UNIVERSITE KASDI MERBAH
OUARGLA

Faculté des mathématiques et sciences de la matière

N° d'ordre :
N° de série :

DEPARTEMENT DE MATHEMATIQUES

MASTER

Spécialité : Mathématiques

Option : probabilite et statistique

Par : Gherair Djemoi

Théme

Modélisation non paramétrique pour les variables aléatoires fonctionnelles cas de données indépendantes

Soutenu publiquement le :01/06/2017

Devant le jury composé de :

Pr.Zibar Said M.A.A université de KASDI Merbah - Ouargla	Président
Ms.Meddi Fatima M.C.B université de KASDI Merbah - Ouargla	Examinateur
Mr.Agoune Rachid M.A.B université de KASDI Merbah - Ouargla	Rapporteur

Année universitaire 2016/2017

Dédication

Je dédie ce travail à :

Mes parents

-A mes frères

et mes surs,et toute la famille

- A mes chers amies

- Je tiens à remercier tous les membres de ma promotion.

-Et a tous mes professeurs

Remerciement

Avant toute considération,je remercie le Grand Dieu le tout puissant qui, m'a aidé pour achever ce travail.

Je tiens tout a remercier premier lieu mon encadreur Monsieur **Agoune Rachid** d'avoir accepté de m'encadrer et pour sa continuité à me soutenir et à m'encourager. Je voudrai aussi le remercier pour sa gentillesse, sa disponibilité et du temps consacré à mon travail.

Nous voulons également remercier : Mr prof : **Said Zibar** pour nous avoir fait l'honneur de présider le jury de notre mémoire Nos remerciements vont également aux : Ms prof : **Meddi Fatima** honorer de leur présence dans ce jury.

Je remercie également les membres du département de Mathématique et Informatique de m'avoir permis de travailler dans de bonnes conditions pendant la réalisation de mon travail.

Merci également a tous les enseignants qui m'ont aidé pendant mon cursus, sans oublier leurs conseils précieux.

Je tiens aussi à remercie mes amis (Nour adinne,Said ,Thabet,Sayah,Abd alkader ,Ahmed,Ali,Brahime,.....)

Je remercie aussi toute personne de prés ou de loin a contribué à la finalisation de ce travail.

Notations et Préliminaires

- p, co :convergence presque complète
- $B(\chi, h)$: boule ouverte du centre χ et rayon h, dans l'espace (E, d).
- $C \;\; ou \;\; C'$:Constantes positives réelles
- $d(,)$:Semi-métrique sur un espace fonctionnel E.
- $E(Y) \;\; ou \;\; EY$: Espérance,pour Y .v.a.r.
- $E(Y/\mathcal{X})$: Espérance conditionnelle de Y sachant $\mathcal{X}$,ou régression
- $\widehat{F}_Y^{\mathcal{X}}(\chi, y) \;\; ou \;\; \widehat{F}_Y^{\chi}(y)$: Distribution conditionnelle de v.a,r. Y sachant v.a.f. $\mathcal{X}$.
- $\phi_\chi(h)$: Mesure de la boule $B(\chi, h)$ par rapport à la loi de probabilité de $\mathcal{X}$
- $1_I(.)$:Fonction d'indicatrice sur un ensemble I
- h :paramétré de lissage ou largeur de fenêtre $h = h(n)$
- H : fonction de noyau
- K : noyau asymétrique
- K_0 : noyau symétrique standard
- $O_{p.c}$:vitesse de convergence presque complète
- S :Sous-ensemble compact de $\mathbb{R}$
- X :Variable aléatoire réelle

- x :Observations de v.a.r
- $\mathcal{X}$:Variable aléatoire fonctionnelle
- χ : :Observations de v.a.f
- r :Opérateur de régression non linéaire
- $\mathcal{X}_i$,$i = 1..n$:Échantillon de v.a.f
- χ_i ,$i = 1..n$:Observations statistiques du v.a.f $\mathcal{X}_i$

Table des matières

Table des figures

Introduction

La statistique fonctionnelle occupe désormais une place importante dans la recherche en statistique. Il s'agit de la modélisation statistique des données qui sont des courbes supposées observées sur toutes leurs trajectoires. Ceci est pratiquement possible en raison de la précision des appareils de mesures modernes et de l'important capacité de stockage qu'offrent les systèmes informatiques actuels. Il est facile d'obtenir une discrétisation très fine d'objets mathématiques tels que des courbes, surfaces,.....

Ce type de variables se retrouve dans de nombreux domaines, comme la météorologie, la chimie quantitative,la biométrie, l'économétrie ou l'imagerie médicale....

La problématique abordée dans cet mémoire L'estimation de la fonction de répartition conditionnelle dans un cadre fonctionnel qui 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. 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 cet estimateur dans les deux cas (i.i.d. et α-mélangeant). La contribution de ce mémoire l'étude de la convergence uniforme presque complété fonctionnel de l'estimateur de la fonction de répartition conditionnelle. La vitesse de convergence de cet estimateur est précisée. Les résultats obtenus sont détaillés dans le chapitre 2 de cet mémoire. Ce sont les premiers résultats uniformes disponibles dans la littérature portant sur l'estimation de la fonction de répartition conditionnelle à une variable fonctionnelle.

En raison de la nouveauté de ce domaine, il faut commencer par clarifier le vocabulaire reliant les statistiques fonctionnelles et non paramétriques (quelles sont les données / va-

riables fonctionnelles ? qu'est-ce qu'un modèle non paramétrique pour un tel ensemble de données ? ...). . et les idées statistiques de base sur les méthodes locales de pondération et leur extension au cas fonctionnel sont exposées. Une attention particulière est accordée à la pondération du noyau.Cela se fait au 1^{ere}chapitre .
le 2^{eme} chapitre on se concentre sur l'étude de la convergence presque complète de la fonction répartition conditionnelle et le 3^{eme} chapitre contient les différentes simulations qui confortent les résultats théoriques obtenus. .

Chapitre 1

principes de bases et Analyse de données fonctionnelles non paramétriques

Le but principal de ce chapitre est de familiariser le lecteur avec des notions tant statistiques fonctionnelles que non paramétriques. d'abord et à cause de la nouveauté de ce domaine de la recherche, nous proposons quelques définitions de base pour clarifier le vocabulaire tant sur des variables fonctionnelles que sur le modèle non paramétrique.
en raison de la nouveauté de ce domaine, il faut commencer par clarifier le Vocabulaire reliant les statistiques fonctionnelles et non paramétriques (ce qui sont données fonctionnelle / variables fonctionnelles ? Qu'est-ce qu'un modale non paramétrique pour un tel ensemble de données ? ...).

1.1 variable aléatoire fonctionnelle

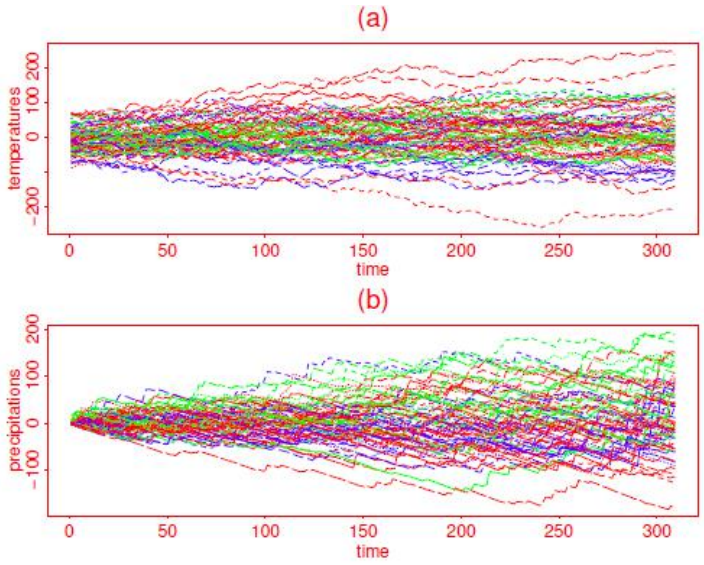
Il y a en réalité un nombre croissant de situations venant des champs différents de sciences appliquées (environmétries, chemometries, la biométrie, la médecine, l'économétrie, .. .) dans lequel les données rassemblées sont des courbes. En effet, le progrès des outils de calcul, tant en termes de mémoire que des capacités informatiques, nous permet de traiter les grands ensembles de données. Particulièrement pour un seul phénomène, nous pouvons observer un très grand ensemble de variables. Par exemple, regardez la situation habituelle suivant où une certaine variable aléatoire peut être observée à plusieurs temps différents dans la gamme (t_{min}, t_{max}).Une observation peut être exprimée par la famille aléatoire $\{X(t_j)\}_{j=1,\ldots,J}$. dans la statistique moderne, quand la grille devient plus en plus

petite , les instants consécutifs sont de plus en plus proche. une façon de le prendre en compte est de considérer les données comme une observation de la famille continue $\mathcal{X} = \{X(t); t \in (t_{min}, t_{max})\}$.

Définition 1 *Une variable aléatoire $\mathcal{X}$ est appelée la variable fonctionnelle (f.v). S'il prend des valeurs dans un espace de dimension infini (ou un espace fonctionnel). une Observation χ de $\mathcal{X}$ est appelé des donnée fonctionnelle.*

Notez que, quand $\mathcal{X}$ (resp χ .) dénote une courbe aléatoire (resp. son observation), nous implicitement faisons l'identification suivant $\mathcal{X} = \{\mathcal{X}(t); t \in T\}$ (resp. $\chi = \{\chi(t); t \in T\}$).Dans cette situation, la caractéristique fonctionnelle vient directement des observations. La situation est une courbe quand la variable est associée à un ensemble unidimensionnel $T \subset \mathbb{R}$ Il est important de faire remarquer que la notion de variable fonctionnelle couvre un plus grand domaine de l'analyse de courbes. Particulièrement une variable fonctionnelle peut être une surface aléatoire, comme par exemple les niveaux gris d'une image ou un vecteur de courbes (et dans ces cas T sont un ensemble bidimensionnel $T \subset \mathbb{R}^2$),ou un autre objet mathématique dimensionnel infini plus compliqué Même si les données réelles utilisées comme des supports partout dans ce travail sont tous les ensembles de données de courbes (c'est-à-dire, un ensemble de données de courbes), toute la méthodologie et des avances théoriques à être présentées plus tard sont potentiellement applicables à une autre sorte de données fonctionnelles.

exemple : tiré du domaine agronomique. il s'agit de mieux comprendre les interactions entre le rendement de blé et les variations climatiques survenues durant la culture. Dans ce but, on dispose de de $n = 198$ parcelles de blé ; pour chaque parcelle i, en plus du rendement Y_i, on a les températures T_i^d et précipitations P_i^d cumulées journalières durant $J = 309$ jours (d'octobre à début Aout) $T_i^d = \{T_i(t_j)\}_{j=1,\ldots,J}$ et $P_i^d = \{P_i(t_j)\}_{j=1,\ldots,J}$ Le graphique suivant représente la version continue $\{T_i = \{T_i(t) : t \in [0; 309]\}\}_{i=1,\ldots,n}$ (resp.$\{P_i = \{P_i(t); t \in [0, 309]\}\}_{i=1,\ldots,n}$) de $\left\{T_i^d\right\}_{i=1,\ldots,n}$ (resp.$\left\{P_i^d\right\}_{i=1,\ldots,n}$



""

FIGURE 1.1 – Courbes de pluviométrie et de température ; relevés journaliers cumulés centrés

une fois de plus, on s'intéresse pour i = 1, . . . , n à la modélisation du lien entre la variable réelle Yi et le couple de variables fonctionnelles (P_i, T_i) (pluviométrie et température).

1.2 données fonctionnels

Depuis l'année quatre-vingt-dix, le nombre croissant de situations ,où des variables fonctionnelles peuvent être observées, a entrainé des développements statistiques différents que nous pourrions à la suite nommer Statistique pour des Variables Fonctionnelles (ou des Données). Nous faisons partie de cette domaine statistique puisque nous proposerons plusieurs méthodes impliquant échantillon fonctionnel statistique $\mathcal{X}_1, ..., \mathcal{X}_n$. Alors, on commence par une définition précise d'un ensemble de données fonctionnel.

Définition 2 . *Un ensemble de données fonctionnel* $\chi_1, ..., \chi_n$ *est l'observation de n variables fonctionnelles* $\mathcal{X}_1, ..., \mathcal{X}_n$ *sont identiquement distribués comme* $\mathcal{X}$.

Cette définition couvre beaucoup des aspects, le plus connu sont les ensembles de données de courbes. Nous n'allons pas examiner comment ces données fonctionnelles ont été collectées , question qui est lié aux problèmes de discrétisation mathématiques

Selon le type des données, une étape préliminaire consiste à les présenter tous les deux d'une façon bien adaptée au traitement fonctionnel Comme nous verrons, si la grille des mesures est assez petite, l'étape premier le plus important implique des techniques d'approximation numériques connues.
Dans d'autres cas standards, des méthodes classiques régularisation peuvent être adaptés.Il arrive qu'en quelques situations on applique de techniques réglables plus sophistiquées, par exemple quand les mesures répétées par sujets sont très peu (des données clairsemées) et/ou avec la grille irrégulière. Ceci est évidemment un domaine de recherche parallèle et complémentaire mais loin de notre but principal qui est le traitement statistique non-paramétriques de données fonctionnelles.Dorénavant, nous allons supposer que nous avons à portée de main un échantillon de données fonctionnelles.

1.3 statistique non-paramétrique pour données fonctionnelles

D'autre part, la statistique non-paramétrique a été beaucoup développée. En effet, depuis le début des années soixante, beaucoup d'attention a été payé au modelage libre (dans la distribution libre que dans le paramètre libre) modèles statistiques et/ou méthodes. La caractéristique fonctionnelle de ces méthodes vient de la nature de l'objet qu'on va évaluer (comme par exemple la fonction de densité, la fonction de régression...) qui n'est pas supposé être paramétrais par un nombre fini de quantités réelles. Dans ce cas là, on parle bien sur de Statistique Non-paramétrique pour laquelle on vas consacrer une bonne partie. il y a tant de façons (différentes) pour définir ce qui est un modèle statistique non-paramétrique dans le contexte dimensionnel fini et la limite entre les modèles non-paramétriques et paramétriques peut parfois sembler peu claire. Nous avons décidé de commencer de la définition suivante de modèle non-paramétrique dans le contexte dimensionnel fini

Définition 3 *soit X un vecteur aléatoire estimé dans $\mathbb{R}^p$ et soit ϕ une fonction définie sur $\mathbb{R}^p$ selon la distribution de X. Un modèle pour l'évaluation de ϕ consiste de présenter quelques contraintes de la forme :*

$$\phi \in C$$

Le modèle est appelé un modèle paramétrique pour l'évaluation de ϕ si C est indexé par un nombre fini des éléments de $\mathbb{R}$.sinon, le modèle s'appelle un modèle non paramétrique.

Notre décision de choisir cette définition a été motivée par le fait qu'elle définit clairement la frontière entre les modèles paramétriques et non paramétriques, et aussi parce que cette définition peut être facilement étendue au cadre fonctionnel.

Définition 4 *soit $\mathcal{Z}$ une variable aléatoire estimée dans un espace dimensionnel infini F et que soit ϕ opérateur définie sur F selon la distribution de $\mathcal{Z}$. Un modèle pour évaluer ϕ consiste à présenter quelques contrainte de la forme :*

$$\phi \in C$$

Le modèle est appelé modèle paramétrique fonctionnel pour l'évaluation de ϕ si C est indexé par un nombre fini des éléments de F. sinon, le modèle est appelé un modèle non-paramétrique fonctionnel.

1.4 Convergence presque complète et inégalités de Bernstein

1.4.1 Convergence presque complète

On dit que la suite de variables aléatoires réelles $(X_n)_{n\in\mathbb{N}}$ converge presque complètement vers une variable aléatoire X lorsque $n \longrightarrow \infty$ (et on note
$lim_{n\longrightarrow\infty}X_n = X$, si et seulement si :

$$\forall \varepsilon \succ 0, \sum_{n\in\mathbb{N}} P\left[|X_n - X| \succ \varepsilon\right] \prec \infty$$

Définition 5 *On dit que la vitesse de convergence presque complète de la suite de variables aléatoires réelles $(X_n)_{n\in\mathbb{N}}$ vers X est d'ordre(U_n) ((U_n) étant une suite numérique déterministe), et on note $X_n = O_{p.co}\ (U_n)$, si et seulement si :*

$$\forall \varepsilon_0 \succ 0, \sum_{n\in\mathbb{N}} P\left[|X_n - X| \succ \varepsilon_0 U_n\right] \prec \infty$$

Notons que la convergence presque complète entraîne à la fois la convergence presque sûre et la convergence en probabilité.

Proposition 1 *si* $\lim_{n\longrightarrow\infty} X_n = X$ *p.co,alors*X_n *converge en probabilité et presque sûrement vers X.*

preuve :

La convergence en probabilité se déduit facilement de la convergence de la série suivante

$$\sum_{n\in\mathbb{N}} P\left[|X_n - X| \succ \varepsilon\right] \prec \infty$$

$(P\left[|X_n - X| \succ \varepsilon\right]$ *,est le terme général d'une série convergente). Le lemme de Borel contelli implique que :*

$$\forall \varepsilon \succ 0, P[\lim_{n\longrightarrow\infty} sup|X_n - X| \succ \varepsilon] = 0$$

De plus, $\lim_{n\longrightarrow\infty} X_n(\omega) \neq X(\omega)$*, implique l'existence de* $\varepsilon \succ 0$*, tel que*

$$, \lim_{n\longrightarrow\infty} sup|X_n(\omega) - X(\omega)| \succ \varepsilon$$

on alors$P\left[\lim_{n\longrightarrow\infty} X_n = X\right] = 1$ *, c'est à dire* $X_n \longrightarrow X$*, P.S .*

en outre le fait que la convergence presque complète est une convergence très forte, elle jouît des propriétés résumées ci dessous.

1.4.2 inégalités de Bernstein

Soit $\{X_n, n \geq 1\}$ une suite de variables aléatoires centrées. Pour démontrer la convergence presque complète, nous avons besoin de trouver des bornes supérieures pour certaines probabilités concernant des sommes de variables aléatoires telles que :

$$P\left(\left|\sum_{i=1}^{n} Z_i\right| > \varepsilon\right)$$

où éventuellement ε décroît avec n. Dans ce contexte, il existe de puissants outils probabilistes appelés inégalités exponentielles. on en trouve différentes versions dans la littérature. Les inégalités diffèrent selon les hypothèses imposés aux variables aléatoires Z_i. Nous en présentons ici celles qu'on appelle inégalités de type Bernstein dont la forme convient le plus à notre travail.

Supposons que $\{X_n, n \geq 1\}$ est une suite de variables aléatoires réelles, indépendantes et centrées.

Proposition 2 *si*

$$\forall m \geq 2, |E(X_i^m)| \leq \left(\frac{m}{2}\right)(a_i)^2 b^{m-2} \quad ,$$

alors

$$\forall \varepsilon \geq P\left[\sum_{i=1}^{n} |X_i| > \varepsilon A_n\right] \leqslant 2exp\left\{\frac{-\varepsilon^2}{2(1+\frac{b\varepsilon}{A_n})}\right\}$$

où $(a_i)_{1\leqslant i\leqslant n}$ *sont des réels positifs,* $b \in \mathbb{R}^+$ *et* $A_n^2 = a_1^2 + a_2^2 + a_3^2 + ... + a_n^2$.

Démonstration. (Bernstein (1946) ; Uspensky (1937) ; Yurinskii (1976))

Corollaire 1 *a) S'il existe une constante positive* $M < \infty$*,telle que* $|X_1| \leq M$*, alors on a*

$$\forall \varepsilon \geq 0, \quad P\left[\sum_{i=1}^{n} |X_i| > \varepsilon n\right] \leqslant 2exp\left\{\frac{-\varepsilon^2 n}{2\sigma^2(1+\frac{M\varepsilon}{\sigma^2})}\right\}$$

où $\sigma^2 = E(X_i^2)$.
b) Supposons que les $(X_i)_{1\leq i\leq n}$ *dépendent de n et que* $\sigma_n^2 = E(X_i^2)$*, s'il existe* $M = M_n < \infty$ *tel que* $|X_1| \leq M$ *si* $\dfrac{M}{\sigma_n^2} \leq C < \infty$ *et si* $u_n = n^{-1}\sigma_n^2 log(n)$ *vérifie* $lim_{n\longrightarrow\infty} = 0$*, alors nous avons*

$$\frac{1}{n}\sum_{i=1}^{n} X_i = O_{aco}(\sqrt{u_n})$$

Démonstration. a) En appliquant la proposition 2 à $a_i^2 = \sigma^2, A_n^2 = n\sigma_n^2$ et b = M nous aboutissons à a).
b) Comme $\frac{Mu_n}{\sigma_n^2}$ tend vers zéro, il suffit de rependre le résultat a) pour $\varepsilon = \varepsilon_0\sqrt{u_n}$ on arrive donc à l'existence d'une constante C' telle que :

$$P\left[\sum_{i=1}^{n} |X_i| > \varepsilon U_n\right] \leqslant 2exp\left\{\frac{-\varepsilon_0^2 log(n)}{2\sigma^2(1+\varepsilon_0\sqrt{\frac{MU_n}{\sigma_n^2}})}\right\}$$

$$\leq 2n^{-C'\varepsilon_0^2}$$

Pour ε_0 bien choisi le terme de droite est le terme général dune série convergente. Ainsi s'achève la preuve de ce corollaire.

Lemme 1 *(inégalité de Hoeffding)*
Soit $X_1, ..., X_n$ *une suite de variables aléatoires indépendantes centrés de même loi telles qu'il existe deux réels positifs* δ_1 *et* δ_2 *vérifiant :*

$$|X_1| \leq \delta_1 \quad et \quad E|X_1|^2 \leq \delta_2$$

Alors, pour tout $\varepsilon \in \left]0, \frac{\delta_2}{\delta_1}\right[$ *on a :*

$$P\left[n^{-1}\left|\sum_{i=1}^{n} X_i\right| > \varepsilon\right] \leq 2e^{\frac{-n\varepsilon^2}{4\delta_2}}$$

1.5 Modèle Non Paramétrique de Régression

Commençons par présenter le modèle fonctionnel non paramétrique de régression. Soit $\mathcal{X}$ une v.a.f. à valeurs dans un espace semi métrique (E, d) ,Y une v.a.r. et $(\mathcal{X}_i, Y_i)_{i=1,...,n}$ n couples identiquement et indépendamment distribués suivant $(\mathcal{X}, Y)$. On s'intéresse alors au modèle :

$$Y_i = r(\mathcal{X}_i) + \varepsilon_i, \qquad i = 1, ..., n \tag{1.1}$$

où les ε_i sont des v.a.r. centrées telles que $E(\varepsilon_i/\mathcal{X}_i) = 0$.

1.6 présentation de l'estimateur à noyau

1.6.1 Cas réel

On suppose que f est continue, et donc qu'elle est la dérivée de F telle que :

$$\forall t \in [0,1] \quad F(t) = \int_0^1 f(x)dx.$$

On peut estimer F de manière empirique par F donnée par :

$$\forall t \in [0,1] \quad \widehat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} Y_i 1_{x_i \leq t},$$

puis par f un taux d'accroissement

$$\widehat{f}(x) = \frac{\widehat{F}(x+h) - \widehat{F}(x-h)}{2h}.$$

On a alors

$$\widehat{f}(x) = \frac{1}{2nh}\sum_{i=1}^{n} Y_i 1_{\{x-h<x_i\leq x+h\}}.$$

$\widehat{f}(x)$ est encore la moyenne locale des Y_i mais sur une fenêtre "glissante" centrée sur x. Si on pose $K(x) = \frac{1}{2}1_{-1<x\leq 1}$ on a :

$$\widehat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} Y_i \frac{1}{h} K\left(\frac{x_i - x}{h}\right). \tag{1.2}$$

Le paramètre $h > 0$ est appelé fenêtre, c'est un paramètre de lissage.
Les méthodes du noyau sont bien connues et utilisées de manière intensive par la communauté d'ions non paramétriques, car elles constituent un moyen utile d'établir une pondération locale.Nous commençons par rappeler quelle est la pondération locale du noyau dans le réel Avant de l'étendre au contexte fonctionnel.

Comme il est bien connu, la pondération locale du noyau est basée sur une fonction de noyau (classiquement désigné par K) et sur un paramètre de lissage, appelé bande passante et habituellement désigné par h. Si x est un nombre réel fixe, la pondération locale du noyau se transforme n Variable aléatoire réelle$\Delta_1, \Delta_2, ...\Delta_n$ tel que

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

L'idée principale de la pondération locale autour de x est d'attribuer à chaque Variable aléatoire réelle X_i un poids prenant en compte la distance entre x et Xi ; Plus X_i est éloigné de x, plus la pondération est petite.

1.6.2 Les conditions sur le noyau K :

K étant le noyau déterminant la forme du voisinage et satisfaisant :
(K.1) $K(x) \geqslant 0$ pour tout x (positivité)
(K.2) $\int K(x)dx = 1$
(K.3)$\int xK(x)dx = 0$ ("symétrie")

1.6.3 Quelques formes des noyaux.

Les noyaux les plus couramment utilisés en pratique sont :
-Noyau de boîte :

$$K(u) = \frac{1}{2}1_{[-1;1]}(u)$$

-Noyau de triangulaire :

$$K(u) = (1 - |u|)1_{[-1;1]}(u)$$

-Noyau de gaussien :

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$$

-Noyau d'Epanechnikov :

$$K(u) = \frac{3}{4}(1 - u^2)1_{[-1;1]}(u)$$

Pour préciser la notion de pondération locale du noyau, considérons le noyau de boîte et réécrivez les Δ_i comme suit :

$$\Delta_i = \frac{1}{h} 1_{[x-h;x+h]}(X_i)$$

Dans cette situation, la caractéristique locale de la pondération semble évidente depuis La variable aléatoire réelle en dehors de la gamme $[x - h; x + h]$ Sont ignorés. En outre, la normalisation $1/h$ est proportionnel à la taille de l'ensemble.

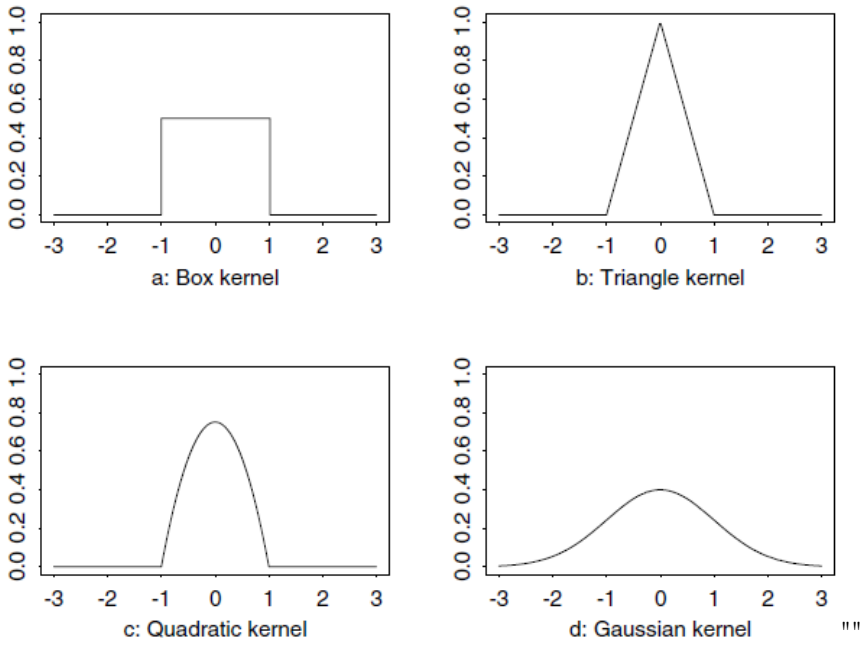


FIGURE 1.2 – formes des noyaux

1.6.4 Cas fonctionnel

L'arrière-plan présenté ci-dessus suffit à introduire le noyau local pondération dans le cas fonctionnel.soit $\mathcal{X}_1, \mathcal{X}_2, ..., \mathcal{X}_n$ variable aléatoire fonctionnelle évalué dans E et soit χ un élément fixe de E. une extension fonctionnelle naïve des idées de pondération locale du noyau multivarié serait de transformer n variable aléatoire fonctionnelle $\mathcal{X}_1, \mathcal{X}_2, ..., \mathcal{X}_n$ dans les quantités n

$$\frac{1}{V(h)} K\left(\frac{d(\chi, \mathcal{X}_i)}{h}\right)$$

où d est semi-métrique sur E, K est un noyau réel (asymétrique). dans cette expression V(h) serait le volume de :

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

notez que les fonctions noyau K à utiliser ici sont nécessairement asymétriques . Par souci de simplicité, dans le reste de ce travail, nous considérerons seulement deux types de noyaux pour pondérer les variables fonctionnelles

Définition 6 *Une fonction K de $\mathbb{R}$ dans $\mathbb{R}^+$ telle que $\int K = 1$ appelé noyau de type I s'il existe deux constantes réelles $0 < C_1 < C_2 < \infty$ tel que :*

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

Une fonction K de $\mathbb{R}$ dans $\mathbb{R}^+$ telle que $\int K = 1$ est appelé un noyau de type II si son support est $[0, 1]$ et si sa dérivée $K^{'}$ existe sur $[0, 1]$ et satisfait pour deux constantes réelles $-\infty < C_2 < C_1 < 0$

$$C_2 \leq K^{'} \leq C_1$$

La première famille de noyaux contient les noyaux discontinus usuels tels que la boîte asymétrique tandis que la deuxième famille contient Les continus (comme le triangle, le quadratique, ...). enfin, pour être en compatible avec cette définition et simplifier notre but, pour la pondération locale Des variables aléatoires réelles, nous considérons simplement le type de noyau suivant.

Définition 7 *Une fonction K de $\mathbb{R}$ dans $\mathbb{R}^+$ telle que $\int K = 1$ avec support compact $[-1, 1]$ et tel que $\forall \mu \in (0, 1), K(\mu) > 0$ est appelé a noyau de type 0.*

1.6.5 Pondération locale et Probabilités de petite boule

Nous pouvons maintenant construire le pont entre la pondération locale et la notion de probabilité de petite boule. pour fixer les idées, considérez le noyau le plus simple parmi ceux de type I, c'est-à-dire le noyau asymétrique.Soit $\mathcal{X}$ un v.a.f. évalué en E et soit χ un élément fixe de E. On peut écrire :

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

pourquoi nous disons des probabilités de petite boule ? En effet, comme nous le verrons la suite, le lissage h de paramètre (également appelé la bande passante) diminue avec la taille de l'échantillon des variables fonctionnelles (plus précisément, les h tend vers zéro lorsque n tend vers ∞). Ainsi, quand nous prenons n très grand, h est proche de zéro et puis $B(\chi,h)$ est considéré comme une petite boule et $P(\mathcal{X} \in B(\chi,h))$ comme une petite probabilité de boule

Désormais, pour tout χ en E et pour tout h réel positif, nous utiliserons le notation :

$$\varphi_\chi(h) = P(\mathcal{X} \in B(\chi,h)) \tag{1.4}$$

Cette notion de probabilité de la petite boule jouera un rôle majeur à la fois à partir de la théorie et des points de vue pratiques. Parce que la notion de boule est fortement Liée au semi-métrique d, le choix de ce semi-métrique deviendra un étape importante.

1.6.6 Quelques théorèmes fondamentales

Parce que les notions de pondération locale du noyau fonctionnel seront au cur de Toutes les méthodes statistiques non paramétriques fonctionnelles à étudier ultérieurement Dans ce mémoire, nous avons décidé de rassembler ici quelques résultats courants communs. nous énoncera deux résultats, selon le fait que le noyau est de type **I** ou **II**, qui peuvent être considérés comme des versions fonctionnelles, les deux lemmes suivants seront utilisés très souvent dans le reste de ce mémoire. Avant de continuer, soit X un v.a.f Prenant ses valeurs dans l'espace semi-métrique (E,d), soit χ un élément fixe de E, soit h un nombre réel positif et soit K une fonction noyau.

Lemme 2 *Si K est un noyau de type* ***I****,alors il existe des constantes réelles finies non négatives C et C*$^{'}$ *tel que :*

$$C\varphi_{\chi}(h) \leqslant EK\left(\frac{d(\chi,\mathcal{X})}{h}\right) \leqslant C^{'}\varphi_{\chi}(h) \tag{1.5}$$

Preuve. Parce que K est un noyau de type I, nous avons par Définition (6).

$$C_1 1_{[0,1]} \leqslant k \leqslant C_2 1_{[0,1]}$$

Ce qui implique directement que :

$$C_1 1_{B(\chi,h)}(\mathcal{X}) \leqslant K\left(\frac{d(\chi,\mathcal{X})}{h}\right) \leqslant C_2 1_{B(\chi,h)}(\mathcal{X})$$

Il suffit d'appliquer (4.3) pour obtenir le résultat revendiqué avec $C = C_1$ et $C^{'} = C_2$.

Lemme 3 *Si K est un noyau de type II et si* $\varphi_{\chi}(.)$ *Vérifie*

$$\exists C_3 > 0, \exists \epsilon_0, \forall \epsilon < \epsilon_0, \int_0^{\epsilon} \varphi_{\chi}(\mu)d\mu > C_3\epsilon\varphi_{\chi}(\epsilon) \tag{1.6}$$

Alors il existe des constantes réelles finies non négatives C et C$^{'}$ *Tel que, pour h assez petit :*

$$C\varphi_{\chi}(h) \leqslant EK\left(\frac{d(\chi,\mathcal{X})}{h}\right) \leqslant C^{'}\varphi_{\chi}(h) \tag{1.7}$$

Preuve. Nous commençons par écrire :

$$EK\left(\frac{d(\chi,\mathcal{X})}{h}\right) = \int_0^1 K(t)dP^{\frac{d(\chi,\mathcal{X})}{h}}(t)$$

et parce que $K^{'}$ Existe, on a $K(t) = K(0) + \int_0^1 K^{'}(t)(u)du$, ce qui implique que

$$\begin{aligned} EK\left(\frac{d(\chi,\mathcal{X})}{h}\right) &= \int_0^1 K(0)dP^{\frac{d(\chi,\mathcal{X})}{h}}(t) + \int_0^1 \left(\int_0^t K^{'}(u)du\right) dP^{\frac{d(\chi,\mathcal{X})}{h}}(t) \\ &= K(0)\varphi_{\chi}(h) + \int_0^1 \left(\int_0^1 K^{'}(u)1_{[u,1]}(t)du\right) dP^{\frac{d(\chi,\mathcal{X})}{h}}(t) \\ &= K(0)\varphi_{\chi}(h) + \int_0^1 \left(\int_0^1 K^{'}(u)1_{[u,1]}(t)du\right) dP^{\frac{d(\chi,\mathcal{X})}{h}}(t) \end{aligned}$$

$$= K(0)\varphi_\chi(h) + \int_0^1 K^{'}(u) P\left(u \leqslant \frac{d(\chi, \mathcal{X})}{h} \leqslant 1\right) du,$$

La dernière équation étant obtenue en appliquant le théorème de Fubini. En utilisant le Le fait que K (1) = 0 nous permet d'écrire :

$$EK\left(\frac{d(\chi, \mathcal{X})}{h}\right) = -\int_0^1 K^{'}(u)\varphi_\chi(hu)du$$

Il suffit d'utiliser (1.6) pour montrer que, pour $h < 0$ et avec $C = -C_3C_1$,

$$EK\left(\frac{d(\chi, \mathcal{X})}{h}\right) \geqslant C\varphi_\chi(h)$$

En ce qui concerne la limite supérieure, il suffit de remarquer que K est borné par Support $[0, 1]$ et les mêmes arguments que pour le lemme (2) peuvent être utilisés par Mettre $C^{'} = Sup_{t\in[0,1]}K(t)$.

1.7 Présentation de l'estimateur de r

On s'intéresse à l'estimation de l'opérateur

$$r(\chi) = E(Y|\mathcal{X} = \chi)$$

L'estimateur que nous étudions dans ce mémoire est une généralisation des estimateurs de type Nadaraya-Watson au cas de variables fonctionnelles introduite par Ferraty et Vieu (2000). Pour tout élément χ de E il s'écrit de la manière suivante :

$$\begin{cases} \widehat{r}(\chi) = \dfrac{\sum_{i=1}^{n} K(h^{-1}d(\chi, \mathcal{X}_i))y_i}{\sum_{i=1}^{n} K(h^{-1}d(\chi, \mathcal{X}_i))} & si \quad \sum_{i=1}^{n} K(h^{-1}d(\chi, \mathcal{X}_i)) \neq 0 \end{cases}$$

1.8 Distribution conditionnement par une v.a.f.

Une autre façon d'étudier le lien entre une v.a.r. Y et une v.a.f.$\mathcal{X}$ est de s'intéresser à la fonction de répartition de Y sachant X définie par :

$$F_Y^{\mathcal{X}}(\chi, y) = P(Y \leq y|\mathcal{X} = \chi)$$

$$= E(1_{(-\infty,y]}(Y)|\mathcal{X} = \chi)$$

$$= r(1_{(-\infty,y]}(Y))$$

Et par analogie avec le contexte de régression fonctionnelle, un noyau naïve conditionnel c.d.f. L'estimateur pourrait être défini comme suit :

$$\widehat{F}_Y^{\mathcal{X}}(\chi, y) = \widehat{r}(1_{(-\infty,y]}(Y)) = \frac{\sum_{i=1}^{n} K(h^{-1}d(\chi, \mathcal{X}_i))1_{(-\infty,y]}(Y_i)}{\sum_{i=1}^{n} K(h^{-1}d(\chi, \mathcal{X}_i))}$$

Enfin, par Roussas, Samanta et Ferraty et Vieu , l'estimateur de la fonction de distribution conditionnelle est donné par :

$$\widehat{F}_Y^{\mathcal{X}}(\chi, y) = \frac{\sum_{i=1}^{n} K(h^{-1}d(\chi, \mathcal{X}_i))H(g^{-1}(y - Y_i))}{\sum_{i=1}^{n} K(h^{-1}d(\chi, \mathcal{X}_i))}, \qquad \forall y \in \mathbb{R}$$

Soit K_0 un noyau symétrique habituel, soit H défini comme suit :

$$\forall \mu \in \mathbb{R} H(\mu) = \int_{-\infty}^{\mu} K_0(\upsilon)d\upsilon$$

considérons K_0 comme un noyau de type 0
en outre, nous pouvons écrire :

$$H(g^{-1}(y - Y_i)) = \begin{cases} 0 \Leftrightarrow y \leq Y_i - g, \\ 1 \Leftrightarrow y \geq Y_i + g. \end{cases}$$

Chapitre 2

Étude asymptotiques

L'étude de la convergence, presque complète des estimateurs à noyau dans le cadre de la variable aléatoire fonctionnelle introduite par Ferraty et Vieu (2000), s'inspire énormément du cas réel exposé ci dessous. Cette convergence implique la convergence en probabilité et la convergence presque sûre.

2.1 Étude de convergence presque complète

Maintenant, nous introduisons les hypothèses de base permettant de donner un théorème général sur la convergence presque complète.

2.1.1 Hypothèses générales

$$(H1) \qquad P(\mathcal{X} \in B(\chi, h)) = \varphi_\chi(h) > 0$$

$$(H2) \qquad \forall (y_1, y_2) \in S \times S, \forall (\chi_1, \chi_2), \in N_\chi \times N_\chi, |F^{\chi_1}(y_1) - F^{\chi_2}(y_2)| \leq C_x \left(d(\chi_1, \chi_2)^{b1} + |y_1 - y_2|^{b2} \right)$$

$$(H3) \qquad \begin{cases} \forall (y_1, y_2) \in \mathbb{R}^2, |H(y_1) - H(y_2)| \leq C|(y_1 - y_2)| \\ \int |t|^{b_2} H^{(1)}(t) dt < \infty \end{cases}$$

$(H4)$ K *est une fonction avec support* $(0,1)$ *telle que* $0 < C_1 < K(t) < C_2 < \infty$

$$(H5) \qquad \lim_{n\longrightarrow\infty} h_k = 0, \lim_{n\longrightarrow\infty} \frac{log(n)}{n\varphi_x(h_k)} = 0$$

$$(H6) \qquad \lim_{n\longrightarrow\infty} h_H = 0 \ \ with \ \ \lim_{n\longrightarrow\infty} n^a h_H = \infty \ \ for \ \ some \ \ a > 0.$$

2.2 Théorie de la convergence

Théorème 8 *pour tout compact S, on a :*

Sous les hypothèses (H1),(H2)et(H3),(H4),nous avons

$$\sup_{y\in S} |\widehat{F}^{\chi}(y) - F^{\chi}(y)| = O(h_k^{b1}) + O(h_H^{b2}) + O\left(\sqrt{\frac{log(n)}{n\varphi_x(h_k)}}\right), \quad p.co \tag{2.1}$$

Preuve de théorème :
pour $i = 1, ..., n$, nous considérons les quantités

$$K_i = K(h_k^{-1} d(\chi, \mathcal{X}_i)), \qquad H_i = H(h_H^{-1}(y - Y_i)$$

soit $\widehat{F}_N^{\chi}(y), \widehat{F}_D^{\chi}(y)$:

$$\widehat{F}_N^{\chi}(y) = \frac{1}{nEK_1} \sum_{i=1}^{n} K_i H_i(y) \qquad \widehat{F}_D^{\chi}(y) = \frac{1}{nEK_1} \sum_{i=1}^{n} K_i$$

Cette preuve est basée sur la décomposition

$$\widehat{F}^{\chi}(y) - F^{\chi}(y) = \frac{1}{\widehat{F}_D^{\chi}} \left\{ \left(\widehat{F}_N^{\chi}(y) - E\widehat{F}_N^{\chi}(y) \right) - \left(F^{\chi}(y) - E\widehat{F}_N^{\chi}(y) \right) \right\} \tag{2.2}$$

$$+ \frac{F^{\chi}(y)}{\widehat{F}_D^{\chi}} \left\{ E\widehat{F}_D^{\chi} - \widehat{F}_D^{\chi} \right\}$$

et sur les résultats intermédiaires suivants :

Lemme 4 *Sous l'hypothèse (H1) et(H5),(H6) nous avons :*

$$\widehat{F}_D^{\chi} - E\widehat{F}_D^{\chi} = O\left(\sqrt{\frac{log(n)}{n\varphi_x(h_k)}}\right), \quad p.co \tag{2.3}$$

Corollaire 2 *Sous l'hypothèse de lemme(4)nous avons :*

$$\sum_{n=1}^{\infty} P\left(\widehat{F}_D^{\chi} < \frac{1}{2}\right) < \infty \tag{2.4}$$

Lemme 5 *Sous l'hypothèse (H1),(H2)et (H4),(H6)nous avons :*

$$\frac{1}{\widehat{F}_D^{\chi}} \sup_{y\in S} |F^{\chi}(y) - E\widehat{F}_N^{\chi}(y)| = O(h_k^{b1}) + O(h_H^{b2}), \quad p.co \tag{2.5}$$

Lemme 6 *Sous l'hypothèse (H1),(H2)et (H4),(H7)nous avons :*

$$\frac{1}{\widehat{F}_D^{\chi}} \sup_{y\in S} |\widehat{F}_N^{\chi}(y) - E\widehat{F}_N^{\chi}(y)| = O\left(\sqrt{\frac{log(n)}{n\varphi_x(h_k)}}\right), \quad p.co \tag{2.6}$$

2.2.1 preuve

Preuve de lemme (4) :

nous avons

$$\widehat{F}_D^{\chi} - E\widehat{F}_D^{\chi} = \frac{1}{nEK_1}\sum_{i=1}^{n} K_i - \frac{1}{nEK_1}\sum_{i=1}^{n} EK_i = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{K_i}{EK_1} - 1\right) =$$

$$\frac{1}{n}\sum_{i=1}^{n}(\Delta_i - 1)$$

d'après de (H1) et (H2),nous pouvons écrire

$$C\varphi_{\chi}(h_k) < EK_1 < C^{'}\varphi_{\chi}(h_k)$$

En utilisant soit le lemme (2) ou (3) pour K est de Type I ou II, nous pouvons obtenir directement cela :

$$|\Delta_i| < C/\varphi_{\chi}(h_k) = \delta_1 \;\; et \;\; E|\Delta_i|^2 < C^{'}/\varphi_{\chi}(h_k) = \delta_2$$

On obtient alors en appliquant une version simplifiée d'une inégalité exponentielle de type Bernstein dans le cas indépendant et hoeffding,du Lemme (1) du Chapitre 1 , d'écrire pour tous $\eta \in (0, \delta_2/\delta_1)$

$$P\left(|\widehat{F}_D^{\chi} - E\widehat{F}_D^{\chi}| > \eta\sqrt{\frac{log(n)}{n\varphi_x(h_k)}}\right) \leq C' exp\left\{-n\eta^2 \frac{log(n)}{\varphi_{\chi}(h_k)} \frac{\varphi_{\chi}(h_k)}{C''}\right\}$$

ce qui équivaut à :

$$P\left(|\widehat{F}_D^{\chi} - E\widehat{F}_D^{\chi}| > \eta\sqrt{\frac{log(n)}{n\varphi_x(h_k)}}\right) \leq C' n^{-C\eta^2}$$

Preuve de corollaire

$$\sum_{n=1}^{\infty} P\left(\widehat{F}_D^{\chi} < \frac{1}{2}\right) \quad tel \quad que \quad \widehat{F}_D^{\chi} \quad < \quad \frac{1}{2} \Rightarrow \quad | \quad \widehat{F}_D^{\chi} - 1 \quad | \quad > \quad \frac{1}{2}$$

$$P\left(|\widehat{F}_D^{\chi}| \leq \frac{1}{2}\right) \leq P\left(|\widehat{F}_D^{\chi} - 1| > \frac{1}{2}\right)$$

en notant que :

$$\widehat{F}_D^{\chi} - 1 = \widehat{F}_D^{\chi} - E\widehat{F}_D^{\chi}$$

et en appliquant le lemme(4) précédent nous pouvons écrire :

$$\sum_{n=1}^{\infty} P\left(\widehat{F}_D^{\chi} < \frac{1}{2}\right) < \infty$$

Preuve de lemme (5)

$$E\widehat{F}_N^{\chi}(y) - F^{\chi}(y) = \frac{1}{EK_1} E(K_1[E(H_1(y)|X) - F^{\chi}(y)]) \tag{2.7}$$

de plus, nous avons

$$E(H_1(y)|X) = \int_{\mathbb{R}} H(h_H^{-1}(y - z)) f^X(z) dz$$

Une intégration par parties, nous donne :

$$E(H_1(y)|X) = \left[H(\frac{y-z}{h_H})F^X(z)\right]_{-\infty}^{+\infty} + \int_{\mathbb{R}} H^{(1)}(\frac{y-z}{h_H})F^X(z)dz.$$
$$= 0 + \int_{\mathbb{R}} H^{(1)}(\frac{y-z}{h_H})F^X(z)dz$$

Le changement de variable $t = \dfrac{y-z}{h_H}$,nous conduit à

$$E(H_1(y)|X) = \int_{\mathbb{R}} H^{(1)}(t)F^X(y - h_H t)dt.$$

Donc, nous avons

$$|E(H_1(y)|X) - F^\chi(y)| \leqslant \int_{\mathbb{R}} H^{(1)}(t)|F^X(y - h_H t) - F^\chi(y)|dt$$

Maintenant, (H2) permet d'écrire :

$$|E(H_1(y)|X) - F^\chi(y)| \leqslant C_\chi \int_{\mathbb{R}} H^{(1)}(t)(h_K^{b_1} + |t|^{b_2} h_H^{b_2}) \tag{2.8}$$

parce que cette inégalité est uniforme sur Y et à cause de (H.3), (2.5) est une conséquence directe de $2.7, 2.8$ et de corollaire 2

Preuve de lemme (6)

S est un compact, il existe donc un recouvrement fini de S tel que :

$$S \subset \bigcup_{K=1}^{S_n} S_K \quad et \quad S_K = (t_K - l_n, t_K + l_n)$$

Prenant $t_y = arg\ min_{t\in\{t_1,\ldots,t_{S_n}\}}|y - t|$, nous avons :

$$\frac{1}{\widehat{F}_D^\chi}\sup_{y\in S}|\widehat{F}_N^\chi(y) - E\widehat{F}_N^\chi(y)| \leq \underbrace{\frac{1}{\widehat{F}_D^\chi}\sup_{y\in S}|\widehat{F}_N^\chi(y) - \widehat{F}_N^\chi(t_y)|}_{T_1} + \underbrace{\frac{1}{\widehat{F}_D^\chi}\sup_{y\in S}|\widehat{F}_N^\chi(t_y) - E\widehat{F}_N^\chi(t_y)|}_{T_2}$$
$$+ \underbrace{\frac{1}{\widehat{F}_D^\chi}\sup_{y\in S}|E\widehat{F}_N^\chi(t_y) - E\widehat{F}_N^\chi(y)|}_{T_3} \tag{2.9}$$

concernant (T_1)

$$\frac{1}{\widehat{F}_D^{\chi}} \sup_{y\in S} |\widehat{F}_N^{\chi}(y) - \widehat{F}_N^{\chi}(t_y)| \leq \frac{1}{\widehat{F}_D^{\chi}} \sup_{y\in S} \frac{1}{nEK_1} \sum_{i=1}^{n} |H_i(y) - H_i(t_y)| K_i,$$

$$\leq \frac{1}{\widehat{F}_D^{\chi}} \sup_{y\in S} \frac{C|y - t_y|}{h_H} \left(\frac{1}{nEK_1} \sum_{i=1}^{n} K_i \right)$$

$$\leq C \frac{l_n}{h_H}. \tag{2.10}$$

la seconde inégalité est obtenue en considérant un argument de Lipschitz alors que le dernier vient de la définition de $\widehat{F}_D^{\chi}$. Prendre maintenant

$$l_n = n^{-a-\frac{1}{2}}$$

et notez que, à cause de (H_6), nous avons :

$$\frac{l_n}{h_H} = O\left(\sqrt{\frac{log(n)}{n\varphi_x(h_k)}} \right)$$

donc, pour une assez grande on peut écrire

$$P\left(\frac{1}{\widehat{F}_D^{\chi}} \sup_{y\in S} |\widehat{F}_N^{\chi}(y) - \widehat{F}_N^{\chi}(t_y)| > \frac{\eta}{3} \sqrt{\frac{log(n)}{n\varphi_x(h_k)}} \right) = 0. \tag{2.11}$$

concernant (T_2)

$$P\left(\sup_{y\in S} |\widehat{F}_N^{\chi}(t_y) - E\widehat{F}_N^{\chi}(t_y)| > \frac{\eta}{3} \sqrt{\frac{log(n)}{n\varphi_x(h_k)}} \right)$$

$$= P\left(\max_{t_y \in \{t_1,\ldots,t_{S_n}\}} |\widehat{F}_N^{\chi}(t_y) - E\widehat{F}_N^{\chi}(t_y)| > \frac{\eta}{3} \sqrt{\frac{log(n)}{n\varphi_x(h_k)}} \right)$$

$$\leq s_n \max_{t_y \in \{t_1,\ldots,t_{S_n}\}} P\left(|\widehat{F}_N^{\chi}(t_y) - E\widehat{F}_N^{\chi}(t_y)| > \frac{\eta}{3} \sqrt{\frac{log(n)}{n\varphi_x(h_k)}} \right)$$

soit

$$A_i = \frac{(K_i H_i(t_y) - E(K_1 H_1(t_y)))}{EK_1}.$$

En utilisant des arguments similaires à ceux de la preuve du lemme (4) et de $H \leq 1$, on en déduit :

$$E|A_i| \leq C/\varphi_x(h_k) \;\; et \;\; EA_i^2 \leq C'/\varphi_x(h_k)$$

. Nous appliquons à nouveau l'inégalité exponentielle de Bernstein pour obtenir

$$P\left(|\widehat{F}_N^{\chi}(y) - E\widehat{F}_N^{\chi}(y)| > \frac{\eta}{3}\sqrt{\frac{log(n)}{n\varphi_x(h_k)}}\right) \leq 2exp\left\{-C\eta^2 log(n)\right\}$$

Ainsi, en choisissant η tel que $C\eta^2 = \frac{3}{2} + 2a$, on obtient

$$\begin{aligned} s_n \max_{t_y \in \{t_1,\ldots,t_{S_n}\}} P\left(|\widehat{F}_N^{\chi}(t_y) - E\widehat{F}_N^{\chi}(t_y)| > \frac{\eta}{3}\sqrt{\frac{log(n)}{n\varphi_x(h_k)}}\right) &\leq Cs_n n^{-3/2-2a} \\ &\leq \frac{C}{l_n} n^{-3/2-2a} \end{aligned}$$

Parce que $l_n = n^{-a-1/2}$, on déduit de l'inégalité précédente que :

$$P\left(\sup_{y\in S}|\widehat{F}_N^{\chi}(t_y) - E\widehat{F}_N^{\chi}(t_y)| > \frac{\eta}{3}\sqrt{\frac{log(n)}{n\varphi_x(h_k)}}\right) \leq Cn^{-1-a}.$$

enfin, en utilisant le corollaire(2), nous obtenons

$$P\left(\frac{1}{\widehat{F}_D^{\chi}}\sup_{y\in S}|\widehat{F}_N^{\chi}(t_y) - E\widehat{F}_N^{\chi}(t_y)| > \frac{\eta}{3}\sqrt{\frac{log(n)}{n\varphi_x(h_k)}}\right) \leq Cn^{-1-a}. \tag{2.12}$$

Concernant (T_3) : à cause de (2.10) nous avons :

$$\sup_{y\in S}|E\widehat{F}_N^{\chi}(y) - E\widehat{F}_N^{\chi}(t_y)| \leq C\frac{l_n}{h_H}.$$

En utilisant des arguments analogues comme pour T_1, nous pouvons montrer pour n assez grand :

$$P\left(\frac{1}{\widehat{F}_D^{\chi}}\sup_{y\in S}|E\widehat{F}_N^{\chi}(y) - E\widehat{F}_N^{\chi}(t_y)| > \frac{\eta}{3}\sqrt{\frac{log(n)}{n\varphi_x(h_k)}}\right) = 0. \tag{2.13}$$

Maintenant, le lemme (6) peut être facilement déduit de équations (2.9), (2.11), (2.12) et (2.13).

Chapitre 3

Simulation

Dans cette section, nous calculons et représentons graphiquement la fonction de Distribution conditionnelle et son estimateur en Distribution conditionnelle vue de les comparer dans des situations simulées .
Dans tout ce qui suit,$\mathcal{X}$ désigne une v.a.f. et Y une variable aléatoire réel. Soit $(\mathcal{X}_i, Y_i)_{i=1,\dots,n}$ n couples identiquement distribués selon la loi de $(\mathcal{X}, Y)$.

3.1 générer les données fonctionnelles

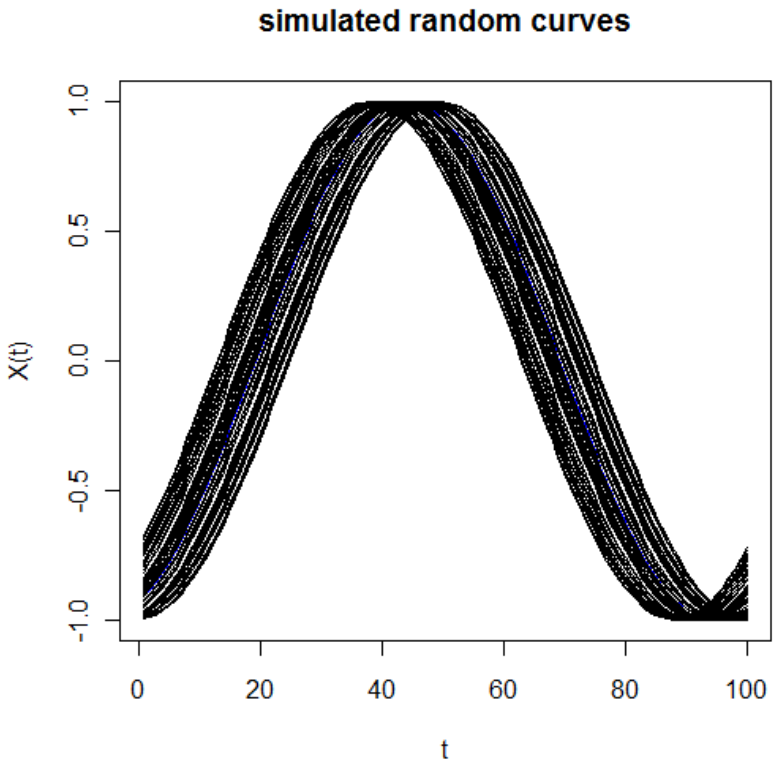
nous simulons 100 variables $w_{0,i}$ uniformément distribuées sur $[0, \pi/4]$,ensuite, nous créons pour chaque $w_{0,i}$,un vecteur $(\mathcal{X}_{0,i}(t_j))_{\{j=1\dots100\}}$ c'est la version discrétisée de la fonction $t \longrightarrow cos(w_{0,i} + \pi(2t-1)/100 - 1)$:

$$\mathcal{X}[i,j] = cos\left(w_{0,i} + \pi\left(\frac{2j}{100} - 1\right)\right)$$

3.1.1 Application dans R :

Pour obtenir Les courbes $\mathcal{X}_i$ qui sont régulièrement sur une grille de discrétisation de 100 points dans l'intervalle $[0, 1]$ dans logiciel R, on suit les étapes suivantes :
$\mathcal{X}$ <- matrix(nrow = 100, ncol=100, byrow=TRUE)
xx <- (1 :100)
w= runif(100,0,pi/4)
for (i in 1 :100) {
for (j in 1 :100) {

$\mathcal{X}$[i,j]=cos(w[i] + pi*(2*j/100-1))
}
}
yy <- $\mathcal{X}$[1,]
plot(xx,yy,type="l",col="blue",xlab="t", ylab="X(t)", main="simulated random curves")
for (j in 2 :100){
lines(xx,$\mathcal{X}$[j,], col="black")
}
on obtient une échantillon de 100 observation illustrées dans la figure 3.1 : représente une



""

FIGURE 3.1 – Les courbes $\mathcal{X}_i$

une observation dans la figure 3.2

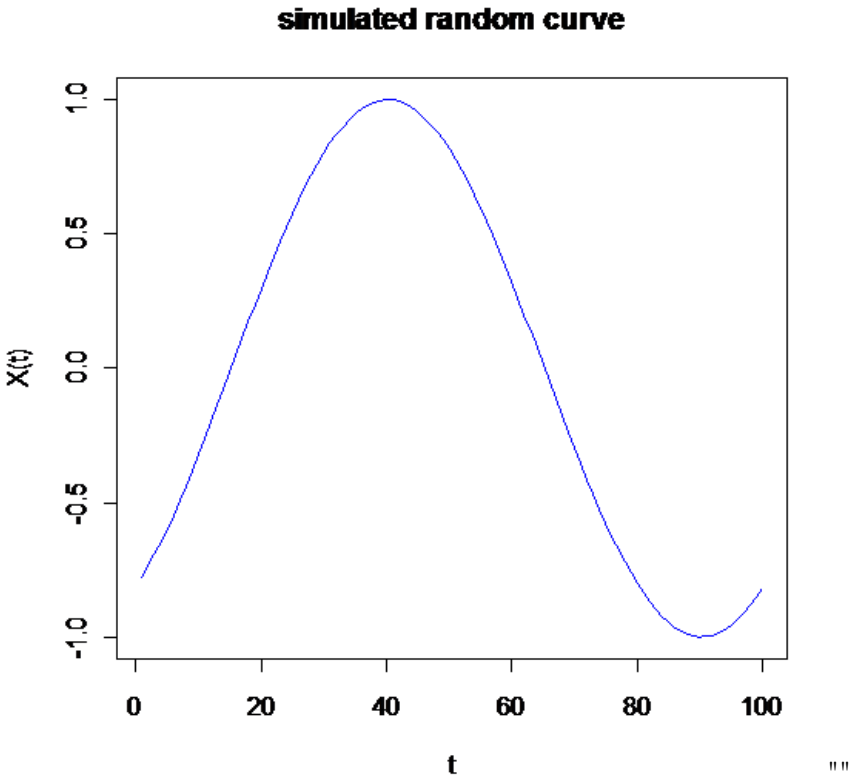


FIGURE 3.2 – Le courbe χ_i

3.2 Algorithme de l'estimateur du fonction de répartition conditionnelle

Soit (X, Y) un couple de v.a.r dont la loi est définie par la densité :

- :choisir la fonction du régression : $r = x^4/4$
- : choisir K la Noyau de gaussien $K = (1/\sqrt{2}\pi)exp(-x^2/2)$
- :considérons un échantillons de $n = 100$ courbe
- :générer $\varepsilon_i \to U(0, 100)$
- :choisir F la fonction de répartition conditionnelle

$$F(x,y) = \begin{cases} 0 & si \quad (x,y) \in]-\infty, 0[\\ 2x^2y - x^2y^2 & si \quad (x,y) \in [0,1] \\ 1 & si \quad (x,y) \in]1, +\infty[\end{cases}$$

3.2.1 le programme sous R

sa fonction de l'estimateur répartition conditionnelle

$$\widehat{F}_Y^{\mathcal{X}}(\chi, y) = \frac{\sum_{i=1}^{n} K(h^{-1}d(\chi, \mathcal{X}_i))H(g^{-1}(y - Y_i))}{\sum_{i=1}^{n} K(h^{-1}d(\chi, \mathcal{X}_i))}, \qquad \forall y \in \mathbb{R}$$

• :le programme sous R est donné par :

$\mathcal{X} < -matrix(nrow = 100, ncol = 100, byrow = TRUE)$

$xx < -(1 : 100)$

$w = runif(100, 0, pi/4)$

$for(i \ \ in \ \ 1 : 100) \ \ \{$
$for(j \ \ in \ \ 1 : 100) \ \ \{$

$\mathcal{X}[i, j] = cos(w[i] + pi * (2 * j/100 - 1))$

$\}$
$\}$

$r = function(t)\{t^4/4\}$

$K = function(x)\{(1/(2 * 3.14)^{0.5}) * exp(-x^2/2)\}$
$H = function(x)\{(x + 1)/2\}$
$Y = r(x[21,]) + runif(100, 0, pi/500)$

$A = 0$
$B = 0$

$for(iin1:100)\{$
$u = abs(x[21,] - x[i,])/0.0099$
$v = abs(1 - Y[i])/0.99$
$A = A + K(u)$
$B = B + (K(u) * H(v))$

$\}$
$f = A/B$

$s = 2 * (x[21,]^2) * Y[21] - (x[21,])^2 * (Y[21])^2$
$plot(f, type = "l", col = "red", xlim = c(0, 100), ylim = c(0, 1), lwd = 1,$
$ylab = "lescourebs", xlab = "t")$

$lines(s, col = "blue", lwd = 2)$
$legend(par('usr')[2], par('usr')[4], xjust = 3, yjust = 1, c("Festimateur", "F"), lwd = c(1, 2))$

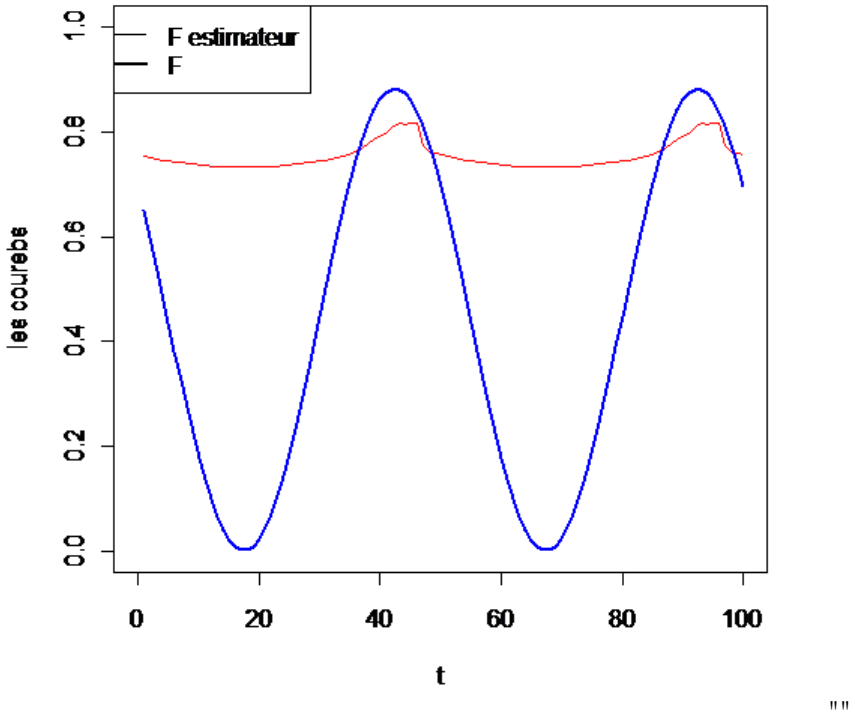


FIGURE 3.3 – Le courbe $\widehat{F}_Y^{\mathcal{X}}(\chi, y)$ *et* $F(\mathcal{X}, Y)$

Graphiques confirment l'existence d'une convergence entre la fonction de répartition conditionnelle $\widehat{F}_Y^{\mathcal{X}}(\chi, y)$ et la fonction $F(\mathcal{X}, Y_i)$, qui montre la bonne performance de l'estimateur $\widehat{F}_Y^{\mathcal{X}}(\chi, y)$

Conclusion

Dans ce travail nous avons présenté d'un pour des outils mathématique existants dans la littérature et qui sont en particulier necesseuré pour la résolution du problème posé pour cette dimension infinie. les vitesses de convergence sont aussi déterminées une application pratique, est réalise pour cela nous avons effectué une simulation pour l'estimateur dans coudre indépendant les résulta de simulations ont validé les résultats obtenus

Bibliographie

[1] Ferraty and Vieu, 2006, Nonparametric functional data analysis, Springer Series in Statistics, New York.

[2] F. Ferraty, A. Laksaci and Ph. Vieu, Estimating some characteristics of the conditional distribution in nonparametric functional models, Prepint (2003).

[3] Amel TADJ,Sur les modèles non paramétriques conditionnels en statistique fonctionnelle

[4] Laurent Delsol,Régression sur variable fonctionnelle : Estimation, , tests de structure et Applications.

[5] Khedidja Kebabi, Estimation non-paramétrique de la fonction de régression : cas dun modèle de censure mixte

[6] Ferraty and Vieu,COURS DE DEA Module STATISTIQUE FONCTIONNELLE : Modéles de régression pour variables aléatoires uni, multi et ∞-dimensionnées

[7] A. Laksaci contribution aux modélles non-paramétrique conditionnelle pour variables explicative fonctionnelle

[8] .Fethi Madani,Aspects théoriques et pratiques dans l'estimation non paramétrique de la densité conditionnelle pour des données fonctionnelles.

الملخص

الدراسة التيتمتناولهافيهذهالمذكرةتتمثلفي تقديرنموذج من النماذجاللامعلمية المشروطة في حالة البيانات الوظيفية والمتمثل في دالةالتوزيعالشرطيةبناء علىطريقةنواةلوظيفةالانحدار،الواردةفي فيراتي وليبكصاسي وآخرون ([1]).مع الأخذ القيم في فضاء الشبه متري . ثمندرسالتقاربشبهالكاملمقدر دالة التوزيعالشرطية. وكمثالعلىذلك،نعطيأمثلةعنالتطبيقاتالموجودةلبياناتالمحاكاة.

Résumé

Dans ce mémoire, nous étudions l'estimation non paramétrique du modèle conditionnel de modèles pour les variables fonctionnelles prenant des valeursdans un espace semi-métrique. Nous avons donné la convergence presque complète de l'estimateur non paramétrique pour la fonction de distribution conditionnelle et construisons l'estimateur par la méthode du noyau pour la fonction de l'estimateur régression, donnée dans Ferraty et al. ([1]) et Laksaci,A titre illustratif, nous donnons des exemples d'applications sur des données simulées.

Abstract

In this memoire we study nonparametric estimation of conditional model of models for functional variables taking values in a semi-metric space. we given the almost complete convergence of nonparametric estimator for conditional distribution function and construct estimator by the kernel method for the generalized regression function, given in Ferraty et al. ([1]) and Laksaci .We illustrates our results by giving an application on simulated samples .