

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université A/Mira de Béjaïa
Faculté des Sciences Exactes
Département de Mathématique

Mémoire de fin de cycle

Mastère en Mathématique

Option

Statistique et Analyse Décisionnelle

Thème

Estimation fonctionnelle de la densité
conditionnelle avec la methode du noyau

Réalisé par :

Mlle ZOUTAT Khadidja.

Mlle FOURAR Nadjette.

présenté devant le jury :

Présidente	M ^r DAHMANI A	U. A/Mira Béjaïa.
Rapporteur	M ^r AIT SAIDI A	U. A/Mira Béjaïa.
Examineur	M ^r RAHMANI S	U. A/Mira Béjaïa.

Béjaïa 2013.

** * * * Remerciements * * * **

Le présent travail est le couronnement d'un dur labeur. L'aboutissement d'énormes efforts déployés pendant les cinq années à l'Université de Bejaia.

Nous remercions dieu, le tout puissant de nous avoir accordé volonté et courage pour accomplir ce travail.

Au terme de cette fructueuse étape de notre formation et fiers de notre acquis, nous pensons à tous les membres du corps scientifique de la faculté des sciences exactes, qu'ils trouvent à travers ce modeste travail, la matérialisation et /ou la concrétisation de cette mosaïque de connaissances qu'ils nous ont transmise durant notre parcours.

Que le Prof AIT SAIDI Ahmed , encadreur digne trouve à travers ces quelques lignes, l'expression de notre profonde gratitude pour avoir accepté de diriger ce travail ainsi que pour l'aide compétente qu'il nous a apportée. Son œil critique nous a été très précieux pour structurer le travail et pour améliorer la qualité des différentes sections.

Nous ne manquerons pas non plus de dire un grand merci aux membres du jury qui ont accepté d'évaluer ce mémoire à sa juste valeur, et de nous faire part de leurs remarques sûrement pertinentes qui contribueront, sans nul doute, au perfectionnement du présent travail.

*** *Dédicaces* ***

À mes très chers parents.

Vous vous êtes dépensés pour moi sans compter. En reconnaissance de tous les sacrifices consentis par vous, pour me permettre d'atteindre cette étape de ma vie, vous avez toute ma tendresse.

À mes adorables grands-parents.

Vous m'avez beaucoup donné. Vous avez toujours été à mes côtés. Sincère gratitude.

À mes frères et sœurs Lyes, Nabila, Rima, Hamza, Karhina, kati. Pour votre profonde passion et soutien tout au long de mes études. Affectueuse reconnaissance.

À mes oncles, tantes, cousins et cousines. Vous avez de près ou de loin contribué à ma formation.

A mes amis (es) ceux avec qui j'ai parcouru un long chemin avec tant de peine et de joie, inquiétude et espoir.

Sans oublier ma chère amie du parcours Khadidja.

À tous ceux dont les noms n'y figurent pas pour une raison ou une autre trouve l'expression de ma profonde gratitude.

Je dédie ce travail.

Mlle FOURAR Nadjette

** * ** Dédicaces * * ***

Je dédie ce modeste travail aux deux personnes les plus chères pour moi dans ce monde : mes deux parents qui n'ont pas cessé d'être à mes cotés dans toutes les étapes que j'ai franchies et qui m'ont aidé pour arriver là où je souhaitais.

À ma défunte grand-mère que dieu lui accorde miséricorde

À mes deux frères Abdenour et Brahime .

À mes deux sœurs Fatima et Mariem que j'adore.

À tous mes amis et amies et plus particulièrement mon binôme Nadjette.

À tous ceux qui m'ont aidé de pré ou de loin dans la réalisation de ce travail.

Et à tous les étudiants et étudiantes de la première promotion de SAD.

Mlle ZOUTAT Khadidja

Table des matières

Table des matières	i
Table des Figures	iii
Introduction Générale	1
1 Présentation des modèles non paramétriques conditionnels et la méthode du noyau	6
1.1 Les modèles non paramétriques conditionnels comme outils d'analyse prévisionnelle	6
1.1.1 La régression classique	7
1.1.2 Les quantiles de régressions	8
1.1.3 La régression modale	9
1.2 Présentation générale de la méthode du noyau	10
1.2.1 Noyaux usuels	10
1.2.2 Type de noyau	11
1.2.3 Utilisation du noyau pour les données fonctionnelles	12
1.3 Probabilité des petites boules	12
1.4 Quelques résultats théoriques de base	13
2 Présentation de l'estimateur à noyau de fonction de densité conditionnelle	14
2.1 Présentation du modèle	14
2.2 Convergence presque complète	15
2.3 Application et commentaire	20
2.3.1 Application à la prédiction	20
2.3.2 Commentaire	27
2.3.3 Sur la mise en œuvre de l'estimateur de la densité conditionnelle . .	27

3 Simulation	28
3.1 Plan de simulation	28
3.2 Algorithme de simulation	29
3.3 Résultats de simulation	30
3.4 Interprétation des résultats	32
Conclusion générale	34
Bibliographie	35

Table des figures

1	Représentaion graphique de la densité conditionnelle $f(y x)$	4
2	Représentaion graphique de la densité conditionnelle estimée $\hat{f}(y x)$	4
3	Comparaison entre $\hat{f}(y x)$ et $f(y x)$	5
3.1	Représentation graphique $f(y x)$	29
3.2	Représentation des courbes Xi	30
3.3	Résultats de la simulation	31
3.4	Représentation de la densité estimée avec la methode du noyau	32

Introduction Générale

De très nombreux travaux concernent l'étude de modèles sur des variables aléatoires réelles, censurées et multi-variées. Ce domaine de la statistique est toujours très étudié. Cependant, les récentes innovations réalisées sur les appareils de mesure et les méthodes d'acquisition ainsi que l'utilisation intensive de moyens informatiques permettent souvent de récolter des données discrétisées sur des grilles de plus en plus fines, ce qui les rend intrinsèquement fonctionnelles. Les courbes de croissance, les enregistrements sonores, les images satellites, les séries chronologiques, les courbes spectrométriques sont des exemples de données de nature fonctionnelle que le statisticien peut être amené à étudier. C'est pourquoi un nouveau champ de la statistique est dédié à l'étude de données fonctionnelles.

Ces dernières années, l'étude des données fonctionnelles s'est fortement développée. Ce type de données permet de considérer des variables comme des fonctions (au cours du temps par exemple), et ainsi de prendre en compte la nature fonctionnelle de la variable (qui correspond par exemple à un processus continu au cours du temps). Les ouvrages de référence de Ramsay et Silverman (2002, 2005) sur les données fonctionnelles permettent au lecteur d'avoir une vue d'ensemble sur le sujet, tant sur le plan théorique que du point de vue des applications.

La première partie du chapitre 1 est consacré aux modèles non paramétriques conditionnels. Cependant, vu l'étendue de la littérature disponible dans ce domaine, nous ne pouvons pas faire un exposé exhaustif. Ainsi, nous allons restreindre notre étude bibliographique aux modèles non paramétriques suivants :

Sur le modèle de régression

Les premiers résultats en statistique non paramétrique fonctionnelle ont été élaborés par Ferraty et Vieu (2000) et ils concernent l'estimation de la fonction de régression à variable explicative de dimension fractale. Ils ont établi la convergence presque complète d'un estimateur à noyau de ce modèle non paramétrique dans le cas i.i.d. En s'inspirant des développements récents de la théorie des probabilités de petites boules, Ferraty et Vieu

(2004) ont généralisé ces derniers résultats au cas α -mélangeant et ils ont exploité l'importance de la modélisation non paramétrique des données fonctionnelles en appliquant leur étude à la discrimination des courbes et à la prévision. Dans le cadre d'observations fonctionnelles α -mélangeantes, Masry (2005) a montré la normalité asymptotique de l'estimateur de Ferraty et Vieu (2004) pour la fonction de régression. Le lecteur peut trouver dans le livre de Ferraty et Vieu (2006), un large éventail des applications de la fonction de régression en statistique fonctionnelle. La convergence en moyenne quadratique a été étudiée par Ferraty et al.(2007). Plus précisément, ils ont explicité le terme asymptotique exact de l'erreur quadratique. Ce résultat a été utilisé par Rachdi et Vieu (2007) pour déterminer un critère de sélection automatique du paramètre de lissage basé sur la validation croisée. L'aversion locale de ce critère a été étudiée par Ben-henni et al.(2007). On trouvera dans cet article une étude comparative entre l'approche locale et globale. Comme travaux bibliographiques récents en régression, nous renvoyons le lecteur à Ferraty et Vieu (2011) ainsi qu'à Delsol(2011). Des résultats sur l'uniforme intégrabilité ont été établis par Delsol(2007,2009) et Delsol et al.(2011). D'autres travaux se sont intéressés à l'estimation de la fonction de régression en utilisant différentes approches : la méthode des k plus proches voisins par Burba et al(2008), les techniques robustes par Azzidine et al.(2008), Attouch et al.(2009) et Crambes et al.(2008), l'estimation par la méthode simplifiée de polynôme local par Barrientos-Marin et al.(2010).

Sur la fonction de répartition conditionnelle

L'estimation de la fonction de répartition conditionnelle dans un cadre fonctionnel a été introduite par Ferraty et al.(2006). Ils ont construit un estimateur à double noyau pour la fonction de répartition conditionnelle et ils ont précisé la vitesse de convergence presque complète de cet estimateur lorsque les observations sont indépendantes et identiquement distribuées. Le cas des observations α -mélangeantes a été étudié par Ferraty et al.(2005). Un exemple d'application sur la prévision via la médiane conditionnelle, ainsi que la détermination d'intervalles de prédiction ont été considérés dans cet article. Plusieurs auteurs ont traité l'estimation de la fonction de répartition conditionnelle comme une étude préliminaire de l'estimation des quantiles conditionnels . Citons par exemple, Ezzahrioui et Ould-Saïd (2005,2006) qui ont étudié la normalité asymptotique de cette estimation dans les deux cas (i.i.d.et α -mélangeant). Une autre méthode d'estimation pour les quantiles conditionnels a été proposée par Laksaci al.(2009). Les résultats asymptotiques de cet article sont la convergence presque complète et la normalité asymptotique dans le cas i.i.d. Nous renvoyons à Cardot et al. (2004) pour une approche linéaire des quantiles

conditionnels en statistique fonctionnelle.

Sur la densité conditionnelle et ses applications

L'estimation de la fonction de densité conditionnelle et ses dérivées, en statistique fonctionnelle, ont été introduites par Ferraty et al.(2006). Ces auteurs ont obtenu la convergence presque complète dans le cas i.i.d. Depuis cet article, une littérature abondante s'est développée sur l'estimation de la densité conditionnelle et ses dérivées, notamment afin de l'utiliser pour estimer le mode conditionnel. En effet, en considérant des observations α -mélangeantes, Ferraty et al.(2005) ont établi la convergence presque complète d'un estimateur à noyau du mode conditionnel défini par la variable aléatoire maximisant la densité conditionnelle. Alternativement, Ezzahrioui et Ould-Saïd (2005,2006) ont estimé le mode conditionnel par le point qui annule la dérivée de l'estimateur à noyau de la densité conditionnelle. Ces derniers se sont concentrés sur la normalité asymptotique de l'estimateur proposé dans les cas (i.i.d.et mixing). La précision des termes dominants de l'erreur quadratique de l'estimateur à noyau de la densité conditionnelle a été obtenue par Laksaci (2007). Nous renvoyons à Laksaci et al.(2010) pour la question du choix du paramètre de lissage dans l'estimation de la densité conditionnelle à variable explicative fonctionnelle. La problématique abordée dans ce travail est l'estimation non paramétrique de la densité conditionnelle en traitant le cas où la variable explicative est fonctionnelle. Nous construisons cette estimation par la méthode du noyau en considérant une suite d'observations indépendantes identiquement distribuées (i.i.d). On établit la convergence presque complète de cette d'estimateur en précisant sa vitesse et à titre illustratif, nous donnerons un exemple d'application sur des données simulées [4].

Youndjé dans ses travaux(pour plus de détail voir thèse Youndjé [6])a donné une présentation assez générale sur l'estimation de la densité conditionnelle quand X et Y sont deux variables réelles dans laquelle il a étudié deux exemples à travers des données simulées. Nous donnerons ici l'exemple où la distribution de Y sachant $X = x$ est uni-modal, et où X est une variable aléatoire qui suit une loi exponentielle de paramètre 1, et $X|Y$ suit une loi normale de moyenne x et de variance 1.

Le graphe qui représente la fonction de densité conditionnelle $f(y|x)$ est le suivant

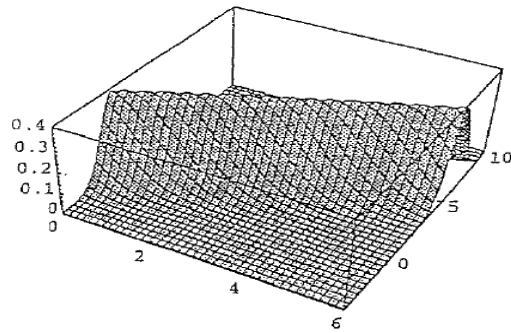


FIG. 1 – Représentaion graphique de la densité conditionnelle $f(y|x)$

Pour estimer la fonction de densité Youndjé a générer un échantillon de taille $n = 50$ du couple aléatoire (X, Y) , a utiliser le noyau parabolique $K_3(u) = \begin{cases} \frac{3}{4}(1 - u^2), & \text{si } |u| \leq 1; \\ 0, & \text{si } |u| > 1. \end{cases}$ et a fixé le paramètre de lissage a $h = 1.00874$.

le graphe de la densité conditionnelle $\hat{f}(y|x)$ estimée est représenté dans la figure suivante :

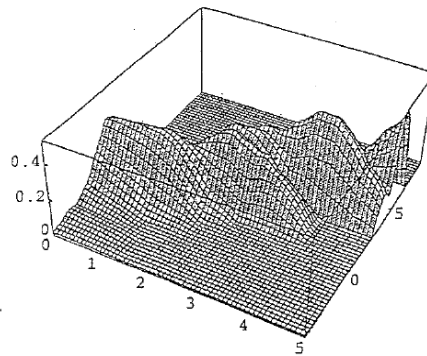


FIG. 2 – Représentaion graphique de la densité conditionnelle estimée $\hat{f}(y|x)$

Afin de montrer la performance de l'estimateur du noyau sur cet exemple, youndjé a comparé les deux fonctions de densités $f(y|x)$ et $\hat{f}(y|x)$ à un point fixe $x = 0.5$ dans le graphe suivant :

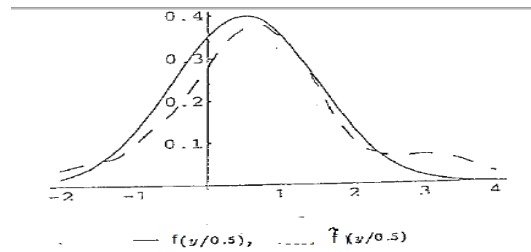


FIG. 3 – Comparaison entre $\hat{f}(y|x)$ et $f(y|x)$

Au cours de notre application, nous essayerons de reprendre le même exemple, mais dans le cas où X appartient à un espace fonctionnel.

Chapitre 1

Présentation des modèles non paramétriques conditionnels et la méthode du noyau

Dans ce chapitre en premier lieu nous allons faire un petit panorama sur les modèles non paramétriques conditionnels : la régression classique, les quantiles de régressions et la régression modale.

En deuxième lieu nous allons présenter le meilleur "généraliste" en matière d'estimation de densité : l'estimateur à noyau. Nous abordons le cas fonctionnel où nous montrons comment introduire le noyau de pondération locale quand on a à faire à des données fonctionnelles.

Définition 1.0.1. Une variable aléatoire est dite fonctionnelle si elle est à valeurs dans un espace de dimension infinie.

1.1 Les modèles non paramétriques conditionnels comme outils d'analyse prévisionnelle

Étudier les liens entre deux variables aléatoires est une question très importante en statistique. D'un point de vue historique, ce problème a été abordé pour la première fois dans un contexte géométrique, par Galileo Galilei en 1632. L'idée principale de ce dernier est d'ajuster un nuage de point par une droite permettant d'interpréter la relation entre des données contaminées. Une formulation mathématique pour ce problème, connu dans la littérature sous le nom de régression linéaire, a été donné par Legendre et Gauss indépendamment, en (1805) et (1809) est basée sur le principe des moindres carrés. En statistique, ce problème peut être modélisé de la manière suivante : supposons qu'on dispose de deux variables aléatoires dépendantes X et Y , la prévision de Y sachant X se fait

à travers X par une application r . Autrement dit, on cherche une fonction r telle que $r(X)$ soit une bonne approximation de Y selon un critère donné. Ainsi, le problème devient la minimisation de la fonction de risque suivante :

$$err(r) = E(l(Y - r(X)))$$

où l est une fonction de perte. Si cette fonction de perte est convexe et admet un minimum unique, alors on peut prendre la quantité

$$\hat{r}(x) = \underset{c \in \mathbb{R}}{\operatorname{arg\,min}} E(l(Y - c) | X = x)$$

comme approximation de Y sachant $X = x$. En conséquence, on peut dire qu'en statistique non paramétrique, les outils naturels pour faire la prévision sont les modèles les plus sollicités tels que la régression classique, la régression modale (mode conditionnel), les quantiles de régression (la médiane conditionnelle)[1]. Dans la suite de ce paragraphe, nous rappelons les définitions de ces outils, en explicitant la fonction de risque associée à chaque modèle et en mettant l'accent sur la diversité et l'étendue de la littérature disponible sur ces modèles.

1.1.1 La régression classique

Ayant observé X , la valeur moyenne de Y est l'approximation la plus utilisée comme prédicateur de Y . La fonction de risque associée à ce prédicateur est

$$l(z) = z^2.$$

Alors, la régression classique est l'unique solution de

$$\hat{r}(x) = E(Y | X = x).$$

Cette estimation est équivalente à l'estimation par la méthode de maximum vraisemblance classique lorsque Y et X sont liées par la relation suivante :

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

où ϵ est de loi normale de variance finie σ^2 et indépendante de X . L'optimalisé de cet outil de prévision dans le cas où le bruit est gaussien la rendu très populaire. Dans notre contexte non paramétrique, les premiers résultats ont été obtenus par Tukey (1961). Tandis que l'estimation par la méthode du noyau a été utilisée pour la première fois en 1964 séparément par Nadaraya et Watson. Cette méthode d'estimation a connu un développement continu. En effet, Devroye (1978) a établi la convergence uniforme presque sûre de cet estimateur. Le taux de convergence optimal pour la régression non paramétrique a été donné par Stone (1980,1982). Collomb (1981, 1983, 1984,1985) apporte une contribution déterminante sur ce modèle. Ces travaux se sont focalisés sur l'utilisation de la régression dans la prévision de séries chronologiques. Les premiers résultats asymptotiques sur l'estimation non paramétrique de la fonction de régression sur les processus α -mélangeants ont été élaborés par Györfi et al.(1989). Dans ce cadre α -mélangeant, Vieu (1991) a donné les termes asymptotiquement exacts de l'erreur quadratique de l'estimateur à noyau de la fonction de régression. Nous renvoyons à Bosq et Lecoutre (1987), Schimek (2000), Sarda et Vieu (2000) pour un large éventail de références. Plus récemment, le modèle de régression non paramétrique a été considéré pour la prévision spatiale (voir de Lu et Chen (2004), Biau et Cadre (2004), Carbon et al.(2007), Li et Tran(2009)).

1.1.2 Les quantiles de régressions

L'inconvénient de la régression classique est que l'estimation de la fonction de régression est sensible aux valeurs aberrantes et peut se montrer inappropriée dans certains cas, comme lorsque la distribution est multimodale ou fortement asymétrique. Ce manque de robustesse peut être résolu par la prévision en utilisant les quantiles conditionnels[4]. Cet outil de prévision est obtenu en introduisant la fonction de risque

$$l(z) = |z| - 2p(z - 1)z.$$

Ainsi le prédicteur est la solution du problème d'optimisation

$$\hat{r}(x) = \min_{c \in \mathbb{R}} E(l(Y - c)|X = x) \quad (1.1)$$

La robustesse de cette méthode provient de la convexité et la bornitude de la fonction de risque. Ces deux propriétés sont indispensables pour la robustesse d'un modèle, la première sert à assurer l'unicité de la solution du problème d'optimisation (1.1), tandis

que la deuxième est utilisée pour réduire l'influence des observations aberrantes sur la prévision [3]. L'utilisation des quantiles de régression et prévision a été étudiée par plusieurs auteurs. A titre d'exemple, Stone (1977) est le premier qui a estimé le quantile conditionnel. Il a établi la convergence en probabilité d'un estimateur basé sur l'estimation empirique de la fonction de répartition conditionnelle. La normalité asymptotique et la convergence uniforme de l'estimateur à noyau de quantiles conditionnels ont été obtenues par Samanta (1989) dans le cas i.i.d. En 1991, Roussas a traité le cas où les observations sont issues d'un processus de Markov. Il a établi la convergence presque sûre d'un estimateur à noyau pour ce paramètre. Berline et al. (1998) donnent un théorème général sur la normalité asymptotique des estimateurs des quantiles conditionnels, indépendamment de la corrélation entre les observations. Zhou et Liang (2003) ont utilisé l'approche ci-dessus (2) pour estimer la médiane conditionnelle. Ils ont montré la normalité asymptotique de l'estimateur construit lorsque les observations sont α -mélangeantes. Une généralisation de cet estimateur a été proposée par Gannoun et al. (2003). Ces derniers ont établi la convergence presque complète et la normalité asymptotique de ces estimateurs. Récemment, Li et Lin (2007) ont rétabli les résultats de Zhou et Liang (2003) pour le cas de variables associées. Nous renvoyons à Lemdani et al. (2009) pour l'estimation des quantiles conditionnels sur des données tronquées.

1.1.3 La régression modale

Un autre prédicateur alternatif à la régression classique est le mode conditionnel. Par définition, le mode conditionnel est la valeur qui maximise la densité conditionnelle. On suppose qu'il existe un compact \mathcal{S} où le mode est unique noté θ . Ce prédicateur est obtenu en considérant :

$$\theta(x) = \underset{y \in \mathcal{S}}{\operatorname{arg\,sup}} f(y|x)$$

où $f(y|x)$ est la densité conditionnelle de Y sachant $X = x$. En utilisant le critère $\operatorname{err}(r)$, nous pouvons remarquer que le mode conditionnel est obtenu en considérant la fonction de risque

$$\begin{cases} l(z)=0 & \text{si } z=0, \\ 1 & \text{sinon.} \end{cases}$$

Un premier estimateur du mode conditionnel fut étudié par Collomd et al.(1987). Ces derniers ont établi la convergence uniforme de cet estimateur. Dans cet article, Collomb a donné un exemple d'un processeur φ -mélangeant où le mode conditionnel prévoit mieux que la régression. Samanta et Thavaneswaran (1990) ont étudié la normalité asymptotique. Motivé par sa supériorité sur la régression en prévision, le mode conditionnel a fait l'objet de plusieurs travaux : Quintela-Del-Rio et Vieu (1997) sur le processus α -mélangeant, Ould-Saïd (1997) pour le cas ergodique, Louani et Ould-Saïd (1999) pour la normalité asymptotique dans le cas des observations fortement mélangeantes. Loannides et Matzner-Lober (2004) proposent un estimateur pour le mode conditionnel lorsque les variables sont entachées d'erreurs. Nous renvoyons à Culd-Lober (2004) propose un estimateur pour le mode conditionnel lorsque les variables sont entachées d'erreurs. Nous renvoyons à Culd-Saïd et Cai (2005) pour le cas des données censurées.

1.2 Présentation générale de la méthode du noyau

Un noyau est une fonction de pondération utilisée dans les techniques d'estimation non paramétrique. Les noyaux interviennent dans l'estimation à noyau pour estimer la densité de probabilité d'une variable aléatoire, elle se base sur un échantillon d'une population statistique et permet d'estimer la densité en tout point du support[8]. En ce sens, cette méthode généralise astucieusement la méthode d'estimation par un histogramme. Cette méthode est aussi appelée méthode de Parzen-Rozenblatt.

1.2.1 Noyaux usuels

Il existe plusieurs types de noyaux, citons les plus utilisés.

$$K_1(u) = \begin{cases} \frac{1}{2}, & \text{si } |u| \leq 1; \\ 0, & \text{si } |u| > 1. \end{cases} \quad (\text{noyau rectangulaire})$$

$$K_2(u) = \begin{cases} (1 - |u|), & \text{si } |u| \leq 1; \\ 0, & \text{si } |u| > 1. \end{cases} \quad (\text{noyau triangulaire})$$

$$K_3(u) = \begin{cases} \frac{3}{4}(1 - u^2), & \text{si } |u| \leq 1; \\ 0, & \text{si } |u| > 1. \end{cases} \quad (\text{noyau parabolique ou d'Epanechnikov});$$

$$K_4(u) = \begin{cases} \frac{15}{16}(1 - u^2)^2 & \text{si } |u| \leq 1; \\ 0, & \text{si } |u| > 1. \end{cases} \quad (\text{noyau quadratique});$$

$$K_5(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \quad (\text{noyau gaussien});$$

$$K_6(u) = \frac{1}{2} \exp(-|u|/\sqrt{2}) \sin(|u|/\sqrt{2} + \pi/4) \quad (\text{noyau de Silverman}).$$

1.2.2 Type de noyau

Nous allons considérer deux sortes de noyaux : noyaux de type I et noyau de type II. La famille du noyau de type I contient les noyaux usuels discontinus, tandis que la seconde famille contient les noyaux standards continus [2].

Définition 1.2.1. Une fonction $K : \mathbb{R} \rightarrow \mathbb{R}^*$ telle que $\int K = 1$ est dite noyau de type I si il existe deux constantes réelles $0 < C_1 < C_2 < \infty$ telle que

$$C_1 1_{[0,1]} \leq K \leq C_2 1_{[0,1]}.$$

Définition 1.2.2. Une fonction $K : \mathbb{R} \rightarrow \mathbb{R}^*$ telle que $\int K = 1$ est dite noyau de type II si son support est $[0, 1]$ et si sa dérivé K' existe sur l'intervall $[0, 1]$ et cette dérivé vérifie la condition suivante :

si il existe deux constante réelles C_1 et C_2 telle que $-\infty < C_2 < C_1 < 0$ alors

$$C_2 \leq K' \leq C_1.$$

Afin de simplifier notre objectif, pour la pondération locale des variables aléatoires réelles nous allons définir le noyau suivant :

Définition 1.2.3. Une fonction $K : \mathbb{R} \rightarrow \mathbb{R}^*$ telle que $\int K = 1$ sur un support compact $[-1,1]$ et $\forall u \in (0, 1), K(u) > 0$ est appelé noyau de type 0 [1].

1.2.3 Utilisation du noyau pour les données fonctionnelles

Soit X_1, X_2, \dots, X_n n v.a.f dans \mathbb{E} et soit x un élément fixe dans \mathbb{E} , l'idée serait de transformer les n v.a.f X_1, X_2, \dots, X_n en n quantités comme suite :

$$\frac{1}{v(h)} K\left(\frac{d(x, X_i)}{h}\right)$$

où d est une semi-métrie sur \mathbb{E} , K est un noyau réel positif et $v(h)$ le volume de

$$B(x, h) = \{\acute{x} \in \mathbb{E}, d(x, \acute{x}) \leq h\}$$

$B(x, h)$ est la boule de centre x et de rayon h associée à la semi-norme. Ce pendant on doit définir $v(h)$, pour cela on passe à la normalisation en utilisant directement la distribution de probabilité des v.a.f. Les variables pondérées locales du noyau fonctionnel sont définies par :

$$\Delta_i = \frac{K\left(\frac{d(x, X_i)}{h}\right)}{E\left(K\left(\frac{d(x, X_i)}{h}\right)\right)} \quad (1.2)$$

Δ_i est une extension du noyau de pondération locale multivariante dans le cadre fonctionnel[2].

1.3 Probabilité des petites boules

Soit X une v.a.f dans \mathbb{E} , x un élément fixe dans \mathbb{E} , et pour mieux fixer les idées, on utilise un noyau asymétrique simple de type I. La relation qui lie la pondération locale et la notion de probabilités petite boule est donnée comme suit :

$$E\left(1_{[0,1]}\left(\frac{d(x, X)}{h}\right)\right) = E(1_{B(x, h)}(X)) = P(X \in B(x, h)) \quad (1.3)$$

Dans la suite de notre travail nous utiliserons, pour tout x dans \mathbb{E} et pour tout h réel positif, la notation suivante :

$$\varphi_x(h) = P(X \in B(x, h)) \quad (1.4)$$

1.4 Quelques résultats théoriques de base

Comme l'idée du noyau du poids local fonctionnel est le foyer de toutes les méthodes non paramétriques fonctionnelles qu'on va étudier, on utilise les deux résultats suivants :

Lemme 1.4.1. *Si K est un noyau de type I, alors il existe deux constantes réelles non négatives C et \acute{C} telles que :*

$$C_{\varphi_x(h)} \leq EK \left(\frac{d(x,X)}{h} \right) \leq \acute{C}_{\varphi_x(h)} \quad (1.5)$$

Lemme 1.4.2. *Si K est un noyau de type II, et si $\varphi_x(\cdot)$ satisfait*

$$\exists C_3 > 0, \forall \epsilon_0 < \epsilon_0 \int_0^{\epsilon} \varphi_x(u) du > C_3 \epsilon \varphi_x(\epsilon) \quad (1.6)$$

et si il existe deux constantes réelles non négatives C et \acute{C} alors :

$$C_{\varphi_x(h)} \leq EK \left(\frac{d(x,X)}{h} \right) \leq \acute{C}_{\varphi_x(h)} \quad (1.7)$$

Preuve. (Voir Ferraty et vieu [2])

Chapitre 2

Présentation de l'estimateur à noyau de fonction de densité conditionnelle

Dans ce chapitre nous allons vous présenter l'estimateur à noyau de la densité conditionnelle quand la variable explicative est fonctionnelle.

Il existe plusieurs estimateurs de la densité conditionnelle tels que : l'estimateur des "points les plus proches", l'estimateur "histogramme", l'estimateur " $\delta - suite$ " et l'estimateur à noyau[7].

Pour notre travail nous nous sommes concentrés sur l'estimation par la méthode du noyau, car l'estimateur à noyau d'une densité est l'un des estimateurs les plus étudiés et les plus performants . Ce travail est basé sur les résultats de Ferraty et Vieu.

2.1 Présentation du modele

Soit $(X_1, Y_1), \dots, (X_n, Y_n)$ un échantillon aléatoire du couple (X, Y) indépendant, identiquement distribué qui est à valeurs dans $\mathcal{H} \times \mathbb{R}$, où \mathcal{H} est un espace de Hilbert muni du produit scalaire $\langle \cdot, \cdot \rangle$, et dont la norme associée est notée $\|\cdot\|$.

l'estimateur à noyau de la densité conditionnelle $f(y|x)$ noté $\hat{f}(y|x)$ est défini par :

$$\hat{f}(y|x) = \frac{\sum_{i=1}^n K\left(\frac{d(X_i, x)}{h}\right) \frac{\partial}{\partial y} H\left(\frac{y - Y_i}{g}\right)}{\sum_{i=1}^n K\left(\frac{d(X_i, x)}{h}\right)}, \forall y \in \mathbb{R} \quad (2.1)$$

où H est définit par

$$\forall u \in \mathbb{R}, H(u) = \int_{-\infty}^u K_0(v) dv$$

et

$$d(X_i, x) = \|X_i - x\|.$$

La fonction K est un noyau de type I ou de type II et la fonction K_0 est un noyau de type 0 et $h = h(n)$ (resp. $g = g(n)$) est une suite de nombres réels positifs qui tend vers zéro lorsque n tend vers l'infini. Il est aussi appelé le paramètre de lissage ou largeur de fenêtre[3].

2.2 Convergence presque complète

Rappel : une suite $(Z_n)_{n \in \mathbb{N}^*}$ de variables aléatoires réelles est dite convergente presque complètement vers la variable aléatoire réelle Z si

$$\forall \epsilon > 0, \sum_{n \geq 1} P(|Z_n - Z| > \epsilon) < \infty. \quad (2.7)$$

Aussi, on dit que la vitesse de convergence presque complète de $(Z_n)_{n \in \mathbb{N}^*}$ vers Z est d'ordre v_n si

$$\exists \epsilon_0 > 0, \sum_{n \geq 1} P(|Z_n - Z| > \epsilon_0 v_n) < \infty. \quad (2.8)$$

et on écrit

$$Z_n - Z = O_{p.co.}(v_n).$$

Tout au long de notre travail, nous noterons par C et C' deux constantes génériques et strictement positives. Afin d'établir la convergence presque complète (p.co.) de notre estimateur on considère les hypothèses suivantes. Soient x (resp. y) un élément de \mathcal{H} (resp. de \mathbb{R}), $\mathbb{N}_x \subset \mathcal{H}$ un voisinage (a neighborhood) de x et \mathcal{S} un sous ensemble compact de \mathbb{R} tels que :

$$P(d(X, x) < h) = \varphi_x(h) > 0, \quad (2.2)$$

$$\exists C > 0, \forall (x, x') \in \mathbb{R} \times \mathbb{R}, |K_0(x) - K_0(x')| \leq C |x - x'|,$$

$$\lim_{n \rightarrow \infty} \frac{\log n}{ng\varphi_x(h)} = 0 \quad \text{et} \quad \exists \zeta > 0, \lim_{n \rightarrow \infty} gn^\zeta = \infty. \quad (2.3)$$

$$\exists C > 0, \exists \varepsilon_0, \forall \varepsilon < \varepsilon_0, \int_0^\varepsilon \varphi_x(u) du > C\varepsilon\varphi_x(\varepsilon), \quad (2.4)$$

$$\begin{cases} \exists C_x > 0 \text{ tel que } \forall (y_1, y_2) \in \mathcal{S}^2, \forall (x_1, x_2) \in N_x \times N_x, \\ |f(y_1|x_1) - f(y_2|x_2)| \leq C_x(d^{\beta_1}(x_1, x_2) + |y_1 - y_2|^{\beta_2}), \beta_1 > 0, \beta_2 > 0. \end{cases} \quad (2.5)$$

Théorème 2.2.1. *Sous les conditions (2.1), (2.2), (2.3), (2.4) et (2.5), nous avons pour tout nombre réel fixé y :*

$$\hat{f}(y|x) - f(y|x) = O(h^{\beta_1}) + O(g^{\beta_2}) + O_{p.co.}\left(\sqrt{\frac{\log n}{ng\varphi_x(h)}}\right). \quad (2.6)$$

Preuve. *La preuve est basée sur la décomposition qui suit*

$$\hat{f}(y|x) - f(y|x) = \frac{(\hat{r}_3(x, y) - E\hat{r}_3(x, y)) - (f(y|x) - E\hat{r}_3(x, y))}{\hat{r}_1(x)} - \frac{f(y|x)}{\hat{r}_1(x)} \{\hat{r}_1(x) - 1\} \quad (2.9).$$

où \hat{r}_1 est défini par

$$\hat{r}_1(x) = \frac{1}{n} \sum_{i=1}^n \Delta_i \quad (2.10)$$

et où

$$\hat{r}_3(x, y) = \hat{r}_1(x) \hat{f}(y|x) = \frac{1}{n} \sum_{i=1}^n \Delta_i \Gamma_i(y) \quad (2.11)$$

avec

$$\Gamma_i(y) = \frac{1}{g} K_0\left(\frac{y - Y_i}{g}\right) \quad (2.12)$$

Ainsi, la preuve est une conséquence directe des résultats qui suivent.

Lemme 2.2.1. *Sous les hypothèses (2.2) et (2.5), lorsque n tend vers l'infini, nous avons :*

$$E\hat{r}_3(x, y) - f(y|x) = O(h^{\beta_1}) + O(g^{\beta_2}). \quad (2.13)$$

Preuve. *Puisque $E\Delta_i = 1$ et puisque K_0 est une fonction intégrable, nous avons :*

$$\begin{aligned} E\hat{r}_3(x, y) - f(y|x) &= E\Delta_1\Gamma_1(y) - f(y|x) \\ &= E(\Delta_1(E(\Gamma_1(y)|X) - f(y|x))) \\ &= E(\Delta_1 \int_{\mathbb{R}} \frac{1}{g} K_0\left(\frac{y-u}{g}\right) f(u|X) du - f(y|x)) \\ &= E(\Delta_1 \int_{\mathbb{R}} \frac{1}{g} K_0\left(\frac{y-u}{g}\right) f(u|X) - f(y|x)) dv \\ &= E(\Delta_1 \int_{\mathbb{R}} \frac{1}{g} K_0(v) (f(y - vg|X) - f(y|x)) dv) \\ &= E(1_{\beta(x,h)}(X) \Delta_1 \int_{\mathbb{R}} \frac{1}{g} K_0(v) (f(y - vg|X) - f(y|x)) dv). \end{aligned} \quad (2.14)$$

Compte tenu du fait que le support $K_0 = [-1, 1]$ et puisque h et g tendent vers zéro, la condition de Hölder (2.5) permet d'écrire que :

$$\sup_{v \in [-1, 1]} |f(y - vg|X) - f(y|x)| = O(h^{\beta_1}) + O(g^{\beta_2}) \quad (2.15)$$

Ainsi, le résultat (2,13) découle directement en combinant (2,14), (2,15) et le fait que $E\Delta_1 = 1$.

Lemme 2.2.2. *Sous les hypothèses du théorème, lorsque n tends vers l'infini, nous avons :*

$$\hat{r}_3(x, y) - E\hat{r}_3(x, y) = O_{p.co.}\left(\sqrt{\frac{\log n}{ng\varphi_x(h)}}\right). \quad (2.16)$$

Preuve. Pour cela, nous utilisons la décomposition suivante :

$$\hat{r}_3(x, y) - E\hat{r}_3(x, y) = \frac{1}{n} \sum_{i=1}^n Z_i, \quad (2.17)$$

où

$$Z_i = (T_i - ET_i) \quad \text{et} \quad T_i = \Delta_i \Gamma_i(y). \quad (2.18)$$

Afin d'appliquer une inégalité de type-Bernstein, nous commençons par montrer

$$|T_i| \leq \frac{C}{g\varphi_x(h)} \quad \text{et} \quad ET_i^2 \leq \frac{C}{g\varphi_x(h)}. \quad (2.19)$$

En utilisant le Lemme 1.4.1 ou le Lemme 1.4.2 et en tenant compte de l'hypothèse (2.4) et du fait que K est de type I ou de type II, nous avons :

$$C\varphi_x(h) \leq EK\left(\frac{d(X_i, x)}{h}\right) \leq C'\varphi_x(h), \quad (2.20)$$

et en utilisant le dernier résultat et puisque K_0 est borné, nous obtenons

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

Le second moment des variables T_i peut être calculé en utilisant l'intégration par changement de variable :

$$\begin{aligned} ET_i^2 &= E\left(\Delta_i^2 \frac{1}{g^2} K_0^2\left(\frac{y - Y_i}{g}\right)\right) = E\left(E\left(\Delta_i^2 \frac{1}{g^2} K_0^2\left(\frac{y - Y_i}{g}\right) \middle| X = x\right)\right) \\ &= \frac{1}{g^2} E\left(\Delta_i^2 \int_{\mathbb{R}} K_0^2\left(\frac{y - u}{g}\right) f(u|X) du\right) \\ &\leq \frac{g}{g^2} E\left(\Delta_i^2 \int_{\mathbb{R}} K_0^2(v) f(y - vg|X) dv\right) \\ &\leq \frac{1}{g} E\Delta_i^2. \end{aligned} \quad (2.21)$$

Puisque $0 < \int K^2 < \infty$, si K est de type I (resp.II) alors $\frac{K^2}{\int K^2}$ est aussi de type I (resp.II). Ainsi, en appliquant le Lemme 1.4.1 ou le Lemme 1.4.2 on trouve

$$C\varphi_x(h) \leq EK^2\left(\frac{d(X_i, x)}{h}\right) \leq C'\varphi_x(h), \quad (2.22)$$

et en utilisant le dernier résultat, on écrit que

$$\frac{C}{\varphi_x(h)} \leq E\Delta_i \leq \frac{C'}{\varphi_x(h)}, \quad (2.23)$$

ce qui implique que

$$ET_i^2 \leq \frac{C}{g\varphi_x(h)} \quad (2.24)$$

En tenant compte de (2.19), on peut appliquer l'inégalité de type-Bernstein-type donné par le corollaire A.9 – i (voir Ferraty et al.[2]), et on obtient :

$$\forall \varepsilon \geq 0, P(|\hat{r}_3(x, y) - E\hat{r}_3(x, y)| > \varepsilon) \leq 2 \exp \frac{\varepsilon^2 ng\varphi_x(h)}{2C'(1 + \varepsilon)} \quad (2.25)$$

Puisque la suite $\frac{\log n}{ng\varphi_x(h)}$ tend vers zéro, en choisissant $\varepsilon = \varepsilon_0 \sqrt{\frac{\log n}{ng\varphi_x(h)}}$ dans le résultat (2.25) nous obtenons directement

$$\begin{aligned} P(|\hat{r}_3(x, y) - E\hat{r}_3(x, y)| > \varepsilon_0 \sqrt{\frac{\log n}{ng\varphi_x(h)}}) &\leq 2 \exp \frac{\varepsilon_0^2 \log n}{2C'(1 + \varepsilon_0 \sqrt{\frac{\log n}{ng\varphi_x(h)}})} \\ &\leq 2n^{-C\varepsilon_0^2}, \end{aligned} \quad (2.26)$$

et il en résulte que pour ε_0 assez large ($\varepsilon_0 > \frac{1}{\sqrt{C}}$) :

$$\sum_{i=1}^{\infty} P(|\hat{r}_3(x, y) - E\hat{r}_3(x, y)| > \varepsilon_0 \sqrt{\frac{\log n}{ng\varphi_x(h)}}) < +\infty. \quad (2.27)$$

Ainsi, la preuve de (2.16) est maintenant achevée.

Lemme 2.2.3. *Sous les hypothèses du théorème, lorsque n tend vers l'infini, nous avons*

$$\hat{r}_1(x) - 1 = O_{p.co}\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right) \quad (2.28)$$

Notons que les dénominateurs introduits dans la décomposition (2.9) sont directement traités en utilisant les lemmes ci-dessus et la proposition A.6-i (voir Ferraty et al. [2]).

Preuve. Ce résultat peut directement se déduire du lemme 2.2.2 en prenant $\Gamma_i(y) = 1$.

2.3 Application et commentaire

2.3.1 Application à la prédiction

Nous commençons par introduire l'ensemble des contraintes qui suivent :

$$\begin{aligned} \exists \xi > 0, \exists ! y_0 \in \mathcal{S}, f(\cdot|x) \text{ est strictement croissante sur } (y_0 - \xi, y_0) \text{ et strictement} \\ \text{décroissante sur } (y_0, y_0 + \xi). \end{aligned} \quad (2.29)$$

Il est clair que si cette dernière hypothèse est satisfaite, le problème de maximisation de $f(y|x)$ sur \mathcal{S} possède y_0 comme une unique solution. Donc sous cette condition, le mode conditionnel $\theta(x)$ peut être défini comme suit :

$$\theta(x) = \arg \sup_{y \in \mathcal{S}} f(y|x). \quad (2.30)$$

Théorème 2.3.1. *Sous les conditions (2.2), (2.3) et en considérant l'hypothèse*

$$\lim_{n \rightarrow +\infty} \frac{\log n}{n\varphi_x(h)} = 0$$

alors si la densité conditionnelle $f(y|x)$ vérifie

$$\lim_{d(x,x') \rightarrow 0} f(y|x) = f(y|x')$$

et

$$\forall y' \in \mathbb{R}, \lim_{|y'-y| \rightarrow 0} f(y'|x) = f(y|x),$$

nous avons

$$\lim_{n \rightarrow +\infty} \hat{\theta}(x) = \theta(x), p.co \quad (2.31)$$

Preuve. La condition (2.29) assure que la densité conditionnelle $f(\cdot|x)$ est continue et strictement croissante sur $(\theta(x) - \xi, \theta(x))$. Ainsi, la fonction $f^{-1}(\cdot|x)$ existe et continue. La propriété de continuité de $f^{-1}(\cdot|x)$ au point $f(\theta(x)|x)$ peut s'écrire comme suit :

$$\forall \epsilon > 0, \exists \delta_1(\epsilon) > 0, \forall y \in (\theta(x) - \xi, \theta(x)) \mid f(y|x) - f(\theta(x)|x) \leq \delta_1(\epsilon) \Rightarrow |y - \theta(x)| \leq \epsilon.$$

Puisque $f(\cdot|x)$ est continue et strictement décroissante sur $(\theta(x), \theta(x) + \xi)$, le même type d'argument peut être utilisé pour arriver à :

$$\forall \epsilon > 0, \exists \delta_2(\epsilon) > 0, \forall y \in (\theta(x), \theta(x) + \xi) \mid f(y|x) - f(\theta(x)|x) \leq \delta_2(\epsilon) \Rightarrow |y - \theta(x)| \leq \epsilon.$$

En combinant les deux résultats, nous obtenons :

$$\forall \epsilon > 0, \exists \delta(\epsilon) > 0, \forall y \in (\theta(x) - \xi, \theta(x) + \xi), \mid f(y|x) - f(\theta(x)|x) \leq \delta(\epsilon) \Rightarrow |y - \theta(x)| \leq \epsilon.$$

Puisque par construction $\hat{\theta}(x) \in (\theta(x) - \xi, \theta(x) + \xi)$, nous avons :

$$\forall \epsilon > 0, \exists \delta(\epsilon) > 0, \mid f(\hat{\theta}(x)|x) - f(\theta(x)|x) \leq \delta(\epsilon) \Rightarrow |y - \theta(x)| \leq \epsilon,$$

il en résulte alors que :

$$\forall \epsilon > 0, \exists \delta(\epsilon) > 0, P[(\hat{\theta}(x) - \theta(x) > \epsilon)] \leq P[\mid f(\hat{\theta}(x)|x) - f(\theta(x)|x) \mid > \delta(\epsilon)] \quad (2.32)$$

D'autre part, il s'ensuit directement des définitions de $\theta(x)$ et de $\hat{\theta}(x)$ que :

$$\begin{aligned} |f(\theta(x)|x) - f(\hat{\theta}(x)|x)| &= |(f(\theta(x)|x) - \hat{f}(\theta(x)|x)) + (\hat{f}(\theta(x)|x) - f(\hat{\theta}(x)|x))| \\ &\leq |(f(\theta(x)|x) - \hat{f}(\theta(x)|x))| + |(\hat{f}(\theta(x)|x) - f(\hat{\theta}(x)|x))| \\ &\leq 2 \sup_{y \in (\theta(x) - \xi, \theta(x) + \xi)} |(f(y|x) - \hat{f}(y|x))|. \end{aligned} \quad (2.33)$$

L'uniforme convergence sur l'ensemble compact $[\theta(x) - \xi, \theta(x) + \xi]$ sur l'estimateur de la densité conditionnelle (voir lemme 2.3.1 suivant) peut être utilisée, donnant directement à partir des deux précédentes inégalités :

$$\forall \epsilon > 0, \sum_{n=1}^{\infty} P(|\hat{\theta}(x) - \theta(x)| > \epsilon) < \infty \quad (2.34)$$

et le résultat recherché (2.34) est maintenant vérifié.

Lemme 2.3.1. *Sous les conditions du théorème 2.3.1, nous avons pour tout sous ensemble compact $\mathcal{S} \subset \mathbb{R}$:*

$$\lim_{n \rightarrow \infty} \sup_{y \in \mathcal{S}} |(f(y|x) - \hat{f}(y|x))| = 0, p.co. \quad (2.35)$$

Preuve. *La preuve est basée sur la décomposition (2.9). Notons que les dénominateurs sont directement traités en utilisant le Lemme 2.2.3 et la Proposition A.6 – i[2]. Notons aussi que le dernier terme de (2.9) est traité en utilisant le Lemme 2.2.3 et le fait que $f(\cdot|x)$ est uniformément bornée sur $y \in \mathcal{S}$ (car elle est continue sur l'ensemble compact \mathcal{S}). Par conséquent, le résultat (2.35) sera une conséquence directe des deux résultats qui suivent :*

$$\lim_{n \rightarrow \infty} \frac{1}{\hat{r}_1(x)} \sup_{y \in \mathcal{S}} |E\hat{r}_3(x, y) - f(y|x)| = 0, p.co. \quad (2.36)$$

et

$$\lim_{n \rightarrow \infty} \frac{1}{\hat{r}_1(x)} \sup_{y \in \mathcal{S}} | \hat{r}_3(x, y) - E\hat{r}_3(x, y) | = 0, p.co. \quad (2.37)$$

Le résultat (2.36) peut être démontré dans ce qui suit. Nous pouvons écrire :

$$| f(y-vg|X) - f(y|x) | \leq | f(y-vg|X) - f(y-vg|x) | + | f(y-vg|x) - f(y|x) |. \quad (2.38)$$

Du fait que \mathcal{S} est compact, la fonction $f(y|x)$ est uniformément continue sur $y \in \mathcal{S}$. Ainsi, en ajoutant le fait que K_0 est de support compact (car il est de type 0), et d'après l'hypothèse sur la densité conditionnelle donnée dans le théorème 2.3.1, nous avons

$$\lim_{n \rightarrow +\infty} \sup_{v \in [-1,1]} \sup_{y \in \mathcal{S}} 1_{B(x,h)}(X) |f(y - vg|X) - f(y - vg|x)| = 0, \quad (2.39)$$

et

$$\lim_{n \rightarrow +\infty} \sup_{v \in [-1,1]} \sup_{y \in \mathcal{S}} |f(y - vg|X) - f(y|x)| = 0, \quad (2.40)$$

ce qui implique que

$$\lim_{n \rightarrow +\infty} \sup_{v \in [-1,1]} \sup_{y \in \mathcal{S}} 1_{B(x,h)}(X) |f(y - vg|X) - f(y|x)| = 0, \quad (2.41)$$

En combinant (2.14) et (2.41) et en tenant compte de la positivité de Δ_1 et de K_0 , nous trouvons :

$$\sup_{y \in \mathcal{S}} |E\hat{r}_3(x, y) - f(y|x)| = o(1). \quad (2.42)$$

Ceci, combiné avec le Lemme 2.2.3 et la Proposition A.6 – i[2], est suffisant pour prouver (2.36) et de finir le traitement de ce terme biais. Il reste à montrer que (2.37) est vraie. En utilisant la compacité de \mathcal{S} , on peut écrire que $\mathcal{S} \subset \cup_{k=1}^n \mathcal{S}_k$ où $\mathcal{S}_k = (t_k - l_n, t_k + l_n)$ et

où l_n et z_n peuvent être choisis tels que :

$$l_n = Cz_n^{-1} \sim Cn^{-2\zeta}. \quad (2.43)$$

En prenant $t_y = \arg \min_{t \in \{t_1, \dots, t_{z_n}\}} |y - t|$, nous avons

$$\frac{1}{\hat{r}_1(x)} \sup_{y \in \mathbb{S}} |\hat{r}_3(x, y) - E\hat{r}_3(x, y)| = A_1 + A_2 + A_3, \quad (2.44)$$

où

$$A_1 = \frac{1}{\hat{r}_1(x)} \sup_{y \in \mathbb{S}} |\hat{r}_3(x, y) - \hat{r}_3(x, t_y)|,$$

$$A_2 = \frac{1}{\hat{r}_1(x)} \sup_{y \in \mathbb{S}} |\hat{r}_3(x, t_y) - E\hat{r}_3(x, t_y)|,$$

$$A_3 = \frac{1}{\hat{r}_1(x)} \sup_{y \in \mathbb{S}} |E\hat{r}_3(x, t_y) - E\hat{r}_3(x, y)|.$$

La condition d'Hölder (2.5) permet d'écrire :

$$\begin{aligned} |\hat{r}_3(x, y) - \hat{r}_3(x, t_y)| &= \frac{1}{n} \sum_{i=1}^n \Delta_i |\Gamma_i(y) - \Gamma_i(t_y)| \\ &\leq \frac{1}{ng} \sum_{i=1}^n \Delta_i \left| K_0\left(\frac{y - Y_i}{g}\right) - K_0\left(\frac{t_y - Y_i}{g}\right) \right| \\ &\leq \frac{C}{ng} \sum_{i=1}^n \Delta_i \left| \frac{y - t_y}{g} \right| \\ &\leq \frac{C\hat{r}_1(x)l_n}{g^2}. \end{aligned} \quad (2.45)$$

Utilisant (2.43) avec $A_1 \leq \frac{C}{(gn^\zeta)^2}$ et la condition sure K_0 donnée en (2.3), on peut écrire

$$\lim_{n \rightarrow +\infty} A_1 = 0. \quad (2.46)$$

En utilisant des arguments similaires, on peut écrire :

$$A_3 \leq \frac{C}{\hat{r}_1(x)(gn^\zeta)^2}, \quad (2.47)$$

et en tenant compte du Lemme 2.2.3 et de la Proposition A.6 – i[2], on trouve :

$$\lim_{n \rightarrow +\infty} A_3 = 0, p.co \quad (2.48)$$

En observant maintenant le terme A_2 , on peut écrire pour tout $\epsilon > 0$:

$$\begin{aligned} P[\sup_{y \in \mathcal{S}} |\hat{r}_3(x, t_y) - E\hat{r}_3(x, t_y)| > \epsilon] &= P(\max_{j=1, \dots, z_n} |\hat{r}_3(x, t_j) - E\hat{r}_3(x, t_j)| > \epsilon) \\ &\leq z_n \max_{j=1, \dots, z_n} P(|\hat{r}_3(x, t_j) - E\hat{r}_3(x, t_j)| > \epsilon) \\ &\leq z_n \max_{j=1, \dots, z_n} P(|\frac{1}{n} \sum_{i=1}^n (U_i - EU_i)| > \epsilon) \end{aligned} \quad (2.49)$$

où

$$U_i = \Delta_i \Gamma_i(t_j). \quad (2.50)$$

En utilisant soit le lemme 1.4.1 ou 1.4.2 et en tenant compte du fait que K est de type I ou II, et puisque $\Gamma_i(y) \leq \frac{C}{g}$, nous avons :

$$|U_i| \leq \frac{C}{g\varphi_x(h)}. \quad (2.51)$$

D'autre part, nous avons en procédant par intégration par changement de variable et en utilisant (2.41)

$$\begin{aligned}
 EU_i^2 &= E(\Delta_i^2 E(\Gamma_i(t_j)|X = x)) \\
 &= \frac{1}{g^2} E(\Delta_i^2 \int_{\mathbb{R}} K_0^2\left(\frac{t_j - u}{g}\right) f(u|x) du) \\
 &= \frac{1}{g} E(\Delta_i^2 \int_{\mathbb{R}} K_0^2(z) f(t_j + zg|x) dz) \\
 &\leq C \frac{1}{g} E(\Delta_i^2 \int_{\mathbb{R}} K_0^2(z) dz f(t_j|x) dz). \tag{2.52}
 \end{aligned}$$

Puisque $f(y|x)$ est bornée (car elle est continue sur l'ensemble compact \mathcal{S}) et en appliquant $E\Delta_i^2 \leq \frac{C}{g\varphi_x(h)}$, nous obtenons :

$$EU_i^2 \leq \frac{C}{g\varphi_x(h)}. \tag{2.53}$$

Puisque les variables U_i sont bornées, nous pouvons appliquer l'inégalité de type-Bernstein donnée par le Corollaire A.9 – i présenté dans [2]. En tenant compte de cette inégalité et d'après (2.49), (2.51) et (2.53) on peut écrire :

$$P(\sup_{y \in \mathcal{S}} |\hat{r}_3(x, t_y) - E\hat{r}_3(x, t_y)| > \epsilon) \leq z_n \exp\{-Cn\epsilon^2 g\varphi_x(h)\}. \tag{2.54}$$

En utilisant (2.43), on trouve :

$$P(\sup_{y \in \mathcal{S}} |\hat{r}_3(x, t_y) - E\hat{r}_3(x, t_y)| > \epsilon) \leq Cn^{2\zeta} \exp(-Cn\epsilon^2 g\varphi_x(h)). \tag{2.55}$$

Puisque $\frac{\log n}{ng\varphi_x(h)}$ tend vers zéro, nous obtenons :

$$\forall \epsilon > 0, \sum_{n=1}^{\infty} P(\sup_{y \in \mathcal{S}} |\hat{r}_3(x, t_y) - E\hat{r}_3(x, t_y)| > \epsilon) < \infty. \tag{2.56}$$

Le dénominateur de A_2 est traité directement en utilisant encore le Lemme 2.2.3 et la Proposition A.6 – i[2]. Ceci est suffisant pour arriver à

$$\lim_{n \rightarrow +\infty} A_2 = 0, p.co. \tag{2.57}$$

Finalement, le résultat (2.37) se déduit de (2.44), (2.46), (2.48) et (2.57). La combinaison de (2.9), (2.36) et (2.37) permet de terminer la preuve du ce Lemme.

2.3.2 Commentaire

Si on prend

$$\varphi_x(h) = Ch + o(h), \quad (2.58)$$

la condition sur $\varphi_x(x)$ donnée en (2.4) n'est pas nécessaire car :

$$\int_0^\varepsilon (Cu + o(u))du > C\varepsilon\varphi_x(\varepsilon) \quad (2.59)$$

et dans ce cas nous avons

$$\hat{f}(y|x) - f(y|x) = O(h^{\beta_1}) + O(g^{\beta_2}) + O_{p.co}\left(\sqrt{\frac{\log n}{ngh}}\right). \quad (2.60)$$

2.3.3 Sur la mise en œuvre de l'estimateur de la densité conditionnelle

La densité conditionnelle est une alternative à l'espérance conditionnelle pour décrire la relation entre deux variables aléatoires. Du point de vue pratique, la plupart des travaux en statistique non paramétrique fonctionnelle traitent la densité conditionnelle comme une étape préliminaire à l'estimation du mode conditionnel (outil de prévision alternative à la régression : voir, par exemple la monographie de Ferraty et Vieu (2006)). Dans cette perspective prédictive, les paramètres h et g sont obtenus en minimisant le critère suivant :

$$\sum_{i=1}^n (Y_i - \hat{\theta}^{-i}(X_i))^2$$

où

$$\hat{\theta}^{-i}(X_i) = \arg \sup_{y \in \mathcal{S}} \hat{f}^{-i}(y|X_i)$$

représente l'estimateur du mode conditionnel basé sur l'ensemble des données à l'exception de (X_i, Y_i) . On peut aussi estimer la densité conditionnelle pour diverses finalités autre que le mode conditionnel (estimation de la fonction de hasard conditionnelle, test de multimodalités,.....).

Chapitre 3

Simulation

Nous présentons dans ce chapitre le travail de simulation effectué pour étayer les différents aspects théoriques abordés dans notre étude. L'expérimentation numérique nous servira en particulier à :

- Étudier la performance de cette méthode ;
- Étudier l'influence de la taille de l'échantillon sur les résultats.

3.1 Plan de simulation

Nous nous contenterons de faire des simulations et d'observer le comportement asymptotique de l'estimateur à noyau calculé à partir d'échantillons simulés. Ceci nous permettra de savoir si empiriquement l'estimateur \hat{f} converge vers f .

Nous utiliserons pour les simulations, des échantillons de lois connues de taille de plus en plus grande $n = (100, 500, 1000 \text{ et } 5000)$.

Nous prenons l'exemple où la densité conditionnelle Y sachant $X = x$ suit une loi normale de norme x et variance 1, telle que la distribution de Y sachant $X = x$ est unimodal

$$f(y|x) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{1}{2}(y-\|x\|)^2}.$$

et supposons que les variables explicatives X_i , sont issues d'un processus stochastique $X = X_i(t) = a_i \cos(2\pi t) + b_i \sin(3\pi t) + c_i(t - 0.45)(t - 0.75) \exp^{(d_i t)}$ pour $t \in [0, 1]$ avec $b_i \rightsquigarrow N(-1; 1)$, $c_i \rightsquigarrow U(1, 5)$ et $d_i \rightsquigarrow U(1, 5)$ [3].

Pour $x = 0.01254$ Le graphique de $f(y|x)$ est représenté dans la figure suivante :

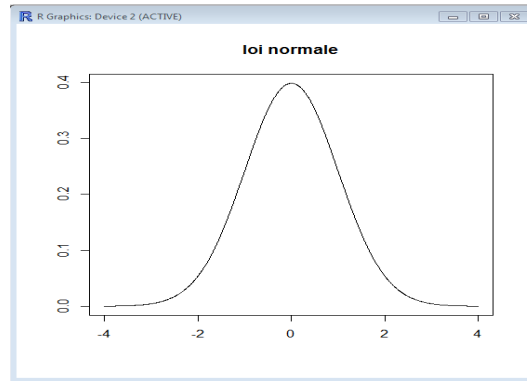


FIG. 3.1 – Représentation graphique $f(y|x)$

3.2 Algorithme de simulation

L'algorithme de simulation que nous avons utilisé comporte quatre phases :

- Simuler un échantillon de taille n ;
- Calculer les paramètres de lissage h et g qu'on fait varier sur un intervalle $[0,1]$ et qui minimise $ASE = \frac{1}{n} \sum_{j=1}^n (f(\hat{y}_j|x) - f(y_j|x))^2$ (Averaged Squared Error) ;
- Construire l'estimateur par la méthode du noyau à partir des observations. Comme le choix du noyau n'a pas d'impact très significatif sur la qualité d'estimation, dans le sens où la fenêtre est bien choisie[2], on se limite qu' à deux sorte de noyau, on utilisera donc le noyau parabolique K sous la condition $0 \leq K(u) \leq 1$ et le noyau $H(u) = \int_{-\infty}^u H_0(v)dv$ tel que H_0 est un un noyau triangulaire de type 0 et $H_0(u) > 0$
- Tracer les deux courbes : la densité théorique f et la densité estimée \hat{f} .

Les simulations et les graphes ont été réalisés à l'aide du logiciel R. Nous avons utilisé la version 3.0.0 pour la programmation.

R est un système d'analyse statistique et graphique créée par Ross Ihaka et Robert Gentleman. Il est à la fois un langage et un logiciel qui comporte de nombreuses fonctions pour les analyses statistiques et graphiques [5].

3.3 Résultats de simulation

On donne les résultats de simulation sous forme de tableaux et de graphes. Les graphes ci-dessous représentent les courbes X_i pour $n=100, 500, 1000$ et 5000 .

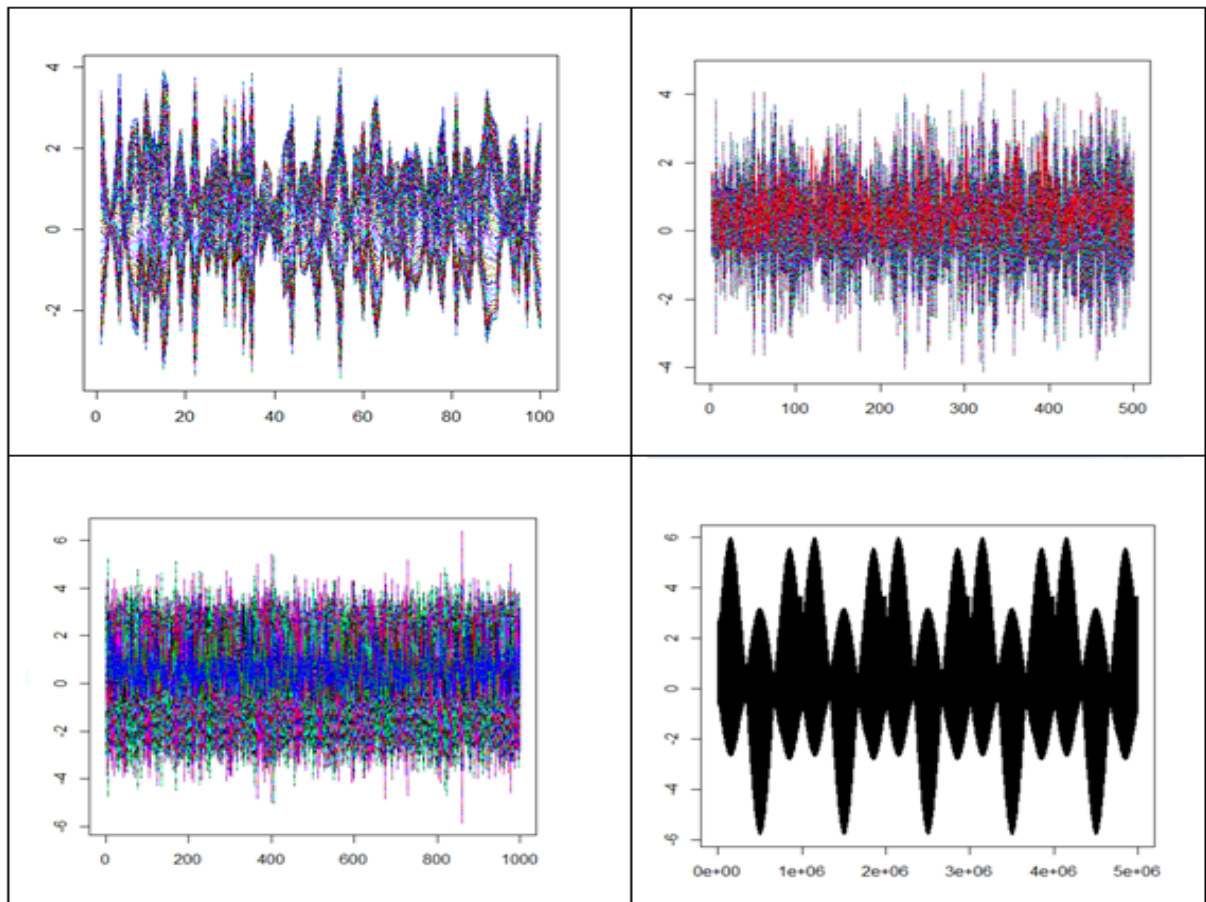


FIG. 3.2 – Représentation des courbes X_i

Les résultats de la simulation sont donnés dans les tableaux ci-dessous

n=100		n=500	
h	0.86	h	0.69
g	0.79	g	0.61
ASE	0.045	ASE	0.026

n=1000		n=5000	
h	0.54	h	0.27
g	0.49	g	0.31
ASE	0.012	ASE	0.0073

FIG. 3.3 – Résultats de la simulation

Dans les graphes ci-dessous on représente la densité conditionnelle estimée et on la compare avec la densité théorique pour tout $n = 100, 500, 1000$ et 5000 .

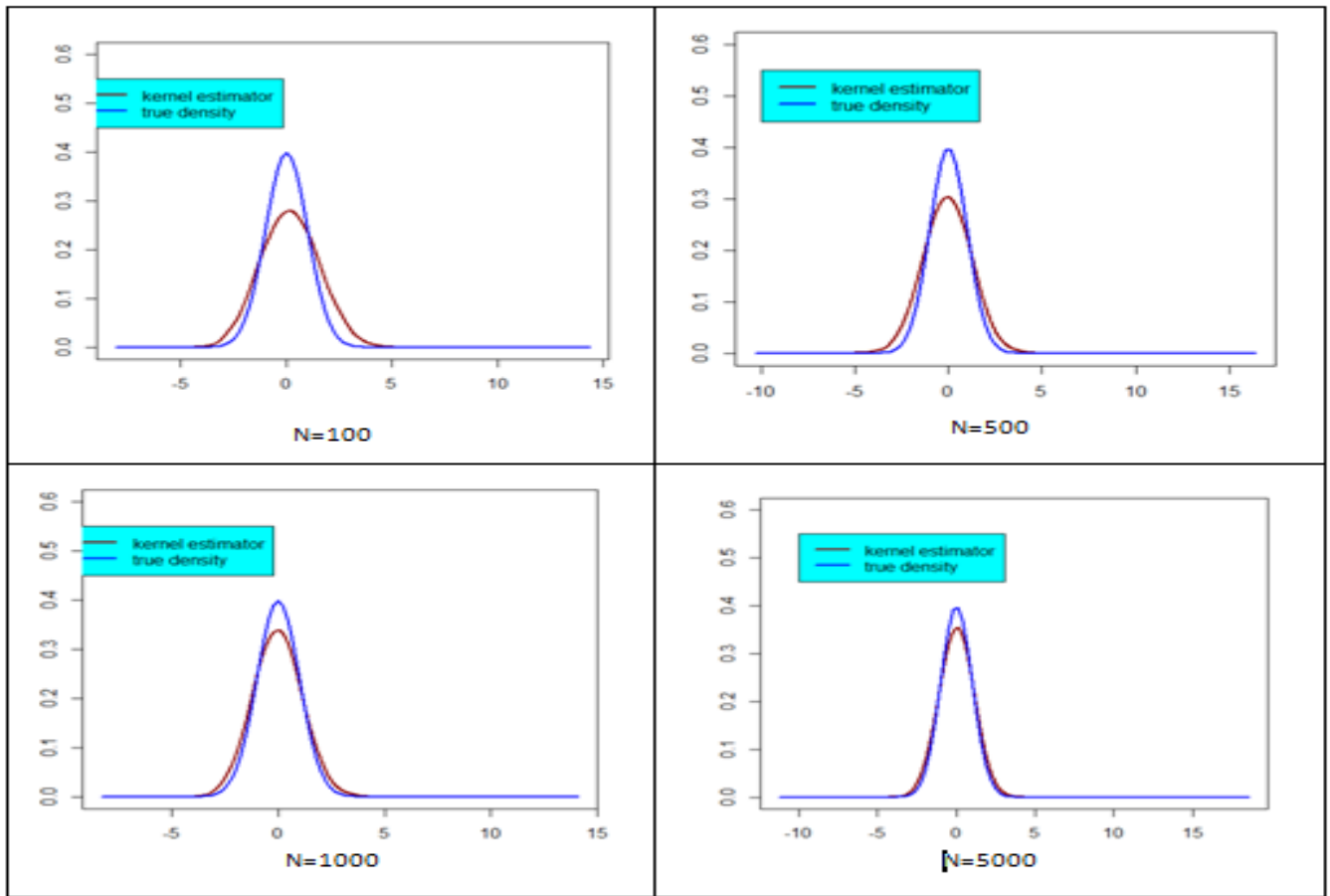


FIG. 3.4 – Représentation de la densité estimée avec la méthode du noyau

3.4 Interprétation des résultats

D'après les graphes et les résultats numériques obtenus de la simulation, on remarque que :

- Plus la taille de l'échantillon est grande plus h et g sont petits, ce qui confirme les hypothèses citées dans le chapitre 2 ;
- plus la taille de l'échantillon est grande, plus ASE est petit et meilleure est la qualité de l'estimation de f .

On conclut donc que la méthode du noyau est un outil très efficace pour estimer la densité conditionnelle dans le cas où les variables explicatives sont fonctionnelles.

Conclusion Générale

Dans ce travail, nous nous sommes intéressées aux aspects théorique et pratique de l'estimation d'une fonction de densité conditionnelle par la méthode du noyau .

Dans un premier temps, nous avons présenté un état de l'art sur les modèles non paramétriques conditionnels et la méthode du noyau pour les variables aléatoires fonctionnelles. Nous abordons le cas fonctionnel où nous avons montré comment introduire le noyau de pondération locale quand on à faire à des données fonctionnelles. Nous avons rappelé les définitions des outils naturels pour faire la prévision tels que la régression classique, la régression modale (mode conditionnel) , les quantiles de régression (la médiane conditionnelle), en explicitant la fonction de risque associé à chaque modèle et en mettant l'accent sur la diversité et l'étendue de la littérature disponible sur ces modèles . Ensuite, nous avons fait une représentation générale de la méthode du noyau en citant quelques noyaux usuels. Ainsi, deux résultats de base associés aux noyaux de type I et de type II ont été énoncés.

Dans un deuxième temps, nous avons procédé à l'étude asymptotique de la fonction de densité conditionnelle. Un estimateur à noyau a été établi dans un cadre indépendant. Les résultats obtenus ont été validés par une simulation. Le choix des paramètres de lissage étant crucial dans le comportement asymptotique de l'estimateur (c'est le cas de la plupart des problèmes d'estimation fonctionnelle), nous avons alors sélectionné les valeurs des paramètres de lissage h et g qui fournissent le plus petit ASE (Averaged squared error) nous permettant d'obtenir le meilleur estimateur de la fonction de densité conditionnelle

$$f(y|x) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{1}{2}(y-\|x\|)^2} .$$

Par ailleurs nous avons montré que l'estimateur de mode conditionnelle peut se déduire de l'estimateur de fonction densité conditionnel.

Bibliographie

- [1] DELSOL Laurent,2008,*Régression sur variable fonctionnelle : Estimation, Tests de structure et Applications*,En vue de l'obtention du DOCTORAT,Université ToulouseIII Paul Sabatier,247p.
- [2] FERRATY F,VIEU P,2005,*Nonparametric Modelling for Functional Data*, Springer.
- [3] LAKSACI A, 2007,Science direct,*Erreur quadratique de l'estimateur à noyau de la densité conditionnelle à variable explicative fonctionnelle*,Université Sidi Bel Abbès Djilali-Liabes,171_75.p.
- [4] TADJ A,2011, *Sur les modèles non paramétriques conditionnels en statistique fonctionnelle*,Thèse, Université de TOULOUSE,146p.
- [5] VINCENT G,2012,*Introduction à la programmation en R*,École d'actuariat, Université Laval,151.
- [6] YOUNDJÉ É,1993, *Estimation non paramétrique de la densité conditionnelle par la méthode du noyau*,Travail de diplôme réalisé en vue d'obtention du grade de docteur,Université de ROUEN.
- [7] YOUNDJÉ É,2011, *Contribution à l'estimation non-paramétrique par la méthode du noyau*,Mémoire d'Habilitation à Diriger des Recherches, Universités de ROUEN et du HAVRE,49p.
- [8] ZOUGAB N,2007, *Etude comparative des méthodes de sélection du paramètre de lissage dans l'estimation de la densité de probabilité par la méthode du noyau*, thèse,Univertité abderrahman Mira,110p.

Résumé

L'objet de ce travail est l'étude de l'estimateur à noyau de la densité conditionnelle d'une variable réelle Y conditionnée par une variable X fonctionnelle. La vitesse de convergence presque complète de cet estimateur a été établie sous certaines conditions, et à titre illustratif, nous avons donné un exemple d'application sur des données simulées.

Mots clés : fonction de densité conditionnelle, estimateur à noyau, estimation non paramétrique, mode conditionnel, convergence presque complète, variable fonctionnelle.