



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

CENTRE UNIVERSITAIRE SALHI AHMED -NAAMA-

INSTITUT DES SCIENCES ET TECHNOLOGIES

DÉPARTEMENT DE MATHÉMATIQUES ET INFORMATIQUE



MÉMOIRE DE FIN D'ÉTUDE
pour obtenir le diplôme de
MASTER EN MATHÉMATIQUES



**Propriétés asymptotiques d'un
estimateur cas des données
vectorielles et fonctionnelles**



Présenté par :
Amraoui WALID

Soutenu devant le jury composé de :

Dr.belguerna Abderrahmane	C-Univ Salhi Ahmed -Naama -	Président
Dr.DAOUDI Hamza	Univ Iben Kheldoun -Tiaret -	Examineur
Dr.kenouza Jamel	C-Univ Salhi Ahmed -Naama -	Encadreur

Année universitaire 2019/2020



Remerciements

Je tiens tout d'abord à remercier les membres du jury pour leur présence, pour leur lecture attentive de ma mémoire ainsi que pour les remarques qu'ils m'adresseront lors de cette soutenance afin d'améliorer mon travail. et remercier ma encadreur vertueux Professeur Dr.kenouza Djamel, superviseur de cette mémoire, pour l'aide qu'il a fournie sur l'année.

Je tiens à le remercier pour sa patience, et surtout pour sa confiance, et pour sa disponibilité et la précision de ses indications. Sans lui ce projet n'aurait pas été mené à terme!

A tous mes enseignants qui m'ont initié aux valeurs authentiques, en signe d'un profond respect et d'un profond amour!!! Merci à vous tous. merci ma reconnaissance à monsieur.

Dédicaces

À la plus belle créature que Dieu a créée sur terre ,
À cet source de tendresse, de patience et de générosité
À ma mère!
à mon père celui qui m'a implanté le courage et l'amour du
travail.
À mes chers frères et soeurs
À tous mes amis et collègues.

Table des matières

1	Présentation	6
1.1	Introduction :	6
1.1.1	Historique de la methode de noyau k . . .	7
1.1.2	Estimation Non-Paramétrique :	13
1.1.3	Statistique des données fonctionnelles :	16
1.1.4	Noyaux	21
1.1.5	Exemples de noyaux	22
1.2	Estimation par la méthode des noyaux	23
1.3	Estimateur à noyau	25
1.4	Propriétés de l'estimateur à noyau	27
1.4.1	Etude du biais	27
1.4.2	Etude de la variance	28
1.4.3	Erreur quadratique moyenne (Mean-Squared Error :MSE)	30
1.4.4	Erreur quadratique moyenne intégrée (Mean-Integrated Squared Error : MISE)	31
1.4.5	Choix théorique optimal du paramètre de lissage :	37
1.4.6	Le choix optimal pratique du paramètre de lissage h :	39
1.4.7	La méthode des moindres carrés avec validation croisée (MC-VC)	42
1.4.8	Le choix optimal du noyau	46
2	l'estimateur dans le cas vectoriel	48
2.1	Hypothèses :	49
2.2	Propriétés asymptotiques :	50
3	l'estimateur dans le cas fonctionnelle	53
3.1	Quelque hypothèses nécessaires :	53
3.2	Type de continuité :	55
3.3	Type de Lipschitz :	55

TABLE DES MATIÈRES

1

3.4	Probabilités des petites boules :	55
3.5	Estimation non paramétrique de la densité conditionnelle	56
3.6	La convergence presque complète de l'estimateur à noyaux	57
3.7	Conclusion :	59

Bibliographie		60
----------------------	--	-----------

Introduction générale

La théorie de l'estimation est une des préoccupations majeures des statisticiens. On trouve dans la littérature deux types d'approches d'estimations de la densité de probabilité : l'approche paramétrique et l'approche non-paramétrique.

L'approche paramétrique suppose que les données sont issues d'une loi de probabilité de forme connue dont seuls les paramètres sont inconnus. Son objectif est de connaître la vraie valeur du paramètre ou plus généralement une fonction de cette valeur. Le principal inconvénient de cette approche est qu'elle nécessite une connaissance préalable du phénomène aléatoire considéré.

Pour pallier les insuffisances et les défauts des familles paramétriques, une seconde approche dite non paramétrique propose de laisser parler les données, sans spécifier au préalable de forme sur f .

L'outil d'estimation non paramétrique nous est fourni par l'histogramme : une fois les données regroupées en classes de valeurs, les fréquences empiriques sont représentées par des aires rectangulaires dont les bases correspondent aux classes elles mêmes. L'histogramme convient bien pour des analyses relativement grossières. Néanmoins, ses discontinuités n'apparaissent pas très naturelles et, ce qui est plus grave, les points tombant près des bords d'une classe et ceux tombant près du milieu ne sont pas différenciés, ceci explique la

variabilité des interprétations statistiques que l'on peut faire d'un histogramme suivant le choix de l'origine et des classes. Pour des densités raisonnablement lisses, l'histogramme apparaît donc comme un estimateur sévèrement limité

Il existe d'autres méthodes non paramétriques plus robustes que la méthode par l'histogramme : la méthode d'estimation par les séries orthogonales et la méthode du noyau.

Nous regarderons brièvement en quoi consiste la méthode d'estimation par l'histogramme et la méthode par les séries orthogonales et en détail l'estimateur par la méthode du noyau vu sa souplesse d'utilisation et ses propriétés de convergence.

Le succès rencontré par l'estimateur à noyau auprès de la communauté des utilisateurs peut essentiellement s'expliquer en trois points :

-D'abord l'expression théorique de l'estimateur est extrêmement simple puisque il s'écrit sous la somme de n variables aléatoires indépendantes et identiquement distribuées

- Ensuite l'estimateur converge vers la densité f en de nombreux sens, en particulier au sens $L1$.

d'autres part, il est convergent dans tous les modes : en probabilité, en moyenne, presque sûrement et presque complètement

- Enfin, l'estimateur à noyau est flexible, dans la mesure où il laisse à l'utilisateur une grande latitude non seulement dans le choix du noyau K , mais encore dans le choix du paramètre de lis-

sage h .

C'est Rosenblatt [19] en 1956, suivi de Parzen [18] en 1962, qui ont proposé une classe d'estimateurs à noyau d'une densité univariée.

Les estimateurs à noyau sont des fonctions de deux paramètres K , appelé noyau, et h dit paramètre de lissage (largeur de fenêtre).

Rosenblatt reprenait l'idée de Fix et Hodges en 1951, qui consistait à estimer la densité en un point, en comptant le nombre d'observations situées dans l'intervalle de longueur $2h$ et centré en ce point.

Avant de construire les estimateurs à noyaux de la densité, déen mesurer les performances théoriques et, le cas échéant, d'identifier le meilleur, il est nécessaire de spécifier un critère d'erreur qui puisse être éventuellement optimisé.

Citons par exemple l'erreur quadratique moyenne intégrée MISE et l'intégrale de l'erreur quadratique ISE.

Les propriétés de convergence de l'estimateur à noyau ont été établies par Parzen [18], Silverman [28] et Nadaraya [16]. Les théorèmes relatifs à l'erreur quadratique moyenne et l'erreur quadratique intégrée moyenne ont été obtenus sous forme élémentaire par Parzen [18]. Enfin, c'est Epanechnikov en 1969 [8] qui s'est rendu compte de l'exis-

tence d'un noyau asymptotiquement optimal K_e . Mais l'erreur quadratique moyenne asymptotiquement intégrée 23

varie peu en fonction du choix de K .

Si le choix du noyau n'est pas un problème dans l'estimation de la densité, il n'en est pas de même pour le choix de la largeur de fenêtre qui ne dépend que de la taille n de l'échantillon. Plusieurs travaux ont montré que les estimateurs peuvent changer dramatiquement pour de petites variations du paramètre de lissage. Actuellement, il n'existe pas de choix optimal pour ce paramètre de lissage. Le choix optimal qui minimise l'erreur relative globale (MISE) dépend de la dérivée seconde de la densité inconnue. Les auteurs se sont alors attachés à introduire des procédures de sélection automatiques et donc moins subjectives que le simple choix à l'oeil. L'étude de ce problème a nourri une littérature abondante, notamment vers le milieu des années quatre-vingt.

Chapitre 1

Présentation

1.1 Introduction :

Dans cet mémoire notre propos sera de présenter les qualités principales du meilleur "généraliste" en matière d'estimation de densité : l'estimateur à noyau. Dans un but didactique, à l'intention des non-spécialistes de l'estimation fonctionnelle, nous avons élagué les considérations mathématiques pour privilégier les idées-forces. Les lecteurs désireux d'entrer dans les techniques de démonstration voudront bien se reporter à la bibliographie. Ce choix nous permettra de présenter au plus grand nombre des résultats extrêmement récents ainsi que des voies de recherche nouvelles, sans désir d'une quelconque exhaustivité. Le problème que nous traitons ici est l'estimation d'une densité de probabilité inconnue f à partir d'un échantillon X_1, X_n de variables aléatoires indépendantes et de même loi à densité f (ce que nous appellerons dans la suite un f -échantillon). Nous aborderons un certain nombre de concepts généraux en estimation fonctionnelle : choix des mesures d'erreur, stabilité relative, problèmes liés

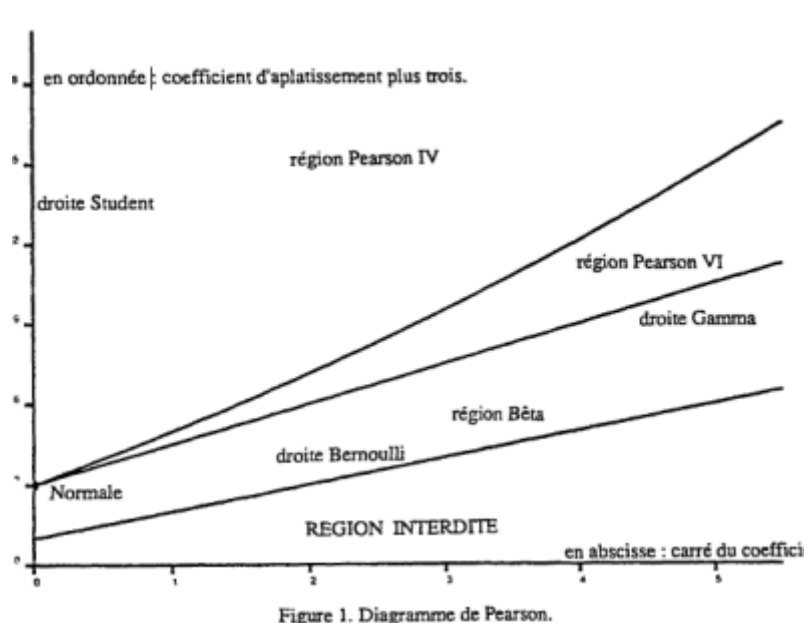
à la taille n de l'échantillon, choix du facteur de lissage et de la fonction noyau.

1.1.1 Historique de la methode de noyau k

C'est à la fin du siècle dernier que Karl Pearson introduisit la célèbre famille de lois de probabilité qui porte son nom et dont chaque élément satisfait à une équation différentielle du type :

$$(1) \quad f'(x) = \frac{(x-a)f(x)}{b_0 + b_1x + b_2x^2} \quad (a, b_0, b_1, b_2) \in \mathbb{R}^4$$

La raison pour laquelle cette famille est encore si populaire de nos jours tient probablement dans deux faits : le premier est qu'elle inclut bon nombre de familles paramétriques usuelles (bt_a , gamma , exponentielle, normale, de Student, de même que des densités moins connues comme les Pearson IV) ; le second est qu'elle permet quasiment tous les mariages possibles entre coefficient d'asymétrie (skewness) et d'aplatissement (kurtosis), et donc de refléter fidèlement l'allure d'une distribution empirique lorsque l'on se limite à ces deux paramètres.



Cette propriété de description est illustrée par la figure 1 : un couple quelconque de réels positifs n'est pas nécessairement la représentation dans le plan d'une loi de probabilité mais, en dehors de la région interdite et de sa frontière (droite Bernoulli) tout point du plan est la représentation d'une loi de Pearson. Malheureusement le modèle de Pearson présente de graves défauts : outre que des modèles simples très courants en soient exclus (Laplace, log-normale, Weibull, Gumbel, logistique) l'équation (1) nous montre les limitations sévères du modèle en ce qui concerne l'existence de modes. Enfin une méthode basée sur les seuls quatre premiers moments (en fonction desquels s'expriment les constantes a, b_0, b_1, b_2) ne peut que se révéler inefficace et non robuste dès que l'on s'écarte quelque peu du modèle. Un exemple d'une telle méthode est celle des moments où l'on substitue les moments empiriques aux moments théoriques dans les expressions des constantes. A moins

d'avoir sur le phénomène aléatoire étudié des informations a priori très précises et indiscutables, le champ d'application d'un modèle paramétrique ne devient satisfaisant que lorsque l'inflation du nombre de paramètres est telle que les méthodes d'estimation du modèle deviennent tout à fait inefficaces. Pour pallier aux insuffisances et aux défauts des familles paramétriques une démarche élémentaire est de faire appel à l'omniprésent histogramme : l'intervalle d'étude, disons $[0, 1]$, est partagé en k sousintervalles de même longueur notés I_1, \dots, I_k et, sur le j^{me} , $f(x)$ est estimé par $k \frac{N_j}{n}$ où N_j est le nombre de points tombés dans I_j .

Nombre de données : 10
Largeur des cellules : 0,2

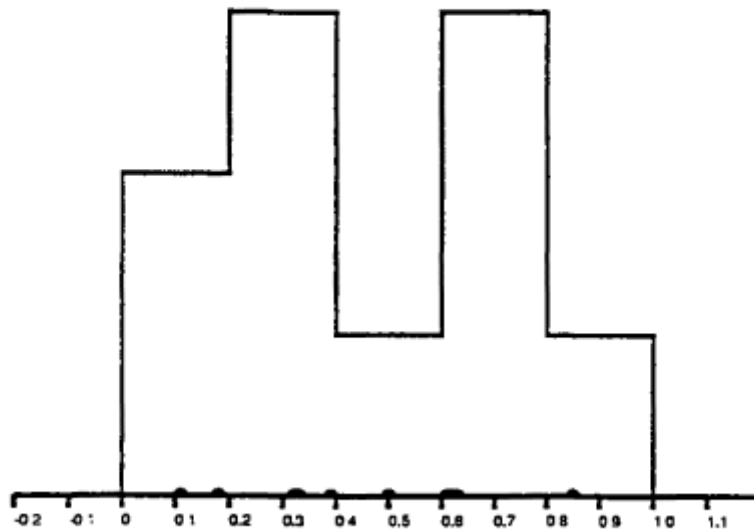


Figure 2. Un histogramme.

L'histogramme convient bien pour des analyses relativement grossières. Néanmoins ses discontinuités n'apparaissent pas très naturelles et, ce qui est plus grave, les points près des bords et les points près du milieu d'un intervalle reçoivent le même poids ; ceci explique la variabilité des interprétations statistiques que l'on peut faire d'un histogramme suivant l'origine choisie. En fait, pour des densités raisonnablement lisses, l'histogramme est un estimateur sévèrement limité. Pour remédier à ce problème, Rosenblatt proposa en 1956 de centrer chaque cellule de l'histogramme au point où l'on estime, introduisant ainsi la première forme

rudimentaire d'estimateur à noyau.

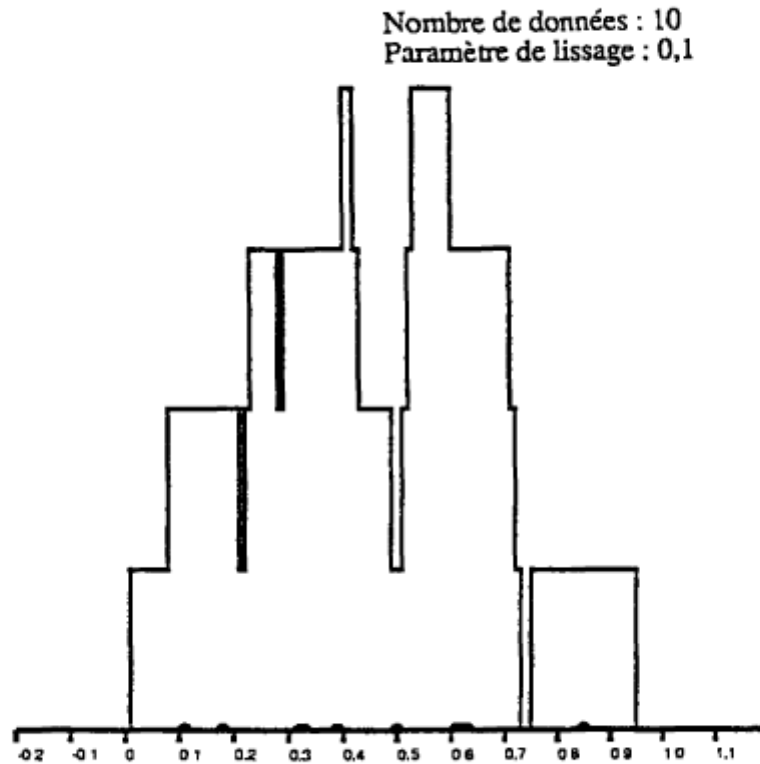


Figure 3. Estimateur à noyau uniforme.

Nous savons maintenant que l'estimateur de Rosenblatt a une meilleure vitesse de convergence que l'histogramme pour la plupart des densités régulières. Il peut s'écrire sous la forme

$$f_n(x) = \frac{1}{nh} \sum -i = 1^n K\left(\frac{x - X_i}{h}\right)$$

où K , appelé le noyau, est la densité uniforme sur $[-1/2; +1/2]$ et où $h > 0$ est la largeur de fenêtre ou constante de lissage. Dès que K est une

densité, f_n en est également une. On peut obtenir des estimateurs plus esthétiques en utilisant des noyaux K lisses (Rosenblatt, 1956 ; Parzen, 1962). Lorsque l'on se limite aux noyaux K positifs les vitesses de convergence varient peu en fonction de K et les critères essentiels de choix du noyau sont la simplicité et la vitesse de calcul d'une part, la régularité de la courbe à obtenir d'autre part. Nous verrons qu'il en est tout autrement lorsque l'on s'autorise à utiliser des noyaux quelconques. Il est facile de vérifier que pour les noyaux usuels et un ensemble de données fixé la loi de densité f_n converge (étroitement) vers la mesure empirique lorsque h tend vers 0 et

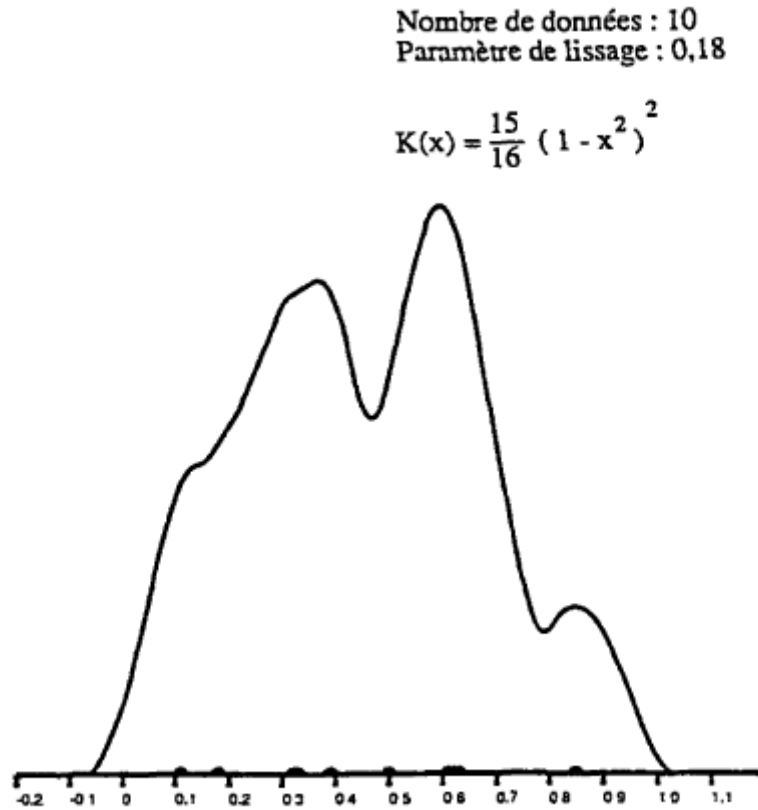


Figure 4. Estimateur à noyau

que f_n tend uniformément vers la fonction nulle lorsque h tend vers l'infini. Le coeur du problème est le choix du facteur de lissage qui, en variant, fait décrire à f_n un ensemble de lois dont les extrêmes sont "proches" de lois discrètes d'un côté, uniformes de l'autre.

1.1.2 Estimation Non-Paramétrique :

Dans ces dernières décennies, une immense innovation sur les appareils de mesure est apparue permettant d'observer plusieurs données de

plus en plus complexes d'une façon continue, tels les indices boursiers, la météorologie, les images satellitaires, la chimie quantitative, la biométrie, l'imagerie médicale... , en raison de la précision des appareils de mesures modernes et de l'importante capacité de stockage qu'offrent les systèmes informatiques actuels, c'est possible à présent d'obtenir une discrétisation très précise de ces objets mathématiques (courbes, images, . . .) pendant toutes leurs trajectoires et qui prennent des valeurs dans des espaces de dimension infinie. Une modélisation statistique est importante pour mieux comprendre le fonctionnement du problème modélisé. En statistiques non paramétriques la performance des outils statistiques se réduit considérablement lorsque la dimension des observations augmente. Ce grand développement technologique a imposé la modernisation des méthodes statistiques comme outils d'analyse et de contrôle. Ainsi, une nouvelle branche de la statistique, dénommée statistique fonctionnelle s'est développée pour traiter des observations comme éléments aléatoires fonctionnels. Les premiers ouvrages de référence en la matière sur le sujet ont été consacrées à l'étude des modèles paramétriques, les monographies de Ramsay et Silverman (1997) pour le cas i.i.d ou pour le cas dépendant Bosq (2000) pour les aspects théoriques. Cependant, la base d'analyse statistique via les modèles linéaires est la préliminaire connaissance de la nature de la covariabilité entre les observations, ce qui est très compliqué à vérifier en statistiques fonctionnelles, par contre dans la statistique classique on dispose

d'outils graphiques, on prend comme exemple le scatter plot qui donne un aperçu sur le rapport entre les observations. Ceci justifie l'intérêt de la modélisation des données fonctionnelles par des méthodes non paramétriques, on peut citer les contributions de Bosq (2000) pour les aspects théoriques, Ferraty et Vieu (2006) pour une étude non paramétrique et Ferraty et Romain (2011) pour des développements récents. Dans le même contexte, nous renvoyons à Ferraty (2010). L'objectif de ce paragraphe est de faire une étude bibliographique sur les modèles non paramétriques conditionnels considérés dans cette thèse. Le traitement non paramétrique des données fonctionnelles est beaucoup plus récent que l'analyse paramétrique. En effet, les premiers résultats ont été obtenus par Gasser et al. (1998). Ils se sont intéressés à l'étrique du mode de la distribution d'une variable fonctionnelle vérifiant une condition fractale. En considérant la même condition fractale Ferraty et Vieu (2000) ont étudié la convergence presque complète d'un estimateur à noyau de la fonction de régression, lorsque les observations sont indépendantes et identiquement distribuées. Dabo-Niang (2002) a obtenu, la convergence presque sûre et la normalité asymptotique d'un estimateur de type histogramme de la densité d'une variable aléatoire dans un espace de dimension infinie. En utilisant la propriété de concentration de la mesure de probabilité de la variable explicative fonctionnelle, Dabo-Niang et Rhomari (2004) ont étudié la convergence en norme L_p de l'estimateur à noyau de la régression non paramétrique. La conver-

gence presque complète pour le cas fortement mélangeant a été étudié par Ferraty et al. (2004). Masry (2005) a montré la normalité asymptotique dans le cas d'observations fonctionnelles -mélangeantes. Les premiers résultats sur les modèles conditionnels ont été obtenus par Ferraty et al.(2006). Ils ont précisé la vitesse de convergence presque complète des estimateurs à noyau pour la fonction de répartition conditionnelle, la densité conditionnelle et ses dérivées, le mode conditionnel et les quantiles conditionnels. Nous renvoyons à Ferraty et Vieu (2006) pour un large éventail d'applications de ces modèles en statistique fonctionnelle. Dabo-Niang et Laksaci (2007) ont ajouté des résultats sur la convergence en norme L_p de l'estimateur à noyau du mode conditionnel dans le cas i.i.d. La détermination des termes dominants de l'erreur quadratique de l'estimateur à noyau de la densité conditionnelle a été obtenue par Laksaci (2007). Ferraty et al.(2008) ont abordé l'estimation de la fonction du hasard conditionnelle et ont établi la convergence presque complète d'un estimateur à noyau de ce modèle non paramétrique

1.1.3 Statistique des données fonctionnelles :

L'analyse des données fonctionnelle est une branche des statistiques vers laquelle l'intérêt de la communauté scientifique se développe de plus en plus, aussi en relation avec le nombre croissant de situations dans lesquelles les scientifiques théoriques et appliqués doivent faire face à des données ayant une nature continue. Le domaine a été popula-

risé en particulier au moyen des livres de Ramsay-Silverman (2002) et (2005) et, au cours des vingt dernières années, Les contributions ont été publiées (voir par exemple Bosq (2000), Ferraty et Vieu (2006) et aussi Hsing et Eubank (2015) pour les monographies générales, Cuevas (2014), Horváth et Rice (2015) et Muller (2005) Pour les études méthodologiques, Bongiorno et al.(2014) pour un ensemble sélectionné d'événements récents). Parallèlement, des méthodologies étendues et variées pour traiter des problèmes ayant une dimension très élevée ont été développés.

Variables et données fonctionnelles :

La statistique pour données fonctionnelles ou analyse des données fonctionnelles étudie des observations qui ne sont pas des variables réelles ou vectorielles mais des courbes aléatoires. En général on veut indiquer que le nombre de variables observées est grand et même parfois beaucoup plus élevé que le nombre des unités statistiques. Dans ce cas également, l'un des principaux intérêts pour les statistiques haute dimension est le nombre croissant de situations impliquant chaque jour un grand nombre Des variables observées. La définition d'une variable aléatoire fonctionnelle comme elle est définie dans les livres de statistique, par exemple, Ferraty et vieu (2006), Bosq et Lecoutre (1987), est une variable aléatoire observée dans plusieurs points d'un intervalle et qui prend des valeurs dans un espace infini, par exemple la variable fonctionnelle $\chi = \{\chi(t); t \in T\}$ avec $T \subset R$ représente une courbe observée sur l'inter-

valle T de R . D'autre part si $T \subset R^2$ notre variable fonctionnelle représente une image. Pour fixer les idées on présente ces définitions.

Définition :

Une variable aléatoire χ est appelée variable fonctionnelle si, elle prend des valeurs dans un espace à dimension infinie. L'observation χ de χ est appelée une donnée fonctionnelle.

Définition :

L'ensemble de données fonctionnelles $X_1; \dots; X_n$ est l'observation de n variables fonctionnelles $X_1; \dots; X_n$ identiquement distribuées comme χ . Nous exposons ci-dessous quelques exemples de Ferraty et Vieu (2002, 2003) sur l'applications statistiques aux problèmes soulevés par l'analyse des données fonctionnelles. Dans un contexte général on parle de deux cas, le cas où les variables fonctionnelles sont indépendantes et dépendantes. Cas des variables fonctionnelles indépendantes : Dans ce cas ils se sont intéressés d'un problème de contrôle de la qualité dans l'industrie alimentaire, on peut trouver ces données sur le site web <http://lib.stat.cmu.edu/> Cet ensemble de données concerne un échantillon de viande finement hachée. L'objectif est d'étudier et mesurer par le spectromètre la teneur en graisse dans les morceaux de viande (figure.1.1).

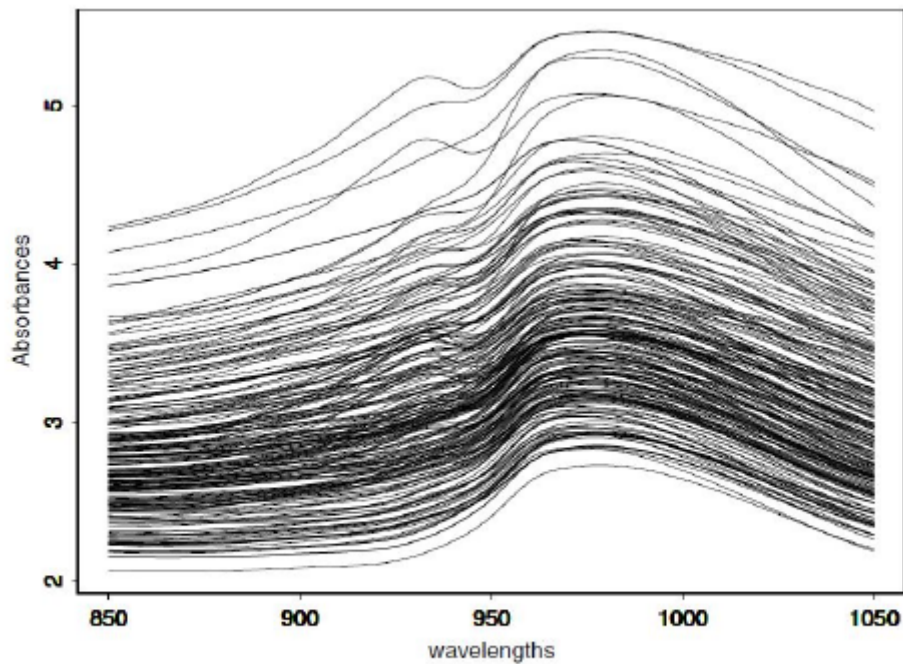


Fig.1.1 Les courbes spectrométriques

Cas des variables fonctionnelles dépendantes : Dans ce cas on prend l'exemple de consommation annuelle d'électricité, aux USA qui donne une série chronologique économique. Les données d'électricité mensuelle consommée par les secteurs résidentiel et commercial à partir de janvier 1973 à février 2001 (338 mois) sont disponibles sur le site web <http://www.economagic.com>, après ils ont décidé de choisir les années passées comme période explicative. Cela signifie que l'ensemble des variables explicatives à inclure dans notre méthode statistique est composé de 28 courbes données qui sont les 28 séries chronologiques annuelles continues. Ces données fonctionnelles dépendantes sont présentées à la Figure.1.2.

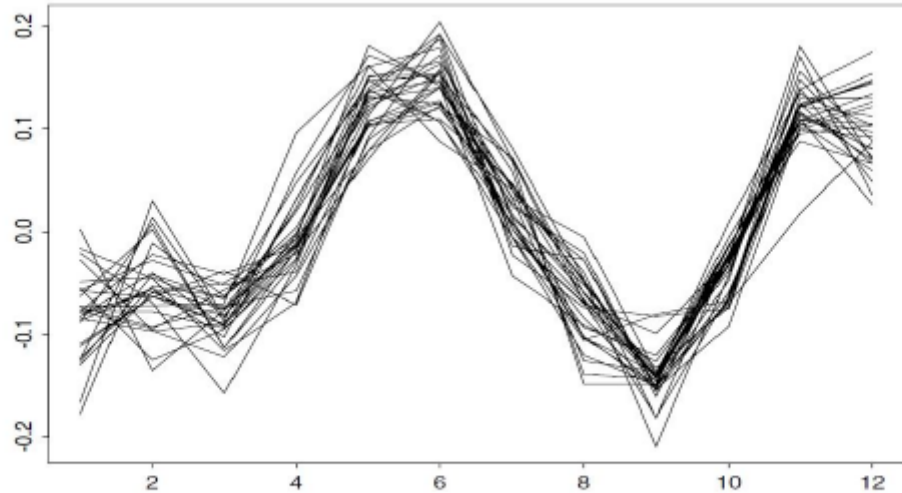


Fig.1.2 Courbes annuelles de consommation d'électricité aux USA

La classification des courbes est un domaine qui a reçu une grande attention ces dernières années, le document de Cholaquidis et al.(2015) propose un algorithme d'agrégation de classification adapté aux données fonctionnelles. Le modèle linéaire fonctionnel a été largement étudié dans la littérature pour les problèmes de régression dont le prédicteur est de dimension infinie. Dans ce problème, deux contributions traitent ce type de modèle dans le contexte plus général où on a observé plus d'une variable fonctionnelle. La modélisation semiparamétrique des données fonctionnelles est un domaine actuellement actif dont les principaux objectifs le compromis entre la grande sensibilité à la dimension des modèles non paramétriques et la flexibilité linéaires (voir Goia et Vieu (2014) pour une courte enquête). Fondamentalement, dans la littérature on peut trouver dif-

férents modèles de différents types d'additifs (Voir par exemple Aneiros-Pérez et Vieu (2008), Ferraty et Vieu (2009)) ou des modèles basés sur les projections (voir par exemple Chen et al.(2011), Ferraty et al.(2013)) et la contribution de Ahmedou et al.(2015) complète cette littérature en proposant une combinaison de ces idées à travers un modèle de régression linéaire fonctionnelle généralisé capable d'intégrer de manière semi-paramétrique additive les informations issues d'un prédicteur fonctionnel et à partir de ses dérivés. Pour la modélisation fonctionnelle de séries temporelles autorégressives, l'extension des processus autorégressifs habituels dans le cadre fonctionnel a été popularisé au cours des dernières années, principalement grâce aux diverses contributions de Bosq (voir par Monographies Bosq (2000), Bosq et Blanke (2007)).

1.1.4 Noyaux

Définissons maintenant plus généralement la notion d'estimateur à noyau :

Définition 1 : Soit $K : \mathbb{R} \rightarrow \mathbb{R}$, on dit que K est un noyau si et seulement si :

$$\int K(u)du = 1$$

* K est dit positif si $K(u) \geq 0 \forall u$.

* K est dit symétrique si $K(u) = K(-u) \forall u$

1.1.5 Exemples de noyaux

Voici quelques exemples de noyaux les plus communément utilisés :

* le noyau rectangulaire : $K(u) = \frac{1}{2}1_{[-1,+1]}(u)$. C'est celui qui donne l'estimateur de type histogramme appelé noyau de Rosenblatt.

*le noyau triangulaire : $K(u) = (1 - |u|)1_{[-1,+1]}(u)$.

*le noyau d'Epanechnikov : $\frac{3}{4}(1 - u^2)1_{[-1,+1]}(u)$

*le noyau de Tukey ou biweight : $K(u) = \frac{15}{16}(1 - u^2)^2 1_{[-1,+1]}(u)$

*le noyau gaussien : $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right), u \in \mathbb{R}$.

Les deux premiers ont l'avantage d'être simples, le noyau triangulaire étant continu partout et conduisant à une estimation 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 dérivables partout. Ainsi il assure le lissage local de la fonction f_n . Ce noyau est d'une forme très proche du noyau Gaussien et il est donc préférable. Notons que plus la valeur de h est élevée plus on élargit la fenêtre, ce qui donne un effet de lissage globale de f_n plus important.

1.2 Estimation par la méthode des noyaux

L'estimation non paramétrique de la fonction de densité peut se voir comme le cumul des fonctions K de chaque observation sur tout le domaine :

Supposons que nous observons n variables aléatoires i.i.d X_1, \dots, X_n de densité f . L'objectif de notre étude est la construction d'un estimateur de f en un point fixe x . Notons $F(x) = P(X_1 \leq x)$ la fonction de répartition de la loi de X_1 . La densité est la dérivée de la fonction de répartition, ce qui permet d'écrire pour tout x :

$$f(x) = F'(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h} = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h}$$

Considérons la fonction de répartition empirique :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 2_{X_i \leq x}, \forall x \in \mathbb{R}$$

La loi des grands nombres permet d'affirmer que F_n est un estimateur de F , c'est-à-dire

$$F_n(x) \xrightarrow{p} F(x)$$

De plus, le théorème de Glivenko-Cantelli nous donne :

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s} 0$$

Il est même possible d'obtenir des intervalles de confiance et de tester l'adéquation des données à différentes lois. Néanmoins, il n'est pas évident d'utiliser F_n pour estimer f . Une des premières idées intuitives est de considérer pour $h > 0$ fixé "petit" :

$$f_n(x) = \frac{F_n(x+h) - F_n(x-h)}{2h} = \frac{1}{2nh} \sum_{i=1}^n 1_{-h \leq X_i - x \leq h}$$

On a alors :

$$E[f_n(x)] = \frac{1}{2h} (E[F_n(x+h)] - E[F_n(x-h)]) = \frac{1}{2} (F(x+h) - F(x-h))$$

$E[f_n(x)]$ tend vers $f(x)$ quand $h \rightarrow 0$. Il faut donc faire dépendre h de la taille de l'échantillon, et le faire tendre vers 0 quand $n \rightarrow \infty$, de sorte que $f_n(x)$ soit un estimateur asymptotiquement sans biais de $f(x)$. L'estimateur f_n reste une fonction en escalier. Pour obtenir quelque chose de plus lisse, on peut remarquer que :

$$\begin{aligned} f_n(x) &= \frac{1}{2nh_n} \sum_{i=1}^n 1_{[x-h_n, x+h_n]}(X_i) = \frac{1}{2} \sum_{i=1}^n 1_{[x-h_n \leq X_i \leq x+h_n]} \\ &= \frac{1}{nh_n} \sum_{i=1}^n \frac{1}{2} 1_{[-1, 1]}\left(\frac{x - X_i}{h_n}\right) = \frac{1}{nh_n} \sum_{i=1}^n k\left(\frac{x - X_i}{h_n}\right) \end{aligned}$$

Cet estimateur appelé estimateur de Rosenblatt, est le premier exemple d'estimateur à noyau construit à l'aide du noyau

$$K(u) = \frac{1}{2} 1_{(-1 < u \leq 1)}$$

1.3 Estimateur à noyau

L'estimateur à noyau est probablement l'estimateur le plus utilisé et certainement le plus étudié mathématiquement, car il possède des propriétés qui le rendent fort intéressant.

Définition 1 : Un estimateur à noyau noté f_n de la fonction f est défini par :

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

où $\{h_n\}_{n \geq 1}$ est une suite de réels positifs appelés paramètres de lissage ou largeur de la fenêtre, qui tend vers 0 quand n tend vers l'infini.

Comme nous allons le voir par la suite, si le noyau K est une fonction de densité alors l'estimateur à noyau f_n est lui aussi une fonction de densité. De plus, ce dernier possède les propriétés de continuité et de différentiabilité. De sorte que si, par exemple, K est la densité normale alors f_n possède des dérivées de tout ordre.

Propriété 1 : Un estimateur à noyau est une densité

Démonstration :

$$\begin{aligned} \int_{-\infty}^{+\infty} f_n(x) dx &= \frac{1}{nh_n} \sum_{i=1}^n \int_{-\infty}^{+\infty} k\left(\frac{x - X_i}{h_n}\right) dx \\ &= \frac{1}{nh_n} \sum_{i=1}^n \int_{-\infty}^{+\infty} k(u) h_n du \end{aligned}$$

(changement de variable $u = \left(\frac{x-X_i}{h_n}\right)$)

$$\frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{+\infty} k(u) du = \frac{1}{n} n = 1$$

Pour mieux saisir l'intuition derrière l'estimateur à noyau, nous avons construit cet estimateur 'a partir de l'équation (1.13) en utilisant un ensemble de données constitué seulement de 7 observations. Le noyau K a été choisi comme étant la densité d'une loi normale de moyenne 0 et de variance 1 et le paramètre de lissage égale à 4. On centre d'abord un noyau individuel sur chacune des 7 observations et la valeur de l'estimateur à noyau $f(z)$ au point z est simplement la somme des ordonnées de chacun des 7 noyaux individuels à ce point x comme représenté à la figure (1.3). Dans une région où l'on a plusieurs observations, la vraie densité a une valeur relativement grande et l'estimateur de la densité, par la méthode du noyau, nous donne effectivement une valeur relativement grande ce qui est observé dans la figure (1.3)

Par exemple si $x = 5$ on a $\hat{f}(x) = 0.03$ qui est égale à la somme des densités des 7 noyaux gaussiens au même point $x = 5$

1.4 Propriétés de l'estimateur à noyau

Nous allons maintenant donner quelques propriétés statistiques élémentaires de l'estimateur de la densité à noyau ainsi que différentes méthodes pour choisir le paramètre de lissage.

1.4.1 Etude du biais

Supposons que l'on dispose d'un échantillon d'observations X_1, \dots, X_n , issu d'une v.a X possédant pour fonction de densité la fonction f que l'on désire estimer. On suppose que f_n est l'estimateur à noyau obtenu en utilisant le noyau K et le paramètre de lissage h et f_n défini par l'équation (1.13). Supposons que

$$K(u) \geq 0, \int K(u)du = 1, \int K(u)udu = 0, \int u^2 K(u)du < \infty$$

et en supposant que la densité de probabilité f admet les deux premières dérivées (continues) nécessaire.

$$E[f_n(x)] = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} E\left[K\left(\frac{x - X_i}{h}\right)\right]$$

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{x - t}{h}\right) f(t) dt$$

$$\frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{x-t}{h}\right) f(t) dt$$

La transformation $z = \frac{x-t}{h}$, i.e. $t = -hz + x$, $|\frac{dz}{dt}| = \frac{1}{h}$

$$E[f_n(x)] = \int_{-\infty}^{+\infty} K(z) f(x - hz) dz$$

Un développement de Taylor de $f(x - hz)$ nous donne :

$$f(x - hz) = f(x) - hz f'(x) + \frac{1}{2} (hz)^2 f''(x) + 0(h^2)$$

$$\begin{aligned} E[f_n(x)] &= \int_{-\infty}^{+\infty} K(z) f(x) dz - \int_{-\infty}^{+\infty} K(z) hz f'(x) dz + \int_{-\infty}^{+\infty} K(z) \frac{(hz)^2}{2} f''(x) dz \\ &= f(x) \int_{-\infty}^{+\infty} K(z) dz - h f'(x) \int_{-\infty}^{+\infty} z K(z) dz + \frac{h^2}{2} f''(x) \int_{-\infty}^{+\infty} z^2 K(z) dz + 0(h^2) \\ &= f(x) + \frac{h^2}{2} k_2 f''(x) + 0(h^2) \end{aligned}$$

$$\text{Biais}(f_n(x)) \approx \frac{h^2}{2} k_2 f''(x) + 0(h^2)$$

Le biais dépend de :

- * h : paramètre de lissage.
- * k_2 : la variance du noyau.
- * $f''(x)$ la seconde dérivée de la fonction de densité au point x .

1.4.2 Etude de la variance

La variance de $f_n(x)$ est donnée par :

$$\text{Var}(f_n(x)) = \text{Var}\left(\frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - X_i}{h}\right)\right)$$

$$= \frac{1}{n^2 h^2} \sum_{i=1}^n \text{var}\left(k\left(\frac{x - X_i}{h}\right)\right)$$

car les $X_i, i=1, \dots, n$, sont i.i.d

$$\begin{aligned} \text{var}\left(k\left(\frac{x - X_i}{h}\right)\right) &= E\left[k\left(\frac{x - X_i}{h}\right)^2\right] - \left(E\left[k\left(\frac{x - X_i}{h}\right)\right]\right)^2 \\ &= \int K\left(\frac{x - t}{h}\right)^2 f(t) dt - \left(\int K\left(\frac{x - t}{h}\right) f(t) dt\right)^2 \\ \text{Var}(fn(x)) &= \frac{1}{n} \int \frac{1}{h^2} K\left(\frac{x - t}{h}\right)^2 f(t) dt - \left(\frac{1}{n} \left(\frac{1}{h} \int K\left(\frac{x - t}{h}\right) f(t) dt\right)\right)^2 \\ &= \frac{1}{n} \int \frac{1}{h^2} k\left(\frac{x - t}{h}\right)^2 f(t) dt - \frac{1}{n} (f(x) + \text{biais}(f_n(x)))^2 \end{aligned}$$

En effectuant le changement de variable suivant $z = \frac{x-t}{h}$, on obtient :

$$\text{Var}(fn(x)) = \frac{1}{nh} \int K(x)^2 f(x-hz) dz - \frac{1}{n} (f(x) + 0(h^2))^2$$

Et en effectuant un développement limité à l'ordre 2, il vient :

$$\begin{aligned} \text{Var}(fn(x)) &= \frac{1}{nh} \int K(x)^2 f(x-hz f'(x) + 0(h)) dz - \frac{1}{n} (f(x) + 0(h^2))^2 \\ \text{Var}(fn(x)) &= \frac{1}{nh} f(x) \int K(x)^2 dz + 0\left(\frac{1}{nh}\right) \end{aligned}$$

d'où :

$$\text{Var}(fn(x)) \approx \frac{1}{nh} f(x) \int K(x)^2 dz$$

Discussion du comportement du biais et de la variance :

*Le biais d'écroit si h diminue mais la variance augmente.

*La variance diminue si h augmente mais le biais

augmente.

*Pour que la variance tende vers zéro, il faut que $nh \rightarrow 1$.

*Plus la courbure de la densité est haute en x , plus le biais est grand.

*La variance est plus grande pour des valeurs plus grandes de la densité.

La figure suivante nous permet de mieux voir le comportement du biais et de la variance.

La variance est représentée par la courbe en pointillé et le biais par la courbe fine, la courbe en gras représente le MSE.

1.4.3 Erreur quadratique moyenne (Mean-Squared Error :MSE)

L'analyse de la performance de l'estimateur à noyau exige la spécification d'un critère d'erreur approprié afin de mesurer l'erreur d'estimation aussi bien qu'en un point que sur l'ensemble des points. Nous étudierons dans un premier temps la proximité de notre estimateur de la densité f_n de la vraie densité f . L'estimateur f_n dépend des données, du noyau K et du paramètre de lissage h . Cette dépendance n'est généralement pas exprimée explicitement. Pour chaque x , $f_n(x)$ peut être considérée comme une variable aléatoire. Lorsque nous considérons l'estimation en un point, une mesure naturelle de la dispersion est l'erreur quadratique moyenne aussi notée MSE qui est :

$$MSE(f_n(x)) = E[(f_n(x) - f(x))^2]$$

Par des propriétés élémentaires de l'espérance et de la variance, on peut montrer que le MSE peut s'écrire.

$$MSE(f_n(x)) = E[f_n(x) - E[f_n(x)]] + (E[f_n(x)] - f(x))^2 \\ (E[f_n(x)] - f(x))^2 + var[f_n(x)]$$

En remplaçant le biais et la variance par leurs valeurs respectives, on trouve :

$$MSE = \frac{1}{4}h^4k_2^2f''(x)^2 + \frac{1}{nh}f(x)j_2 + o(h^4 + \frac{1}{nh})$$

Où $k_2 = \int zK(z)dz$ et $j_2 = \int K(z)^2dz$.

Asymptotiquement, on a :

$$AMSE = \frac{1}{4}h^4k_2^2f''(x)^2 + \frac{1}{nh}f(x)j_2$$

1.4.4 Erreur quadratique moyenne intégrée (Mean-Integrated Squared Error : MISE)

Cependant, Il peut être intéressant d'avoir une mesure globale de la précision de f_n comme estimateur de f au lieu d'avoir une mesure de précision en un point donné. Cette mesure globale qu'on note MISE est définie par :

$$MISE(f_n(x)) = E[\int f_n(x) - f(x)^2 dx]$$

On peut montrer que :

$$MISE(f_n(x)) = \int E[(f_n(x) - f(x))^2] dx \\ = \int MSE(f_n(x)) dx \\ = \int (E[f_n(x)] - f(x))^2 dx + \int Var[f_n(x)] dx$$

$$MISE(f_n(x)) = \frac{1}{4} k_2^2 b(f) + \frac{1}{nh} j_2 + o(h^4) + \frac{1}{nh}$$

Où : $\beta(f) = \int f''(x)^2 dx$

Asymptotiquement, on a :

$$AMISE(f_n(x)) = \frac{1}{4} h^4 k_2^2 \beta(\mathbf{f}) + \frac{1}{nh} j_2$$

Si l'on compare les deux composantes du MISE : on observe l'un des problèmes fondamentaux dans l'estimation de densité. éliminer le biais, on choisit h petit alors que l'intégrale de la variance devient grande. Par contre si on choisit h grand de façon à réduire l'intégrale de la variance, on introduit systématiquement un biais dans l'estimation. Donc, le choix du paramètre de lissage h doit être un compromis entre le biais et la variance. Nous verrons un peu plus loin quelques méthodes permettant de choisir le paramètre de lissage de façon optimale. La plupart de ces méthodes ont pour critère la minimisation d'une certaine mesure de dispersion comme le MSE, ou le MISE. Nous verrons par la suite que le choix du noyau n'a pas vraiment d'impact sur l'estimateur de la densité.

Théorème : Li and Racine (2007)

Si on suppose que $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ quand $n \rightarrow \infty$,

$$\hat{f}_n(x) \rightarrow_p f(x)$$

Démonstration : D'après l'inégalité de Cheby-

chev on a :

$$\begin{aligned} P(|f_n(x) - f(x)| > \epsilon) &\leq \frac{E(f_n(x) - f(x))^2}{\epsilon^2} \\ &= \frac{1}{\epsilon^2} h^4 k_2^2 f''(x)^2 + \frac{1}{nh\epsilon^2} f(x) j_2 \\ &\rightarrow 0 \end{aligned}$$

où la deuxième ligne est donnée par la formule

Lemme 1. Lyapounov : Supposons que pour tout

n, w_{n1}, \dots, w_{nn} sont des v.a indépendantes, tel que

$E[w_{ni}] = 0$ et $\sigma_{ni}^2 = E[w_{ni}^2] < \infty$.

Soit $s_n^2 = \sum_{i=1}^n \sigma_{ni}^2$. Pour un certain $\delta > 0$, si la condition suivante est satisfaite :

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{s_n^{2+\delta}} E[|w_{ni}|^{2+\delta}] = 0$$

alors :

$$\sum_{i=1}^n w_{ni}/s_n \rightarrow_d N(0, 1)$$

La condition (1.17) est appelée la condition de Lyapounov

Lemme : La condition de Lyapounov est satisfaite

quand w_{n1}, \dots, w_{nn} sont i.i.d, $\sigma_n^2 = E[w - ni^2] > 0$ et

$\lim_{n \rightarrow \infty} E[|w_{ni}|^{2+\delta}]/n^{\delta/2} = 0$ pour un certain $\delta > 0$

Démonstration :

Si les v.a sont i.i.d, on a

$$\begin{aligned} s_n^2 &= n\sigma_n^2, \text{ et} \\ s_n^{2+\delta} &= (n^{1/2}\sigma_n)^{2+\delta} = n^{1+\delta/2}\sigma_n^{2+\delta} \end{aligned}$$

De plus :

$$\sum_{i=1}^n \frac{1}{s_n^{2+n}} E[|w_n i|^{2+\delta}] = \sigma_n^{-2-\delta_n-1-\delta/2} \sum_{i=1}^n E[|w_n i|^{2+\delta}]$$

$$\sigma_n^{-2-\delta_n-1-\delta/2} E[|w_n i|^{2+\delta}]$$

La condition de Lyapounov est satisfaite si $n^{-\delta/2} E[|w_n i|^{2+\delta}] \rightarrow 0$. Supposons que $\lim_{n \rightarrow \infty} \sigma_n^2$ existe dans le cas de données i.i.d, le résultat de Lyapounov (CLT) peut être indiqué comme suit :

Corollaire : Supposons que pour tout n , $w_n 1, \dots, w_n n$ sont i.i.d, avec $E[w_n i] = 0$, $\lim_{n \rightarrow \infty} E[w_n i^2] > 0$ et est finie, et $\lim_{n \rightarrow \infty} E[|w_n i|^{2+\delta}]/n^{\delta/2} = 0$ pour un certains $\delta > 0$, alors

$$\frac{1}{n^{1/2}} \sum_{i=1}^n w_n i \sim N(0, \lim_{n \rightarrow \infty} E[w_n i^2])$$

On démontre maintenant la normalité asymptotique de l'estimateur de densité à noyau.

Pagan and Ullah (1999) supposons que $nh \rightarrow \infty$ et $1/(nh)^{1/2} h^2 \rightarrow 0$ et $f(x) > 0$, on a

$$(nh)^{1/2} (f_n(x) - f(x)) \rightarrow_d N(0, f(x) \int K^2(u) du)$$

Démonstration :

$$(nh)^{1/2} (f_n(x) - f(x)) = \frac{1}{n^{1/2}} \sum_{i=1}^n \left(\frac{1}{h^{1/2}} K\left(\frac{X_i - x}{h}\right) - \frac{1}{n^{1/2}} E\left[K\left(\frac{X_i - x}{h}\right)\right] \right)$$

$$+ (nh)^{1/2} \left(\frac{1}{h} E\left[K\left(\frac{X_i - x}{h}\right)\right] - f(x) \right)$$

D'après ce qu'on a vu précédemment on :

$$\frac{1}{h}E[K(\frac{X_i - x}{h})] - f(x) = o(h^2).$$

On définit :

$$w_{ni} = \frac{1}{h^{1/2}}K(\frac{X_i - x}{h}) - \frac{1}{h^{1/2}}[K(\frac{X_i - x}{h})]$$

$$(nh)^{1/2}(f_n(x) - f(x)) = \frac{1}{n^{1/2}} \sum_{i=1}^n w_{ni} + O_p((nh)^{1/2}h^2)$$

$$\frac{1}{n^{1/2}} \sum_{i=1}^n w_{ni} + o_p(1)$$

car on par hypothèse $(nh)^{1/2}h^2 \rightarrow 0$ On vérifie maintenant les conditions du corollaire 1 Par définition de w_{ni} , $E[w_{ni}] = 0$ et

$$E[w_{ni}^2] = \frac{1}{h}E[K^2(\frac{X_i - x}{h})] - \frac{1}{h}(E[K(\frac{X_i - x}{h})])^2$$

or

$$E[K(\frac{X_i - x}{h})] = h \int K(u)f(x + uh)du$$

$$= O(h)$$

, et

$$\frac{1}{h}E[K^2(\frac{X_i - x}{h})] = h \int K(u)^2 f(x + uh)du$$

$$\rightarrow f(x) \int K^2(u)du$$

les résultats (1.24) et (1.26) impliquent que

$$\lim_{n \rightarrow \infty} E[w_{ni}^2] = f(x) \int K^2(u)du$$

Pour finir nous montrons que $E[|w_{ni}^{2+\delta}|]/n^\delta \rightarrow 0$ On utilise l'inégalité suivante afin de développer $E[|w_{ni}^{2+\delta}|]$. Pour

$r > 0$,

$$E\left[\sum_{i=1}^m |X_i|^r\right] \leq c_r \sum_{i=1}^m E[|X_i|^r]$$

où $c_r = 1$ quand $r \leq 1$, et $c_r = m^{r-1}$ quand $r \geq 1$.

D'après l'inégalité (1.27)

$$E[|w_{ni}^{2+\delta}|] \leq 2^{1+\delta} \left(\frac{1}{h^{1+\delta/2}} E\left[\left|K\left(\frac{X_i - x}{h}\right)\right|^{2+\delta}\right] + \frac{1}{h^{1+\delta/2}} E\left[\left|K\left(\frac{X_i - x}{h}\right)\right|^{2+\delta}\right] \right)$$

par (1.24)

$$\frac{1}{h^{1+\delta/2}} E\left[\left|K\left(\frac{X_i - x}{h}\right)\right|^{2+\delta}\right] = O(h^{1+\delta/2})$$

et

$$\begin{aligned} \frac{1}{h^{1+\delta/2}} E\left[\left|K\left(\frac{X_i - x}{h}\right)\right|^{2+\delta}\right] &= \frac{1}{h^{1+\delta/2}} \int \left|K\left(\frac{X_i - x}{h}\right)\right|^{2+\delta} f(u) du \\ &= \frac{1}{h^{\delta/2}} \int |K(u)|^{2+\delta} f(x + uh) du \\ &= O\left(\frac{1}{h^{\delta/2}}\right) \end{aligned}$$

Par conséquent,

$$E[|w_{ni}^{2+\delta}|]/n^{\delta/2} = O\left(\left(\frac{1}{nh^{1/2}}\right)^\delta\right)$$

D'après les hypothèses du théorème 2, $nh \rightarrow \infty$
 et, $(1/nh^{1/2}) \rightarrow 0$, on a le résultat suivant $(nh^{1/2})(f_n(x) - f(x)) \rightarrow^d N(0, f(x) \in K^2(u)du)$

1.4.5 Choix théorique optimal du paramètre de lissage :

Pour le paramètre de lissage on fait la distinction entre

*h paramètre de lissage constant (ou global).

*h(x) paramètre de lissage variable (local).

Ces choix différents du paramètre de lissage résultent des estimateurs à noyau suivants :

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

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

Le choix h(x) implique qu'un noyau différent est utilisé en chaque point. Nous allons décrire des choix théoriques optimaux des paramètres de lissage h et h(x). Un critère approprié pour sélectionner un paramètre de lissage constant h est la MISE. Le paramètre de lissage optimal est la valeur de h qui minimise la MISE :

$$\frac{dMISE(f_n)}{dh} = h^3 k_2^2 \beta(f) - \frac{1}{nh^2} j_2$$

$$h_{opt} = \left(\frac{1}{n} \frac{\gamma(K)}{\beta(f)} \right)^{1/5}$$

où $\gamma(K) = j_2 k_2^{-2}$ Un critère approprié pour sélectionner un paramètre de lissage variable (local) h(x) est la mesure de performance locale MSE.

$$\frac{dMSE}{dh} = h^3 k_2^2 f''(x)^2 - \frac{1}{nh^2} f(x) j_2$$

d'où

$$h_{opt}(x) = \left(\frac{1}{n} \frac{f(x)\gamma(K)}{\beta(f)} \right)^{1/5}$$

Remarque : Les choix h_{opt} et $h_{opt}(x)$ sont des choix théoriques, qui ne sont pas utilisables en pratique car ils dépendent des quantités inconnues f' et f'' .
Exemple 2. On utilise les mêmes données de l'exemple 1 et on fait varier le paramètre de lissage pour estimer la densité par la méthode du noyau, et on obtient les résultats représentés dans les trois figures suivantes :

Le choix d'une valeur de h trop grande conduit à une courbe trop lisse. La courbe estimée ne traduit pas suffisamment les variations de la vraie distribution. (fig 1.5) Par contre, en choisissant un paramètre de lissage très petit, l'allure de la distribution change. Il s'agit d'une distribution sur-estimée (fig 1.6)

L'estimation de la densité nécessite également le choix adéquat de la fenêtre h , et pour cette valeur idéale du paramètre de lissage, nous obtenons une allure qui suit parfaitement la vraie distribution. (fig 1.7) Les courbes obtenues dans cet exemple illustrent à quel point les formes estimées sont différentes en fonction de l'ordre de grandeur du paramètre de lissage. La principale difficulté repose sur le choix optimal de la fenêtre h .

1.4.6 Le choix optimal pratique du paramètre de lissage h :

Cas d'une distribution normale

Rappelons l'expression pour le paramètre de lissage optimal constant :

$$h_{opt} = \left(\frac{1}{n} \frac{\gamma(K)}{\beta(f)} \right)^{1/5}$$

Supposons que f appartient à une famille de distributions normales $N(\mu; \sigma^2)$, de moyenne μ et variance σ^2 inconnues, alors

$$f(x) = \frac{1}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right)$$

avec $\varphi = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, la densité de probabilité normale réduite.

$$f''(x) = \frac{1}{\sigma^3} \varphi''\left(\frac{x - \mu}{\sigma}\right)$$

La quantité inconnue $\beta(f)$ s'écrit alors

$$\beta(f) = \int (f''(x))^2 dx = \frac{1}{\sigma^6} \int \left\{ \varphi''\left(\frac{x - \mu}{\sigma}\right) \right\}^2 dx = \frac{1}{\sigma^5} \int \{ \varphi''(v) \}^2 dv$$

$$\text{et } \varphi(v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2}$$

$$\text{d'où } \varphi'(v) = -\frac{v}{\sqrt{2\pi}} e^{-v^2/2}$$

$$\text{et } \varphi''(v) = \frac{v}{\sqrt{2\pi}} (v^2 - 1) e^{-v^2/2}$$

$$\begin{aligned} \beta(f) &= \frac{1}{\sigma^5} \frac{1}{2\pi} \left\{ \int_{-\infty}^{+\infty} v^4 e^{-v^2} dv - 2 \int_{-\infty}^{+\infty} v^2 e^{-v^2} dv + \int_{-\infty}^{+\infty} e^{-v^2} dv \right\} \\ &= \frac{1}{\sigma^5} \frac{1}{2\pi} \left\{ -\frac{1}{2} \int_{-\infty}^{+\infty} v^2 e^{-v^2} dv + \int_{-\infty}^{+\infty} e^{-v^2} dv \right\} \end{aligned}$$

$$\text{Posons } \mu = \sqrt{2v} \rightarrow d\mu = \sqrt{2} dv$$

$$\begin{aligned} &= \frac{1}{\sigma^5} \frac{1}{2\pi} \left\{ -\frac{1}{2} \int_{-\infty}^{+\infty} \frac{u^2}{2} e^{-u^2/2} \frac{du}{\sqrt{2}} + \frac{1}{\sqrt{2}} \int_{-\infty}^{+\infty} e^{-u^2/2} du \right\} \\ &= \frac{1}{\sigma^5} \frac{1}{2\pi} \left\{ -\frac{1}{4} \sqrt{\pi} + \sqrt{\pi} \cdot 1 \right\} \\ &= \frac{1}{\sigma^5} \frac{1}{2\pi} \frac{3}{4} \sqrt{\pi} = \frac{1}{\sigma^5} \frac{3}{8\sqrt{\pi}} \end{aligned}$$

L'expression du paramètre de lissage optimal asymptotique devient

$$h_{AMISE} = \left\{ \frac{8\sqrt{\pi}\gamma(K)}{3} \right\}^{1/5} \sigma n^{-1/5}$$

lissage du type "normale référence" est défini par

$$\hat{h}_{NR} = \left\{ \frac{8\sqrt{\pi}\gamma(K)}{3} \right\}^{1/5} \hat{\sigma} n^{-1/5}$$

Où $\hat{\sigma}$ est un estimateur de σ .

Quelques choix possibles pour $\hat{\sigma}$ sont données ci dessous :

* L'estimateur sans biais de la variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

L'écart interquartile empirique standardisé :

$$\frac{Eies}{\phi^{-1}(\frac{3}{4}) - \phi^{-1}(\frac{1}{4})} \equiv \frac{R}{\phi^{-1}(\frac{3}{4}) - \phi^{-1}(\frac{1}{4})} \simeq \frac{R}{1.349}$$

où $\phi(\cdot)$ est la fonction de répartition de la loi normale réduite. Remarquons que $\phi^{-1}(\frac{3}{4}) - \phi^{-1}(\frac{1}{4})$ est l'écart interquartile d'une variable aléatoire normale réduite. La motivation pour la standardisation utilisant cette quantité est simple :

Si $X \sim N(\mu, \sigma^2)$, alors $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$ et

$$\begin{aligned} P\{\phi^{-1}(\frac{1}{4}) \leq Z \leq \phi^{-1}(\frac{3}{4})\} &= 0.50 \\ \Leftrightarrow P\{\phi^{-1}(\frac{1}{4}) \leq \frac{X-\mu}{\sigma} \leq \phi^{-1}(\frac{3}{4})\} &= 0.50 \\ P\{\sigma\phi^{-1}(\frac{1}{4}) + \mu \leq X \leq \sigma\phi^{-1}(\frac{3}{4}) + \mu\} &= 0.50 \end{aligned}$$

L'écart interquartile de X est

$$F^{-1}(\frac{3}{4}) - F^{-1}(\frac{1}{4}) = \sigma[\phi^{-1}(\frac{3}{4}) - \phi^{-1}(\frac{1}{4})]$$

ce qui justifie l'estimateur proposé. On propose d'utiliser le minimum entre S et $R/(1.349)$, lissage suivant :

$$\hat{h}_{NR} = \left\{ \frac{8\sqrt{\pi}\gamma(K)}{3} \right\}^{1/5} \min\left(S, \frac{R}{1.349}\right) n^{-1/5}$$

Pour quelques noyaux, on donne les expressions de \hat{h}_{NR}

Exemple 3 : on a pris comme échantillon 100 réalisations d'une loi normale $N(0,1)$. Le premier graphe de la figure(1.8) en rouge représente la densité de l'estimateur à noyau de f en utilisant le h_{NR} , le graphe bleu est la densité d'estimateur à noyau de f avec un paramètre de lissage h choisi par défaut.

1.4.7 La méthode des moindres carrés avec validation croisée (MC-VC)

La méthode de validation croisée (en anglais : cross-validation) du type moindres carrés a été introduite par Rudemo (1982) et Bowman(1984). Cette méthode permet d'obtenir un paramètre de lissage simple et attrayant. La méthode est motivée par la décomposition suivante de l'erreur quadratique moyenne intégrée $MISE\{f_n(\cdot)\}$ de l'estimateur à noyau :

$$MISE\{f_n(\cdot)\} = E\left[\int f_n(x)f(x)^2\right]$$

$$E\left[\int \hat{f}^2 dx\right] + 2E\left[\int f_n(x)f(x)\right] + \int f^2 dx$$

Remarquons que le terme $\int f^2(x)dx$ ne dépend pas de h , et donc minimiser $MISE\{f_n(\cdot)\}$ par rapport à h est équivalent à minimiser

$$MISE\{\hat{f}\} - \int f^2(x)dx = E\left[\int f_n^2 + 2 \int f_n(x)f(x)\right]dx$$

L'expression à droite de cette équation est inconnue car elle dépend de la densité f . Un estimateur pour $\int f_n(x)f(x)dx$ est donné par

$$\frac{1}{n} \sum_{i=1}^n f_{n, -i}(X_i)$$

où

$$f_{n, -i}(x) = \frac{1}{n-1} \sum_{j=1, j \neq i}^n K_h(x_i - X_j)$$

est l'estimateur à noyau basé sur l'échantillon "réduit" $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$, où l'observation X_i a été supprimée. On appelle cet estimateur le "leave - one - out estimator". Le terme "validation croisé" vient du fait qu'une partie de l'échantillon est utilisée pour obtenir l'information sur une autre partie : les observations $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$ sont utilisées pour avoir une idée de $f(X_i)$. L'estimateur (1.30) est un estimateur sans biais de $E\{\int f_n(x)f(x)dx\}$. En effet,

$$E\left\{\frac{1}{n} \sum_{i=1}^n f_{n, -i}(X_i)\right\} = \frac{1}{n} \sum_{i=1}^n E\{f_{n, -i}(X_i)\}$$

$$\begin{aligned} E\{f_{n, -i}(X_i)\} &= \frac{1}{n-1} \sum_{j=1, j \neq i}^n E\{K_n(X_i - X_j)\} \\ &= E\{K_h(X_1 - X_2)\} \\ &= \int \int K_h(x - y) f(x) f(y) dx dy \\ &= \int \left\{ \int K_h(x - y) f(y) dy \right\} f(x) dx \end{aligned}$$

$$\begin{aligned}
&= \int E\{f_n(x)\}f(x)dx \\
&= E\left\{\int f_n(x)f(x)dx\right\}
\end{aligned}$$

Un estimateur sans biais pour la quantité

$$MISE\{f_n(x)\} - \int f^2(x)dx = E\left[\int f_n^2(x)dx + 2 \int f_n(x)f(x)dx\right]$$

est donné par

$$LSCV(h) = \int f_n^2(x)dx - \frac{2}{n} \sum_{i=1}^n f_{n,-i}(X_i)$$

Cette quantité est appelée la quantité de validation croisée (LSCV), LSCV veut dire en anglais "Least-square cross-validation" Le paramètre de lissage du type "validation croisée" est la valeur de h qui minimise cette quantité, c'est-à-dire

$$\hat{h}_{LSCV} = \operatorname{argmin}_h LSCV(h)$$

La méthode du maximum de vraisemblance avec validation croisée (MV-VC) Les paramètres d'un modèle statistique peuvent être estimés par la méthode du maximum de vraisemblance. Dans cette méthode, on retient les paramètres qui maximisent la fonction de vraisemblance ou, plus souvent le logarithme de la fonction de vraisemblance. Par exemple, pour l'estimation d'un paramètre inconnu d'une densité, l'estimateur du maximum de vraisemblance serait :

$$\hat{\theta} = \operatorname{argmax}_{\theta} l(\theta|x)$$

où θ est un vecteur de paramètres, x est un vecteur d'observations, et

$$l(\theta|x) = \log \prod_{i=1}^n f(x_i|\theta) = \sum_{i=1}^n \log f(x_i|\theta)$$

Dans le contexte non paramétrique, la forme de f dépend des observations et l'équation (1.31) n'est pas directement applicable. HABBEMA et al.(1974) et DUIN(1976) ont donc suggéré de considérer la méthode de validation croisée pour estimer la fonction de vraisemblance. Supposons que l'on dispose d'une nouvelle donnée y indépendante des données déjà disponibles x . En considérant x fixe, la fonction de log-vraisemblance de h se note de la façon suivante :

$$l(h|y) = \log[\hat{f}(y|h, x)]$$

où \hat{f} est une densité de noyau basée sur les observations x . Comme on ne dispose généralement pas d'un échantillon supplémentaire, la technique de validation croisée est nécessaire. On remplace donc

(1.32) par :

$$l(h|x_i) = \log[\hat{f}_i(x_i|h)]$$

Ceci peut être fait pour chacune des observations. En prenant la moyenne et en utilisant (1.30), on peut obtenir après quelques manipulations mathématiques l'expression suivante pour la log-vraisemblance avec validation croisée :

$$MVVC_f = \frac{1}{n} \sum_{i=1}^n \log \left[\sum_{j \neq i} K\left(\frac{x_i - x_j}{h}\right) \right] + \log \left[\frac{1}{h(n-1)} \right]$$

L'estimation du paramètre de lissage optimal par la méthode du maximum de vraisemblance avec validation croisée est obtenue en maximisant cette expression.

1.4.8 Le choix optimal du noyau

Supposons que l'on a choisi le paramètre de lissage h de telle sorte que le MISE soit minimum. Comme on l'a montré avant, ce paramètre de lissage optimal h_{opt} est égal à

$$h_{opt} = \left(\frac{1}{n} \frac{\gamma(K)}{\beta(f)} \right)^{1/5}$$

On remarque que le h_{opt} dépend de la densité f qui est inconnue. Néanmoins, nous pouvons tirer des conclusions. On remarque que h_{opt} tend vers zéro mais de façon très lente quand n augmente. En substituant h_{opt} dans la formule (1.16), on montre alors que si h est choisi de façon optimale la valeur appropriée pour le MISE sera :

$$MISE_{opt}(f_n) = \frac{5}{4} \left(\frac{\beta(f) j_2^4 k_2^2}{n^4} \right)^{1/5}$$

Pour minimiser le MISE, il faut choisir le noyau K qui minimise la valeur $C(K)$ tel que :

$$C(K) = (j_2^2 k_2^2)^{1/5}$$

Hodges et Lehmann (1956) ont montré que ce problème de minimisation est résolu en choisissant $K(z)$ égale à

$$K(z) = \begin{cases} \frac{3}{4\sqrt{5}}(1 - \frac{1}{5}z^2) & \text{si } |z| \leq \sqrt{5} \\ 0 & \text{sinon} \end{cases}$$

On le note K_{EP} parce que le noyau est appelé le noyau dépanechnikov. Nous pouvons donc considérer l'efficacité d'un noyau K (notée $eff(K)$) quelque en le comparant à K_{EP} puisque ce dernier minimise le MISE si h est choisi de façon optimale.

$$Eff(K) = \left(\frac{MISE_{opt}(\hat{f})_{using K_{EP}}}{MISE_{opt}(\hat{f})_{using K}} \right)^{5/4}$$

Le tableau suivant donne quelques noyaux et leurs efficacités respectives.

Chapitre 2

l'estimateur dans le cas vectoriel

Modèle

Soient X et Y deux variables aléatoires définies sur l'espace de probabilité $(\Omega; \mathcal{A}; \mathbb{P})$ à valeurs dans $\mathbb{R}^p \times \mathbb{R}$. Soit $(X_1; Y_1); \dots; (X_n; Y_n)$ une suite des observations identiquement distribuées que $(X; Y)$. On fixe un point $(x_1; \dots; x_p) \in \mathbb{R}^p$, on désigne par $F_x = F^{(x_1, \dots, x_p)}$ la fonction de répartition conditionnelle de Y sachant $X = x$. L'estimateur à noyau de la fonction de répartition conditionnelle, noté $\hat{F}^x = \hat{F}^{(x_1, \dots, x_p)}$ est défini par :

$$\hat{F}^{(x_1, \dots, x_p)}(y) = \frac{\sum_{i=1}^n K(h_K^{-1}((x_1, \dots, x_p) - X_i))H(h_H^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}((x_1, \dots, x_p) - X_i))} \quad \forall y \in \mathbb{R} \quad (3.1)$$

où K est un noyau, H est une fonction de répartition et $h_K = h_{K;n}$ (resp. $h_H = h_{H;n}$) est une suite de réels positifs. Notons que cet estimateur est une généralisation à la dimension p de l'estimateur (2.1).

Avant d'estimer la densité conditionnelle, on estime la densité de la variable explicative et la den-

sité conjointe des variable (X; Y) dans IR^{p+1} , respectivement par ;

$$F(x_1; \dots; x_p) = \frac{1}{nh_K^P} \sum_{i=1}^n K_i(x_1; \dots; x_p)$$

et

$$g_n^{(i)}(x_1; \dots; x_p, y) = \frac{1}{nh_H^j h_K^P} \sum_{i=1}^n K_i(x_1; \dots; x_p) H_i^{(i)}(y),$$

où

$$K_i(x_1; \dots; x_p) = K(h_K^{-1}((x_1, \dots, x_p) - X_i)) \text{ et } H_i^j(y) = H^{(j)}((h_K^{-1}(y - Y_i))).$$

Ainsi, l'estimateur de la densité conditionnelle noté $\hat{f}^{(x_1; \dots; x_p)}$ est le rapport entre $g_n^{(1)}(x_1; \dots; x_p; y)$ et $f_n(x_1; \dots; x_p)$.

2.1 Hypothèses :

Afin d'établir les propriétés asymptotique de l'estimateur $\hat{h}^{(x_1; \dots; x_p)}$ on introduit les hypothèses suivantes :

(H1) La fonction de répartition conditionnelle est $k+1$ fois continûment dérivables autour de S.
 $\forall y \in S; F^{(x_1; \dots; x_p)}(y) < 1$

(H2) la densité de la variable explicative est strictement positive et de classe C^k au voisinage de $(x_1; \dots; x_p)$,

(H3) La fonction F est strictement croissante, de dérivée bornée et d'ordre k tels que

$$\forall (y_1; y_2) \in IR^2; \quad |F^{(j)}(y_1) - F^{(j)}(y_2)| \leq A|y_1 - y_2|; \quad j = 0, 1.$$

(H4) K un noyau de I_{Rp} d'ordre k , borné et intégrable d'intégrale égale à un

$$(H5) \lim_{n \rightarrow \infty} h_K = 0 \text{ et } \lim_{n \rightarrow \infty} \frac{\log n}{nh_H^j h_K^p} = 0, \quad j = 0, 1$$

$$(H6) \lim_{n \rightarrow \infty} h_H = 0 \text{ et } \lim_{n \rightarrow \infty} n^\beta h_H = \infty \quad \forall \beta > 0$$

2.2 Propriétés asymptotiques :

On établit le résultat suivant

Théorème 3.1 : Soit $(X_1; Y_1); \dots; (X_n; Y_n)$ un n -échantillon de $(X; Y)$, alors, sous les hypothèses (H1)-(H6), on a :

$$\sup_{y \in S} |\hat{F}^{(x_1, \dots, x_2)}(y) - F^{(x_1, \dots, x_2)}(y)| = O(h_K^k) + O(h_H^k) + O\left(\sqrt{\frac{\log n}{nh_H h_p^K}}\right) \quad p.co.$$

et si la suite $(X_i; Y_i)_{i=1, \dots, n}$ est algébriquement α -échangeante, tel que le couple $(X_i; X_j)$ a une densité continue notée f_{ij} , pour tout $i \neq j$, et que

$$\exists \eta > 0, An^{\frac{4-a+3\beta}{a+1} + \eta} \leq h_H h_K^P \quad \text{et} \quad h_K^P \leq A'n^{\frac{1}{1-a}}$$

avec $a > \frac{(6-3\beta) - \sqrt{(6-3\beta)^2 - 12(1+\beta)}}{2}$ alors,

$$\sup_{y \in S} |\hat{F}^{(x_1, \dots, x_2)}(y) - f^{(x_1, \dots, x_2)}(y)| = O(h_K^k) + O(h_H^k) + O\left(\sqrt{\frac{\log n}{nh_H h_p^K}}\right) \quad p.co.$$

Preuve du théorème 3.1 : Par une démonstration analogue à celle utilisée dans le cas uni-variée on montre que notre résultat est issu des trois résul-

tats préliminaire suivants :

$$\sup_{y \in S} |\hat{F}^{(x_1, \dots, x_p)}(y) - F^{(x_1, \dots, x_p)}(y)| = O(h_K^k) + O(h_H^k) + O\left(\sqrt{\frac{\log n}{nh_H h_p^k}}\right) \quad p.co.$$

$$\sup_{y \in S} |\hat{f}^{(x_1, \dots, x_p)}(y) - f^{(x_1, \dots, x_p)}(y)| = O(h_K^k) + O(h_H^k) + O\left(\sqrt{\frac{\log n}{nh_H h_p^k}}\right) \quad p.co.$$

$$\exists \delta > 0 \text{ tel que } \sum_{n=1}^{\infty} IP\{ \inf_{y \in S} |1 - \hat{F}^{(x_1, \dots, x_p)}(y)| < \delta \} < \infty$$

La démonstration de (3 :5) est identique à celle du lemme 2.4 du chapitre précédent. Tandis que la démonstration des équations (3 :3) et (3 :4) repose sur les décompositions suivantes

$$\hat{F}^x(y) - F^x(y) = \frac{1}{f_n(x)} \{ (g_n(x, y) - IE g_n(x, y)) - (F^x(y) f(x) - IE g_n(x, y)) \} + \frac{F^x(y)}{f_n(x)}$$

et

$$\hat{f}^x(y) - f^x(y) = \frac{1}{f_n(x)} \{ (g_n^{(1)}(x, y) - IE g_n^{(1)}(x, y)) - (f^x(y) f(x) - IE g_n^{(1)}(x, y)) \} + \frac{f^x(y)}{f_n(x)}$$

Par ailleurs, la preuve du théorème est basée sur les résultats suivants.

Lemme 3.1 : [5] Sous les hypothèses du théorème 3.1, on a

$$f_n(x_1, \dots, x_p) - f(x_1, \dots, x_p) = O(h_K^k) + O\left(\sqrt{\frac{\log n}{nh_k^p}}\right), p.co.$$

$$\exists \delta > 0 \text{ tel que } \sum_{n=1}^{\infty} IP(f_n(x_1, \dots, x_p) < \delta) < \infty$$

Lemme 3.2 Sous les mêmes conditions du lemme précédent, on a :

$$\sup_{y \in S} |F^{(x_1, \dots, x_p)}(y) f(x_1, \dots, x_p) - IE g_n(x_1, \dots, x_p, y)| = O(h_K^k) + O(h_H^k). \quad (3.10)$$

$$\sup_{y \in S} |f^{(x_1, \dots, x_p)}(y) f(x_1, \dots, x_p) - I E g_n^{(1)}(x_1, \dots, x_p, y)| = O(h_K^k) + O(h_H^k). \quad (3.11)$$

Lemme 3.3 Sous les hypothèses (H1) - (H4) et (H6). On a,

$$\frac{1}{f_n(x_1, \dots, x_p)} \sup_{y \in S} |g_n(x_1, \dots, x_p, y) - I E g_n(x_1, \dots, x_p, y)| = O\left(\sqrt{\frac{\log n}{n h_H h_p^K}}\right) \quad p.$$

$$\frac{1}{f_n(x_1, \dots, x_p)} \sup_{y \in S} |g_n^{(1)}(x_1, \dots, x_p, y) - I E g_n^{(1)}(x_1, \dots, x_p, y)| = O\left(\sqrt{\frac{\log n}{n h_H h_H h_p^K}}\right)$$

et si la suite $(X_i; Y_i)_{i=1; \dots; n}$ est algébriquement α mélangante, tel que le couple $(X_i; X_j)$ a une densité continue notée f_{ij} , pour tout $i \neq j$, et que

$$\exists \eta > 0, A_n^{\frac{4-a+3\beta}{a+1} + \eta} \leq h_H h_K^p \quad \text{et} \quad h_K^p \leq A' n^{\frac{1}{1-\alpha}}$$

avec $a > \frac{(6-3\beta) - \sqrt{(6-3\beta)^2 - 12(1+\beta)}}{2}$. Alors

$$\frac{1}{f_n(x_1, \dots, x_p)} \sup_{y \in S} |g_n(x_1, \dots, x_p, y) - I E g_n(x_1, \dots, x_p, y)| = O\left(\sqrt{\frac{\log n}{n h_H h_p^K}}\right) \quad p.$$

$$\frac{1}{f_n(x_1, \dots, x_p)} \sup_{y \in S} |g_n^{(1)}(x_1, \dots, x_p, y) - I E g_n^{(1)}(x_1, \dots, x_p, y)| = O\left(\sqrt{\frac{\log n}{n h_H h_H h_p^K}}\right)$$

Chapitre 3

l'estimateur dans le cas fonctionnelle

3.1 Quelques hypothèses nécessaires :

Considérons le couple de variables aléatoire $(\chi ; Y)$ où Y est à valeurs dans \mathbb{R} et χ est à valeurs dans \mathbb{R}^P . On a l'opérateur de régression r (qui est un opérateur non linéaire de E dans \mathbb{R}) et nous considérerons les modèles suivants :

$$r \in C_E^0, \quad (2.1)$$

où :

$$C_E^0 = \{f : E \longrightarrow \mathbb{R}, \lim_{d(\chi, \chi') \rightarrow 0} f(\chi') = f(\chi)\}$$

(2.2) et $\exists > 0$ telque

$$r \in Lip_{E, \beta}$$

où

$$Lip_{E, \beta} = \{f : E \longrightarrow \mathbb{R}, \exists C \in \mathbb{R}_*^+, \forall \chi' \in E, |f(\chi) - f(\chi')| < C d(\chi, \chi')^\beta\}$$

On a l'opérateur F_Y^χ (qui est un opérateur non linéaire de $E \rightarrow \mathbb{R}$ dans \mathbb{R}) peut être défini comme suit :

$$F_Y^\chi \in C_0^{E \times \mathbb{R}} \cap S_{cdf}^\chi \quad (2.4)$$

lorsque S_{cdf}^X est :

$S_{cdf}^X = \{f : E \times R \rightarrow, f(\chi, \cdot), \text{est un c.d.f strictement croissant}\}$. et

$$C_{E \times IR}^0 = \left\{ \begin{array}{l} f : E \times IR \rightarrow IR, \forall \chi' \in N_\chi \\ \lim_{d(\chi, \chi') \rightarrow 0} = f(\chi, y) \\ \text{et } \forall y' \in IR, \lim_{|y - y'| \rightarrow 0} f(\chi, y') = f(\chi, y) \end{array} \right.$$

et on a :

$$C_{E \times IR}^0 = \left\{ \begin{array}{l} f : E \times IR \rightarrow IR, \\ \forall (\chi_1, \chi_2) \in N_x^2, \forall (y_1, y_2) \in S^2 \\ |f(\chi_1, y_1) - f(\chi_2, y_2)| \leq C(d(\chi_1, \chi_2)^\beta + |y_1 - y_2|)^\beta \end{array} \right.$$

Où, il existe $\beta > 0$ tel que :

$$F_Y^X \in Lip_{E \times IR, \beta} \cap S_{\chi cdf}^X \quad (2.5)$$

On a l'opérateur f_Y^X (qui est un opérateur non linéaire de $E \times IR$ dans IR) Rappelons que S est un compact fixe $S \in R$. Nous commençons par introduire l'ensemble suivant de contraintes :

$$S_{dens}^X = \left\{ \begin{array}{l} f : E \times IR \rightarrow IR, \\ \exists \xi > 0, \exists ! y_0 \in S, f(\chi, \cdot) \text{ est stricte croissante sur } (y_0 - \xi, y_0) \\ \text{et stricte dcroissante sur } (y_0, y_0 + \xi) \end{array} \right.$$

alors on a :

$$f_{\chi Y} \in C_{E \times IR}^0 \cap S_{dens}^X \quad (2.6)$$

où, il existe $\beta > 0$ tel que :

$$f_{\chi Y} \in Lip_{E \times IR, \beta} \cap S_{dens}^X \quad (2.7)$$

3.2 Type de continuité :

Définition 2.1 : Les modèles fonctionnels non paramétriques (2.1) et (2.4) et (2.6) seront appelés modèles non paramétriques fonctionnels de type de continuité, parce que la propriété de continuité est leur caractéristique fonctionnelle commune principale.

3.3 Type de Lipschitz :

Définition 2.2 : Les modèles fonctionnels non paramétriques (2.3), (2.5) et (2.7) seront appelés modèles non paramétriques fonctionnels de type de Lipschitz pour la même raison

3.4 Probabilités des petites boules :

Soit \mathcal{X} un v.a.f. a valeur dans E et soit χ un élément fixe de E . Nous pouvons écrire :

$$E(1_{[0,1]}(\frac{d(\chi, \mathcal{X})}{h})) = E(1_{B(\chi, h)}(\mathcal{X})) = P(\mathcal{X} \in B(\chi, h)). \quad (2.8)$$

telque

$$B(\chi, h) = \{\chi' \in E, d(\chi, \chi') \leq h\}$$

qui est la boule, par rapport à la topologie induite par d , centrée à χ et de rayon h (où d est une semi-métrique sur E).

nous disons des probabilités de petite boule, car le paramètre de lissage h (également appelé fenêtre) diminue avec la taille de l'échantillon des variables fonctionnelles (plus précisément, h tend à zéro lorsque n a tendance à ∞). Ainsi, lorsque

nous prenons une très grande quantité, h est proche de zéro, puis $B(\chi; h)$ est considéré comme une petite boule et $P(\mathcal{X} \in B(\chi; h))$ comme une petite probabilité de boule. Nous utiliserons pour tout χ en E et pour tout réel h , la notation :

$$\phi_x(h) = P(\mathcal{X} \in B(\chi, h)) \quad (2.9)$$

où :

$$\forall \epsilon > 0, P(\mathcal{X} \in B(\chi, \epsilon)) = \phi_x(\epsilon) \quad (2.10)$$

3.5 Estimation non paramétrique de la densité conditionnelle

On sait que la fonction de densité conditionnelle peut être obtenue en dérivant le conditionnel c.d.f. (Voir 1.9). Comme nous avons l'estimateur $\hat{F}_y^{\mathcal{X}}(\chi; y)$ de $F_y^{\mathcal{X}}(\chi; y)$ nous proposons l'estimation suivante :

$$\hat{F}_y^{\mathcal{X}}(\chi; y) = \frac{d}{du} \hat{F}_y^{\mathcal{X}}(\chi; y) \quad (2.11)$$

en supposant la différentiabilité de H , nous avons

$$\frac{d}{dy} \hat{F}_y^{\mathcal{X}}(\chi; y) = \frac{\sum_{i=1}^n K(h^{-1}d(\chi, \mathcal{X}_i)) \frac{d}{dy} H(g^{-1}(y - Y_i))}{\sum_{i=1}^n K(h^{-1}d(\chi, \mathcal{X}_i))}$$

alors

$$\hat{F}_y^{\mathcal{X}}(\chi; y) = \frac{\sum_{i=1}^n K(h^{-1}d(\chi, \mathcal{X}_i)) \frac{1}{g} H'(g^{-1}(y - Y_i))}{\sum_{i=1}^n K(h^{-1}d(\chi, \mathcal{X}_i))} \quad (2.12)$$

plus généralement, nous pouvons affirmer pour n'importe quel noyau K_0 la définition suivante :

$$\hat{F}_y^{\mathcal{X}}(\chi; y) = \frac{\sum_{i=1}^n K(h^{-1}d(\chi, \mathcal{X}_i)) \frac{1}{g} H'(g^{-1}(y - Y_i))}{\sum_{i=1}^n K(h^{-1}d(\chi, \mathcal{X}_i))} \quad (2.13)$$

présentée par Ferraty (voir [1]).

3.6 La convergence presque complète de l'estimateur a noyaux

La condition de continuité de Hölder :

La condition de continuité de Hölder suivante doit être ajoutée pour l'estimation de densité conditionnelle du noyau $\hat{F}_y^{\mathcal{X}}(x; y)$ définie par (2.13) :

$$\begin{cases} \exists C > \infty \in IR \times IR, |K_0(x) - K_0(x')| \leq C|x - x'|, \\ \lim_{n \rightarrow \infty} \frac{\log n}{n g \phi_x(h)} = 0 \text{ et } \exists \zeta > 0, \lim_{n \rightarrow \infty} g n \zeta = \infty \end{cases} \quad (2.14)$$

Proposition 2.1 On suppose que $\lim_{n \rightarrow \infty} U_n = 0$; $X_n = O_{p.co.}(u_n)$, $\lim_{n \rightarrow \infty} Y_n = l_Y$, où l_Y est un nombre réel déterministe. Nous avons :

$$X_n Y_n = O_{p.co.}(u_n)$$

Lemme 2.1 : Selon les hypothèses (2.10) et (2.11), on a :

$$\hat{r}_1(\chi) = O_{p.co.}\left(\sqrt{\frac{\log n}{n \varphi_x(h)}}\right)$$

Preuve. Voir [1] **Lemme 2.2** : Si K est un noyau de type I, il existe des constantes réelles C et C' positifs tel que :

$$C \varphi_x(h) \leq EK\left(\frac{d(\chi, \mathcal{X})}{h}\right) \leq C' \varphi(h) \quad (2.15)$$

Preuve. Voir [1] **Lemme 2.3** Si K est un noyau de type II et φ_x vérifie que :

$$\exists \epsilon_0, \forall \epsilon < \epsilon_0, \int_0^\epsilon \varphi_x(u) du > C_3 \in \varphi_x(\epsilon) \quad (2.16)$$

Alors : il existe des constantes réelles positifs C et C'' telles que, pour h suffisamment petit :

$$C \varphi_x(h) \leq EK\left(\frac{d(\chi, \mathcal{X})}{h}\right) \leq C'' \varphi(h) \quad (2.17)$$

3.6. La convergence presque complète de l'estimateur a noyau 58

Preuve. Voir [1] **Corollaire 1** (L'équation de Bernstein) Si $\exists M < \infty$; et on note $\sigma^2 = EZ_1^2$, tel que les Z_i sont des v.a.r. indépendants, nous avons

$$\forall \epsilon \geq 0, P(|\sum_{i=1}^n Z_i| > \epsilon n) \leq 2exp\{-\frac{\epsilon^2 n}{2\sigma^2(1 + \epsilon\frac{M}{\sigma^2})}\}$$

Théorème 2.1 Sous le modèle de type-continuité défini par (2.6) et (2.10), et si l'estimation du noyau satisfait (2.11), (2.12) et la condition de continuité de Hölder(2.17), nous avons pour tout sous-ensemble compact $S \subset \mathbb{R}$:

$$\lim_{n \rightarrow \infty} \sup_{y \in S} |f_Y^X(y) - \hat{f}_Y^X(y)| = 0, p.co. \quad (2.18)$$

qui a été présenté par Ferraty (voir[1])

3.7 Conclusion :

La conclusion résumera les principales contributions de ce travail et Comparera brièvement les résultats obtenus avec ceux existant actuellement dans la littérature.

L'objectif de ce travail a été d'étudier la convergence presque complète de l'estimateur de noyau des données vectorielle et fonctionnelle dans un cadre non paramétrique.

. On peut catégoriser nos hypothèses en trois types, des hypothèses structurales, des hypothèses de régularité pour la variable explicative et des hypothèses techniques.

Dans cette étude on considère les variables fonctionnelles et vectorielles. Les techniques du cadre fonctionnel et vectoriel pour l'estimation de la fonction de densité conditionnelle et la fonction de hasard conditionnelle pour les observations aléatoirement l'estimateur à noyau, les techniques étudiées dans ce mémoire présentent des avantages évidents, notre estimateur présente une erreur (bais) moins réduite, Les techniques des données F et V sont couramment rencontrées dans la littérature économique et statistique, nous avons établi les conditions qui garantissent que l'estimateur est asymptotiquement convergent approche a grandement simplifiée les preuves de théorèmes dans la thèse. L'aspect non paramétrique est bien exploité dans ce travail par les hypothèses données.

Bibliographie

Bibliographie

- [1] f. ferraty et p. vieu : Nonparametric functional data analysis : Theory and practice. Springer, 2006.
- [2] g. collomb, w. hardle et s. hassani : A note on prediction via estimation of the conditional mode function. *Journal of Statistical Planning and Inference*, 15 :227-236, 1987.
- [3] g. collomb, s. hassani, p. sarda et p. vieu : estimation non paramétrique de la fonction de hasard pour des observations dépendantes. *statistique et analyse des données*, 10(3) :42-49, 1985.
- [4] bashtannyk, d.m and hyndman, R.J. (2001). Bandwidth selection for kernel conditional density estimation. *Comput. Statist. Data Anal.*, 36, Pages 279-298.
- [5] Dabo-Niang, S. and Laksaci, A. (2007). Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle. *Pub. Inst. Stat. Univ. Paris*, 3, Pages 27-42.
- [6] Collomb, G. Estimation nonparamétrique de la régression (in french). PhD Université Paul Sabatier, Toulouse (1976).
- [7] Hsu, P.L. And Robbins, H. (1974). complete Convergence And The Law Of Large Numbers. *proc. Nat. Acad. Sci. Usa* 33 25-31.
- [8] Lagodowski, Z.A. And Rychlik, Z. (1985). complete Convergence And Convergence Rates For Randomly Indexed Sums Of Random Variables With Multidimensional Indices. *bull. Acad. Sci. Polon. Ser. Math.* 32 219-223.

- [9] Szynal, D. (1972). on Almost Complete Convergence For The Sum Of A Random Number Of Independent Random Variables. *bull. Acad. Sci. Polon. Ser. Math. Astronom. Phys.* 20 571-574.
- [10] Hsu, P.L. And Robbins, H. (1974). complete Convergence And The Law Of Large Numbers. *proc. Nat. Acad. Sci. Usa* 33 25-31.
- [11] Gut, A. (1993). complete Convergence And Ces Summation For I.I.D. Random Variables. *probab. Th. Rel. Fields* 97 169-178.
- [12] Erdos, P. (1950). remark On My Paper "On A Theorem Of Hsu And Robbins". *ann. Math. Statist.* 21 138.
- [13] Erdős, P. (1949). on A Theorem Of Hsu And Robbins. *ann. Math. Statist.* 20 286-291.
- [14] Asmussen, S. And Kurtz, T.G. (1980). necessary And Sufficient Conditions For Complete Convergence In The Law Of Large Numbers. *ann. Probab.* 8 176-182.
- [15] Gut, A. (1993). complete Convergence And Ces Summation For I.I.D. Random Variables. *probab. Th. Rel. Fields* 97 169-178.
- [16] Attouch, M., Laksaci, A., Ould-Saïd, E. (2009). Asymptotic Distribution of Robust Estimator for Functional Nonparametric Models. *Communications in Statistics : Theory and Methods.* 38, 1317-1335.
- [17] Bosq, D., Lecoutre, J. P.(1987). Théorie de l'estimation fonctionnelle. *ECONOMICA*, Paris.
- [18] Burba, F., Ferraty, F., Vieu, P. (2008). Convergence de l'estimateur à noyau des k plus proches voisins en régression fonctionnelle non-paramétrique. *C. R. Acad. Sci. Paris.* 346, 339-342.

- [19] Carbon, M., Francq, C., Tran, L. T. (2007). Kernel regression estimation for random fields. *J. Statist. Plann. Inference.* 137, 778-798.
- [20] Cardot, H., Crambes, C., Sarda, P. (2004). Spline estimation of conditional quantiles for functional covariates, *C. R. Math. Acad. Sci. Paris.* 339, 141-144.
- [21] Collomb, G. (1983). Méthodes non paramétriques en régression, analyse de séries temporelles, prédiction et discrimination. Doctorat d'état, Toulouse
- [22] Collomb, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau. *Z. W. Gebiete.* 66, 441-460.
- [23] Collomb, G. (1985). Nonparametric regression : an up to date bibliography *Statistics.* 16, 309-324.
- [24] Deheuvels, P., Mason, D. (2004). General asymptotic confidence bands based on kernel type function estimators. *Statistical Inference for Stochastic Processes,* 7, 225-277.
- [25] Delsol, L. (2007). Régression non paramétrique fonctionnelle : expression asymptotique des moments, *Ann. I.S.U.P. Vol LI,* 3, 43-67.
- [26] Devroye, L. (1978). The uniform convergence of the Nadaraya-Watson regression function estimate. *Canad. J. Statist.* 6, No.2, 179-191.
- [27] Ezzahrioui, M. (2007). Préviation dans les modèles conditionnels en dimension infinie. PhD Thesis. Univ. du Littoral Côte d'Opale.
- [28] Ferraty, F., Goia, A., Vieu, P. (2002). Functional Nonparametric Model for Time Series : A Fractal Approach for Dimension Reduction. *TEST,* 11, 317-344.

-
- [29] Ferraty F. Laksaci, A. Tadj, A. Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference.* 140, 335-352.
- [30] Ferraty, F., Laksaci, A., Tadj, A. Vieu, P. (2011). Kernel regression with functional response. *Electronic Journal of Statistics.* 5, 159-171.