] \rightarrow K(1/2) - \int_0^{1/2} K'(s) \tau_x(s) ds > 0.$$

Ainsi, pour n assez grand, il existe une constante $C' > 0$, telle que :

$$\mathbb{E}[\widehat{F}_D^x] \geq C' \text{ pour tout } h_K \in (a_n, b_n).$$

d'où,

$$\sup_{h_K \in (a_n, b_n)} \left| \mathbb{E}[\widehat{F}_D^x] - \widehat{F}_D^x \right| \geq \frac{C'}{2}.$$

En choisissant $C = C'/2$, nous obtenons :

$$\mathbb{P} \left(\inf_{h_K \in (a_n, b_n)} \widehat{F}_D^x \leq C \right) \leq \mathbb{P} \left(\sup_{h_K \in (a_n, b_n)} \left| \mathbb{E}[\widehat{F}_D^x] - \widehat{F}_D^x \right| \geq C \right),$$

et le lemme 4.2 mène au résultat désiré. □

Preuve du lemme 4.4. : Notons que :

$$\begin{aligned} \widehat{B}(x) = \mathbb{E} \left[\widehat{F}_D^x \right] &= \mathbb{E} [\widehat{g}(x)] - m(x) \mathbb{E} \left[\widehat{F}_D^x \right] = \frac{1}{\phi_x(h_K)} \mathbb{E} [K_1(x) [\mathbb{E}(Y|X) - m(x)]] \\ &= \frac{1}{\phi_x(h_K)} \mathbb{E} [K_1(x) [m(X) - m(x)]] . \end{aligned}$$

Alors, les conditions (5.2) et (4.10) permettent de déduire que :

$$\mathbb{1}_{B(x, h_K/2)} |m(X) - m(x)| \leq C(h_K/2)^b.$$

Maintenant, en utilisant (4.6), nous pouvons écrire que :

$$\left| \widehat{B}(x) \right| \leq C(h_K)^b.$$

Finalement, nous obtenons :

$$\sup_{a_n \leq h_K \leq b_n} \left| \widehat{B}(x) \right| \leq Cb_n^b.$$

□

Preuve du lemme 4.5. : Comme dans la preuve du lemme 4.2, il suffit de montrer qu'il existe $\eta'_0 > 0$ tel que pour un $b_0 > 0$:

$$\sum_n \mathbb{P} \left\{ \sup_{a_n \leq h_K \leq b_0} \sqrt{\frac{n\phi_x(a_n)}{\log n}} |\widehat{g}(x) - \mathbb{E}[\widehat{g}(x)]| \geq \eta'_0 \right\} < \infty.$$

Nous utilisons les mêmes notations que pour la démonstration du lemme 4.2, et nous considérons le processus empirique :

$$\alpha'_n(K) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i K_i - \mathbb{E}[Y_i K_i]).$$

définissons ensuite la classe de fonction :

$$\mathcal{G}'_{K,j} = \left\{ (z, y) \mapsto yK(\gamma^{-1}d(x, z)) \text{ pour } h_{K,j-1} \leq \gamma \leq h_{K,j} \right\},$$

qui admet une fonction enveloppe $F_{K,j}(\cdot, \cdot)$ telle que :

$$F_{K,j}(z, y) \leq Cy \mathbb{1}_B(x, h_{K,j}/2)(z).$$

En utilisant (4.5), (4.9), (4.10) et (4.13) nous obtenons :

$$\mathbb{E} [F_{K,j}^p(X, Y)] \leq C^p \phi_x(h_{K,j}/2) \quad \text{et} \quad \mathbb{E} [F^2(X, Y)] = O(\phi_x(h_{K,j}/2)) \quad (\text{B.2})$$

Maintenant en utilisant les idées que celles de la preuve du lemme 4.2, nous avons :

$$\beta'_n = \mathbb{E} \left[\left\| \sqrt{n} \alpha'_n(K) \right\|_{\mathcal{G}'_{K,j}} \right] = O \left(\sqrt{n \phi_x(h_{K,j}/2)} \right),$$

et en appliquant l'inégalité de Bernstein au processus empirique $\alpha'_n(K)$, nous obtenons :

$$\begin{aligned} & \mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k} \alpha'_k(K)\|_{\mathcal{G}'_{K,j}} > \eta'_0 \sqrt{n \phi_x(h_{K,j}/2) \log n} \right\} \\ & \leq \mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k} \alpha_k(K)\|_{\mathcal{G}'_{K,j}} > \beta'_n + t \right\} \\ & \leq n^{-C' \eta_0'^2}. \end{aligned}$$

Ce qui donne :

$$\begin{aligned} & \mathbb{P} \left\{ \sup_{a_n \leq h_K \leq b_0} \sqrt{\frac{n \phi_x(h_K)}{\log n}} |\widehat{g}(x) - \mathbb{E}[\widehat{g}(x)]| \geq \eta'_0 \right\} \\ & \leq L(n) \mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k} \alpha'_k(K)\|_{\mathcal{G}'_{K,j}} > \eta'_0 \sqrt{n \phi_x(h_{K,j}/2) \log n} \right\} \\ & \leq \log(n) n^{-C' \eta_0'^2}, \end{aligned}$$

et un choix approprié de η'_0 mène à :

$$\sup_{a_n \leq h_K \leq b_n} |\widehat{g}(x) - \mathbb{E}[\widehat{g}(x)]| = O_{p.co.} \left(\sqrt{\frac{\log n}{n \phi(a_n)}} \right).$$

□

B.1.2 Preuves des Sections 4.4, 4.5 et 4.6.

Preuve du lemme 4.7. : De la même façon que pour la preuve du lemme 4.4, nous avons :

$$\begin{aligned} \widetilde{B}_n(x, y) &= \frac{\mathbb{E}[\widehat{F}_N^x(y)] - F^x(y) \mathbb{E}[F_D^x]}{\mathbb{E}[F_D^x]} \\ &= \frac{1}{\phi_x(h_K) \mathbb{E}[F_D^x]} \mathbb{E}[K_1(x) [\mathbb{E}[H_1(y)|X] - F^x(y)]]. \end{aligned}$$

Nous obtenons ainsi :

$$|\mathbb{E}[H_1(y)|X] - F^x(y)| \leq \int_0^1 K(t) |F^X(y - h_H t) - F^x(y)| dt.$$

En utilisant maintenant (4.10) et (4.15), nous obtenons :

$$\widetilde{B}(x, y) \leq C \left(h_K^{\beta_1} + h_H^{\beta_2} \right), \quad (\text{B.3})$$

Ce qui implique finalement que

$$\sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} \left| \tilde{B}(x, y) \right| \leq C (b_n^{\beta_1} + d_n^{\beta_2}).$$

□

Preuve du lemme 4.8. : La preuve de ce lemme est similaire à celle du lemme 4.2. Elle va donc être présentée brièvement et les mêmes notations vont être adoptées. Plus précisément, cette démonstration est basée sur l'application de l'inégalité de Bernstein au processus empirique $\alpha_n(K, H)$, défini par :

$$\alpha_n(K, H) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (K_i H_i - \mathbb{E}[K_i H_i]),$$

pour lequel nous pouvons écrire :

$$\widehat{F}_N^x(y) - \mathbb{E} \left[\widehat{F}_N^x(y) \right] = \frac{1}{\sqrt{n} \phi_x(h_K)} \alpha_n(K, H).$$

Comme pour le lemme 4.2, nous définissons la classe de fonctions suivante :

$$\mathcal{G}_{K,H,j} = \left\{ (z, t) \mapsto K(\gamma^{-1}d(x, z)) H(\delta^{-1}(y - t)) \text{ pour } h_{k,j-1} \leq \gamma \leq h_{k,j} \text{ et } c_n \leq \delta \leq d_n \right\}.$$

D'après (4.11) et (4.16), $\mathcal{G}_{K,H,j}$ est une classe de fonctions mesurable ponctuellement. De plus, en utilisant le fait que $H < 1$, nous avons que l'enveloppe $G_{K,H,j}$ de $\mathcal{G}_{K,H,j}$ est telle que :

$$G_{K,H,j}(z, t) \leq C \mathbb{I}_B(x, h_{K,j}/2)(z).$$

Ce qui implique que :

$$\sigma''^2 := \mathbb{E} [G_{K,H,j}^2(X, Y)] = O(\phi_x(h_{K,j}/2)) \text{ et } \mathbb{E} [G_{K,j}^p(X)] \leq C^p \sigma''^2.$$

Maintenant en combinant les lemmes A.1 et A.2 avec (4.16) et (4.17), nous obtenons :

$$\beta_n'' := \mathbb{E} \left[\|\sqrt{n} \alpha_n(K, H)\|_{\mathcal{G}_{K,H,j}} \right] = O(\sqrt{n \phi_x(h_{K,j}/2)}).$$

Ainsi, pour $\epsilon_0 > 0$ et $t = \epsilon_0/2 \sqrt{n \phi_x(h_{K,j}/2) \log n}$:

$$\begin{aligned} & \mathbb{P} \left\{ \sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} \sqrt{\frac{n \phi_x(h_K)}{\log n}} \left| \widehat{F}_N^x(y) - \mathbb{E} \left[\widehat{F}_N^x(y) \right] \right| \geq \epsilon_0 \right\} \\ & \leq L(n) \mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k} \alpha_k(K, H)\|_{\mathcal{G}_{K,H,j}} > \epsilon_0 \sqrt{n \phi_x(h_{K,j}/2) \log n} \right\} \\ & \leq L(n) \mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k} \alpha_k(K, H)\|_{\mathcal{G}_{K,H,j}} > \beta_n'' + t \right\} \\ & \leq \log(n) n^{-C' \epsilon_0^2}. \end{aligned}$$

Ainsi, il existe un $\epsilon_0 > 0$ tel que :

$$\sum_n \mathbb{P} \left\{ \sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} \sqrt{\frac{n\phi_x(a_n)}{\log n}} \left| \widehat{F}_N^x(y) - \mathbb{E} \left[\widehat{F}_N^x(y) \right] \right| \geq \epsilon_0 \right\} < \infty.$$

Ce qui termine la preuve de ce lemme. \square

Preuve du lemme 4.10. : La condition (4.20) implique que :

$$|\mathbb{E} [H'_1(y)|X] - f^x(y)| \leq C_x \int_0^{1/2} K(t) \left(h_K^{\beta_1} + |t|^{\beta_2} h_H^{\beta_2} \right) dt. \quad (\text{B.4})$$

Ainsi en utilisant (4.5) et (4.6), nous obtenons :

$$\begin{aligned} \ddot{B}(x, y) &= \frac{\mathbb{E} \left[\widehat{f}_N^x(y) \right] - f^x(y) \mathbb{E} [F_D^x]}{\mathbb{E} [F_D^x]} \\ &= \frac{1}{\phi_x(h_K) \mathbb{E} [F_D^x]} \mathbb{E} [K_1(x) [\mathbb{E} [H'_1(y)|X] - f^x(y)]] \\ &\leq C \left(h_K^{\beta_1} + h_H^{\beta_2} \right) \end{aligned}$$

Ce qui implique que :

$$\sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} |\ddot{B}(x, y)| \leq C \left(b_n^{\beta_1} + d_n^{\beta_2} \right),$$

\square

Preuve du lemme 4.11. : Cette démonstration suit les mêmes étapes que celles du lemme 4.2 et va donc être présentée d'une façon plus synthétique. En utilisant les mêmes notations, nous pouvons écrire :

$$\widehat{f}_N^x(y) - \mathbb{E} [\widehat{f}_N^x(y)] = \frac{1}{\sqrt{n} h_H \phi_x(h_K)} \alpha_n(K, H')$$

où

$$\alpha_n(K, H') = \frac{1}{\sqrt{n}} \sum_{i=1}^n (K_i H'_i - \mathbb{E} [K_i H'_i])$$

est le processus empirique basé sur les couples d'observations $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$.

Pour $h_{H,k} = 2^k c_n$, nous considérons la classe de fonctions suivante :

$$\mathcal{G}_{K,k,j} = \left\{ (z, t) \mapsto K(\gamma^{-1}d(x, z)) K(\delta^{-1}(y - t)) \text{ pour } h_{K,j-1} \leq a \leq h_{K,j} \text{ et } h_{H,k-1} \leq \delta \leq h_{H,k} \right\}.$$

La fonction enveloppe de $\mathcal{G}_{K,k,j}$ est notée par $G_{K,k,j}$ et vérifie la condition suivante :

$$G(z, t) \leq C \mathbb{I}_{B(x, h_{K,j}/2)}(z) \mathbb{I}_{(y-h_{H,k}/2, y)}(t).$$

Sous les hypothèses (4.5), (4.9) et (4.10) nous obtenons :

$$\sigma'''^2 := \mathbb{E} [G_{K,H,j}^2(X, Y)] = O(h_{H,j} \phi_x(h_{K,j}/2)), \quad \mathbb{E} [G_{K,j}^p(X)] \leq C^p \sigma'''^2$$

et

$$\beta_n''' = \mathbb{E} [\|\sqrt{n} \alpha_n(K, H)\|_{\mathcal{G}_{K,H,j}}] = O\left(\sqrt{nh_{H,j} \phi_x(h_{K,j}/2)}\right).$$

Ainsi, en appliquant l'inégalité de Bernstein au processus empirique $\alpha_n(K, H')$ avec :

$$t''' = \epsilon'_0/2 \sqrt{nh_{H,k} \phi_x(h_{K,j}/2) \log n} \text{ pour un } \epsilon'_0 > 0,$$

nous obtenons que :

$$\begin{aligned} \mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k} \alpha_k(K, H')\|_{\mathcal{G}_{K,H,j}} > \epsilon'_0 \sqrt{n \phi_x(h_{K,j}/2) \log n} \right\} \\ \leq \mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k} \alpha_k(K, H')\|_{\mathcal{G}_{K,H,j}} > \beta_n''' + t''' \right\} \\ \leq n^{-C' \epsilon_0'^2}. \end{aligned}$$

Puisque $b_n \rightarrow 0$ et $d_n \rightarrow 0$, il existe des constantes positives b_0 et d_0 telles que :

$$\sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} |\widehat{f}_N^x(y) - E \widehat{f}_N^x(y)| \leq \sup_{a_n \leq h_K \leq b_0} \sup_{c_n \leq h_H \leq d_0} |\widehat{f}_N^x(y) - E \widehat{f}_N^x(y)|.$$

Soient $L(K, n) = \max\{j : h_{K,j} \leq 2b_0\}$ et $L(H', n) = \max\{j : h_{H,k,j} \leq 2d_0\}$. Alors, sous les conditions (4.13) et (4.21) nous avons :

$$L(K, n) \leq 2 \log n \quad \text{et} \quad L(H', n) \leq 2 \log n.$$

Nous obtenons finalement :

$$\begin{aligned} \mathbb{P} \left\{ \sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} \sqrt{\frac{nc_n \phi_x(a_n)}{\log n}} \left| \widehat{f}_N^x(y) - \mathbb{E} [\widehat{f}_N^x(y)] \right| \geq \epsilon'_0 \right\} \\ \leq \mathbb{P} \left\{ \sup_{a_n \leq h_K \leq b_n} \sup_{c_n \leq h_H \leq d_n} \sqrt{\frac{nh_H \phi_x(h_K)}{\log n}} \left| \widehat{f}_N^x(y) - \mathbb{E} [\widehat{f}_N^x(y)] \right| \geq \epsilon_0 \right\} \\ \leq L(K, n) L(H', n) \mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k} \alpha_k(K, H')\|_{\mathcal{G}_{K,H,j}} > \epsilon_0 \sqrt{nh_{H,k} \phi_x(h_{K,j}/2) \log n} \right\} \\ \leq n^{-C' \epsilon_0'^2} \log^2 n, \end{aligned}$$

ce qui achève la preuve du lemme 4.11. □

Preuve du corolaire 4.13. : Notons que pour tout $h_K \in (a_n, b_n)$ et $h_H \in (c_n, d_n)$, nous avons :

$$|1 - \widehat{F}^x(y)| \leq \frac{(1 - F^x(y))}{2} \text{ implique que } |\widehat{F}^x(y) - F^x(y)| \geq \frac{(1 - F^x(y))}{2}.$$

Ainsi :

$$\begin{aligned} & \sum_n \mathbb{P} \left(\inf_{h_K \in (a_n, b_n), h_H \in (c_n, d_n)} |1 - \widehat{F}^x(y)| \leq \frac{1 - F^x(y)}{2} \right) \\ & \leq \sum_n \mathbb{P} \left(\sup_{h_K \in (a_n, b_n), h_H \in (c_n, d_n)} |\widehat{F}^x(y) - F^x(y)| \geq \frac{1 - F^x(y)}{2} \right) < \infty. \end{aligned}$$

Le résultat recherché est obtenu maintenant en appliquant le théorème 4.6. \square

B.2 Démonstrations relatives au chapitre 5

B.2.1 Preuves de la section 5.3

Pour des raisons de concision, nous n'allons détailler que la démonstration concernant l'estimateur de la fonction de répartition conditionnelle. Les démonstrations des autres estimateurs suivent les mêmes lignes, et sont donc présentées sous une façon plus synthétique.

Preuve du théorème 5.1. : Commençons par montrer la première assertion de (5.1). Notons pour cela :

$$z_n = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^{\beta_1} + \sqrt{\frac{\ln n}{k_{1,n}}} \right\}$$

et écrivons pour $\alpha \in (0, 1)$,

$$\begin{aligned} \sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{F}^x(y) - F^x(y)| &= \sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{F}^x(y) - F^x(y)| \mathbb{1} \left\{ \phi_x^{-1} \left(\alpha \frac{k_{1,n}}{n} \right) \leq H_{k,x} \leq \phi_x^{-1} \left(\frac{k_{2,n}}{\alpha n} \right) \right\} \\ &+ \sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{F}^x(y) - F^x(y)| \mathbb{1} \left\{ H_{k,x} \notin \left(\phi_x^{-1} \left(\frac{\alpha k_{1,n}}{n} \right), \phi_x^{-1} \left(\frac{k_{2,n}}{\alpha n} \right) \right) \right\}. \end{aligned}$$

Ainsi, pour tout $\epsilon > 0$, nous avons :

$$\begin{aligned} & \mathbb{P} \left\{ \sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{F}^x(y) - F^x(y)| \geq \epsilon z_n \right\} \\ & \leq \mathbb{P} \left\{ \sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{F}^x(y) - F^x(y)| \mathbb{1}_{\left\{ \phi_x^{-1}\left(\frac{\alpha k_{1,n}}{n}\right) \leq H_{k,x} \leq \phi_x^{-1}\left(\frac{k_{2,n}}{n\alpha}\right) \right\}} \geq \frac{\epsilon z_n}{2} \right\} \\ & \quad + \mathbb{P} \left\{ H_{k,x} \notin \left(\phi_x^{-1}\left(\frac{\alpha k_{1,n}}{n}\right), \phi_x^{-1}\left(\frac{k_{2,n}}{n\alpha}\right) \right) \right\}. \end{aligned}$$

Donc pour montrer que

$$\sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{F}^x(y) - F^x(y)| = O_{p.co.}(z_n),$$

Il suffit de prouver les trois résultats suivants :

$$\sum_n \sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P} \left\{ H_{k,x} \leq \phi_x^{-1}\left(\frac{\alpha k_{1,n}}{n}\right) \right\} < \infty, \quad (\text{B.5})$$

$$\sum_n \sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P} \left\{ H_{k,x} \geq \phi_x^{-1}\left(\frac{k_{2,n}}{n\alpha}\right) \right\} < \infty, \quad (\text{B.6})$$

$$\sup_{\phi_x^{-1}\left(\frac{k_{1,n}}{n}\right) \leq h \leq \phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)} |\widetilde{F}^x(y) - F^x(y)| = O_{p.co.}(z_n), \quad (\text{B.7})$$

où

$$\widetilde{F}^x(y) = \frac{\sum_{i=1}^n K\{h^{-1}d(x, X_i)\} \mathbb{1}_{\{Y_i \leq y\}}}{\sum_{i=1}^n K\{h^{-1}d(x, X_i)\}}.$$

La preuve de (B.7) n'est pas présentée ici car elle suit, pas à pas, les mêmes étapes que le preuve du théorème 4.1. Les assertions (B.5) et (B.6) requièrent beaucoup plus d'attention. Elles seront donc présentées avec beaucoup plus de détails.

L'outil clé pour nos démonstrations, est la version de l'inégalité de Chernoff suivante, dont la preuve ser donnée à la toute fin de l'annexe.

Lemme B.1. — Soient U_1, \dots, U_n des variables aléatoires de Bernoulli indépendantes avec $\mathbb{P}(U_i = 1) = p$ pour tout $i \in \{1, \dots, n\}$. Posons $U = X_1 + \dots + X_n$ et $\mu = pn$. Alors, pour tout $\omega > 0$, nous avons :

$$\mathbb{P} \{U \geq (1 + \omega)\mu\} \leq \exp \{-\mu \min(\omega^2, \omega) / 4\} \quad (\text{B.8})$$

et si $\omega \in (0, 1)$, nous avons :

$$\mathbb{P} \{U \leq (1 - \omega)\mu\} \leq \exp \{-\mu (\omega^2/2)\}. \quad (\text{B.9})$$

Ainsi, en utilisant le lemme B.1, nous pouvons écrire :

$$\begin{aligned} \mathbb{P} \left\{ H_{k,x} \leq \phi_x^{-1} \left(\frac{\alpha k_{1,n}}{n} \right) \right\} &= \mathbb{P} \left\{ \sum_{i=1}^n \mathbb{1}_{B(x, \phi_x^{-1}(\frac{\alpha k_{1,n}}{n}))} > k \right\} \\ &= \mathbb{P} \left\{ \sum_{i=1}^n \mathbb{1}_{B(x, \phi_x^{-1}(\frac{\alpha k_{1,n}}{n}))} > \frac{k}{\alpha k_{1,n}} \alpha k_{1,n} \right\} \\ &\leq \exp \{-(k - \alpha k_{1,n})/4\}. \end{aligned}$$

Ce qui implique que :

$$\sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P} \left\{ H_{k,x} \leq \phi_x^{-1} \left(\frac{\alpha k_{1,n}}{n} \right) \right\} \leq k_{2,n} \exp \{-(1 - \alpha)k_{1,n}/4\} \leq n^{1 - \{(1-\alpha)/4\} \frac{k_{1,n}}{\ln n}}.$$

D'une manière similaire, nous avons :

$$\mathbb{P} \left\{ H_{k,x} \geq \phi_x^{-1} \left(\frac{k_{2,n}}{\alpha n} \right) \right\} \leq \exp \left\{ -\frac{(k_{2,n} - \alpha k)^2}{2\alpha k_{2,n}} \right\}.$$

Il s'en suit que

$$\begin{aligned} \sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P} \left\{ H_{k,x} \geq \phi_x^{-1} \left(\frac{k_{2,n}}{\alpha n} \right) \right\} &\leq k_{2,n} \exp \{-(1 - \alpha)k_{1,n}/2\alpha\} \\ &\leq n^{1 - \{(1-\alpha)/2\alpha\} \frac{k_{2,n}}{\ln n}}. \end{aligned}$$

Et comme $k_{1,n}/\ln n \rightarrow \infty$ nous obtenons finalement (B.5) et (B.6). La démonstration de (5.1) est à présent finie.

Les autres assertions du théorème 5.1 peuvent être démontrées de la même façon. \square

B.2.2 Preuves de la section 5.4

Preuve du corollaire 5.3. : Soit $(\tilde{k}, \tilde{\ell})$ un couple aléatoire à valeurs dans $(k_1, k_2) \times (\ell_1, \ell_2)$, et notons par \hat{F}^x , l'estimateur défini dans (5.4), mais en utilisant les nombres

de voisins aléatoires \tilde{k} et $\tilde{\ell}$. Comme une conséquence directe des résultats d'uniformité sur le nombre de voisins de (5.1), nous avons :

$$|\widehat{F}^x(y) - F^x(y)| = O \left\{ \phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^{\beta_1} \right\} + O_{p.co.} \left(\sqrt{\frac{\ln n}{k_{1,n}}} \right). \quad (\text{B.10})$$

Il suffit de remarquer que le corolaire 5.3 est un cas particulier de (B.10) quand le couple $(\tilde{k}, \tilde{\ell})$ est le paramètre de validation croisée obtenu en minimisant le critère CV_{GG} . \square

Preuve des corolaires 5.2, 5.4 et 5.5. : De la même façon, l'on peut obtenir des résultats du même genre que ceux de (B.10) pour d'autres estimateurs k NN impliquant des nombres de voisins aléatoires, dont les corolaires 5.2, 5.4 et 5.5 suivent directement. \square

Preuve des inégalités de type Chernoff dans le lemme B.1. : Soit $0 \leq \tau \leq 1 - p$. Par l'inégalité de Markov nous obtenons, pour tout $\zeta > 0$,

$$\mathbb{P}\{U \geq n(p + \tau)\} \leq \frac{\mathbb{E}(e^{\zeta U})}{e^{\zeta n(p + \tau)}} \leq \frac{(pe^{\zeta} + 1 - p)^n}{e^{\zeta n(p + \tau)}}.$$

En prenant $\zeta = \ln\{(1 - p)(p + \tau)/(p(1 - p - \tau))\}$, nous arrivons à la forme la plus usuelle de l'inégalité de Chernoff.

$$\forall \tau \in [0, 1 - p] \quad \mathbb{P}\{U \geq n(p + \tau)\} \leq \left(\frac{p}{p + \tau} \right)^{n(p + \tau)} \left(\frac{1 - p}{1 - p - \tau} \right)^{n(1 - p - \tau)}. \quad (\text{B.11})$$

Cette inégalité est le point crucial pour prouver les deux résultats (B.8) et (B.9) établis au lemme B.1. \square

Preuve de (B.8). : Soit $\omega > 0$. Le fait d'appliquer le résultat (B.11) avec $\tau = p\omega$ donne :

$$\mathbb{P}\{U \geq \mu(1 + \omega)\} \leq \left(\frac{1}{1 + \omega} \right)^{n\mu(1 + \omega)} \left(\frac{1 - p}{1 - p - p\omega} \right)^{n(1 - p - p\omega)}. \quad (\text{B.12})$$

Par une simple expansion de Taylor d'ordre deux, nous pouvons aussi montrer que :

$$\ln \left\{ (1 + \omega)^{\mu(1 + \omega)} \left(\frac{1 - p - p\omega}{1 - p} \right)^{1 - p - p\omega} \right\} \geq \frac{p}{4} \min(\omega^2, \omega). \quad (\text{B.13})$$

Enfin, le résultat (B.8) recherché, vient directement de (B.12) et (B.13). \square

Preuve de (B.9). . . En appliquant le résultat (B.11) aux variables $V_i = -U_i$ et en prenant $\tau = p\omega$, nous avons directement que

$$\mathbb{P}\{U \leq (1 - \omega)\mu\} \leq \frac{e^{-\mu\omega}}{(1 - \omega)^{\mu(1-\omega)}}. \quad (\text{B.14})$$

Rappelons que pour avoir l'assertion (B.9), nous supposons que $0 < \omega < 1$. Par de simples arguments analytiques nous pouvons voir que

$$\ln(1 - \omega) \geq \frac{\omega^2/2 - \omega}{1 - \omega}. \quad (\text{B.15})$$

Enfin, the le résultat (B.9) recherché vient directement de(B.14) et (B.15). \square

Références bibliographiques

- Aneiros G. et Vieu P. (2014). Variable selection in infinite-dimensional problems. *Statistics & Probability Letters*, 94 :12–20.
- Aneiros G. et Vieu P. (2015). Partial linear modelling with multi-functional covariates. *Computational Statistics*, 30(3) :647–671.
- Aneiros-Pérez G. et Vieu P. (2006). Semi-functional partial linear regression. *Statistics & Probability Letters*, 76(11) :1102–1110.
- Aneiros-Pérez G. et Vieu P. (2008). Nonparametric time series prediction : A semi-functional partial linear modeling. *Journal of Multivariate Analysis*, 99(5) :834–857.
- Aneiros-Pérez G. et Vieu P. (2011). Automatic estimation procedure in partial linear model with functional data. *Statistical Papers*, 52(4) :751–771.
- Baïllo A. et Grané A. (2009). Local linear regression for functional predictor and scalar response. *Journal of Multivariate Analysis*, 100(1) :102–111.
- Banon G. (1978). Nonparametric identification for diffusion processes. *SIAM Journal on Control and Optimization*, 16(3) :380–395.
- Beirlant J., Berlinet A., et Biau G. (2008). Higher order estimation at lebesgue points. *Annals of the Institute of Statistical Mathematics*, 60(3) :651–677.
- Benhenni K., Ferraty F., Rachdi M., et Vieu P. (2007). Local smoothing regression with functional data. *Computational Statistics*, 22(3) :353–369.

- Biau G., Cérou F., et Guyader A. (2010). Rates of convergence of the functional-nearest neighbor estimate. *IEEE Transactions on Information Theory*, 56(4) :2034–2040.
- Bogachev V.I. (1998). *Gaussian measures*, volume 62. American Mathematical Society Providence.
- Bongiorno E., Goia A., Salinelli E., et Vieu P. (2014). An overview of iwfos'2014. *Contributions in Infinite-Dimensional Statistics and Related Topics, Esculapio, Bologna*, pages 1–6.
- Bosq D. (2000). Linear processes in function spaces : Theory and applications, volume 149 of lecture notes in statistics.
- Bouraine M., Saidi A.A., Ferraty F., et Vieu P. (2010). Choix optimal de l'indice multi-fonctionnel : méthode de validation croisée, r. *Roumaine Math. Pures et Appl*, 50(5) :355–367.
- Bouzebda S., Elhattab I., et collab. (2011). Uniform-in-bandwidth consistency for kernel-type estimators of shannon's entropy. *Electronic Journal of Statistics*, 5 :440–459.
- Bro R. (1999). Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. *Chemometrics and Intelligent Laboratory Systems*, 46(2) :133–147.
- Burba F., Ferraty F., et Vieu P. (2009). k-nearest neighbour method in functional nonparametric regression. *Journal of Nonparametric Statistics*, 21(4) :453–469.
- Cacoullos T. (1966). Estimation of a multivariate density. *Annals of the Institute of Statistical Mathematics*, 18(1) :179–189.
- Cérou F. et Guyader A. (2006). Nearest neighbor classification in infinite dimension. *ESAIM : Probability and Statistics*, 10 :340–355.
- Chacón J.E. et Rodríguez-Casal A. (2010). A note on the universal consistency of the kernel distribution function estimator. *Statistics & Probability Letters*, 80(17) :1414–1419.

- Chagny G. et Roche A. (2016). Adaptive estimation in the functional nonparametric regression model. *Journal of Multivariate Analysis*, 146 :105–118.
- Chagny G., Roche A., et collab. (2014). Adaptive and minimax estimation of the cumulative distribution function given a functional covariate. *Electronic Journal of Statistics*, 8(2) :2352–2404.
- Chen C.P. et Zhang C.Y. (2014). Data-intensive applications, challenges, techniques and technologies : A survey on big data. *Information Sciences*, 275 :314–347.
- Chen D., Hall P., Müller H.G., et collab. (2011). Single and multiple index functional regression models with nonparametric link. *The Annals of Statistics*, 39(3) :1720–1747.
- Collomb G. (1980). Estimation de la régression par la méthode des k points les plus proches avec noyau : quelques propriétés de convergence ponctuelle. Dans *Statistique non Paramétrique Asymptotique*, pages 159–175. Springer.
- Collomb G. (1981). Estimation non-paramétrique de la régression : revue bibliographique. *International Statistical Review/Revue Internationale de Statistique*, pages 75–93.
- Cover T. (1968). Estimation by the nearest neighbor rule. *IEEE Transactions on Information Theory*, 14(1) :50–55.
- Cuevas A. (2014). A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*, 147 :1–23.
- Dabo-Niang S. (2002). Sur l'estimation fonctionnelle en dimension infinie : Application aux diffusions. *Thèse de Doctorat, Université de Paris*, 6.
- Dabo-Niang S. et Rhomari N. (2009). Kernel regression estimation in a banach space. *Journal of Statistical Planning and Inference*, 139(4) :1421–1434.
- Davidian M., Lin X., et Wang J.L. (2004). Introduction (emerging issues in longitudinal and functional data analysis). *Statistica Sinica*, pages 613–614.
- De Gooijer J.G. et Gannoun A. (2000). Nonparametric conditional predictive regions for time series. *Computational Statistics & Data Analysis*, 33(3) :259–275.

- Deheuvels P. et Ouadah S. (2013). Uniform-in-bandwidth functional limit laws. *Journal of Theoretical Probability*, 26(3) :697–721.
- Demongeot J., Laksaci A., Madani F., et Rachdi M. (2011). A fast functional locally modeled conditional density and mode for functional time-series. Dans *Recent advances in functional data analysis and related topics*, pages 85–90. Springer.
- Demongeot J., Laksaci A., Madani F., et Rachdi M. (2013). Functional data : local linear estimation of the conditional density and its application. *Statistics*, 47(1) :26–44.
- Demongeot J., Laksaci A., Naceri A., et Rachdi M. (2016). Estimation locale linéaire de la fonction de régression pour des variables hilbertiennes. *Comptes Rendus Mathématique*, 354(8) :847–850.
- Deville J.C. (1974). Méthodes statistiques et numériques de l'analyse harmonique. Dans *Annales de l'INSEE*, pages 3–101. JSTOR.
- Devroye L., Györfi L., Krzyżak A., et Lugosi G. (1994). On the strong universal consistency of nearest neighbor regression function estimates. *The Annals of Statistics*, pages 1371–1385.
- Devroye L. et Wagner T.J. (1982). 8 nearest neighbor methods in discrimination. *Handbook of Statistics*, 2 :193–197.
- Devroye L.P. et Wagner T.J. (1977). The strong uniform consistency of nearest neighbor density estimates. *The Annals of Statistics*, pages 536–540.
- Dony J., Einmahl U., et collab. (2009). Uniform in bandwidth consistency of kernel regression estimators at a fixed point. Dans *High Dimensional Probability V : The Luminy Volume*, pages 308–325. Institute of Mathematical Statistics.
- Dony J. et Mason D.M. (2010). Uniform in bandwidth consistency of kernel estimators of the tail index. *Extremes*, 13(3) :353–371.
- Dony J., Mason D.M., et collab. (2008). Uniform in bandwidth consistency of conditional u-statistics. *Bernoulli*, 14(4) :1108–1133.
- Einmahl U., Mason D.M., et collab. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics*, 33(3) :1380–1403.

- Ezzahrioui M. et Ould-Said E. (2008). Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *Journal of Nonparametric Statistics*, 20(1) :3–18.
- Ferraty F., Laksaci A., Tadj A., et Vieu P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *Journal of statistical planning and inference*, 140(2) :335–352.
- Ferraty F., Laksaci A., Tadj A., Vieu P., et collab. (2011). Kernel regression with functional response. *Electronic Journal of Statistics*, 5 :159–171.
- Ferraty F., Laksaci A., et Vieu P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statistical Inference for Stochastic Processes*, 9(1) :47–76.
- Ferraty F., Mas A., et Vieu P. (2007). Nonparametric regression on functional data : inference and practical aspects. *Australian & New Zealand Journal of Statistics*, 49(3) :267–286.
- Ferraty F., Rahbi A., et Vieu P. (2008). Estimation non-paramétrique de la fonction de hasard avec variable explicative fonctionnelle. *Revue Roumaine de Mathématiques Pures et Appliquées*, 53(1) :1–18.
- Ferraty F. et Vieu P. (2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*, 330(2) :139–142.
- Ferraty F. et Vieu P. (2006). *Nonparametric functional data analysis : theory and practice*. Springer Science & Business Media.
- Frank L.E. et Friedman J.H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, 35(2) :109–135.
- Gasser T., Hall P., et Presnell B. (1998). Nonparametric estimation of the mode of a distribution of random curves. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 60(4) :681–691.
- Geenens G. (2015). Moments, errors, asymptotic normality and large deviation principle in nonparametric functional regression. *Statistics & Probability Letters*, 107 :369–377.

- Geenens G. et collab. (2011). Curse of dimensionality and related issues in nonparametric functional regression. *Statistics Surveys*, 5 :30–43.
- Goia A. et Vieu P. (2015). A partitioned single functional index model. *Computational Statistics*, 30(3) :673–692.
- Goia A. et Vieu P. (2016). An introduction to recent advances in high/infinite dimensional statistics. *Journal of Multivariate Analysis*, 146 :1–6.
- Györfi L., Kohler M., Krzyzak A., et Walk H. (2006). *A distribution-free theory of nonparametric regression*. Springer Science & Business Media.
- Haan L. (1971). A form of regular variation and its application to the domain of attraction of the double exponential distribution. *Probability Theory and Related Fields*, 17(3) :241–258.
- Härdle W. et Marron J.S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *The Annals of Statistics*, pages 1465–1481.
- Hoel P.G., Port S.C., et Stone C.J. (1972). *Introduction to statistical theory*. Houghton Mifflin Co., Collier-Macmillan Ltd.
- Hoffmann-Jørgensen J. (1974). Sums of independent banach space valued random variables. *Studia Mathematica*, 52(2) :159–186.
- Horváth L. et Kokoszka P. (2012). *Inference for functional data with applications*, volume 200. Springer Science & Business Media.
- Hsing T. et Eubank R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons.
- Kara L.Z., Laksaci A., Rachdi M., et Vieu P. (2017). Uniform in bandwidth consistency for various kernel estimators involving functional data. *Journal of Nonparametric Statistics*.
- Kudraszow N.L. et Vieu P. (2013). Uniform consistency of knn regressors for functional variables. *Statistics & Probability Letters*, 83(8) :1863–1870.
- Laksaci A., Madani F., et Rachdi M. (2013). Kernel conditional density estimation when the regressor is valued in a semi-metric space. *Communications in Statistics-Theory and Methods*, 42(19) :3544–3570.

- Laloë T. (2008). A k-nearest neighbor approach for functional regression. *Statistics and Probability Letters*, 78(10) :1189–1193.
- Li J. (1985). Strong convergence rates of error probability estimation in the nearest neighbor discrimination rule. *Journal of Mathematics (Wuhan)*, 5 :113–118.
- Li W.V. et Shao Q.M. (2001). Gaussian processes : inequalities, small ball probabilities and applications. *Handbook of Statistics*, 19 :533–597.
- Lian H. (2011). Functional partial linear model. *Journal of Nonparametric Statistics*, 23(1) :115–128.
- Lian H. et collab. (2011). Convergence of functional k-nearest neighbor regression estimate with functional responses. *Electronic Journal of Statistics*, 5 :31–40.
- Liptser R.S. et Shiryaev A.N. (1977). *Statistics of random processes*, vol. 1.
- Manteiga W.G. et Vieu P. (2007). Statistics for functional data. *Computational Statistics & Data Analysis*, 51(10) :4788–4792.
- Mas A. (2012). Lower bound in regression for functional data by small ball probability representation in hilbert space. *Electron. J. Statist*, 6 :1745–1778.
- Mason D.M. et Swanepoel J.W. (2011). A general result on the uniform in bandwidth consistency of kernel-type function estimators. *Test*, 20(1) :72–94.
- Masry E. (2005). Nonparametric regression estimation for dependent functional data : asymptotic normality. *Stochastic Processes and their Applications*, 115(1) :155–177.
- Moore D.S. et Yackel J.W. (1977). Consistency properties of nearest neighbor density function estimators. *The Annals of Statistics*, pages 143–154.
- Nadaraya E.A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1) :141–142.
- Ouadah S. (2013a). Uniform-in-bandwidth kernel estimation for censored data. *Journal of Statistical Planning and Inference*, 143(8) :1273–1284.
- Ouadah S. (2013b). Uniform-in-bandwidth nearest-neighbor density estimation. *Statistics & Probability Letters*, 83(8) :1835–1843.

- Parzen E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3) :1065–1076.
- Quintela-Del-Río A. (2008). Hazard function given a functional variable : Non-parametric estimation under strong mixing conditions. *Journal of Nonparametric Statistics*, 20(5) :413–430.
- Rachdi M., Laksaci A., Demongeot J., Abdali A., et Madani F. (2014). Theoretical and practical aspects of the quadratic error in the local linear estimation of the conditional density for functional data. *Computational Statistics & Data Analysis*, 73 :53–68.
- Rachdi M. et Vieu P. (2007). Nonparametric regression for functional data : automatic smoothing parameter selection. *Journal of Statistical Planning and Inference*, 137(9) :2784–2801.
- Ramsay J.O. (1982). When the data are functions. *Psychometrika*, 47(4) :379–396.
- Ramsay J.O. et Dalzell C.J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 539–572.
- Ramsay J.O., Hooker G., et Graves S. (2009). *Functional data analysis with R and MATLAB*. Springer Science & Business Media.
- Ramsay J.O. et Silverman B.W. (1997). *Functional data analysis*.
- Ramsay J.O. et Silverman B.W. (2002). *Applied functional data analysis : methods and case studies*, volume 77. Citeseer.
- Ramsay J.O. et Silverman B.W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer, New York.
- Rao C.R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, 14(1) :1–17.
- Rosenblatt M. (1969). Conditional probability density and regression estimators. *Multivariate analysis II*, 25 :31.
- Rosenblatt M. et collab. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3) :832–837.

- Roussas G.G. et collab. (1969). Nonparametric estimation of the transition distribution function of a markov process. *The Annals of Mathematical Statistics*, 40(4) :1386–1400.
- Sarda P. et Vieu P. (2000). Kernel regression. *Smoothing and Regression : Approaches, Computation, and Application*, pages 43–70.
- Schimek M. (2012). *Smoothing and Regression : Approaches, Computation, and Application*. Wiley Series in Probability and Statistics. Wiley. ISBN 9781118150641.
- Schimek M.G. (2013). *Smoothing and regression : approaches, computation, and application*. John Wiley & Sons.
- Schimek M.G. et Turlach B.A. (1998). Additive and generalized additive models : a survey. cahier de recherche, Discussion Papers, Interdisciplinary Research Project 373 : Quantification and Simulation of Economic Processes.
- Scott D.W. (2015). *Multivariate density estimation : theory, practice, and visualization*. John Wiley & Sons.
- Shang H.L. (2014). Bayesian bandwidth estimation for a functional nonparametric regression model with mixed types of regressors and unknown error density. *Journal of Nonparametric Statistics*, 26(3) :599–615.
- Stone C.J. (1977). Consistent nonparametric regression. *The annals of statistics*, pages 595–620.
- Stone C.J. (1982). Optimal global rates of convergence for nonparametric regression. *The annals of statistics*, pages 1040–1053.
- Stone C.J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistics*, pages 1285–1297.
- Stone C.J. (1985). Additive regression and other nonparametric models. *The annals of Statistics*, pages 689–705.
- Tran T.N., Wehrens R., et Buydens L.M. (2006). Knn-kernel density-based clustering for high-dimensional multivariate data. *Computational Statistics & Data Analysis*, 51(2) :513–525.

- Tucker L.R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika*, 23(1) :19–23.
- Tukey J.W. et collab. (1961). Curves as parameters, and touch estimation. Dans *Proceedings of the 4th Berkeley Symposium*, volume 31, pages 681–694.
- Valderrama M.J. (2007). An overview to modelling functional data. *Computational Statistics*, 22(3) :331–334.
- Van Der Vaart A. et Wellner J.A. (1996). *Weak Convergence and Empirical Processes*. Springer.
- Van Der Vaart A. et Wellner J.A. (2011). A local maximal inequality under uniform entropy. *Electronic Journal of Statistics*, 5(2011) :192.
- Varron D. (2012). Some uniform in bandwidth functional results for the tail uniform empirical and quantile processes. *arXiv preprint arXiv :1201.5517*.
- Varron D. et Van Keilegom I. (2011). Uniform in bandwidth exact rates for a class of kernel estimators. *Annals of the Institute of Statistical Mathematics*, 63(6) :1077–1102.
- Wand M.P. et Jones M.C. (1994). *Kernel smoothing*. Crc Press.
- Wang J.F., Ma W.M., Fan G.L., et Wen L.M. (2015). Local linear quantile regression with truncated and dependent data. *Statistics & Probability Letters*, 96 :232–240.
- Watson G.S. (1964). Smooth regression analysis. *Sankhyā : The Indian Journal of Statistics, Series A*, pages 359–372.
- Youndjé E., Sarda P., et Vieu P. (1993). Kernel estimate of conditional density - bandwidth selection for dependent data. *Comptes rendus de l'academie des sciences serie I-MATHEMATIQUE*, 316(9) :935–938.
- Zhang J.T. (2013). *Analysis of variance for functional data*. CRC Press.

Nous nous proposons, dans ce travail de recherche, d'établir la convergence presque complète uniforme sur les paramètres de lissage de quelques estimateurs à noyaux non paramétriques sur un échantillon i.i.d d'une variable explicative fonctionnelle afin d'estimer une variable réponse réelle. Les résultats obtenus nous permettent de traiter d'estimateurs de type k NN et d'en développer des outils de sélection automatique du nombre de voisins optimal par validation croisée.

Nous traitons en dernière partie deux exemples d'applications de ces outils pour illustrer leur simplicité d'implémentation et leur bon comportement.

In this work, we propose to establish the uniform in bandwidth almost complete consistency of some nonparametric estimators, dealing with an i.i.d sample of a functional regressor to estimate a real response variable. The obtained results allow us to deal with k NN type estimators and then to develop optimal number of neighbours' automatic selection tools with cross validation.

We process at last two application's examples about those tools to illustrate their simplicity of implementation and their good behaviour.

نقترح في هذه الأطروحة دراسة التقارب شبه الكامل المنتظم على معامِل السّلاسة لبعض المقدّرات من نوع التّواة الكلاسيكية غير البارامترية على عيّنة مستقلة و متطابقة التّوزيع من متغيّر تفسيري غير متتهي الأبعاد لتقدير متغيّر جوابي حقيقي. التّناجِح الحصل عليها تمكّننا بمعالجة المقدّرات من نوع k ن و إعداد أدوات الإختيار الأوتوماتيكي لأمثل عدد الحيران بواسطة طريقة المصادقة المتقاطعة.

ندرس في الفرع الأخير أمثلة لتطبيق هذه الأدوات لتوضيح بصاطة تنفيذها و حسن سيرها.