Ce travaille est dédiée à toute ma famille...

et mes meilleurs amis...

Remarcimemt

Il est naturel de remercier à la fin d'un tel travail tous ceux qui, plus ou moins directement, ont contribué à le rendre possible.

Un remerciement trés particulier va à Monsieur "Fethi MADANI" pour l'attention qu'il a porté à mon travail. Ces conseils et idées ont été précieux et ont guidé ma recherche au cours de mon travail. Je souhaite tout d'abord lui exprimer ma profonde gratitude.

Je voudrais aussi remercier chaleureusement chacun des membres du jury qui me font le grand honneur d'y participer.

Je remercie sincérement Monsieur "Djellouli Ghouti" pour l'honneur qu'il me fait en présidant le jury.

Je remercie vivement Monsieur "Kandouci Abdeldjabar" pour la confiance dont il me fait preuve en faisant parties de ce jury.

Je voudrais aussi remercier Madame "Mokhtari Fatiha" pour l'intérêt qu'elle a bien voulu accorder à mon travail en acceptant de participer au jury.

Je remercie également Melle "Rouane Rachida" pour son aide et les services qu'il m'a rendus. Il m'a toujours encouragé et précieusement conseillé.

Je tiens à exprimer aussi ma reconnaissance à tous mes enseignants.

Je remercie trés amicalement tous mes amis et d'ailleurs de leurs sympathie et leurs aides de prés ou de loin.

Enfin, j'adresse mes remerciements à l'ensembles des personnes présentes à cette soutenance.

Table des matières

La déscription du mémoire

La particularité de la statistique non-paramétrique est que le paramètre inconnu qu'on cherche à detecter, à estimer ou à classifier n'est pas supposé d'appartenir à une famille indicée par un petit nombre de paramètres réels.

En général, dans la théorie non-paramétrique on suppose que le nombre de paramètres qui décrivent la loi des observations est une fonction croissant du nombre d'observations, ou encore que le nombre de paramètres est infini.

La tradiction de considérer le problème de l'estimation statistique comme celui d'estimation d'un nombre fini de paramètre remonte à Fisher. Cependant, les modèles paramétriques ne fournissent qu'une aproximation, souvent imprécise, de la structure statistique sous-jacente. Les modèles statistiques qui expliquent plus précisement les données sont souvent plus complexes : les inconnues de ces modèles sont, en général, des fonction possédant certaines propriétés de régularité. Le problème de l'estimation non-paramètrique consiste à estimer, à partir des observation, une fonction inconnue, élément d'une certaine classe fonctionnelle assez massive.

La théorie de l'estimation non-paramétrique s'est développée considérablement ces deux derniéres décennies.

Les progrés de l'outil informatique tant en capacitéde mémoire qu'en calcul permettent d'enregistrer des données de plus volumineuses. Ainsi, un trés grand nombre de variables peuvent être obsevées pour l'étude d'un même phénomène. C'est notamment le cas lorsqu'on dipose d'une famille de variables $\{X(\theta)\}_{\theta\in\Theta}$ indexée par un paramètre θ variant dans un espace "continu" Θ (par exemples, $\Theta \subset \mathbb{R}$ ou $\Theta \subset \mathbb{R}^p$ ou encore $\Theta \subset H$ avec H un hibertien,...).

Bien évidemment, il est techniquement impossible de mesurer $X(\theta)$ pour tout θ dans Θ. Néanmoins, on dispose d'une discrétisation suffisamment fine $\{\theta_i\}_{i=1,\dots,I}$ de Θ pour pouvoir considérer que le comportement de $\{X(\theta)\}_{\theta\in\Theta}$

est "proche" de celui de $\{X(\theta_i)\}_{i=1,\dots,I}$. Une telle famille $X = \{X(\theta)\}_{\theta\in\Theta}$ de variable aléatoire sera appelée variable aléatoire fonctinnelle.

Dans ce mémoire, on c'est intéressé à l'estimation non-paramétrique de la dencité conditionnelle pour des données fonctionnelles. Pour cela, notre travail est structure de la manière suivantes :

Chapitre 1

Introduction générale

1.1 Donnée fonctionelles

La modélisation statistique de données fonctionnelles est liée à l'étude d'ensembles de données dans les quels les observations peuvent être assimilées à des courbes ou à des surfaces. L'importance de ce sujet est motivé par le développement des outils informatiques et leurs capacités de stockage. De ce faite un énorme potentiel en terme d'applications c'est considérablement développé traitant ce type de données issues de domaines variés : tel que la météorologie, la médecine et dans des nombreux autres domaines. Ainsi, de part la richesse en applications et les problèmes théoriques soulevés, cette branche de la statistique est devenue le centre d'intérêt de plusieurs études, tant sur le plan pratique que sur le plan théorique.

Définition 1.1.1. *On appelle modéle fonctionnel, tout modèle prenant en compte au moins une variable aléatoire fonctionnelle (v.a.f).*

Définition 1.1.2. *Un modèle fonctionnel est dit paramétrique si C indexable par un nombre fini de paramètres appartenant à f, ou C n'est qu'un sous-ensemble de F l'ensemble des fonctions définies sur l'espace fonctionnel f et à valeurs dans l'espace F. Un modèle fonctionnel est dit non-paramétrique dans le cas contraire.*

Historiquement, les premiers résultats considérant des observations sous forme de trajectoires ont été obtenus par Obhukov(1960), Holmstrom (1961) en climatologie, Deville (1974) en démographie, Molenaar et Boomsma (1987) puis Kirkpatrick et Heckman (1989) en génétique. Dans ces derniers temps la modélisation par la régression (paramétrique ou non-paramétrique) a fait l'objet d'une littérature trés dense dans ce domaine de la statistique fonctionnelle. En effet, dans le contexte paramétrique, la contribution de Ramsay et Silverman (2002,2005) offre un large éventail de méthode statistique pour des variables fonctionnelles, en particulier pour des variables indépendantes. Tandis que le monographe de Bosq (2000) traite des modèles fonctionnels linèaires dépendants (modèles d'autorégression). Cardot et al. (1999) ont construit un estimateur pour le modèle de régression linèaire Hilbertien en utilisant l'analyse en composantes principales fonctionnelles. Cet estimateur est construit à l'aide des propriétés spectrales de la version empirique de l'opérateur de variance-covriance de la variable explicative fonctionnelle. La convergence presque complète, ainsi que la convergence en norme L_2 pour une version régularisée ont été établi pour cet estimateur par les mêmes auteurs (2000). Cardot et al. (2004), ont introduit un estimateur pour les quantiles conditionnels vu comme forme linèaire continue définie sur un espace de Hilbert. En ce qui concerne le cas non-paramétrique, les premiers résultats sont ceux de Ferraty et Vieu (2000), étendus en (2002) au cas non standard de la régression telle la prévision en séries chronologiques.

En utilisant le développement récent de la théorie des probabilitès des petites boules, Ferraty et Vieu (2003) ont obtenu des résultats trés fins et qui offrent une solution originale pour le problème de flèau de la dimension.

Niang et Rhomari (2004) ont utilisé cet estimateur de la régression pour établir la convergence en norme Lp, une application à la discrimination des courbes a été proposée par ces auteurs.

Masry (2005), a étudié la normalité asymptotique de l'estimateur de la régression dans le cas dépendant. L'ouvrage de Ferraty et Vieu (2006) constitue une contribution de référence en la matière et fournit de plus amples détails concernant le comportement asymptotique des estimateurs à noyau pour plusieurs modèles fonctionnels à savoir la régression, le mode (avec ou sans conditionnement),la médiane (avec ou sans conditionnement) ainsi que les quantiles conditionnels. Notons, que cet ouvrage contient aussi de nombreux exemples de problèmes concret pour lesquels les données sont de nature fonctionnelle. Nous renvoyons a Ferraty et al (2007) pour des résultats précurseurs sur la vitesse de convergence asymptotiquement exacte en norme L_2. En utilisant ce résultat, Rachdi et Vieu (2007) construit un critére de choix automatique du paramétre de lissage pour l'estimateur à noyau de la fonction de régression à co-variables fonctionnelles. Récemment, Delsol (2007) généralise au cas dépendant les résultats de Ferraty et al. (2007) en donnant les termes asymptotiquemant dominants des moments centrés et des erreurs Lq de l'estimateur à noyau de la fonction de régression. Benhenni et al. (2007) ont établi la consistance de l'estimateur de la fonction de régression pour des co-variables fonctionnelles sous conditions de longue mémoire. Tandis que Burba et al. (2008) se sont intéressés à l'estimation de la régression via a méthode des k plus proches voisins (kNN), ils ont obtenu la convergence presque compléte de l'estimateur construit.

Un des premiers articles sur les modèles nonparamétriques conditionnels avec variables explicatives fonctionnelles f$\hat{u}$t celu de Ferraty et al.(2006). Nous renvoyonsàFerraty et al. (2005) pour l'utilisation de ces modèles conditionnels fonctionnels à la prévision des séries temporels fonctionnelles. Ezzahrioui et Ould-Saïd (2006, 2008a, 2008b, 2008c) et Niang et Laksaci (2007, 2008) ont une contribution déterminante dans ce domaine. Le lecteur pourra également se rapporter au travail de Manteiga et Vieu (2007) pour une bibliographie récente et exhaustive en la matière. La littérature sur la statistique fonctionnelle robuste est trés restreinte. En effet, les premiers résultats conséquents dans le domaine ont été fournis par Cadre (2001). Ce dernier étudié l'estimation de la médiane de la distribution d'une variable aléatoire à valeurs dans un espace de Banach. Azzeddine et al. (2008) ont étudié la convergence presque compléte d'une famille d'estimateurs robustes basée sur la mèthode du noyau, en considérant des observations indépendantes. La vitesse de convergence en norme Lp fait l'objet d'un travail de Crambes et al. (2008) en considérant les deux types d'observations indépendantes et mélangeantes.

1.2 Champs d'application des données fonctionnelles

En industrie alimentaire

Cet exemple concerne des données spectrométriques couramment utilisées en chimie quantitative. On s'intèresse à un échantillon de 215 morceaux de viandes finiment hachées contenant chacun un certain taux de matière grasse (lipide). Pour chaque morceau, on obcerve le spectre dans le proche infrarouge pour 100 longueurs d'onde $\{\lambda_j\}_{j=1,\ldots,100}$ réparties entre 850 et 1050 nanométres avec un pas constant. On observe donc pour chaque morceau i, une famille "discrète" $X_i^d = \{X_i(\lambda_j)\}_{j=1,\ldots,100}$ que l'on peut considérer comme une version discrétisée de la variable fonctionnelle $X_i = \{X_i(\lambda)\}_{\lambda\in[850,1050]}$. On parle dans ce cas de courbe spectrométrique. Leurgans et al. (1993) conforte cette idée en remarquant que "the spectra observed are to all intents and purpoese functional observations" alors que la représentation des 215 spectres ci-dessous fait apparaitre clairement l'aspect fonctionnel :

Un autre exemple de variables aléatoires fonctionnelles dépendantes portant sur l'étude de phénomènes liés à l'environnement est le problème de pollution. Il s'agit d'étudier la courbe de concentration d'ozone au Pôle Nord sur quatre années successives (de 2000 à 2004). L'objectif est de prévoir la concentration de l'ozone dans une journée étant donné la courbe de concentration de l'ozone de la veille. En procédant par un découpage journalier de la courbe de concentration annuelle de l'ozone, on obtient les courbes représentées dans Figure ??. Notons que plusieurs auteurs se sont intéressés aux phénoménes liés à l'environnement, on peut citer entre autres, Damon et Guillas (2002), Aneiros-Perez et al. (2004), Cardot et al. (2004, 2006), Meiring (2005).
Bref, de nombreux autres domaines d'application où l'on peut être confronté à des données de natures fonctionnelles existent et/sinon affluent. Vu le grand nombre des exemples que l'on peut citer.

Consommation d'électricité aux USA

dans le cadre des données dépendantes, on peut considérer l'exemple d'une série chronologique qui concerne la consommation annuelle d'électricité, aux USA, par des secteurs résidentiels et commerciaux de janvier 1973 jusqu'en février 2001 (338 mois). Le but de cette étude est de prévoir la consommation d'électricité de l'année suivante sachant la consommation d'électricité de toute l'année précédente. L'échantillon se compose de 28 données comme le montre la Figure 1.3. Cette série chronologique peut être regardée comme étant un ensemble de données fonctionnelles dépendantes (c'està-dire, une population de 28 courbes : chaque année correspond à 1 courbe).

Chapitre 2

Sur l'estimation non paramétrique (méthode du noyau)

La statistique non paramétrique regroupe l'ensemble des méthodes statistiques qui permettent de tirer de l'information pertinente de données sans faire l'hypothèses que la loi de probabilité de ces observations appartient à une famille paramétrée connue.

2.1 Estimation à noyau de la densité

Définition 2.1.1. *L'origine des la méthode des noyaux est due à Rosenblatt (1956). Celui-ci a propsé une sorte d'histogramme mobile ou la fenêtre de comptage des observation se déplace avec la valeur de x. La densité en x est estimée par la fréquence relative des observations dans l'intervalle $[x-h, x+h]$, donc centrè sur x, divisée naturellement par longeur de l'intervalle $2h$. On appelle h la longeur de fenêtre (bien que cette largeur soit en fait égale à $2h$). Pour des raisons qui apparaitront plus loin nous écrivons l'estimation ainsi obtenue à partir des observation $x_1, x_2, ..., x_n$ sous la forme suivante(conservant encore la même notation $\hat{f}_n$) :*

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

ou

$$\mathbf{K}(u) = \frac{1}{n} \quad si \quad u \in [-1, +1] \qquad et \quad 0 \quad sinon.$$

En effet $x_i \in [x-h, x+h]$ si et seulement si $\frac{x-x_i}{h} \in [-1, +1]$ et x_i est alors comptabilisé$\frac{1}{2}$. Ainsi $\sum_{i=1}^{n} K(\frac{x-x_i}{h})$ est égal au nombre d'observation tombant dans $[x-h, x+h]$ divisé par 2 pour obtenir la division de la fréquence relative par 2h. Comme K est discontinue en $+1$ et -1, $\hat{f}_n(x)$ présente des petits sauts de discontinunuité aux points $x_1 \pm h$, $x_2 \pm h$,..., $x_n \pm h$. Parzen (1962) a proposé une généralisation de l'idée de Rosenblattant, entre autres, de lisser davantage l'estimation. A la fonction K ci-dessus on substitue une fonction que l'on pourra choisir continue ou dérivable partout, propriété qui se transfère à la fonction $\hat{f}_n$. En d'autres termes on fera entrer ou sortir les points x_i quand on déplace la fenêtre. Toute fois la fonction K est soumise aux conditions suivantes :

- **K** est positive (ou nulle).

- **K** est paire.

- $\displaystyle\int_R K(u) = 1.$

Une telle fonction est alors appelée *noyau*. La première condition garantit que le poids $K(\frac{x-x_i}{h})$ de chaque observation x_i reste positif ou nul, la deuxième que ce poids soit identique autour de x, tandis que la troisième condition est une normalisation des poids de façon que $\hat{f}_n$ soit bien une densité. En effet, avec le changement de variable $u = \frac{x-x_i}{h}$. on obtient :

$$\int_{\mathbb{R}} \hat{f}_n(x)dx = \frac{1}{nh}\sum_{i=1}^{n}\int_{\mathbb{R}} K\left(\frac{x-x_i}{h}\right)dx = \frac{h}{nh}\sum_{i=1}^{n}\int_{\mathbb{R}} K(u)du = \frac{1}{n}\sum_{i=1}^{n} 1 = 1.$$

Notons qu'une fonction noyau est, en fait, une fonction de densité symètrique autour de zéro, donc de moyenne nule (si elle existe).
Comme :

$$\frac{1}{n}\int_R \mathbf{K}\left(\frac{x-x_i}{h}\right)dx = \frac{1}{n}\int_R \mathbf{K}(u)du = \frac{1}{n},$$

on peut donner une interprétation concréte de $\hat{f}_n$. Supposons, pour fixer les idées, que K ait pour support $[-1, +1]$. Alors $\hat{f}_n$ est obtenue en remplaçant chaque observation x_i par une même petite «densité» (son aire étant réduite à $\frac{1}{n}$ de support $[x_i - h, x_i + h]$, puis en sommant ces petites densités. Cette vision correspond à un principe général de lissage de données discrétes qui consiste à faire «bouger» chaque donée pour lui substituer un élément continu. En pratique on impose comme condition supplémentaire que K décroisse de part et d'autre de zéro, dans l'idée naturelle de donner un poids plus faible aux observations au fur et à mesure qu'elles s'éloignent du centre de la fenêtre x.

Ainsi les noyaux les plus usuels sont :

$K(u) = \frac{1}{2}$ si $u \in [-1, 1]$ noyau de Rosenblatt.

$K(u) = 1 - |u|$ si $u \in [-1, 1]$ noyau de triangulaire.

$K(u) = \frac{3}{4}(1 - u^2)$ si $u \in [-1, 1]$ noyau de d'Epanechnikov.

$K(u) = \frac{15}{16}(1 - u^2)^2$ si $u \in [-1, 1]$ noyau de Tukey ou biweight.

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

Les deux premiers ont l'avantage d'être simples, le noyau triangulaire étant continu partout et conduisant à une estimation $\hat{f_n}$ continue. Le troisième doit sa notoriété à une propriété d'optimalité théorique mais sans grand intérêt pratique . Le quatrième est, à notre sens, le plus intéressant car donnant une estimation dérivable partout, tout en étant simple à mettre en oeuvre. En fait il s'agit du noyau le plus simple parmi les noyaux de forme polynomiale d'erivables partout. Ainsi il assure le lissage local de la fonction $\hat{f_n}$. Ce noyau est d'une forme trés proche du noyau gaussien et est donc préférable, ce dernier ayant un coût de calcul plus élevé du fait de son support infini (la «largeur de fenêtre» h devenant conventionnellement l'écart-type de la loi de Gauss). Notons que plus la valeur de h est élevée plus on élargit la fenêtre, ce qui a un effet de lissage global de $\hat{f_n}$ plus important. Ceci est à rapprocher du choix de la largeur des intervalles pour l'histogramme.Et choix de la fenêtre pour l'estimation a noyau.

2.1.1 Propriétés asymptotiques des estimateurs à noyaux

Il existe beaucoup de résultats à n fini et l'on doit se satisfaire de résultats asymptotiques. Reprenons l'expression générale d'un estimateur à noyau pour un échantillon aléatoire $X_1, X_2, ..., X_n$:

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

Pour calculer le biais et la variance en un point x fixé posons :

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

Ainsi la variable aléatoire $\hat{f}_n$ est la moyenne des Z_i qui, en tant que fonctions respectives des X_i, sont des variables aléatoires i.i.d.. Soit $Z = \frac{1}{n}\mathbf{K}(\frac{x-X}{n})$ la v.a. symbolisant la loi commune aux Z_i comme X symbolise la loi mêre des X_i de densité f.

calcul le biais

On a :

$$\begin{aligned}
\mathbb{E}(\hat{f}_n(x)) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i\right) = \mathbb{E}(Z) = \frac{1}{h}\mathbb{E}\left(\mathbf{K}\left(\frac{x-X}{h}\right)\right) \\
&= \frac{1}{h}\int_{\mathbb{R}} \mathbf{K}\left(\frac{x-t}{h}\right) f(t)dt \\
&= \int_{\mathbb{R}} \mathbf{K}(u) f(x+uh)du \quad \textit{en posant } u = \frac{x-t}{h}.
\end{aligned}$$

Comme $\int_{\mathbb{R}} \mathbf{K}(u)du = 1$ le biais peut s'écrire :

$$\mathbb{E}(\hat{f}_n(x)) - f(x) = \int_{\mathbb{R}} \mathbf{K}(u)[f(x+uh) - f(x)]du.$$

On voit que le biais résulte de l'écart de la valeur de la densité dans la fenêtre centrée sur x par rapport à sa valeur en x même. Si f était constante dans la fenêtre le biais serait nul, et de même si f était parfaitement linéaire en raison de la parité du noyau K. Comme pour l'histogramme le biais ne dépend pas de la taille de l'échantillon et ne peut être réduit à zéro qu'en faisant tendre h vers zéro. Prenons un développement de Taylor de f au voisinage de x :

$$f(x+uh) = f(x) + uhf'(x) + \frac{u^2h^2}{2}f''(x) + o(h^2)$$

Le biais s'écrit :

$$\begin{aligned}
\mathbb{E}(\hat{f}_n(x)) - f(x) &= hf'(x)\int_{\mathbb{R}} u\mathbf{K}(u)du + \frac{h^2}{2}f''(x)\int_{\mathbb{R}} u^2\mathbf{K}(u)du + o(h^2) \\
&= \frac{h^2}{2}f''(x)\int_{\mathbb{R}} u^2\mathbf{K}(u)du + o(h^2)
\end{aligned}$$

puisque K est paire. Pour h petit le biais dépend donc de $f''(x)$ et du moment d'ordre 2 du noyau. Le biais est du signe de f"(x) : si f est concave en x le biais est négatif, si elle est convexe le biais est positif. En particulier si x est un point où f est à un maximum le biais sera n'egatif. On sous-estime donc (en moyenne) la hauteur du maximum, ce que l'on peut comprendre intuitivement : la densité au voisinage de x étant plus faible il y a n'ecessairement un déficit de points dans la fenêtre. A l'inverse les minima éventuels seront surestimés. Par conséquent la méthode tend à écrêter les creux et les pics de la densité ce qui est un inconvénient majeur.

calcul de la variance

On a :

$$V(\hat{f}_n(x)) = V\left(\frac{1}{n}\sum_{i=1}^{n} Z_i\right) = \frac{1}{n}V(Z) = \frac{1}{n}\left\{\mathbb{E}(Z^2) - [\mathbb{E}(\mathbb{Z})]^2\right\}$$

avec :

$$\begin{aligned}\frac{1}{n}\mathbb{E}(Z^2) &= \frac{1}{nh^2}\int_{\mathbb{R}}\left[\mathbf{K}\left(\frac{x-t}{h}\right)\right]^2 f(t)dt. \\ &= \frac{1}{nh}\int_{\mathbb{R}}[\mathbf{K}(u)]^2 f(x+uh)du \quad (\textit{en posant } u = \frac{x-t}{h})\end{aligned}$$

et :

$$\frac{1}{h}[\mathbb{E}(Z)]^2 = \frac{1}{n}\left[\int_{\mathbb{R}}\mathbf{K}(u)f(x+uh)du\right]^2.$$

Alors que le terme $\frac{1}{n}[\mathbb{E}(Z)]^2$ tend bien vers zéro quand $n \to \infty$ on voit que le terme $\frac{1}{n}(\mathbb{E}(Z)^2)$ ne tend vers zéro que si $nh \to \infty$.

Par conséquent, pour que $\hat{f}(x)$ converge vers $f(x)$ en moyenne quadratique les mêmes conditions sont n'ecessaires que pour l'histogramme :

$n \to \infty$,$h \to 0$, $nh \to \infty$ le terme $\frac{1}{n}[\mathbb{E}(Z)]^2$ est d'ordre $\frac{1}{n}$,ce que l'on note $o(\frac{1}{h})$. En utilisant le d'eveloppement de Taylor :

$f(x+uh) = f(x) + uhf'(x) + o(h)$, on obtient :

$$\frac{1}{n}\mathbb{E}(Z^2) = \frac{1}{nh}f(x)\int_{\mathbb{R}}[\mathbf{K}(\mathbf{u})]^2 du + o(\frac{1}{n})$$

d'où :

$$V\left(\hat{f}_n(x)\right) = \frac{1}{nh}f(x)\int_{\mathbb{R}}[\mathbf{K}(\mathbf{u})]^2 du + o(\frac{1}{n})$$

Finalement l'erreur quadratique moyenne en x fixé est :

$$MSE(\hat{f}_n) = \frac{h^4}{4}[f''(x)]^2 \left[\int_{\mathbb{R}} u^2\mathbf{K}(u)du\right]^2 + \frac{f(x)}{nh}\int_{\mathbb{R}}[\mathbf{K}(u)]^2du + o(h^4) + o(\frac{1}{n})$$

Faisant abstraction des termes $o(h_4) + o(\frac{1}{n})$ négligeables dans les conditions de convergence, on voit que plus la largeur de fenêtre h est faible plus le biais diminue mais plus la variance augmente et, inversement, l'élargissement de la fenêtre augmente le biais et diminue la variance. Il existe un optimum (mais valable uniquement pour le point x) qui, comme pour l'histogramme, est obtenu en d'erivant par rapport à h, soit :

$$h_{opt} = \left[\frac{f(x)\int_{\mathbb{R}}[\mathbf{K}(u)]^2du}{[f''(x)]^2[\int_{\mathbb{R}} u^2\mathbf{K}(u)du]^2}\right]^{1\backslash 5} n^{-1\backslash 5}$$

et, en substituant h_{opt} dans la formule de l'expression asymptotique de l'MSE,celle-ci prend la forme $k(x)\nu(\mathbf{K})n^{-4\backslash 5}$ où $\nu(\mathbf{K})$ est une expression qui ne d'epend que du choix du noyau et $k(x)$ est fonction de $f(x)$ et de $f''(x)$. Ainsi la convergence est plus rapide que pour l'histogramme, étant d'ordre $n^{-4\backslash 5}$ lieu de $n^{-2\backslash 3}$.
Jusqu'à présent nous avons raisonné à x fixé. Il est clair que ce qui nous intéresse est de connaître le comportement de l'estimateur $\hat{f}_n(x)$ de la fonction f globalement sur tout $\mathbb{R}$. Pour cela on considère, pour une réalisation donnée, son écart à f intégré sur tout $\mathbb{R}$ ce qui conduit, en prenant l'espérance mathématique de cet écart intégré, au critère d'erreur quadratique intégrée moyenne :

$$MISE(\hat{f}_n) = \mathbb{E}\left(\int_{\mathbb{R}}\left[\hat{f}_n(x) - f(x)\right]^2 dx\right).$$

Celle-ci se calcule aisément à partir des résultats précédents car, étant donné les conditions de régularité imposées à f et à $\mathbf{K}$, il est licite d'intervertir les intégrations (l'une explicite, l'autre implicite dans le calcul de l'espérance mathématique) ce qui conduit à intégrer l'expression de MISE. en x fixé :

$$MISE(\hat{f}_n) = \int_{\mathbb{R}}\mathbb{E}\left(\left[\hat{f}_n(x) - f(x)\right]^2 dx\right) = \int_{\mathbb{R}} eqim(\hat{f}_n(x))dx.$$

D'où

$$MISE(\hat{f}_n) = \frac{h^4}{4}\int_{\mathbb{R}}[f''(x)]^2dx\left[\int_{\mathbb{R}} u^2\mathbf{K}(u)\right]^2 + \frac{1}{nh}\int_{\mathbb{R}}[\mathbf{K}(u)]^2du + o(h^4) + o(\frac{1}{n})$$

Comme précédemment on trouve un h optimal qui est en $n^{-1\backslash 5}$ et une MISE de la forme $g(f'')\nu(\mathbf{K})n^{-4\backslash 5}$. Le même critère aurait pu être appliqué à l'histogramme, la vitesse de convergence étant également conservée en $n^{-2\backslash 3}$.

2.2 Estimateur à noyaux de la densité conditionnelle

Dans ce pragraphe, nous présentons un estimateur à noyau de la dérivée d'ordre j de la densité conditionnelle .

Cet estimateur $\hat{f}^x(j)$ de f^x est donnée par :

$$\hat{f}^{x(j)}(y) = \frac{h_H^{-j-1}\sum_{i=1}^n K\left(\frac{d(x,X_i)}{h_K}\right) H^{(j+1)}\left(\frac{y-Y_i}{h_K}\right)}{\sum_{i=1}^n K\left(\frac{d(x,X_i)}{h_K}\right)}, \ \forall y \in \mathbb{R}$$

.

Notons que, cet estimateur est analogue à celui introduit par Rosenblatt (1969) dans le cas où X est une variable aléatoire réelle. Il est aussi largement étudié depuis ce temps (cf.Youndjé, 1996, Madani 2012).

2.3 Estimation par polynôme locaux

L'etimation non paramétrique par polynôme locaux découle de l'estimation non paramétrique à noyau, et plus précisement de l'estimastion de nadaraya-waston.

Si $k \geq o$ l'estimateur de Nadaraya-waston f_n^{nw} vérifie :

$$f_n^{nw}(x) = \arg\min_{\theta\in\mathbb{R}} \sum_{i=1}^n (y_i - \theta)^2 \mathbf{K}\left(\frac{X_i - x}{h}\right)$$

Donc, f_n^{nw} est obtenu par approximation des moindre carrée localement constante des valeurs Y_i.
Plus généralement définissons l'approximation des moindre carrée localement polynomiale si

$$f \in \sum(\beta, L), \beta > 1, \ell = [\beta]$$

alors pour z suffisamment voisine de x.

$$f(z) \approx f(x) + \grave{f}(z-x) + ... + \frac{f^{(\ell)}(x)}{\ell!}(z-x)^{\ell} = \theta^T(x)U\left(\frac{z-x}{h}\right)$$

où

$$U(u) = (1, u, \frac{u^2}{2!}, ..., \frac{u^{\ell l}}{\ell})^T$$

$$\theta(x) = (f(x), f'(x)h, f''(x)h^2, ..., f^{(\ell)}h^{h^{(\ell)}})^T$$

Définition 2.3.1. *Soient $K : \mathbb{R} \to \mathbb{R}$ un noyau,$h > 0$ une fenétre et $\ell \geq 0$ un entier. Le vecteur $\hat{\theta}_n(x) \in \mathbb{R}^{\ell+1}$ défini par*

$$\hat{\theta}_n(x) = \arg \min_{\theta \in \mathbb{R}^{\ell+1}} \sum_{i=1}^{n} \left[Y_i - \theta^T U\left(\frac{X_i - 1}{h}\right)\right]^2 K\left(\frac{X_i - x}{h}\right)$$

est appelé "estimateur localement polynomial d'ordre ℓ" de θ. La statistique

$$\hat{f}_n(x) = U^T(o)\hat{\theta}_n(x)$$

est l'estimateur polynomial d'ordre ℓ de $f(x)$.
Fan (1993) étudie l'optimalité des vitesses dans cas variables alétoires x réelles. Dans le cas de la dimension infinie ($\dim(H) = +\infty$) un estimateur à proposé par Ferraty. Vieu (2006) Fenaty. Mas ,Vieu(2006) on montré la normalité asympototique de cet estimateur de m par une autre méthode et ensuite.

2.3.1 Expression matricielle

On observe un ensemble de points $(x_i, y_i)_{1 \leq i \leq n}$. Les valeurs de f_n^{NW}. aux point x_i s'obtinnent par lissage des valeurs y_i de la façon suivant

$$\begin{pmatrix} \hat{r}_n^{NW}(x_1) \\ \vdots \\ \hat{r}_n^{NW}(x_n) \end{pmatrix} = \begin{pmatrix} \frac{K_h(0)}{\sum_\iota K_h(x_\iota - x_1)} & \cdots & \frac{K_h(x_1 - x_n)}{\sum_\iota K_h(x_\iota - x_1)} \\ \vdots & \frac{K_h(x_i - x_j)}{\sum_\iota K_h(x_\iota - x_j)} & \vdots \\ \frac{K_h(x_n - x_1)}{\sum_\iota K_h(x_\iota - x_1)} & \cdots & \frac{K_h(0)}{\sum_\iota K_h(x_\iota - x_n)} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

2.4 Quelques résultats

- Dans Gilles R. Ducharme, Mariem Mint El Mouvid(2001) L'estimateur local linéaire de fonction de répartition conditionnel est représentée par à rapport de deux U-statistiques. En utilisant les résultats de la théorie des U-statistiques, ils présentent deux résulats sur la convergence ponctuelle et uniforme de l'estimateur en question. Ces résultats l'ont permet de démontrer la convergence presque sûre de l'estimateur linéaire local de quantiles conditionnels.

- Amparo Baìllo (2008) a introduit une nouvelle technique de régression non paramétrique dans le contexte de la covariable fonctionnelle et la réponse scalaire. il propose un estimateur de régression locale linéaire et étudie son comportement asymptotique. Son échantillon fini est comparé avec un estimateur à noyau de la régression de type Nadayara-Watson et avec l'estimateur de régression linéaire à travers une étude de simulation avec la méthode Monte Carlo. Il a conclu que l'estimateur locale linéaire de la regression est plus performant que l'estimateur à noyau dans le sens que l'erreur quadratique moyenne est inférieure.

- André Mas(2009) aborde le problème de la régression non paramétrique locale avec des variables fonctionnelles. En premier lieu, il propose un estimateur de la fonction de régression inconnue. La construction de cet estimateur est lié à la résolution d'un problème inverse linéaire.

  En utilisant une méthode classique de décomposition, il établit une borne pour l'erreur quadratique moyenne. Cette borne dépend de la probabilité des petites boules de la variable explicative qui est supposé appartenir á la classe des fonctions gamma.

- J. Barrientos-Marin (2009), son travail traite de la régression non paramétrique d'une variable réponse scalaire sur une variable explicative fonctionnelle. Il propose de modéliser le comportement locale de l'opérateur de régression (i.e le lien entre la variable réponse scalaire et la variable explicative fonctionnelle). Après l'écriture explicite de l'estimateur, une étude des propriétées asymptotiques est instalées et justifier par des simulation à travers des données simulés.

Chapitre 3

Estimation local linéaire de la densité conditionnelle.

Dans cet chapitre, nous introduisons une nouvelle alternative d'estimation non paramétrique de la densité conditionnelle de variable de réponse scalaire donnée une variable aléatoire à valeurs dans un espace semi-métrique. Sous certaines conditions générales, nous établissons les consistances presque complets ponctuelles et uniformes avec les vitesses de convergence.

3.1 Introduction

Dans ce paragraphe, on présente une nouvelle technique d'estimation non paramétrique IL s'agit de la modélisation polynomiale locale de la fonction de la densité conditionnelle lorsque la variable explicative est de type fonctionnel. Il est bien connu qu'un lissage polynomiale locale a plusieurs avantages par rapport à la méthode du noyau, à savoir les propriétés de la polarisation.

Notons que, les questions dans les espaces de dimension éventuellement infinie sont particulièrement intéressante, à la fois pour les problémes fondamentaux qu'ils formulent, mais aussi pour de nombreuses applications qu'elles ouvrent. En outre, la méthode du noyau est connu d'être un cas particulier de la méthode de polynômes locaux.

Nous signalons que, la densité conditionnelle joue un rôle important dans la prédiction non paramétrique, il existe plusieurs outils en statistique non paramétrique, tels que le mode conditionnel, la médiane conditionnelle ou les quantiles conditionnels, qui sont basées sur l'estimateur prééliminaire du "paramètre" fonctionnelle proposée dans cet chapitre. Dans la statistique non paramétrique fonctionnelle, les premiers résultats sur la consistance presque complète ont été obtenus par Fan, J. and Gijbels, I. (1996), pour l'estimation des densité conditionnelles / distribution lorsque les données sont indépendantes et identiquement distribuées. La cas de mélange fort a été étudié par Ferraty, F. et al (2008)

D'autre part, la convergence en norme L^p de l'estimateur du noyau du mode conditionnel a été traité par Ezzahrioui, M et Ould-Saïd, E (2008), et certains propriétees asymptotiques du quantile conditionnel et estimateurs de mode par Fan, J. and Gijbels, I. (1996).

Pour certains progrés les plus récents dans les statistiques non paramétriques pour les données fonctionnelles nous référons à Benhenni, K et al (2007), Baìllo, A. and Grané, A. (2009) et Youndjé, E. (1993).
Dans ce travail, nous introduisons une nouvelle estimation non paramétrique de la densité conditionnelle des données fonctionnelles. Notre estimation est basée sur l'approche linéaire local. Notons que l'estimateur linéaire local de la densité conditionnelle a été largement étudiée, lorsque le la variable explicative prend ces valeurs dans un espace de dimension finie.
Nous nous intéressons à prouver, sous certaines conditions générales, la convergence presque complète avec vitesse de convergence. L'intérêt de la consistance uniforme vient principalement du fait que l'exécution ponctuelle de tous les estimateurs n'est pas suffisante pour quantifier son efficacité, mais, une certaine stabilité est nécessaire.

Signalons ici que, dans la statistique fonctionnelle, la convergence uniforme n'est pas un prolongement direct des résultats ponctuelles précédentes, mais elle nécessite quelques outils supplémentaires et conditions. finalemant nous mettrons l'accent sur les conséquences des résultats précédents àl'estimation du mode conditionnel.

3.2 Modèle

Nous introduisons n paires de variables aléatoires (X_i, Y_i) pour $i = 1, ..., n$ du couple (X, Y) à valeur dans $F \times R$, où F est un espace semi-métrique équipé de la semi-métrique d.

En outre, nous supposons qu'il existe une version régulière de la probabilité conditionnelle de Y donné X, ce qui est absolument continue par rapport à la mesure de Lebesgue sur R et a densité bornée, notée f_x. Lissage polynomiale locale est basée sur l'hypothèse ce paramètre fonctionnel est assez lisse pour être localement bien approchée par un polynôme. Dans la statistique fonctionnelle, il existe plusieurs moyens pour étendre les idées linéaires locales.

Ici, nous adoptons une modélisation fonctionnelle locale , c'est à dire nous estimons la densité conditionnel f_x par $\hat{a}$ qui est obtenue en minimisant la quantité suivante :

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} (h_H^{-1} H(h_H^{-1}(y - Y_i)) - a - b\beta(X_i, x))^2 K(h_H^{-1}\delta(x, X_i)) \tag{3.1}$$

où $\beta(.,.)$ est une fonction connue à partir de $\mathcal{F}^2$ dans $\mathbb{R}$ telle que, $\forall \xi \in \mathcal{F}, \beta(\xi, \xi) = 0$, avec K et H sont des noyaux et $h_K = h_{K,n}$(resp. $h_K = h_{K,n}$) est choisi comme une suite de réels positifs $\delta(.,.)$ est une fonction de $F \times F$ tel que $d(.,.) = |\delta(.,.)|$. De toute évidence, par une simple algèbre, on obtient explicitement la définition suivante de $\hat{f}^x$:

$$\hat{f}^x = \frac{\sum_{i,j=1}^{n} W_{ij}(x) H(h_H^{-1}(y - Y_i))}{h_H \sum_{i,j=1}^{n} W_{ij}(x)} \tag{3.2}$$

où

$$W_{ij}(x) = \beta(X_i, x)(\beta(X_i, x) - \beta(X_j, x))K(h_K^{-1}\delta(x, X_i))K(h_K^{-1}\delta(x, X_j)).$$

Remarque 3.2.1. *Évidemment, si $b = 0$, alors on obtient à partir de (3.1) l'estimateur Nadaraya-Watson a étudié, dans le cas fonctionnel.*

La minimisation de (3.1) peut être réalisée par un $\hat{b}$ "ondulée" qui oblige $\hat{f}^x$ de s'adapter à tous les points de données dans un voisinage de x .

3.3 convergence presque complète

Dans ce qui suit x désigne un point fixe dans F, N_x désigne un quartier fixe de x, $S_{\mathbb{R}}$ être un compact fixe de $\mathbb{R}$, et $\phi_x(r_1, r_2) = \mathbb{P}(r_2 \leq \delta(X, x) \leq r_1)$.

Notez que notre modèle non paramétrique sera assez générale dans le sens que nous avons juste besoin des hypothèses suivantes :

(H1) pour toute $r > 0, \phi_x(r) := \phi_x(-r, r) > 0$

(H2) f^x La densité conditionnelle est telle que :il existe $b_1 > 0, b_2 > 0, \forall (y_1, y_2) \in S_{\mathbb{R}^2}$ et $\forall (x_1, x_2) \in N_x \times N_x$

$$|f^{x_1}(y_1) - f^{x_2}(y_2)| \leq C_x \left(d^{b_1}(x_1, x_2) + |y_1 - y_2|^{b_2} \right),$$

où C_x est une constante positive dépendant de x.

(H3) $\beta(.,.)$ la fonction est tel que :

$$\forall x, x' \in \mathcal{F}, C_1 d(x, x') \leq |\beta(x, x')| \leq C_2 d(x, x'), \ o, C_1 > 0, C_2 > 0.$$

(H4) K est une fonction positive, différentiables avec support $[-1, 1]$

(H5) H est une fonction positive bornée, lipschitzienne en continu, de telle sorte que :

$$\int |t|^{b_2} H(t) dt < \infty \ \ et \ \ \int H^2(t) dt < \infty.$$

(H6) La bande passante h_K satisfait : il existe un n_0 entier, tels que

$$\forall n > n_0, -\frac{1}{\phi_x(h_K)} \int_{-1}^{1} \phi_x(z h_K, h_K) \frac{d}{dz}(z^2 K(z)) dz > C_3 > 0$$

et

$$h_K \int_{\mathbf{B}(x, h_K)} \beta(u, x) dP(u) = o \left(\int_{B(x, h_K)} \beta^2(u, x) dP(u) \right)$$

où $B(x, r) = x' \in \mathcal{F}(x, x') \leq r$ désigne la boule fermée de centre x et de rayon r, et $dP(x)$ est la distribution cumulative de X.

(H7) Le paramètre de lissage h_H satisfait :

$$\lim_{n \to \infty} n \gamma h_H = \infty \ \ pour \ \ certains \ \gamma > 0 \ \ et \ \ \lim_{n \to \infty} \frac{\ln n}{n h_K \phi_x(h_K)} = 0.$$

Remarquez que ces conditions sont trés standard dans ce contexte. Les conditions (H1), (H3), et (H6) sont les mêmes. Hypothèses (H2) est une condition

de régularité qui caractérise l'espace fonctionnel de notre modèle et il est nécessaire d'évaluer le terme de biais dans les résultats asymptotiques de ce travaille. Les hypothèses (H5) et (H7) sont des conditions techniques et sont, aussi, similaires.
Le théorème suivant donne la convergence presque-complet (a.co.) de $\hat{f}_x$.

Théorème 3.3.1. *Sous les hypothèses (H1), (H2), (H3), (H4), (H5), (H6) et (H7), nous avoir que :*

$$\sup_{y\in S_{\mathbb{R}}} |\hat{f}^x(y) - f^x(y)| = o(h_K^{b_1} + h_H^{b_2}) + o\left(\sqrt{\frac{\ln n}{nh_H\phi_x(h_K)}}\right), \quad a.co.$$

Nous remarquons que la démonstration du théoréme 3.3.1 est une conséquence directe de la décomposition :

$$\forall y \in S_{\mathbb{R}}, \hat{f}^x(y) - f^x(y) = \frac{1}{\hat{f}_D^x}\left\{(\hat{f}_N^x(y) - \mathbb{E}[\hat{f}_N^x(y)]) - (f^x(y) - \mathbb{E}[\hat{f}_N^x(y)])\right\} + \frac{f^x(y)}{\hat{f}_D^x}(1-\hat{f}_D^x), \tag{3.3}$$

où

$$\hat{f}_N^x(y) = \frac{1}{n(n-1)h_H\mathbb{E}[W_{12}(x)]}\sum_{i\neq j} W_{ij}(x)H(h_H^{-1}(y-Y_i)).$$

et

$$\hat{f}_D^x = \frac{1}{n(n-1)h_H\mathbb{E}[W_{12}(x)]}\sum_{i\neq j} W_{ij}(x)$$

Lemme 3.3.1. *Sous les hypothèses (H1), (H3), (H4) et (H6), nous avons ce qui suit :*

$$1 - \hat{f}_D^x = o\left(\sqrt{\frac{\ln n}{n\phi_x(h_K)}}\right), a.co$$

et

$$\exists \delta > 0, \quad \textit{de telle sorte que} \sum_{i=1}^{\infty} \mathbb{P}(\hat{f}_D^x < \delta) < \infty.$$

Lemme 3.3.2. *Sous les hypothèses (H1), (H2) et (H4), on obtient :*

$$\sup_{y\in S_{\mathbb{R}}} |f^x(y) - \mathbb{E}[\hat{f}_N^x(y)]| = o(h_K^{b_1} + h_H^{b_2})$$

Lemme 3.3.3. *Sous les hypothèses du théorème 4.2.1, nous obtenons :*

$$\sup_{y\in S_{\mathbb{R}}} |\hat{f}_N^x(y) - \mathbb{E}[\hat{f}_N^x(y)] = o\left(\sqrt{\frac{\ln n}{nh_H\phi_x(h_K)}}\right), a.co$$

3.4 convergence uniforme presque complète

Cette section est consacrée à la version uniforme du théorème 4.2.1. Plus précisément, notre objectif est d'établir la convergence uniforme presque complète de $\hat{f}^x$ sur un sous-ensemble de F de $\mathcal{F}$, comme

$$S_{\mathcal{F}} \subset \bigcup_{k=1}^{d_n} B(x_k, r_n),$$

où $x_k \vdash \mathcal{F}$ et rn (resp. d_n) est une suite de nombres réels positifs.

Dans la pratique, la consistance uniforme a une grande importance car elle nous permet de faire la prédiction, même si les données ne sont pas parfaitement respectées. Par ailleurs, les résultats de la convergence uniforme sont des outils indispensables pour certains problèmes de données non paramètrique comme pour le choix du paramètre de lissage. Il est intèressant de noter que, dans le cas multivarié, la consistance uniforme est une extension standard de la consistance ponctuelle, cependant, dans notre cas fonctionnel, quelques outils supplèmentaires et des conditions topologiques sont nécessaires. Ainsi, en plus des conditions présentées dans la section précédente, nous avons besoin des hypothèses suivantes :

(U1) Il existe une fonction $\phi(.)$ différentiable, tel que :

$$\forall x \in S_{\mathcal{F}}, 0 < C\phi(h) \leq \phi_x(h) \leq C'\phi(h) < \infty \; et \; \exists \eta_0 > 0, \forall \eta < \eta_0, \phi'(\eta) < C,$$

où C et C' sont des constantes strictement positives et où ϕ' représente la dérivée première de ϕ.

(U2) La densité conditionnelle f^x satisfait, pour une constante strictement positive C, ce qui suit :

$$\forall (y_1, y_2) \in S_{\mathbb{R}} \times S_{\mathbb{R}}, \forall (x_1, x_2) \in S_{\mathcal{F}} \times S_{\mathcal{F}}$$

$$|f^{x_1}(y_1) - f^{x_2}(y_2)| \leq C(d^{b_1}(x_1, x_2) + |y_1 - y_2|^{b_2}).$$

(U3) La fonction $\beta(.,.)$ satisfait $(H3)$ et, pour une constante strictement positive C' , la condition de Lipschitz suivante :

$$\forall (x_1, x_2) \in S_{\mathcal{F}} \times S_{\mathcal{F}} |\beta(x_1, x') - \beta(x_2, x')| \leq Cd(x_1, x_2)$$

(U4) le noyau K satisfait (H4) et, pour certains constante C strictement positive, la condition de Lipschitz suivante :

$$|K(x) - K(y)| \leq C||x| - |y||$$

(U5) pour certains $\gamma \in (0,1)$, $\lim\limits_{n\to\infty} n^{\gamma} h_H = \infty$, *et* $r_n = o\left(\frac{\ln n}{n}\right)$ la séquence d_n satisfait :

$$\frac{(\ln n)^2}{n^{1-\gamma}\phi(h_K)} < \ln d_n < \frac{n^{1-\gamma}\phi(h_K)}{\ln n},$$

et

$$\sum_{n=1}^{\infty} n^{(3\gamma+1)|2} d_n^{1-\beta} < \infty$$

pour certains $\beta > 1$.

Notez que les conditions $(U1)$ et $(U2)$ sont, respectivement, les versions uniformes $(H1)$ et $(H2)$. En effet, les conditions $(U1)$ et $(U5)$ sont liés àla structure topologique de la variable fonctionnelle. Par conséquent, comme dans le cas ponctuel, le choix de la structure topologique, contrôlé ici par le biais de la fonction $\delta(.,.)$, Joue un rôle crucial.Donc, un bon choix de cette fonction améliore la vitesse de convergence de l'estimateur.Plus précisément, nous verrons par la suite, qu'un bon semi-métrique est que l'augmentation de la concentration de la mesure de probabilité des variables X fonctionnels ainsi que minimise d_n. Il faut noter que les deux conditions $(U1)$ et $(U2)$ sont vérifiées pour plusieurs-processus en temps continu.

Théorème 3.4.1. *Sous les hypothèses* $(U1)$, $(U2)$, $(U3)$, $(U4)$, $(U5)$, $(H5)$ *et* $(H6)$, *nous avoir que :*

$$\sup_{x\in S_{\mathcal{F}}} \sup_{x\in S_{\mathbb{R}}} |\hat{f}^x(y) - f^x(x)| = o(h_H^{b_1}) + o(h_H^{b_2}) + o_{a.co.}\left(\sqrt{\frac{\ln d_n}{n^{1-\gamma}\phi(h_K)}}\right) \tag{3.4}$$

Lemme 3.4.1. *Sous les hypothèses* $(U1)$, $(U3)$, $(U4)$, $(U5)$ *et* $(H6)$, *on obtient que :*

$$\sup_{x\in S_{\mathcal{F}}} |\hat{f}_D^x - 1| = o_{a.co.}\left(\sqrt{\frac{\ln d_n}{n\phi(h_K)}}\right)$$

Preuve :

En effet,

$$\hat{f}_D^x = T_1 \left[\underbrace{\left(\frac{1}{n}\sum_{j=1}^{n} \frac{K_j(x)}{\phi_x(h_K)}\right)}_{S_1(x)} \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} \frac{K_i(x)\beta_i^2(x)}{h_K^2\phi_x(h_K)}\right)}_{S2(x)} \right.$$

$$\left. - \underbrace{\left(\frac{1}{n}\sum_{j=1}^{n}\frac{K_j(x)\beta_j(x)}{h_K\phi_x(h_K)}\right)}_{S3(x)} \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{K_i(x)\beta_i(x)}{h_K\phi_x(h_K)}\right)}_{S4(x)} \right]$$

et de la même façon, tout ce qu'il reste à montrer les convergences sont uniformes suivantes :

$$\sup_{x\in S_{\mathcal{F}}} |S_k(x) - \mathbb{E}[S_k(x)]| = o\left(\sqrt{\frac{\ln nd_n}{n\phi_x(h_K)}}\right), \text{ a.co. pour } k = 2,3,4, \qquad (3.5)$$

et $\sup_{x\in S_{\mathcal{F}}} |\mathbb{E}[S_2(x)]\mathbb{E}[S_4(x)]\mathbb{E}[S_2(x)S_4(x)] - var[S_3(x)]| = o\left(\sqrt{\frac{\ln nd_n}{n\phi_x(h_K)}}\right)$, $a.co.$

et , également que, de façon uniforme sur $x \in S_{\mathcal{F}}$: $T_1 = o(1)$ et $|\mathbb{E}[S_k(x)]| = o(1), pour\ k = 2,3,4.$

De toute évidence, les deux dernières équations sont des conséquences directes de l'Assomption (U1) . En effet, en notant : $j(x) = \arg\min_{j\in\{1,2,\ldots,d_n\}} |\delta(x,x_k)|$,nous considérons la décomposition suivante :

$$\sup_{x\in S_{\mathcal{F}}} |S_k(x) - \mathbb{E}[S_k(x)]| \leq \underbrace{\sup_{x\in S_{\mathcal{F}}} |S_k(x) - S_k(x_{j(x)})|}_{F_1^k}$$

$$+ \underbrace{\sup_{x\in S_{\mathcal{F}}} |S_k(x_{j(x)}) - \mathbb{E}[S_k(x_{j(x)})]|}_{F_2^k} + \underbrace{\sup_{x\in S_{\mathcal{F}}} |\mathbb{E}[S_k(x_{j(x)})] - \mathbb{E}[S_k(x)]|}_{F_3^k}.$$

Nous avons , ensuite , d' évaluer chaque terme $F_j^k, j = 2,3,4.$ depuis F_1^k et F_3^k avoir presque le même traitement, nous allons examiner les deux points suivants : Traitement des termes F_1^k et F_3^k . Tout d'abord, nous analysons le premier terme F_1^k pour $k = 2,3,4$. Comme K est pris en charge $[-1,1]$, nous pouvons écrire pour tout $k = 2, 3, 4$, que :

$$\begin{aligned}
F_1^k &\leq \frac{1}{nh_K^{k-2}\phi_x(h_K)} \sup_{x\in S_{\mathcal{F}}} \sum_{i=1}^{n} |K_i(x)\beta_i^{k-2}(x)1_{B(x,h_K)}(X_i) \\
&\quad -K_i(x_{j(x)})\beta_i^{k-2}(x_{j(x)})1_{B(x,h_K)}(X_i)| \\
&\leq \frac{C(k-2)}{nh_K^{k_2}\phi_x(h_K)} \sup_{x\in S_{\mathcal{F}}} \sum_{i=1}^{n} K_i(x)1_{B(x,h_K)}(X_i) \\
&\quad +\frac{1}{nh_K^{k-2}\phi_x(h_K)} \sup_{x\in S_{\mathcal{F}}} \sum_{i=1}^{n} \beta_i^{k-2}((x_{j(x)}))1_{B(x_{j(x)},h_K)}(X_i) \\
&\quad \times|K_i(x)1_{B(x,h_K)}(X_i) - K_i(x_{j(x)})|.
\end{aligned}$$

$$\begin{aligned}
F_1^k &\leq \frac{1}{nh_K^{k-2}\phi_x(h_K)} \sup_{x\in S_{\mathcal{F}}} \sum_{i=1}^{n} |K_i(x)\beta_i^{k-2}(x)1_{B(x,h_K)}(X_i) \\
&\quad -K_i(x_{j(x)})\beta_i^{k-2}(x_{j(x)})1_{B(x,h_K)}(X_i)| \\
&\leq \frac{C(k-2)}{nh_K^{k_2}\phi_x(h_K)} \sup_{x\in S_{\mathcal{F}}} \sum_{i=1}^{n} K_i(x)1_{B(x,h_K)}(X_i) \\
&\quad \times|\beta_i^{k-2}(x) - \beta_i^{k-2}(x_{j(x)})1_{B(x_{j(x)})}, h_K)(X_i)| \\
&\quad +\frac{1}{nh_K^{k-2}\phi_x(h_K)} \sup_{x\in S_{\mathcal{F}}} \sum_{i=1}^{n} \beta_i^{k-2}((x_{j(x)}))1_{B(x_{j(x)},h_K)}(X_i) \\
&\quad \times|K_i(x)1_{B(x,h_K)}(X_i) - K_i(x_{j(x)})|.
\end{aligned}$$

La condition de Lipschitz sur K nous permet d'écrire directement

$$\begin{aligned}
&1_{B(x_{j(x)},h_K)}(X_i)|K_i(x)1_{B(x,h_K)}(X_i) - K_i(x_{j(x)})| \\
&\leq C_\epsilon 1_{B(x,h_K)\bigcap B(x_{j(x)},h_K)}(X_i)C1_{B(x_{j(x)},h_K)\bigcap \overline{B(x,h_K)}}(X_i).
\end{aligned}$$

De la même manière, la condition de Lipschitz sur β donne

$$1_{B(x_{j(x)},h_K)}(X_i)|\beta_i^2(x) - \beta_i^2(x_{j(x)})1_{B(x_{j(x)})}, h_K)(X_i)|$$

$$\leq \epsilon h_K 1_{B(x,h_K)\bigcap B(x_{j(x)},h_K)}(X_i)h_K^2 1_{B(x,h_K)\bigcap \overline{B(x_{j(x)},h_K)}}(X_i).$$

ce qui implique que, pour $k = 3, 4$

$$1_{B(x,h_K)}(X_i)\beta_i^{k-2}(x) - \beta_i^{k-2}(x_{j(x)})1_{B(x_{j(x)})}, h_K)(X_i)$$

$$\leq \epsilon h_K^{k-3} 1_{B(x,h_K)\bigcap B(x_{j(x)},h_K)}(X_i)h_K^{k-2} 1_{B(x,h_K)\bigcap \overline{B(x_{j(x)},h_K)}}(X_i).$$

ainsi,

$$F_1^K \leq c \sup_{x \in S_{\mathcal{F}}} (F_{11}^k + F_{12} + F_{13}^k + F_{14}),$$

où

$$F_{11} = \frac{C(k-2)}{n\phi_x(h_K)} \sum_{i=1}^{n} 1_{B(x,h_K) \bigcap \overline{B(x_{j(x)},h_K)}}(X_i).$$

$$F_{12} = \frac{C\epsilon}{n\phi_x(h_K)} \sum_{i=1}^{n} 1_{B(x,h_K) \bigcap B(x_{j(x)},h_K)}(X_i).$$

$$F_{13} = \frac{C(k-2)\epsilon}{nh_K\phi_x(h_K)} \sum_{i=1}^{n} 1_{B(x,h_K) \bigcap B(x_{j(x)},h_K)}(X_i).$$

$$F_{14} = \frac{C}{n\phi_x(h_K)} \sum_{i=1}^{n} 1_{B(x_{j(x)},h_K) \bigcap \overline{B(x,h_K)}}(X_i).$$

Maintenant, pour évaluer ces termes F_{11}^K,F_{12},F_{13}^k et F_{14}^k, nous appliquons une inégalité norme pour des sommes de variables aléatoires bornées avec Z_i est identifié de telle sorte que :

$$Z_i = \begin{cases} \frac{1}{\phi_x(h_K)} \sup_{x \in S_{\mathcal{F}}} [1_{B(x,h_K) \bigcap \overline{B(x_{j(x)},h_K)}}(X_i)] \quad pour F_{11}^K \\ \\ \frac{\epsilon}{\phi_x h_K(h_K)} \sup_{x \in S_{\mathcal{F}}} [1_{B(x,h_K) \bigcap B(x_{j(x)},h_K)}(X_i)] \quad pour\ F_{12}\ F_{13}^K \\ \\ \frac{1}{\phi_x(h_K)} \sup_{x \in S_{\mathcal{F}}} [1_{B(x_{j(x)},h_K) \bigcap \overline{B(x,h_K)}}(X_i)] \quad pour\ F_{14} \end{cases}$$

De toute évidence, dans la deuxième partie de $(U1)$, nous avons pour le premier et le dernier cas :

$$Z_1 = o\left(\frac{1}{\phi(h_K)}\right), \mathbb{E}[Z_1] = o\left(\frac{\epsilon}{\phi(h_K)}\right)\ et var(Z_1) = o\left(\frac{\epsilon}{(\phi(h_K))^2}\right)$$

Ainsi, nous obtenons :

$$F_{11}^k = o\left(\frac{\epsilon}{\phi(h_K)}\right) + o_{a.co.}\left(\sqrt{\frac{\epsilon \ln n}{n\phi(h_K)^2}}\right).$$

De la même façon, l'hypothèse $(U5)$ permet d'obtenir, pour F_{12} ou F_{13}^k cas

$$Z_1 = o\left(\frac{\epsilon}{h_K\phi(h_K)}\right) = \mathbb{E}[Z_1] = o\left(\frac{\epsilon}{h_K}\right)\ et\ var(Z_1) = o\left(\frac{\epsilon}{h_K^2\phi(h_K)}\right),$$

ce qui implique que :

$$F_{12}^k = o_{a.co.}\left(\sqrt{\frac{\ln d_n}{n\phi(h_K)}}\right).$$

Pour réaliser l'étude de la F_1 terme, il suffit de mettre l'ensemble des résultats intermédiaires et à utiliser (U_5) pour obtenir :

$$F_1^k = o_{a.co}\left(\frac{\ln d_n}{n\phi(h_K)}\right). \tag{3.6}$$

En outre, depuis :

$$F_3^k \leq \left[\sup_{x\in S\mathcal{F}} |S_k(x) - S_k(x_{j(x)})|\right]$$

nous avons également :

$$F_3^k = o\left(\sqrt{\frac{\ln d_n}{n\phi(h_K)}}\right).$$

Le traitement de la terme F_2^k . Pour tout $\eta > 0$, on a que :
$\mathbb{P}\left(F_2^k > \eta\sqrt{\frac{\ln d_n}{n\phi(h_K)}}\right) = \mathbb{P}\left(\max_{j\in 1,\ldots,d_n} |S_k(x_j(x)) - \mathbb{E}[S_k(x_j(x))]| > \eta\sqrt{\frac{\ln d_n}{n\phi(h_K)}}\right)$

$$\leq d_n \max_{j\in 1,\ldots,d_n} \mathbb{P}|S_k(x_j) - \mathbb{E}[S_k(x_j)]| > \eta\sqrt{\frac{\ln d_n}{n\phi(h_K)}}$$

définir

$$\Delta_{ki} = \frac{1}{nh_K^{k-2}\phi(h_K)}\left(K_i(x_k)\beta_i^{k-2}(x_k) - \mathbb{E}[K_i(x_k)\beta_i^{k-2}(x_k)]\right), pour k = 2,3,4.$$

nous obtenons pour tout $j = 1, ..., d_n$ et $i = 1, ..., n$ qui :

$$\mathbb{E} = |\Delta_{ki}|^m = o(\phi(h_K)^{-m+1}), pour k = 2,3,4.$$

Ainsi, on peut appliquer une inégalité de type Bernstein pour obtenir directement :

$$\mathbb{P} = |S_i(x_k) - \mathbb{E}[S_i(x_k)]| > \eta\sqrt{\frac{\ln d_n}{n\phi(h_K)}}$$

$$\mathbb{P} = \left(\frac{1}{n}|\sum_{i=1}^{n}\Delta_{\iota ki}| > \eta\sqrt{\frac{\ln d_n}{n\phi(h_K)}}\right)$$

$$\leq 2\exp\{-C\eta^2 \ln d_n\}$$

Ainsi, en choisissant η telle que $C\eta^2$, on obtient :

$$d_n \max_{k\in 1,\ldots,d_n} \mathbb{P} = |S_i(x_k) - \mathbb{E}[S_i(x_k)]| > \eta\sqrt{\frac{\ln d_n}{n\phi(h_K)}} \leq C' d_n^{1-\beta} \tag{3.7}$$

depuis $\sum_{i=1}^{\infty} d_n^{1-\beta}$, nous obtenons que :

$$F_2 = o_{a.co}\left(\sqrt{\frac{\ln d_n}{n\phi(h_K)}}\right).$$

■

Corollaire 3.4.1. *Sous les hypothèses du lemme précidant, nous avons ce qui suit :*

$$\sum_{n=1}^{\infty} \mathbb{P}(\inf_{x\in S_{\mathcal{F}}} |\hat{f}_D^x < \frac{1}{2}|) < \infty.$$

Preuve :

De toute évidence, nous avons ce qui suit :

$$\inf_{x\in S_{\mathcal{F}}} \hat{f}_D(x) \leq \frac{1}{2} \Rightarrow \exists x \in S_{\mathcal{F}},\ \textit{de telle sorte que}\ 1-\hat{f}_D(x) \geq \frac{1}{2} \Rightarrow \sup_{x\in S_{\mathcal{F}}} |1-\hat{f}_D(x)| \geq \frac{1}{2}.$$

on obtient :

$$\mathbb{P}\left(\inf_{x\in S_{\mathcal{F}}} \hat{f}_D(x) \leq \frac{1}{2}\right) \leq \left(\mathbb{P} \sup_{x\in S_{\mathcal{F}}} |1-\hat{f}_D(x)| \geq \frac{1}{2}.\right)$$

En conséquence :

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\inf_{x\in S_{\mathcal{F}}} |\hat{f}(x)| < \frac{1}{2}\right) < \infty.$$

■

Lemme 3.4.2. *Sous les hypothèses* $(U1)$, $(U2)$ *et* $(H5)$, *on obtient ce qui suit :*

$$\sup_{x\in S_{\mathcal{F}}} \sup_{x\in S_{\mathbb{R}}} |f^x(y) - \mathbb{E}[\hat{f}_N^x(y)]| = o(h_H^{b_1}) + o(h_H^{b_2})$$

Preuve :

Il suffit de combiner les preuves des lemmes précédents, et en supposant que l'état du Lipschitz uniformément sur (x, y) dans $S_{\mathcal{F}} \times S_{\mathbb{R}}$.

■

Lemme 3.4.3. *Sous les hypothèses du théorème précidant, nous obtenons que :*

$$\sup_{x\in S_{\mathcal{F}}}\sup_{x\in S_{\mathbb{R}}}|\hat{f}_N^x(y)-\mathbb{E}[\hat{f}_N^x(y)]| = o_{a.co.}\left(\sqrt{\frac{\ln d_n}{n^{1-\gamma}\phi(h_K)}}\right)$$

Preuve :

La preuve de ce lemme suit les étapes que pour prouver le lemme précédant, où $S_2(x)$ et $S_4(x)$, sont remplacées par :

$$\begin{cases} T_2^x(y) &= \frac{1}{n}\sum_{j=1}^n \frac{K_i(x)H_j(y)}{h_H\phi_x(h_K)} \\ T_3^x(y) &= \frac{1}{n}\sum_{j=1}^n \frac{K_j(x)\beta_j H_j(y)}{h_K h_H\phi_x(h_K)} \\ T_4^x(y) &= \frac{1}{n}\sum_{j=1}^n \frac{K_j(x)\beta_j^2 H_j(y)}{h_K^2 h_H\phi_x(h_K)} \end{cases}$$

Pour ce faire, nous gardons les notations utilisées précédemment, à savoir, les définitions de $j(x)$, t_y et ι_n. La preuve est basée sur la décomposition suivante qui sera utilisée pour les trois termes :

$$\begin{aligned} |T_i^x(y)-\mathbb{E}[T_i^x(y)]| \leq & \underbrace{\sup_{x\in S_{\mathcal{F}}}\sup_{x\in S_{\mathbb{R}}}|T_i^x(y)-T_i^{x_{j(x)}}(y)|}_{E1} \\ & + \underbrace{\sup_{x\in S_{\mathcal{F}}}\sup_{x\in S_{\mathbb{R}}}|T_i^{x_{j(x)}}(y)-T_i^{x_{j(x)}}(t_y)|}_{E2} \\ & + \underbrace{\sup x\in S_{\mathcal{F}} \sup x\in S_{\mathbb{R}}|T_i^{x_{j(x)}}(t_y)-\mathbb{E}[T_i^{x_{j(x)}}(t_y)]|}_{E3} \\ & + \underbrace{\sup_{x\in S_{\mathcal{F}}}\sup_{x\in S_{\mathbb{R}}}|\mathbb{E}[T_i^{x_{j(x)}}(t_y)]-\mathbb{E}[T_i^{x_{j(x)}}(y)]|}_{E4} \\ & + \underbrace{\sup_{x\in S_{\mathcal{F}}}\sup_{x\in S_{\mathbb{R}}}|\mathbb{E}[T_i^{x_{j(x)}}(y)]-\mathbb{E}[T_i^x(y)]|}_{E5}. \end{aligned}$$

De même que pour l'étude de la F_1 terme, nous obtenons

$$E_1 = o_{a.co.}\left(\sqrt{\frac{\ln d_n}{n^{-1-\gamma}\phi(h_K)}}\right) \ et\ E_5 = o\left(\sqrt{\frac{\ln d_n}{n^{-1-\gamma}\phi(h_K)}}\right) \tag{3.8}$$

Concernant le E_2 terme, en utilisant l'état de la Lipschitz sur le noyau H, on peut écrire :

$$T_i^{x_{j(x)}}(y) - T_i^{x_{j(x)}}(t_y) \leq C\frac{1}{nh_K^{\iota}\phi(h_K)}\sum_{i'=1}^{n} k_{i'}(x_{j(x)})\beta_{i'}^{\iota}(x_{j(x)})|H_{i'}(y) - H_{i'}(t_y)| \leq \frac{\iota_n}{h_H^2}S_i(x_{j(x)}),$$

$S_i(.)$ pour $i = 2, 3, 4$, Ainsi, en utilisant les faits que $\lim_{n\to+\infty} n^{\gamma}h_K = \infty$ et $\iota_n = n^{-\frac{3}{2}-\frac{1}{2}}$, nous obtenons :

$$E2 = o_{a.co.}\left(\sqrt{\frac{\ln n}{n^{1-\gamma}\phi(h_K)}}\right) \ et\ E2 = o\left(\sqrt{\frac{\ln n}{n^{1-\gamma}\phi(h_K)}}\right). \tag{3.9}$$

Enfin, pour toute $\eta > 0$,

$$\mathbb{P}\left(E3 > \sqrt{\frac{\ln n}{nh_H\phi(h_K)}}\right).$$

$$= \mathbb{P}\left(\max_{j\in\{1,2,\ldots s_n\}}\max_{k\in\{1,\ldots,d_n\}}|T_i^{x_k}(t_j) - \mathbb{E}[T_i^{x_k}(t_j)]| > \eta\sqrt{\frac{\ln n}{nh_H\phi(h_K)}}\right)$$

$$\leq s_n d_n \max_{j\in\{1,2,\ldots s_n\}}\max_{k\in\{1,\ldots,d_n\}}\mathbb{P}\left(|T_i^{x_k}(t_j) - \mathbb{E}[T_i^{x_k}(t_j)]| > \eta\sqrt{\frac{\ln n}{nh_H\phi(h_K)}}\right)$$

Cette dernière probabilité peut être traitée à l'aide de l'inégalité de Bernstein classique, avec $a_n = (h_H\phi_x(h_K))^{\frac{-1}{2}}$ .Rappelons que, le choix d'un est motivé par moment d'ordre m de $Z_i^{\iota,k}$,en fin , on a

$$\forall j \leq s_n, \mathbb{P}\left(|T_i^{x_k}(t_j) - \mathbb{E}[T_i^{x_k}(t_j)]| > \eta\sqrt{\frac{\ln n}{nh_H\phi(h_K)}}\right) \leq 2\exp\{-C\eta^2\ln d_n\}.$$

Par conséquent, depuis $s_n = o(\iota_n^{-1}) = o(n^{\frac{3}{2}\gamma+\frac{1}{2}})$,et en choisissant $C\eta^2 = \beta$ on :

$$s_n d_n \max_{j\in\{1,2,\ldots s_n\}}\max_{k\in\{1,\ldots,d_n\}}\mathbb{P}\left(|T_i^{x_k}(t_j) - \mathbb{E}[T_i^{x_k}(t_j)]| > \eta\sqrt{\frac{\ln n}{nh_H\phi(h_K)}}\right) \leq C's_n d_n^{1-C\eta^2}.$$

En utilisant le fait que $\lim_{n\to\infty} n^{\gamma} h_H = \infty$ et la seconde partie de la condition $(U5)$,on obtient :

$$E3 = o_{a.co.}\left(\sqrt{\frac{\ln d_n}{n^{1-\gamma}\phi(h_K)}}\right). \tag{3.10}$$

■

3.5 Application

Etudions maintenant la convergence presque complète de l'estimateur à noyau du mode conditionnel de Y sachant $X = x$, noté $\theta(x)$, uniformément sur un sous-ensemble fixe "compact" $S_{\mathcal{F}}$ de $\mathcal{F}$. Pour ce faire, nous supposons que $\theta(x)$ satisfait, le $S_{\mathcal{F}}$, la propriété d'unicité uniforme suivant :

(U6) $\forall \epsilon_0 > 0, \exists \eta > 0, \forall r : S \to S_{\mathbb{R}}$, nous avons ce qui suit :

$$\sup_{x\in S_{\mathcal{F}}} |\theta(x) - r(x)| \geq \epsilon_0 \Rightarrow \sup_{x\in S_{\mathcal{F}}} |f^x(r(x)) - f^x(\theta(x))| \geq \eta.$$

En outre, nous supposons, également, qu'il existe une $j > 1$ tel que $\forall x \in S_{\mathcal{F}}$, la fonction f^x est J-fois continûment différentiable sur l'intérieur (SIR) par rapport à y et entier que :

$$\text{(U7)} \quad \begin{cases} f^{x(\iota)}(\theta(x)) = 0, \text{ si } 1 \leq \iota < j, \\ \\ \text{et } f^{x(j)}(.) \text{ est uniformément continue sur } \mathbb{R} \\ \\ \text{de telle sorte que } f^{x(j)}(\theta(x))| > C > 0, \end{cases}$$

où $f^{x(j)}$ désigne la dérivée d'ordre j du f^x de densité conditionnelle. Nous estimons que le mode conditionnel $\theta(x)$ par la variable aléatoire $\hat{\theta}(x)$ définie par :

$$\hat{\theta}(x) = \arg\sup_{y\in S_{\setminus}} \hat{f}^x(y).$$

Ainsi, d'aprés le théorème 3.4.1, on déduit le corollaire suivant.

Corollaire 3.5.1. *Sous les hypothèses du théorème 3.4.1, et si la densité conditionnelle f_x satisfait les hypothèses $(H9)$ et $(H10)$, nous obtenons :*

$$\sup_{x\in S_{\mathcal{F}}} |\hat{\theta}(x) - \theta(x)|^j = o(h_H^{b_1}) + o(h_H^{b_2}) + o_{a.co.}\left(\sqrt{\frac{\ln d_n}{n^{1-\gamma}\phi(h_K)}}\right)$$

conclusion

nous avons abordé une étude globale de l'estimation de la densité conditionnelle quand la variable explicative est fonctionnelle. LA méthode d'estimation concerne la méthode d'estimation par polynômes locaux.

Nous avons considéré une approche pour l'estimation de la densité conditionnelle quand les données sont fonctionnelles. L'estimateur proposé est une généralisation, au cas fonctionnel, de l'estimateur local linéaire introduit par Fan et Gijbels (1996). Comme résultats asymptotiques, nous avons établi la vitesse de convergence presque complète (ponctuelle et uniforme).

Les expressions des vitesses de convergence obtenues ont la même forme que l'estimateur à noyau, dont les deux structures fonctionnelles sont bien exploitées. Plus précisément, la dimensionalité du modèle dans la partie biais, alors que la dimensionalité de l'espace fonctionnel de la variable explicative a été explicitée dans la partie dispersion.

Bibliographie

[1] Fethi-Madani.(2012) Functional data : local linear estimation of the conditional density and its application **6**, Pages 2–8.

[2] Barrientos-Marin, J. (2007). Some Practical Problems of Recent Nonparametric Procedures : Testing, Estimation, and Application. PhD thesis from the Alicante University (Spain).

[3] Barrientos-Marin, J., Ferraty, F. and Vieu, P. (2010). Locally Modelled Regression and Functional Data. *J. of Nonparametric Statistics*, **22**, No. 5, Pages 617–632.

[4] Benhenni, K., Ferraty, F., Rachdi, M. and Vieu, P. (2007). Local smoothing regression with functional data. *Computational Statistics*, **22**, No. 3, Pages 353–369.

[5] Bosq, D. (2000). *Linear Processes in Function Spaces : Theory and applications.* Lecture Notes in Statistics, **149**, Springer.

[6] Baìllo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response, *Journal of Multivariate Analysis*, **100**, Pages 102–111.

[7] Cai, T.-T. and Hall, P. (2006). Prediction in functional linear regression, *Annals of Statistics*, **34**, Pages 2159–2179.

[8] Chu, C.-K. and Marron, J.-S. (1991). Choosing a kernel regression estimator. With comments and a rejoinder by the authors. *Statist. Sci.*, **6**, Pages 404–436.

[9] Ezzahrioui, M. and Ould-Saïd, E. (2008). Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *J. Nonparametr. Stat.*, **20**, Pages 3–18.

[10] Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, **87**, Pages 998–1004.

[11] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications.* London, Chapman & Hall.

[12] Fan, J. and Yim, T.-H. (2004). A cross-validation method for estimating conditional densities. *Biometrika*, **91**, Pages 819–834.

[13] Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *Journal of statistical planning and inference*, **140**, Pages 335–352.

[14] Ferraty, F., Laksaci, A. and Vieu, P. (2005). Functional times series prediction via conditional mode. *C. R., Math., Acad. Sci. Paris*, **340**, Pages 389–392.

[15] Ferraty, F., Laksaci, A. and Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat. Inference Stoch. Process.,* **9**, Pages 47–76.

[16] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis. Theory and Practice.* Springer Series in Statistics. New York.

[17] Ferraty, F., Van Keilegom, I. and Vieu, P. (2008). On the validity of the bootstrap in nonparametric functional regression. *Scandinavian J. of Statist.*, **37**, No. 2, Pages 286–306.

[18] Hyndman, R. and Yao, Q. (2002). Nonparametric estimation and symmetry tests for conditional density functions. *J. Nonparametr. Stat.*, **14**, Pages 259–278.

[19] Müller, H.-G. and Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Stat.*, **33**, No. 2, Pages 774–805.

[20] Ould-Saïd, E. and Cai, Z. (2005). Strong uniform consistency of nonparametric estimation of the censored conditional mode function. *J. Nonparametr. Stat.,* **17**, Pages 797–806.

[21] Rachdi, M. and Vieu, P. (2007). Nonparametric regression for functional data : automatic smoothing parameter selection. *Journal of Statistical Planning and Inference,* **137**, Pages 2784–2801.

[22] Ramsay, J.-O. and Silverman, B.-W. (1997). *Functional data analysis.* Springer Series in Statistics. New York.

[23] Ramsay, J. O. and Silverman, B. W. (2002). *Applied functional data analysis. Methods and case studies.* Springer Series in Statistics. New York.

[24] Sarda, P. and Vieu, P. (2000). *Kernel Regression.* Pages 43–70, Wiley, New York.

[25] Vieu, P. (1996). A note on density mode estimation. *Statist. Probab. Lett.,* **26**, Pages 297–307.

[26] Youndjé, E. (1993). Estimation non paramétrique de la densité conditionnelle par la méthode du noyau. PhD thesis (in French), Rouen University.