

Dédicace

A toute ma famille

Remerciement

Je remercie en premier lieu, notre dieu qui ma donner la force pour effectuer ce présent travail.

En second lieu, mes remerciements vont à Monsieur Fethi Madani pour son aide et ses conseils, en saluant en lui son savoir faire, sa compétence et ces connaissances dont il nous a fait en profiter.

Mes remerciement envers Monsieur Azzouz Abdelhalim, qui ma fait l'honneur de bien vouloir présider le jury de ce mémoire.

Je me retourne vers *M^{elle}* Fatima Benziadi, qui a bien voulu accepter de juger ce travail et de faire partie du jury.

Mes remerciement aussi pour Monsieur kandouci Abdeldjebbar, qui a bien voulu contribuer et examiner ce travail et de faire partie du jury.

A tout mes enseignants de département de Mathématiques et Informatique. Et à tous les étudiantes et étudiants de Mathématique. Enfin, à tous les personnes qui m'ont aidé à réaliser ce modeste travail.

Table des matières

Introduction générale	4
1 Une littérature sur l'analyse des données fonctionnelles	5
1.1 Risque quadratique moyen (Mean Square error)	6
1.2 Risque quadratique moyenne intégré (Mean integrated squared error)	7
1.3 La validation croisée	7
2 Le choix optimal de la largeur de fenêtre (cadre réel)	9
2.1 Introduction	9
2.2 Le critère de sélection des paramètres de lissage	11
2.3 Optimalité asymptotique	13
3 Estimateur à noyau de la densité conditionnelle (données fonctionnelles)	17
3.1 Critère de choix local et global pour la sélection de paramètres de lissage	18
3.2 Résultats principaux	21
3.2.1 Les hypothèses	21
3.2.2 Résultats pour les deux critères de choix global et local	23
3.2.3 Application sur des données simulés	33

Introduction

La statistique non paramétrique connaît un grand essor chez de nombreux auteurs et dans différents domaines. En effet, celle-ci possède un champ d'application très large permettant, ainsi, l'explication de certains phénomènes mal modélisés jusqu'à présent, tels que les séries chronologiques, et prédire les réalisations futures. Il faut mentionner, par ailleurs, que les progrès atteints dans les procédés de recueil de données ont permis d'offrir la possibilité aux statisticiens de disposer de plus en plus souvent d'observations de variables dites fonctionnelles, c'est-à-dire de courbes. Ces données sont modélisées comme étant des réalisations d'une variable aléatoire prenant ses valeurs dans un espace abstrait de dimension éventuellement infinie. Dans ce travail, nous nous intéressons à l'estimation non paramétrique de la densité conditionnelle, pour des variables aléatoires fonctionnelles. Dans ce but notre travail est présenté de la manière suivante : Un chapitre introductif, qui présente une littérature des données fonctionnelles et quelques définitions liés à la statistique non paramétrique. Dans le chapitre 2, nous exposons des résultats asymptotiques de l'estimateur à noyau de la densité conditionnelle quand les variables sont réelles. Le chapitre 3, présente l'essentielle de notre travail où on propose deux critères de choix (un choix global et un autre local) automatique des paramètres de lissage afin de rendre efficace notre estimateur. Ce chapitre sera illustré par un exemple pour comparer ces deux critères de choix.

Chapitre 1

Une littérature sur l'analyse des données fonctionnelles

Au cours de ces dernières années, la branche de la statistique consacrée à l'analyse des données fonctionnelles a connu un réel essor tant en termes des développements théoriques et méthodologiques que de la diversification des domaines d'application. Ceci revient aux progrès qu'a connus l'outil informatique au niveau des capacités de stockage qui permettent d'enregistrer des données de plus en plus volumineuses. Ainsi, un très grand nombre de variables peuvent être observées pour l'étude d'un même phénomène. Une fois la réalité des variables fonctionnelles présentée, on s'intéresse aux aspects de modélisation les concernant. Dans ce but, nous donnons quelques définitions permettant de fixer un vocabulaire. Rappelons, tout d'abord, qu'une variable aléatoire fonctionnelle est tout simplement une variable aléatoire à valeurs dans un espace de dimension éventuellement infinie que nous noterons \mathcal{F} . Par exemple, cet espace \mathcal{F} peut être un espace de fonctions, d'opérateur linéaires, ... Selon la terminologie en vigueur dans la littérature, on parle aussi bien de variables aléatoires fonctionnelles que de données fonctionnelles, ce qui englobe notamment tout ce qui concerne l'analyse statistique de courbes.

Définition 1.0.1 *On appelle modèle fonctionnel, tout modèle prenant en compte au moins une variable aléatoire fonctionnelle (v.a.f).*

Définition 1.0.2 *Un modèle fonctionnel est dit paramétrique si C est indexable par un nombre fini de paramètres appartenant à \mathcal{F} , où C n'est qu'un sous-ensemble de $\mathbb{F}_{\mathcal{F}}^{\mathcal{F}'}$ ($\mathbb{F}_{\mathcal{F}}^{\mathcal{F}'}$ l'ensemble des fonctions définies sur l'espace fonctionnel \mathcal{F} et à valeurs dans l'espace \mathcal{F}'). Un modèle fonctionnel est dit non-paramétrique dans le cas contraire*

De très nombreux travaux ont été dédiés à l'étude de modèles impliquant des variables aléatoires multivariées et c'est un domaine de la statistique qui 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 ne sont que quelques exemples illustrant le grand nombre et la diversité des données de nature fonctionnelle auxquelles le statisticien peut être confronté. C'est une des raisons pour lesquelles un nouveau champ de la statistique, dédié à l'étude de données fonctionnelles, a suscité un fort engouement au début des années quatre-vingt, sous l'impulsion, notamment, des travaux de Grenander (1981), Dauxois et al. (1982) et Ramsay (1982). Il a été popularisé par Ramsay et Silverman (1997), puis par les différents ouvrages de Bosq (2000), Ramsay et Silverman (2002), Ramsay et Silverman (2005), et Ferraty et Vieu (2006). C'est un domaine de la Statistique en plein essor comme en témoignent les hors-séries qui lui sont ou seront consacrés dans des revues reconnues comme *Statistica Sinica* (2004), *Computational Statistics, and Data Analysis* (2006), *Computational Statistics* (2007), *Journal of Multivariate Analysis* (2008). De plus, même si les données dont dispose le statisticien ne sont pas de nature fonctionnelle, il peut être amené à étudier des variables fonctionnelles construites à partir de son échantillon initial.

1.1 Risque quadratique moyen (Mean Square error)

Soit $x_0 \in \mathbb{R}$ fixé. On définit le risque quadratique moyen de \widehat{f}_n au point x_0 par

$$MSE = MSE(x_0) = [(\widehat{f}_n(x_0) - f(x_0))^2],$$

Où \mathbf{E}_f désigne l'espérance par rapport à la loi de (X_1, \dots, X_n) :

$$\mathbf{E}_f[(\widehat{f}_n(x_0) - f(x_0))^2] = \int \dots \int (\widehat{f}_n(x_0, x_1, \dots, x_n) - f(x_0))^2 \prod_{i=1}^n [f(x_i) dx_i]$$

Il est évident que

$$MSE = b^2(x_0) + \sigma^2(x_0),$$

Où

$$b(x_0) = \mathbf{E}_f[\widehat{f}_n(x_0)] - f(x_0)$$

Et

$$\sigma^2(x_0) = \mathbf{E}_f\left[\left(\widehat{f}_n(x_0) - E_f[\widehat{f}_n(x_0)]\right)^2\right]$$

Remarque 1.1.1 Les quantités $b(x_0)$ et $\sigma^2(x_0)$ sont appelées respectivement biais et variance de l'estimateur \widehat{f}_n au point x_0 .

1.2 Risque quadratique moyenne intégré (Mean integrated squared error)

Nous avons précédemment présenter le risque quadratique moyen de l'estimateur à noyau de la densité, il est également intéressant d'évaluer le risque global de \widehat{f}_n , pour cela on introduit le risque quadratique intégré da \widehat{f}_n :

$$MISE = \mathbf{E}_p \int (\widehat{f}_n(x) - f(x))^2 dx,$$

Où MISE signifie en anglais "Mean Integrated Squared Error". D'après le théorème de Tonelli-Fubini on voit que

$$MISE = \int MSE(x)dx = \int b^2(x)dx + \int \sigma^2(x)dx. \quad (1.1)$$

Cette décomposition du risque quadratique intégré en somme de deux termes, le terme de biais et le terme de variance.

1.3 La validation croisée

La validation croisée(" cross-validation ") est une méthode d'estimation de fiabilité d'un modèle fondé sur une technique d'échantillonnage. En fait, il y a au moins trois techniques

de validation croisée : " holdout method ", " k -fold cross-validation " et " leave-one-out cross validation " (LOOCV).

- La première méthode est très simple, il suffit de diviser l'échantillon de taille n en échantillon d'apprentissage ($> 60\%$ de l'échantillon) et échantillon de test. Le modèle est bâti sur l'échantillon d'apprentissage et validé sur l'échantillon de test. L'erreur est estimée en calculant l'erreur quadratique moyenne.

- Dans la seconde, on divise k fois l'échantillon, puis on sélectionne un des k échantillons comme ensemble de validation et les $(k - 1)$ autres échantillons constitueront l'ensemble d'apprentissage. On calcule comme dans la première méthode l'erreur quadratique moyenne. Puis on répète l'opération en sélectionnant un autre échantillon de validation parmi les $(k - 1)$ échantillons qui n'ont pas encore été utilisés pour la validation du modèle. L'opération se répète ainsi k fois pour qu'en fin de compte chaque sous-échantillon ait été utilisé exactement une fois comme ensemble de validation. La moyenne des k erreurs quadratiques moyennes est enfin calculée pour estimer l'erreur de prédiction.

- La troisième méthode est un cas particulier de la deuxième méthode où $k = n$, c'est-à-dire que l'on apprend sur $(n - 1)$ observations puis on valide le modèle sur la n ème observation et l'on répète cette opération n fois .

Chapitre 2

Le choix optimal de la largeur de fenêtre (cadre réel)

Dans ce chapitre, on présente des résultats d'optimalité sur le choix du paramètre de lissage basé sur la méthode de validation croisée lorsqu'on estime la densité conditionnelle de Y sachant X par la méthode de noyau dans le cas réel.

2.1 Introduction

Soit $(X_1, Y_1), \dots, (X_n, Y_n)$ un n -échantillon i.i.d d'un couple aléatoire (X, Y) sur \mathbb{R}^2 de densité g et soit f la densité marginale de X . Nous nous intéressons ici au problème de l'estimation de la densité conditionnelle de Y sachant $X = x$ définie par :

$$\varphi(y/x) = \begin{cases} \frac{g(x, y)}{f(x)} & \text{si } f(x) \neq 0 \\ 0 & \text{sinon.} \end{cases}$$

à partir de $(X_1, Y_1), \dots, (X_n, Y_n)$. Estimer la densité conditionnelle de Y sachant X peut aider à construire un prédicteur de Y_{n+1} sachant X_{n+1} . Le problème de la prévision de Y_{n+1} sachant X_{n+1} est usuellement traité en utilisant un estimateur de l'espérance conditionnelle $\mathbb{E}[Y/X = x]$, (voir pour ce sujet, Watson (1964), Collomb (1976) Härdle et Marron (1985), Györfi et al.(1989),...) Mais dans certains cas (lorsque la densité conditionnelle n'est pas unimodale par exemple) il est préférable d'utiliser un estimateur d'un paramètre caractérisant la loi de Y

sachant X différent de la régression.

Dans cet ordre d'idée, Collomb et al.(1987) utilise un estimateur du densité conditionnel (qui se déduit de manière naturelle de l'estimateur de la densité conditionnelle) pour prévoir Y dans le cas où la distribution comporte plusieurs pics strictement plus important que les autres. Gannoun (1990) et Martins (1992) utilisent un estimateur de la médiane conditionnelle pour prévoir Y . L'estimateur à noyau de φ que nous allons étudier dans ce travail est défini par :

$$\varphi_h(y/x) = \begin{cases} \frac{g_h(x, y)}{f_h(x)} & \text{si } f_h(x) \neq 0 \\ 0 & \text{sinon,} \end{cases} \quad (2.1)$$

où g_h, f_h sont les estimateurs respectifs de Parzen-Rosenblatt (1962, 1956) de g et de f définis par :

$$\begin{cases} g_h(x, y) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) K\left(\frac{y - Y_i}{h}\right), \\ f_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \end{cases} \quad (2.2)$$

K étant une application de \mathbb{R} dans \mathbb{R} et $h = h(x) \in \mathbb{R}_*^+$ la largeur de fenêtre. Certaines propriétés de convergence de l'estimateur φ_h se déduisent de manière naturelle des propriétés des estimateurs de densités g_h et f_h . Citons les articles de Roussas (1967), Rosenblatt (1969), Delecoix (1975), pour des résultats de convergence ponctuelle et la normalité asymptotique, Ioannides et Roussas (1987), Carbon (1988) pour les propriétés de convergence uniforme presque sûre et Silveira (1991) pour les propriétés de convergence L^1 .

Comme pour la plupart des estimateurs fonctionnels, l'utilisation de φ_h nécessite le choix d'un paramètre de lissage h et ce choix est d'une importance capitale pour sa performance. En effet considérons par exemple la mesure d'erreur quadratique intégrée définie par :

$$ISE(\varphi_h, \varphi) = d_I(\varphi_h, \varphi) = \int \int (\varphi_h(y/x) - \varphi(y/x))^2 W(x) W'(y) f(x) dx dy \quad (2.3)$$

où W et W' sont des fonctions de poids positives. Il est démontré dans [Vieu, 1991] que sous certaines hypothèses de régularité on a

$$ISE(\varphi_h, \varphi) = \frac{C_1}{nh^2} + C'_1 h^4 + o\left(\frac{1}{nh^2} + h^4\right), \quad (2.4)$$

où C_1 et C'_1 sont deux constantes dépendant des fonctions f et g . La formule (2.4) nous indique que dans le but de minimiser $ISE(\varphi_h, \varphi)$, la largeur de fenêtre ne doit pas être choisie trop grande puisque dans ce cas elle augmentera la composante proportionnelle à h^4 , ni trop petite car elle augmentera la composante proportionnelle à h^{-2} . La fenêtre optimale théorique est celle qui minimise les deux premiers termes de $ISE(\varphi_h, \varphi)$ c'est-à-dire

$$h_0 = \left[\frac{2C'_1}{C_1} \right]^{-\frac{1}{6}} n^{-\frac{1}{6}} \quad (2.5)$$

Mais cette fenêtre n'est pas calculable dans des applications pratiques car elle dépend des quantités C_1 et C'_1 qui dépendent à leur tour de la fonction inconnue φ . Nous avons proposé une méthode pour sélectionner une fenêtre qui soit asymptotiquement "aussi bonne" que h_0 . Cette méthode s'inspire des idées de validation croisée qui ont été proposées dans d'autres problèmes d'estimation fonctionnelle, citons en exemple [Marron, 1987] pour la densité.

2.2 Le critère de selection des paramètres de lissage

la règle de sélection que nous proposons ici est motivée par les considérations intuitives suivantes :

Soit h_0 la largeur da fenêtre optimal, c'est-à-dire celle qui minimise les 2 premiers termes de (2.4). Cette fenêtre n'est pas calculable dans la pratique car elle dépend des fonctions inconnues f et g . Nous allons chercher "à l'approximer". Pour cela nous décomposons $ISE(\varphi_h, \varphi)$ comme suit :

$$ISE(\varphi_h, \varphi) = A + B - 2C \quad (2.6)$$

où

$$\begin{aligned} A &= \int \int \varphi_h^2(y/x) W(x) W'(y) f(x) dx dy, \\ B &= \int \int \varphi^2(y/x) W(x) W'(y) f(x) dx dy, \\ C &= \int \int \varphi_h(y/x) \varphi(y/x) W(x) W'(y) f(x) dx dy. \end{aligned}$$

Puisque B est indépendant de h , choisir h minimisant $ISE(\varphi_h, \varphi)$ revient à choisir h minimisant $A - 2C$. Puisque C s'écrit aussi

$$\begin{aligned} C &= \int \int \varphi_h(y/x) W(x) W'(y) g(x, y) dx dy, \\ &= \mathbb{E}_{(X, Y)} \left[\varphi_h(Y|X) W(X) W'(Y) \right], \end{aligned}$$

en utilisant alors l'idée de Rudemo(1982), on peut "approximer" l'intégrale par somme discrète. Ainsi C peut-être approximé par

$$\begin{aligned} \bar{C} &= \frac{1}{n} \sum_{i=1}^n \varphi_h^{-i}(Y_i|X_i) W(X_i) W'(Y_i), \\ &= \frac{1}{n} \sum_{i=1}^n \frac{g_h^{-i}(X_i, Y_i)}{f_h^{-i}(X_i)} W(X_i) W'(Y_i) \end{aligned}$$

avec

$$g_h^{-i}(x, y) = \frac{1}{(n-1)h^2} \sum_{j \neq i} K\left(\frac{x - X_j}{h}\right) K\left(\frac{y - Y_j}{h}\right) \quad (2.7)$$

$$f_h^{-i}(x) = \frac{1}{(n-1)h} \sum_{j \neq i} K\left(\frac{x - X_j}{h}\right), \quad (2.8)$$

De même, A peut-être s'écrire

$$A = \mathbb{E}_X [H_h(X) W(X)],$$

où

$$H_h(x) = \int \varphi_h^2(y/x) W'(y) dy$$

Ainsi A peut-être approximé par

$$\begin{aligned} \bar{A} &= \frac{1}{n} \sum_{i=1}^n H_h^{-i}(X_i) W(X_i), \\ &= \frac{1}{n} \sum_{i=1}^n \frac{g_h^{-i}(X_i, Y_i)}{f_h^{-i}(X_i)} W(X_i) W'(Y_i) \\ &= \frac{1}{n} \sum_{i=1}^n \int \frac{g_h^{-i^2}(X_i, y)}{f_h^{-i^2}(X_i)} W'(y) dy W(X_i), \end{aligned}$$

Finalement la règle de sélection de la largeur de fenêtre est obtenue en choisissant h minimisant le critère suivant :

$$\begin{aligned} CV(h) &= \frac{1}{n} \sum_{i=1}^n \int \left(\frac{g_h^{-i}(X_i, y)}{f_h^{-i}(X_i)} \right)^2 W'(y) dy W(X_i) \\ &\quad - \frac{2}{n} \sum_{i=1}^n \frac{g_h^{-i}(X_i, Y_i)}{f_h^{-i}(X_i)} W(X_i) W'(Y_i) \end{aligned}$$

2.3 Optimalité asymptotique

Afin d'étudier l'optimalité asymptotique nous introduisons les hypothèses suivantes : les fonctions de poids W et W' sont bornées à supports compacts de supports respectifs S_W , $S_{W'}$ et que S_W° et $S_{W'}^\circ$ sont non-vides. Par la suite des hypothèses sur la largeur de fenêtre, le noyau K et la loi du couple (X, Y) . La largeur de fenêtre est sélectionnée dans

$$H_n = [An^{-\frac{1}{6}-\varepsilon}, Bn^{-\frac{1}{6}+\varepsilon}], \quad 0 < \varepsilon < \frac{1}{6}, \quad 0 < A < B < \infty \quad (H.1)$$

$$\text{le noyau } K \text{ est lipschitzien d'ordre 1, à support compact, symétrique et } \int K = 1; \quad (H.2)$$

$$\text{Les fonctions } g, f, \varphi \text{ sont de classe } \mathcal{C}^2 \text{ et lipschitziennes d'ordre 1;} \quad (H.3)$$

Il existe des constantes $\gamma > 0, \Gamma > 0$ telles que

$$\begin{aligned} (i) \quad & \forall x \in S_W \quad f(x) \geq \gamma, \\ (ii) \quad & \forall (x, y) \in S_W \times S_{W'} \quad g(x, y) \geq \Gamma; \end{aligned} \quad (\text{H.4})$$

Les fonctions g, f sont bornées. (H.5)

Théorème 2.3.1 *Soit d l'une des distances d_I, d_A , Sous les hypothèses (H.1)-(H.5) on a :*

$$\frac{d(\varphi_{\hat{h}}, \varphi)}{d(\varphi_{h_0}, \varphi)} \rightarrow 1 \quad p.s \quad (2.9)$$

où

$$h_0 = \arg \inf_{h \in H_n} d(\varphi_h, \varphi),$$

et

$$\hat{h} = \arg \inf_{h \in H_n} CV(h).$$

Dans ce qui suit , on note d_A^*, d_I^* , et d_M^* sont des critères d'erreur supplémentaires

Lemme 2.3.1 *Sous les hypothèses (H.1)-(H.5) on a*

$$(i) \quad \sup_{h \in H_n} \left| \frac{d_A^*(\varphi_h, \varphi) - d_M^*(\varphi_h, \varphi)}{d_M^*(\varphi_h, \varphi)} \right| \rightarrow 0 \quad p.s$$

$$(ii) \quad \sup_{h \in H_n} \left| \frac{d_I^*(\varphi_h, \varphi) - d_M^*(\varphi_h, \varphi)}{d_M^*(\varphi_h, \varphi)} \right| \rightarrow 0 \quad p.s$$

Nous considérons par ailleurs le critère d'erreur,

$$\bar{d}_A(\varphi_h, \varphi) = \frac{1}{n} \sum_{i=1}^n \left(\frac{g_h^{-i}(X_i, Y_i)}{f_h^{-i}(X_i)} - \frac{g(X_i, Y_i)}{f(X_i)} \right)^2 \frac{W(X_i)W'(Y_i)f(X_i)}{g(X_i, Y_i)} \quad (2.10)$$

On a le résultat suivant.

Lemme 2.3.2 *Sous les hypothèses (H.1)-(H.5) on a*

$$\sup_{h \in H_n} \left| \frac{\bar{d}_A(\varphi_h, \varphi) - d_M^*(\varphi_h, \varphi)}{d_M^*(\varphi_h, \varphi)} \right| \rightarrow 0 \quad p.s$$

Nous rappelons dans la proposition (2.3.1) ci-dessous des résultats de convergence uniforme qui seront utiles par la suite.

Proposition 2.3.1 *(voir. Härdle et Marron (1985) lemme (1) et (5.5))*

Sous les hypothèses (H.1)-(H.3) on a

(i) *Pour tout compact S de \mathbb{R}^2 ,*

$$\sup_{h \in H_n} \sup_{(x,y) \in S} |g_h(x, y) - g(x, y)| \rightarrow 0 \quad p.s,$$

$$\sup_{i=1, \dots, n} \sup_{h \in H_n} \sup_{(x,y) \in S} |g_h^{-i}(x, y) - g(x, y)| \rightarrow 0 \quad p.s.$$

(ii) *Pour tout compact S' de \mathbb{R} ,*

$$\sup_{h \in H_n} \sup_{x \in S'} |f_h(x) - f(x)| \rightarrow 0 \quad p.s,$$

$$\sup_{i=1, \dots, n} \sup_{h \in H_n} \sup_{x \in S'} |f_h^{-i}(x) - f(x)| \rightarrow 0 \quad p.s.$$

Proposition 2.3.2 *(voir. Härdle et Vieu (1991) lemme 8)*

Sous les hypothèses du théorème (2.3.1) il existe des constantes positives C_2, C'_2 telles que,

$$\bar{d}_A(f_h, f) = \frac{C_2}{nh} + C'_2 h^4 + o\left(\frac{1}{nh} + h^4\right) \quad p.s.$$

Une conséquence de la proposition (2.3.2) est donnée dans le corollaire suivant.

Corollaire 2.3.1 *Sous les hypothèses du théorème (2.3.1) on a :*

$$\sup_{h \in H_n} \frac{\bar{d}_A(f_h, f)}{d_M^*(\varphi_h, \varphi)} \leq C.$$

Lemme 2.3.3 *Sous les hypothèses du théorème (2.3.1) on a :*

$$(i) \quad \sup_{i=1,\dots,n} \sup_{h \in H_n} \sup_{(x,y) \in S_W \times S_{W'}} \left| \frac{g_h^{-i}(x,y) - g_{h^*}^{-i}(x,y)}{d_M^*(\varphi_h, \varphi)} \right| \rightarrow 0 \quad p.s.,$$

$$(ii) \quad \sup_{i=1,\dots,n} \sup_{h \in H_n} \sup_{x \in S_W} \left| \frac{f_h^{-i}(x) - f_{h^*}^{-i}(x)}{d_M^*(\varphi_h, \varphi)} \right| \rightarrow 0 \quad p.s.$$

Remarque 2.3.1 *Pour le détail des démonstrations de ces résultats , consulter (la thèse de Youndjé [19])*

Chapitre 3

Estimateur à noyau de la densité conditionnelle (données fonctionnelles)

L'estimation de la densité conditionnelle est une technique statistique qui permet une meilleure compréhension de la relation entre une variable réponse et un ensemble de variables on comparaison avec des méthodes de régression usuelles. Par conséquent, cette technique est d'une grande importance dans de nombreux domaines scientifiques où les connaissances sur les moyens conditionnelles, obtenus par des méthodes de régression ne sont pas suffisant pour tirer des conclusions valables sur le problème. En outre, la fonction de la densité conditionnelle intervient dans de nombreux domaines. Une des applications les plus utiles est la prévision d'une série temporaires, tel que le taux d'inflation. Rappelons que, si $g(x, y)$ désigne la densité conjointe de (X, Y) et $h(x)$ désigne la densité marginale de X , alors la densité conditionnelle de Y sachant $X = x$ est obtenue par $f(x, y) = g(x, y) / h(x)$. Notons également que l'estimation de la densité conditionnelle est, à certains égards, une généralisation à la fois de la régression non paramétrique et l'estimation de la densité univariée standard. L'estimateur à noyau de la densité conditionnelle a été considéré par Rosenblatt (1969) qui a étudié la problème de l'estimation de la densité de Y sachant $X = x$, où X est un variable aléatoire unidimensionnel. En outre, le problème de l'estimation de densité conditionnelle semble s'être ignoré jusqu'à ce qu'il soit revu avec certains estimateurs améliorés (cf. Hyndman et al. (1996)). En effet, la modification suivante de la forme de l'estimateur de Rosenblatt a été pris en compte :

$$\widehat{f}_{(a,b)}(x, y) = \frac{b^{-1} \sum_{j=1}^n K(a^{-1}||x - X_j||_x) K(b^{-1}||y - Y_j||_y)}{\sum_{j=1}^n K(a^{-1}||x - X_j||_x)} \quad (3.1)$$

où $(X_1, Y_1), \dots, (X_n, Y_n)$ un échantillon d'observations indépendantes de la distribution de (X, Y) et $\|\cdot\|_x$ et $\|\cdot\|_y$ sont des métriques dans l'espace des valeurs de X et Y , respectivement.

Le noyau $K(u)$ est supposé satisfaire certaines conditions. Le choix de $K(u)$ est défini en termes de la fonction de densité de probabilité univariée et unimodale. Par ailleurs, Youndjé (1993 et 1996), Hyndman et al. (1996) et d'autres donnent le biais, variance et l'erreur quadratique (MSE) et quelques propriétés de la convergence de l'estimateur et proposent également un estimateur de noyau alternatif avec une erreur quadratique moyenne assez petite de l'estimateur standard. D'autre part, nous ne pouvons pas continuer sans citer le travail de Fan et al. (1996) qui a proposé un estimateur de la densité conditionnelle en généralisant l'estimateur de Rosenblatt et en utilisant des techniques de polynômes locaux. Ensuite, Hyndman et Yao (1998) ont introduit deux autres estimateurs paramétriques locaux qui améliorent les estimateurs donnés par Fan et al. (1996). Stone (1994), quant à lui, a suivi une voie différentes en utilisant des produits tensoriels de splines polynomiales pour obtenir des estimateurs log de la densité conditionnelles. Pour d'autres études sur l'estimation non paramétrique de la densité conditionnelle nous pouvons voir aussi Gannoun (1990), Youndjé (1993 et 1996), Hall et al. (1999), Härdle et al. (1991), Bashtannyk et Hyndman (2001), Gannoun et al. (2003). Dans ce travail nous nous intéressons à l'estimation non paramétrique de la densité conditionnelle pour une variable explicative fonctionnelle conditionné par une variable réponse prend ces valeurs dans un espace semi-métrique. Cet estimateur dépend d'un paramètre de lissage qui contrôle son comportement asymptotique est qui est le centre d'intérêt de ce travail de mémoire. En premier lieu nous commençons par la construction d'une procédure de choix automatique global et local basé sur la technique de la validation croisée. Par la suite, nous montrons à travers une étude théorique et un exemple de simulation que le choix local est plus précis que le choix global.

3.1 Critère de choix local et global pour la sélection de paramètres de lissage

soient $(X_i, Y_i)_{1 \leq i \leq n}$ une suite des observations i.i.d de (X, Y) à valeurs dans $\mathcal{F} \times \mathbb{R}$ où (\mathcal{F}, d) est un espace semi-métrique équipé de la semi métrique.

Nous supposons qu'il existe une version régulière de la probabilité conditionnelle de Y sachant

X , absolument continue par rapport à la mesure de Lebesgue dans \mathbb{R} . soit $f(x, \cdot)$ la densité de la probabilité conditionnelle de Y sachant $X = x \in \mathcal{F}$, que nous allons estimer. Pour cela, on définit l'estimateur à noyau $\hat{f}_{(a,b)}$ de f comme (1), en considérons deux noyaux différents :

$$\forall x \in \mathcal{F} \text{ et } \forall y \in \mathbb{R}, \hat{f}_{(a,b)}(x, y) = \frac{b^{-1} \sum_{i=1}^n K(a^{-1}d(x, X_i))H(b^{-1}(y - Y_i))}{\sum_{i=1}^n K(a^{-1}d(x, X_i))} \quad (3.2)$$

où K est un noyau et $a = a_{K,n}$ (resp, $b = b_{H,n}$) est une suite de nombres réels positifs. Notons que l'estimateur (3.2) a été utilisé par Roussas (1968) dans le cas réel et par Ferraty et al.(2006) dans le cas fonctionnel. Notre but, est le choix automatique des paramètres de lissage (a, b) dans l'estimation non paramétrique de la densité conditionnelle. Pour cela, nous proposons une règle de choix automatique basé sur la technique de validation croisé classique "leave-out-one" et nous étudions l'optimalité asymptotique des deux critères dans le sens moyenne carré. En effet, avec la majorité des travaux antérieur sur la sélection du paramètre de lissage, notre règle est basée sur la minimisation de l'erreur quadratique intégré qui est mesuré par la mesure de probabilité, $dP_X(\mathbf{x})$, de la variable fonctionnelle \mathbf{X} et quelques fonctions de poids non-négatives W_1 et W_2 :

$$d_1(\hat{f}_{(a,b)}, f) = \int \int \left(\hat{f}_{(a,b)}(x, y) - f(x, y) \right)^2 W_1(x)W_2(y) dP_X(x)dy \quad (3.3)$$

l'approximation discrète de (3,2) est l'erreur quadratique moyenne donnée par :

$$d_2(\hat{f}_{(a,b)}, f) = \frac{1}{n} \sum_{i=1}^n \left(\hat{f}_{(a,b)}(X_i, Y_i) - f(X_i, Y_i) \right)^2 \frac{W_1(X_i)W_2(Y_i)}{f(X_i, Y_i)} \quad (3.4)$$

ou, aussi, l'erreur quadratique moyenne intégrée (MISE) :

$$d_3(\hat{f}_{(a,b)}, f) = \int \int \mathbb{E} \left(\hat{f}_{(a,b)}(x, y) - f(x, y) \right)^2 W_1(x)W_2(y) dP_X(x)dy \quad (3.5)$$

Cependant, ces fonctions de poids dépendent de la densité conditionnelle f , aussi le paramètre de lissage qui minimise ces erreurs n'est pas calculable dans la pratique. Ainsi, nous devons

trouver d'autres fonctions de poids qui sont asymptotiquement équivalentes aux distances quadratiques (3.3), (3.4) et (3.5). Suivant les mêmes idées de Youndjé (1996), nous pouvons écrire :

$$d_1(\widehat{f}_{(a,b)}, f) = A + B - 2C$$

où

$$\begin{aligned} A &= \int \int \widehat{f}_{(a,b)}^2(x, y) W_1(x) W_2(y) dP_X(x) dy, \\ B &= \int \int \mathbf{f}^2(x, y) W_1(x) W_2(y) dP_X(x) dy, \\ C &= \int \int \widehat{f}_{(a,b)}(x, y) f(x, y) W_1(x) W_2(y) dP_X(x) dy. \end{aligned}$$

Puisque le deuxième terme B est indépendant de (a, b) , le problème de minimiser le d_1 est équivalent à minimiser $A - 2C$. Une manière simple de construire une procédure de sélection automatique des largeurs de fenêtre optimales (a, b) par rapport à la mesure d'erreur d_1 est d'estimer les quantités A et C . Pour ce but, comme mentionné ci-dessus, nous adoptons la technique standard "leave one-out" comme dans Rudemo (1982) pour l'évaluation de la densité de probabilité et [Rachdi et Vieu (2007)] pour l'évaluation de l'opérateur de régression, en considérant le critères suivant :

$$\begin{aligned} GCV(a, b) &= \frac{1}{n} \sum_{i=1}^n W_1(X_i) \int \widehat{f}_{(a,b)}^{-i^2}(X_i, y) W_2(y) dy \\ &\quad - \frac{2}{n} \sum_{i=1}^n \widehat{f}_{(a,b)}^{-i}(X_i, Y_i) W_1(X_i) W_2(Y_i) \end{aligned} \quad (3.6)$$

et respectivement, pour un $y \in \mathbb{R}$:

$$\begin{aligned} LCV_{x,y}(a, b) &= \frac{1}{n} \sum_{i=1}^n W_{1,x}(X_i) \int \widehat{f}_{(a,b)}^{-i^2}(X_i, z) W_{2,y}(z) dz \\ &\quad - \frac{2}{n} \sum_{i=1}^n \widehat{f}_{(a,b)}^{-i}(X_i, Y_i) W_{1,x}(X_i) W_{2,y}(Y_i) \end{aligned} \quad (3.7)$$

où $W_{2,x}$ (respectivement $W_{2,y}$) est une fonction de poids locale positive autour de x (respectivement y), et pour tout $i = 1, \dots, n$:

$$\widehat{f}_{(a,b)}^{-i}(x, y) = \frac{b^{-1} \sum_{j \neq i}^n K(a^{-1}d(x, X_j))H(b^{-1}(y - Y_j))}{\sum_{j \neq i}^n K(a^{-1}d(x, X_j))} \quad (3.8)$$

Ces critères sont obtenues en utilisant le développement suivant :

$$\begin{aligned} C &= \int \int \widehat{f}_{(a,b)}(x, y) f(x, y) W_1(x) W_2(y) dP_X(x) dy \\ &= \int \int \widehat{f}_{(a,b)}(x, y) W_1(x) W_2(y) dP_{Y|X=x}(y) dP_X(x). \\ &= \int \int \widehat{f}_{(a,b)}(x, y) W_1(x) W_2(y) dP_{(X,Y)}(x, y) \\ &= \mathbb{E}_{(X,Y)} \left(\widehat{f}_{(a,b)}(X, Y) W_1(X) W_2(Y) \right) \end{aligned}$$

et

$$A = \mathbb{E}_X \left(\int \widehat{f}_{(a,b)}^2(X, y) W_1(X) W_2(y) dy \right)$$

où \mathbb{E}_Z dénote l'espérance par rapport à la distribution de la variable aléatoire Z .

Finalement, procédure globale (respectivement, local) de validation croisée consiste en choix de la largeur de fenêtre (a, b) qui réduisent au minimum $GCV(a, b)$ (respectivement, $LCV_{x,y}(a, b)$) sur un ensemble donné $H_n \subset \mathbb{R}^{+2}$ (respectivement, $H_n(x, y) \subset \mathbb{R}^{+2}$).

3.2 Résultats principaux

3.2.1 Les hypothèses

Pour déduire l'optimalité asymptotique des paramètres de lissage choisis par la règle GCV (respectivement, $LCV_{x,y}$), nous supposons que la fonction de poids W_1 (respectivement W_2) est bornée avec support dans un sous-ensemble S_X de \mathcal{F} (respectivement sur un sous-ensemble compact S_Y de \mathbb{R}) et la densité conditionnelle $f(\cdot, \cdot)$ est bornée sur $S_X \times S_Y$. Dans la suite, nous noterons par C et C' deux constantes strictement positives et nous introduisons les hypothèses

suivantes :

Les fonctions de poids sont prises, pour chaque courbe x , tel que pour un certain w réel positif :

$$w = a^\nu \text{ pour } 0 < \nu < 1 \text{ et } W_{1,x} \text{ est li et supported a } \mathcal{B}(x, w) \quad (3.9)$$

où $\mathcal{B}(x, h)$ dénote la boule fermée de centre x et de rayon $h > 0$,

$$\forall x \in S_X, 0 < C\phi(h) \leq \mathbb{P}(X \in \mathcal{B}(x, h)) \leq C'\phi(h) \quad (3.10)$$

où $\phi(h)$ est une fonction réel positive tels que $\lim_{h \rightarrow 0} \phi(h) = 0$.

il existe quelques constantes strictement positives b_1, b_2 et δ , tels que :

$\forall (x_0, y_0) \in S_X \times S_Y, \forall (x_1, x_2) \in S_X \times S_X, \text{ et } \forall (y_1, y_2) \in S_Y \times S_Y$, nous avons :

$$f(x_0, y_0) > \delta \text{ et } |f(x_1, y_1) - f(x_2, y_2)| \leq C \left(d^{\mathbf{d}_1}(x_1, x_2) + |y_1, y_2|^{b_2} \right) \quad (3.11)$$

Le noyau K est borné et Lipschitzien dans son support $(0, 1)$, il existe quelques constantes positives C et C' tels que :

$$0 < C < K(t) < C' < \infty \quad (3.12)$$

et si $K(1) = 0$, le noyau K doit remplir la condition supplémentaire $-\infty < C < K'(t) < C' < 0$, où K' est la première dérivée du K .

Le noyau H est borné et continûment lipschitzienne, tel que :

$$\int |t|^{b_2} H(t) dt < \infty \text{ et } \int H^2(t) dt < \infty \quad (3.13)$$

la fonction ϕ est telle que :

$$\left\{ \begin{array}{l} \exists C > 0, \exists \eta_0 > 0, \forall \eta < \eta_0, \phi'(\eta) < C, \\ \text{et si } K(1) = 0, \text{ la fonction } \phi(\cdot) \text{ doit to fulfill la condition supplémentaire :} \\ \exists C > 0, \exists \eta_0 > 0, \forall_0 < \eta < \eta_0, \int_0^\eta \phi(u) du > C\eta\phi(\eta), \end{array} \right. \quad (3.14)$$

Pour n assez grand, ϵ -entropy de Kolmogorov de S_X dénoté par le ψ_{S_X} (cf. par exemple, Kolmogorov et Tikhomiros (1959) et Theodoros et Yannis (1997)) satisfait, pour certains γ dans $(0, 1)$:

$$\sum_{n=1}^{\infty} n^{(3\gamma+1)/2} \exp \left\{ (1-\beta) \psi_{S_X} \left(\frac{\log n}{n} \right) \right\} < \infty \text{ pour quelques } \beta > 1 \quad (3.15)$$

et pour tous $(a, b) \in H_n$ nous avons :

$$\lim_{n \rightarrow +\infty} n^\gamma b = \infty \quad \text{et} \quad \phi(a) \geq C n^{-\delta} \text{ for some } \delta \in (0, 2 - 2\gamma) \quad (3.16)$$

3.2.2 Résultats pour les deux critères de choix global et local

Théorème 3.2.1 *Sous les hypothèses (3.8) - (3.15), si l'ensemble H_n des paramètres de lissage (a, b) est fini avec :*

$$\#(H_n) = O(n^\alpha) \text{ pour un certain } \alpha > 0, \text{ o } \# \text{ denote la cardinal} \quad (3.17)$$

nous avons, pour $k = 1, 2, 3$, que :

$$\frac{d_k(\widehat{f}_{(a_1, b_1)}, \mathbf{f})}{d_k(\widehat{f}_{(a_0, b_0)}, \mathbf{f})} \rightarrow 1 \text{ presque srement (p.s.) quand } n \rightarrow +\infty \quad (3.18)$$

où

$$(a_0, b_0) = (a_{0_{K,n}}, b_{0_{H,n}}) = \arg \inf_{(a,b) \in H_n} d_k(\widehat{f}_{(a,b)}, \mathbf{f})$$

et

$$(a_1, b_1) = (a_{1_{K,n}}, b_{1_{H,n}}) = \arg \inf_{(a,b) \in H_n} GCV(a, b)$$

dans le context local, nous supposons que (3.13) est vérifié pour $S_X = \mathbf{B}(x, w)$ et nous déduisent les mêmes résultats d'optimalité, pour le critère local.

Proof of theorem 3.2.1.

On considère la décomposition suivante ,pour tout $(x, y) \in S_X \times S_Y$, et pour tout $(a, b) \in H_n$:

$$\begin{aligned} \left(\widehat{f}_{(a,b)}(x, y) - f(x, y)\right)^2 &= \left(\widehat{f}_N(x, y) - f(x, y)\widehat{f}_D(x)\right)^2 \\ &+ 2\left(1 - \widehat{f}_D(x)\right)\widehat{f}_D(x)\left(\widehat{f}_{(a,b)}(x, y) - f(x, y)\right)^2 \\ &+ \left(1 - \widehat{f}_D(x)\right)^2\left(\widehat{f}_{(a,b)}(x, y) - f(x, y)\right)^2 \end{aligned}$$

où

$$\widehat{f}_D(x) = \frac{1}{n\mathbb{E}[K(a^{-1}d(x, X))]} \sum_{i=1}^n K(a^{-1}d(x, X_i)) \quad (D_1)$$

$$\begin{aligned} \widehat{f}_N(x, y) &= \frac{1}{n\mathbb{E}[K(a^{-1}d(x, X))]} \sum_{i=1}^n K(a^{-1}d(x, X_i))H(b^{-1}(y - Y_i)) \\ &= \widehat{f}_D(x)\widehat{f}_{(a,b)}(x, y) \quad (D_2) \end{aligned}$$

De la convergence uniforme de $\widehat{f}_D(x)$ vers $\mathbb{E}\left[\widehat{f}_D(x)\right] = 1$ (voir Ferraty et al., 2010), on a :

$$\begin{aligned} \forall k = 1, 2, 3, \quad d_k(\widehat{f}_{(a,b)}(x, y) - f(x, y)) &= d_k(\widehat{f}_N(x, y), f(x, y)\widehat{f}_D(x)) \\ &+ \circ_{p.s.}(d_k(\widehat{f}_{(a,b)}(x, y) - f(x, y))) \end{aligned}$$

$$\forall k = 1, 2, 3, \quad d_k(\widehat{f}_{(a,b)}(x, y) - f(x, y)) = d_k(\widehat{f}_N(x, y), f(x, y)\widehat{f}_D(x)) + \circ_{p.s.}(d_k(\widehat{f}_{(a,b)}(x, y) - f(x, y)))$$

Il suffit par la suite de montrer le résultat réclamé pour d_k^* qui est défini par :

$$d_k^*(\widehat{f}_{(a,b)} - f) = d_k(\widehat{f}_N(x, y), f(x, y)\widehat{f}_D(x)) \text{ pour tout } k = 1, 2, 3.$$

Notons que, en suivant les mêmes étapes utilisé dans Rachdi et Vieu (2007), pour montrer les lemmes 5, 6 et 7, on notant :

$$K^*(x, X_i) = \frac{1}{n\mathbb{E}[K(a^{-1}d(x, X))]} \left(K(a^{-1}d(x, X_i))H(b^{-1}(y - Y_i)) \right. \\ \left. - bf(x, y)K(a^{-1}d(x - X_i)) \right) \quad \text{On obtient}$$

$$\sup_{h \in \hat{H}_n} \left(\frac{|d_k^*(\hat{f}_{(a,b),f}) - d_l^*(\hat{f}_{(a,b),f})|}{d_3^*(\hat{f}_{(a,b),f})} \right) \rightarrow 0 \text{ p.s, pour tout } k \neq l$$

De l'autre côté, on introduit la mesure d'erreur d_5 définie par

$$d_5(\hat{f}_{(a,b)}, f) = \frac{1}{n} \sum_{i=1}^n \left(\hat{f}_{(a,b)}^{-i}(X_i, Y_i) - f(X_i, Y_i) \right)^2 \frac{W_1(X_i)W_2(Y_i)}{f(X_i, Y_i)}$$

où

$$\hat{f}_{(a,b)}^{-i}(X_i, Y_i) = \frac{\hat{f}_N^{-i}(X_i, Y_i)}{\hat{f}_D^{-i}(X_i)}$$

Avec

$$\hat{f}_N^{-i}(x, y) = \frac{1}{nb\mathbb{E}[K(a^{-1}d(x, X))]} \sum_{j \neq i}^n K(a^{-1}d(x, X_j))H(b^{-1}(y - Y_j))$$

et

$$\hat{f}_D^{-i}(x) = \frac{1}{nb\mathbb{E}[K(a^{-1}d(x, X))]} \sum_{j \neq i}^n K(a^{-1}d(x, X_j))$$

Remarquons que

$$GCV(a, b) - d_5(\hat{f}_{(a,b)}, f) = CT(a, b) + T$$

où

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

et

$$T = \frac{1}{n} \sum_{i=1}^n f(X_i, Y_i) W_1(X_i) W_2(Y_i)$$

Donc ,la preuve de ce théorème est complet si on peut prouver que d_5 est asymptotiquement équivalent à d_3 et que :

$$\sup_{h \in H_N} \left| \frac{CT(a, b)}{d_3(\widehat{f}_{(a,b)}, f)} \right| \rightarrow 0, \text{ p.s comme } n \rightarrow +\infty$$

La démonstration du théorème 3.2.1 est achevée quand les lemmes 3.2.1, 3.2.2 et 3.2.3 sont démontrés.

Lemme 3.2.1 *Sous les hypothèses (2.9),(2.10),(2.11)et(2.13), on a*

$$d_3(\widehat{f}_{(a,b),f}) \geq C \frac{1}{nb\phi(a)}$$

Dans ce qui suit ,on note pour tout $i = 1, \dots, n$, par

$$K_i = K(a^{-1}d(x, X_i)) \text{ et } H_i(y) = H(b^{-1}(y - Y_i))$$

Preuve du lemme 3.2.1 Il est claire que :

$$d_3(\widehat{f}_{(a,b)}, f) \geq \int \int \mathbf{var} \left[\widehat{f}_{(a,b)}(x, y) \right] W_1(x) W_2(y) dP_x(x) dy$$

Par conséquent , il suffit d'évaluer le terme variance $\mathbf{var} \left[\widehat{f}_{(a,b)}(x, y) \right]$, par l'utilisation des techniques computationnelles comme dans Laksaci (2007) et du fait que, $\mathbb{E}(\widehat{f}_D(x)) = 1$,on obtient :

$$\begin{aligned} \mathbf{var} \left[\widehat{f}_{(a,b)}(x, y) \right] &= \mathbf{var} \left[\widehat{f}_N(x, y) \right] - 2\mathbb{E}(\widehat{f}_N(x, y)) \mathbf{Cov}(\widehat{f}_N(x, y), \widehat{f}_N(x)) \\ &+ \left(\mathbb{E}(\widehat{f}_N(x, y)) \right)^2 (\mathbf{var} \widehat{f}_D(x)) + o\left(\frac{1}{nb\phi(a)}\right) \end{aligned} \quad \dots(D_3)$$

D'après les définitions (D_2) et (D_3) de l'estimateur \widehat{f}_N et \widehat{f}_D ,on obtient

$$\mathbf{var} \left(\widehat{f}_N(x, y) \right) = \frac{1}{n(b\mathbb{E}[K(a^{-1}d(x, X))])^2} \mathbf{var}(K_1(x)H_1(y))$$

$$\text{Cov}(\widehat{f}_N(x, y), \widehat{f}_D(x)) = \frac{1}{n(b\mathbb{E}[K(a^{-1}d(x, X))])^2} \text{Cov}(K_1(x)H_1(y), K_1(x))$$

et

$$\text{Var}(\widehat{f}_D(x)) = \frac{1}{n(b\mathbb{E}[K(a^{-1}d(x, X))])^2} \text{Var}(K_1(x))$$

De plus, pour tout $i, j = 1, 2$, on que :

$$\mathbb{E}[K_1^i(x)H_1^j(y)] = b\mathbb{E}[K_1^i(x)f(X_1, y)] \int H^j(t)dt + o(b\mathbb{E}[K_1(x)])$$

et

$$0 < C\phi(a) \leq \mathbb{E}[K_1(x)] \leq C'\phi(a)$$

En comparant asymptotiquement les trois quantités dans, on peut voir que le premier terme est mené .D'où ,par on obtient :

$$\text{Var}[\widehat{f}_{(a,b)}(x, y)] = \frac{1}{n(b\mathbb{E}[K(a^{-1}d(x, X))])^2} \mathbb{E}[K_1^2(x)f(X_1, y)] \int H^2(t)dt + o\left(\frac{1}{nb\phi(a)}\right)$$

En outre ,en utilisant le fait que la densité conditionnelle ne disparaît pas sur un voisin de $S_X \times S_Y$,on obtient :

$$d_3(\widehat{f}_{(a,b)}, f) \geq C \frac{1}{nb\phi(a)}$$

la preuve est complète .

Lemme 3.2.2 *Sous les hypothèses (2.9),(2.10),(2.11),(2.13)et(2.16), on obtient*

$$\sup_{(a,b) \in H_n} |nb\phi(a)CT(a, b)| \rightarrow 0, p.s \text{ commen } \rightarrow \infty$$

Preuve du lemme 3.2.2 Par la définition de $CT(a, b) \in H_n$:

$$\begin{aligned} |CT(a, b)| &= \left| \frac{1}{n} \sum_{i=1}^n \widehat{f}_{(a,b)}^{-i^2}(X_i, y) W_2(y) W_1(X_i) dy - \frac{2}{n} \sum_{i=1}^n \widehat{f}_{(a,b)}^{-i^2}(X_i, Y_i) \frac{W_1(X_i) W_2(Y_i)}{f(X_i, Y_i)} \right| \\ &= \left| \sum_{i=1}^n \frac{1}{n \widehat{f}^{-i^2}(x)} \left[\int \widehat{f}_N^{-i^2}(X_i, y) W_2(y) W_1(X_i) dy - 2 \widehat{f}_N^{-i^2}(X_i, Y_i) \frac{W_1(X_i) W_2(Y_i)}{f(X_i, Y_i)} \right] \right| \end{aligned}$$

cela a combiné avec la consistance uniforme de $\widehat{f}_D^{-i}(x)$ de 1 (voir Ferraty et al., 2008), est assez prouver que , quand $n \rightarrow \infty$:

$$\sup_{(a,b) \in H_n} b\phi(a) \left| \sum_{i=1}^n \left[\int \widehat{f}_N^{-i^2}(X_i, y) W_1(X_i) dy - 2\widehat{f}_N^{-i^2}(X_i, Y_i) \frac{W_1(X_i)W_2(Y_i)}{f(X_i, Y_i)} \right] \right| \rightarrow 0, p.s.$$

pour faire ceci ,on peut utiliser des arguments similaires que dans le cas réel (voir Youndjé, 1996). En effet ,soit pour tout $1 \leq i, j, k \leq n$:

$$\begin{aligned} a_{ij} &= K_j^2(X_i)W_1(X_i) \\ b_{ij} &= \int H_j^2(y)W_2(y)dy - H_j^2(Y_i) \frac{W_2(Y_i)}{f(X_i, Y_i)} \\ U_{ij} &= a_{ij}b_{ij} \\ c_{ijk} &= K_j(X_i)K_k(X_i)W_1(X_i) \\ d_{ijk} &= \int H_j(y)H_k(y)W_2(y)dy - H_j(Y_i)H_k(Y_i) \frac{W_2(Y_i)}{f(X_i, Y_i)} \\ V_{ijk} &= c_{ijk}d_{ijk} \end{aligned}$$

Maintenant,nous avons examiné les limites suivantes

$$\sup_{(a,b) \in H_n} \left| \frac{b\phi(a)}{(n-1)^2} \sum_{i \neq j} U_{ij} \right| \rightarrow 0, p.s.$$

et

$$\sup_{(a,b) \in H_n} \left| \frac{b\phi(a)}{(n-1)^2} \sum_{i \neq j \neq k \neq i} V_{ijk} \right| \rightarrow 0, p.s.$$

la preuve des deux limites précitées suit en adoptant les mêmes étapes de la preuve du lemme (1.2.3) dans Rachdi et Vieu (2007). Remarquons que ,par le lemme de Borel-Cantelli ,il est assez de montrer qu'il $\exists \xi, \xi' > 0$ tel que pour tout $p \in \mathbb{N}^*$ il existe deux constantes C, C' afin que :

$$\mathbb{E} \left(n^{-2} b\phi(a) \sum_{i \neq j} U_{ij} \right)^{2p} \leq C n^{-\xi p} \quad (p_1)$$

et

$$\mathbb{E}\left(n^{-2}b\phi(a) \sum_{i \neq j \neq k \neq i} V_{ijk}\right)^{2p} \leq C'n^{-\xi'p} \quad (p_2)$$

pour prouvé (p₁) on a :

$$\mathbb{E}\left(n^{-2}b\phi(a) \sum_{i \neq j} U_{ij}\right)^{2p} = (n^{-2}b\phi(a))^{2p} \sum_{i_1 \neq j_1} \dots \sum_{i_{2p} \neq j_{2p}} \mathbb{E}(U_{i_1 j_1} \dots U_{i_{2p} j_{2p}})$$

pour calculer ces quantités ,nous avons montrés que :

$$\mathbb{E}(U_{i_1 j_1} \dots U_{i_{2p} j_{2p}}) = 0, \text{ si } m > 2p$$

Où m denote le cardinal de l'ensemble $\{i_1, j_1, \dots, i_{2p}, j_{2p}\}$. De l'autr côté,observons que,pour tout i, j :

$$\begin{aligned} \mathbb{E}[b_{ij}|X_1, \dots, X_n] &= \int \int b^{-2} H^2(b^{-1}(y-z)) f(X_j, z) W_2(y) dy dz \\ &- \int \int b^{-2} H^2(b^{-1}(y-z)) \frac{f(X_i, y) f(X_j, z)}{f(X_i, y)} W_2(y) dy dz \\ &= 0 \quad (p_3) \end{aligned}$$

et donc,nous pouvons écrire

$$\mathbb{E}[U_{ij}|X_1, \dots, X_n] = a_{ij} \mathbb{E}[b_{ij}|X_1, \dots, X_n] = 0$$

et

$$\mathbb{E}[U_{ij}|X_i] = \mathbb{E}[\mathbb{E}[U_{ij}|X_1, \dots, X_n]|X_i] = 0$$

Donc,si $m > 2p$ il $\exists a \in \{1, \dots, 2p\}$ tel que i_a (ou j_a) paraît seulement une fois dans $\{i_1, j_1, \dots, i_{2p}, j_{2p}\}$, par conséquent,il suffit de calculer en conditionnent respectivement X_{i_a} (ou X_{j_a} pour montrer p_3 . De l'autre côté ,pour $2 \leq m \leq 2p$ on déduit par (10) et (20) que :

$$|\mathbb{E}(U_{i_1 j_1} \dots U_{i_{2p} j_{2p}})| \leq \frac{G}{b^{4p} \phi(a)^{4p}} \mathbb{E}\left[\prod_{i,j=1}^m K^{\tau_{ij}}(a^{-1}d(X_i, Y_i)) \prod_{i=1}^m W_1^{\beta_i}(X_i)\right]$$

Où $\beta_i = 0, 1$, et $\sum_{i,j} \tau_{ij} = 2p$ et pour tout $i = 1, \dots, m$ il $\exists j$ tel que $\tau_{ij} \neq 0$ et $\beta_i = 1$ ou $\beta_j = 1$. Les arguments similaires comme ceux invoqués pour prouver le lemme 9 Rachdi et Vieu (2007) peuvent être utilisés, lesquels permettent d'obtenir que pour tout

$$m = 2, \dots, 2p, |\mathbb{E}(U_{i_1 j_1} \dots U_{i_{2p} j_{2p}})| \leq \frac{C}{b^{4p} (\phi(a))^{4p}} \phi^{m/2}(a)$$

Par conséquent, par (14) :

$$\begin{aligned} \mathbb{E} \left(n^{-2} b \phi(a) \sum_{i \neq j} U_{ij} \right)^{2p} &\leq C \sum_{M+2}^{2P} n^{m-4p} b^{-2p} (\phi(a))^{\frac{m}{2}-2p} \\ &= C \frac{1}{(nb\phi(a))^{2p}} (\phi(a))^p \sum_{m=2}^{2p} \left(n \sqrt{\phi(a)} \right)^{m-2p} \\ &\leq C \frac{1}{(nb\phi(a))^{2p}} (\phi(a))^p \\ &\leq C \frac{1}{(n^{1-\gamma})^{2p}} (\phi(a))^p \end{aligned}$$

Il suffit maintenant combiner p_3 avec l'hypothèse (16), pour avoir (16), pour avoir p_1

Concernant p_2 , on utilise des arguments analogues comme la preuve de p_1 . En dénotant m' le cardinal de l'ensemble $\{i_1, j_1, k_1, \dots, i_{2p}, j_{2p}, k_{2p}\}$ et puisque $\mathbb{E}[d_{ijk} | X_1, \dots, X_n] = 0$ on déduit que, si $m' > 3p$ on a :

$$\mathbb{E}(V_{i_1 j_1 k_1} \dots V_{i_{2p} j_{2p} k_{2p}}) = 0$$

De plus, quand $3 \leq m' \leq 3p$ la modification simple de la preuve du lemme ... dans Youndjé (1996), donne :

$$\mathbb{E}[V_{i_1 j_1 k_1} \dots V_{i_{2p} j_{2p} k_{2p}}] \leq \frac{C}{b^{4p} (a)^{4p}} (b\phi(a))^{m'/2}$$

Cela même pour écrire directement sous (2.12) :

$$\begin{aligned} \mathbb{E} \left(n^{-2} b\phi(a) \sum_{i \neq j \neq k \neq i} V_{ijk} \right)^{2p} &\leq C \sum_{m'=3}^{3P} n^{m'} b^{-4p} (b\phi(a))^{\frac{m'}{2}} (\phi(a))^{-4p} (n^{-2} b\phi(a))^{2p} \\ &= C \frac{1}{((n(b\phi(a))^{(1/2)})^p)} \sum_{m'=3}^{3p} (n(b\phi(a))^{1/2})^{m'-3p} \\ &\leq C \frac{1}{n^{(1-\gamma/2)p} \phi(a)^{(p/2)}} \end{aligned}$$

Encore une fois ,on utilise (2.14) pour compléter la preuve de ce lemme

Lemme 3.2.3 *Sous les hypothèses (2.9),(2.10),(2.11),(2.13)et(2.16),on a*

$$\sup_{(a,b) \in H_n} \left| \frac{d_5(\widehat{f}_{(a,b)}, f) - d_3^*(\widehat{f}_{(a,b)}, f)}{d_3^*(\widehat{f}_{(a,b)}, f)} \right| \rightarrow 0, p.s \text{ comme } n \rightarrow +\infty$$

Preuve du lemme 3.2.3 Par l'utilisation des mêmes arguments comme ceux utilisés au commencement du la preuve du théorème,et en intriduisant :

$$d_5^*(\widehat{f}_{(a,b)}, f) = \frac{1}{n} \sum_{i=1}^n \left(\widehat{f}_N^{-i}(X_i, Y_i) - f(X_i, Y_i) \widehat{f}_D^{-i}(X_i) \right)^2 \frac{W_1(X_i)W_2(Y_i)}{f(X_i, Y_i)}$$

on peut écrire

$$d_5(\widehat{f}_{(a,b)}, f) = d_5^*(\widehat{f}_{(a,b)}, f) + o(d_5(\widehat{f}_{(a,b)}, f))$$

Par conséquent ,à cause de l'équivalence asymptotique entre d_3 et d_2^* la preuve de ce lemme sera complété dès que nous montrons que

$$\sup_{(a,b) \in H_n} \left| \frac{d_5(\widehat{f}_{(a,b)}, f) - d_2^*(\widehat{f}_{(a,b)}, f)}{d_3^*(\widehat{f}_{(a,b)}, f)} \right| \rightarrow 0, p.s$$

TO do that , on considère la décomposition suivante

$$\begin{aligned} \left(\widehat{f}_N^{-i} - f \widehat{f}_D \right)^2 &= \left(\widehat{f}_N - \widehat{f}_D \right)^2 + 2 \left(\widehat{f}_N^{-i} - f \widehat{f}_N + f(\widehat{f}_D - \widehat{f}_D^{-i}) \right) \left(\widehat{f}_N - f \widehat{f}_D \right) \\ &+ \left(\widehat{f}_N^{-i} - \widehat{f}_N + f(\widehat{f}_D - \widehat{f}_D^{-i}) \right)^2 \end{aligned}$$

En outre ,observant que ,pour tout $i = 1, \dots, n$,on a :

$$\begin{aligned} \widehat{f}_N^{-i}(x, y) - \widehat{f}_N(x, y) &= \frac{1}{n-1} \widehat{f}_N(x, y) \\ &- \frac{1}{(n-1)b\mathbb{E}[K(a^{-1}d(x, X))]} K\left(\frac{d(x, X_i)}{a}\right) H\left(\frac{y, Y_i}{b}\right) \end{aligned}$$

et

$$\widehat{f}_D^{-i}(x) - \widehat{f}_D(x) = \frac{1}{n-1} \widehat{f}_D(x) - \frac{1}{(n-1)b\mathbb{E}[K(a^{-1}d(x, X))]} K\left(\frac{d(x, X_i)}{a}\right)$$

Depuis K est une fonction bornée ,on déduit par la consistance uniforme de \widehat{f}_D et \widehat{f}_N que

$$|\widehat{f}_N^{-i}(x, y) - \widehat{f}_N(x, y)| \leq \frac{C}{(n-1)b\mathbb{E}[K(a^{-1}d(x, X))]} \leq \frac{C}{(n-1)b\phi(a)}, p, s$$

et

$$|\widehat{f}_D^{-i}(x) - \widehat{f}_D(x)| \leq \frac{C}{(n-1)b\mathbb{E}[K(a^{-1}d(x, X))]} \leq \frac{C}{(n-1)b\phi(a)}, p, s$$

D'où ,ou obtient :

$$\begin{aligned} \left| d_5^*(\widehat{f}_{(a,b)}, f) - d_2^*(\widehat{f}_{(a,b)}, f) \right| &\leq \frac{C}{(n-1)b\phi(a)} \sup_{(x,y) \in S_X \times S_Y} \left| \widehat{f}_N(x, y) - f(x, y) \widehat{f}_D \right| \\ &+ \frac{C}{(n-1)^2 b^2 (\phi(a))^2} \end{aligned}$$

En combinant ce dernier résultat avec le lemme nous obtenons le résultat réclamé. ■

Théorème 3.2.2 *Sous les hypothèses (1.8) - (1.14), si l'ensemble $H_n(x, y)$ des largeurs de fenêtre (a, b) est fini avec :*

$$\#(H_n(x, y)) = \mathcal{O}(n^{\alpha(x,y)}) \text{ pour un certain } \alpha(x, y) > 0, \quad (3.19)$$

then, we have for $k = 1, 2, 3$, that :

$$\frac{d_k(\widehat{f}_{(a_1, b_1)}, f)}{d_k(\widehat{f}_{(a_0, b_0)}, f)} \rightarrow 1 \text{ presque srement (p.s.) comme } n \rightarrow +\infty \quad (3.20)$$

comme

$$(a_0, b_0) = (a_{0_{K,n}}, b_{0_{H,n}}) = \arg \inf_{(a,b) \in H_n(\mathbf{x}, \mathbf{y})} d_k(\widehat{f}_{(a,b)}, f)$$

et

$$(a_1, b_1) = (a_{1_{K,n}}, b_{1_{H,n}}) = \arg \inf_{(a,b) \in H_n(x,y)} LCV_{x,y}(a, b)$$

3.2.3 Application sur des données simulés

Le but principal de cette section est de montrer la performance de notre étude théorique sur un échantillon fini. Notre premier objectif est d'évaluer la performance de la méthode de lissage global et le second est de la comparer avec le lissage local. Pour ces objectifs, nous considérons que le modèle non paramétrique fonctionnelle suivante :

$$Y_i = r(X_i) + \varepsilon_i, \text{ pour } i = 1, \dots, n \quad (3.21)$$

où les ε_i sont générés de façon indépendante selon la distribution $\mathcal{N}(0, 1)$. l'échantillon des variables explicatives fonctionnelles X_i pour $i = 1, \dots, n$, qui est supposé pour être indépendant ε_i pour $i = 1, \dots, n$, est généré selon l'expression suivante :

$$X_i(t) = a_i \sin(4(b_i - t)) + b_i + \eta_{i,t}, \quad \forall t \in [0, 1] \text{ et } i = 1, 2, \dots, n$$

où b_i (respectivement, $\eta_{i,t}$) est $\mathcal{N}(0, 3)$, (respectivement, $\mathcal{N}(0, 0, 5)$), tandis que les n variables aléatoires a_i sont générées par $\mathcal{N}(4, 3)$. D'autre part, la réponse scalaire Y_i , définie par (22), est calculée par l'opérateur suivant :

$$r(x) = \int_0^1 \frac{dt}{1 + |x(t)|}$$

Signalons que, avec cette définition du modèle, la densité conditionnelle de Y sachant $X = x$ peut être explicitement définie par :

$$f(x, y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y - r(x))^2\right)$$

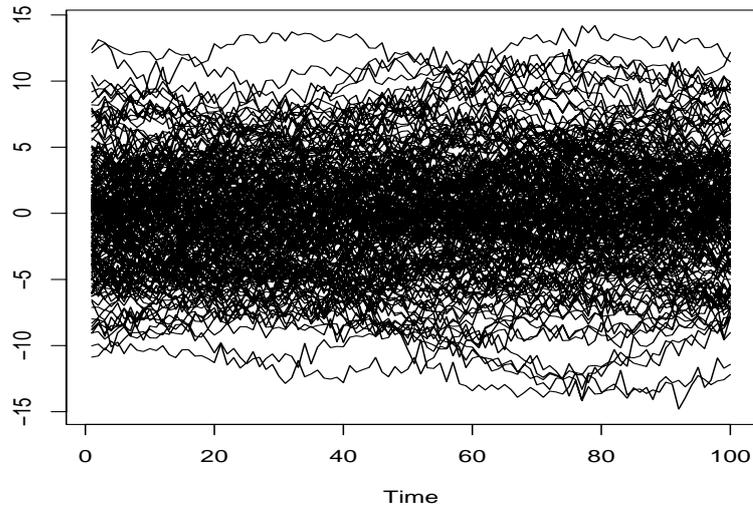


FIGURE 3.1 – A sample of 200 irregular curves

Afin de vérifier l'efficacité de cette méthode de selection globale des échantillons finis, nous comparons l'erreur quadratique moyenne (d_2) de la largeur de fenêtre globale de validation croisée (a^{GCV}, b^{GCV}) à celle donnée par le lissage global (a^{Gd_2}, b^{Gd_2}). Notant que l'utilisation de d_2 comme critère d'exactitude, est motivée par le fait qu'il est plus facile traiter erreur quadratique moyenne d'un point de vue computationnelle et elle est asymptotiquement équivalente à d_1 et à d_3 (cf. preuve de lemme 3.2.3). Pour des raisons pratiques, nous sélectionnons les paramètres (a_{GCV}, b_{GCV}) et (a^{Gd_2}, b^{Gd_2}) sur un ensemble fini H_n défini par (a_q, b_q) où a_q (respectivement, b_q) est le quantile d'ordre q du vecteur de toutes les distances entre les courbes (respectivement, entre les variables de réponses). Concernant les fonctions de poids, W_1 et W_2 , nous nous rappelons que ces fonctions sont présentées pour réduire les effets de frontière (où bord) par leurs appuis, mais le rôle de leurs expressions ne sont pas très déterminant dans la pratique comme il est précisé par Härdle and Marron (1985). Dans notre étude de simulation, nous prenons

$$W_1(t) = \begin{cases} 1 & \text{si } \min_{i=1, \dots, n} d(t, X_i) < a \\ 0 & \text{sinon} \end{cases}$$

$$W_2(z) = \begin{cases} 1 & \text{si } z \in [\min\{Y_i, i = 1, \dots, n\} \times 0.9, \max\{Y_i, i = 1, \dots, n\} \times 1.1] \\ 0 & \text{sinon} \end{cases}$$

Le noyau K (respectivement, H) est choisi pour être quadratique sur $(0, 1)$ (respectivement, sur $(-1, 1)$). Un autre aspect important laissant assurer un bon comportement de la méthode, est

n	50	100	150	200	250
$d_2(\widehat{f}_{(a^{GCV}, b^{GCV})}, f)$	0.1183	0.0850	0.0774	0.0395	0.0262
$d_2(\widehat{f}_{(a^{Gd_2}, b^{Gd_2})}, f)$	0.593	0.04284	0.0392	0.03333	0.02361
$\frac{d_2(\widehat{f}_{(a^{Gd_2}, b^{Gd_2})}, f)}{d_2(\widehat{f}_{(a^{GCV}, b^{GCV})}, f)}$	1.9949	1.9859	1.9744	1.1851	1.1096

utilisation d'un semi-métrique qui est bien adapté au genre de données lesquelles nous devons traiter. Les résultats des quatre dimensions de l'échantillon sélectionnées, le $n \in 50, 100, 200, 250$, sont recueillis dans le tableau 2,1. Puis, il peut voir que le tableau 2,1 montre le bon comportement de notre procédure fonctionnelle, dans ce sens, les erreurs d_2 a calculé à l'aide des paramètres donnés par les sélecteurs de largeur de fenêtre est très proche du minimum d'erreurs. Ce fait est illustré dans la dernière rangée par le rapport $d_2(\widehat{f}_{(a^{GCV}, b^{GCV})}, f) / d_2(\widehat{f}_{(a^{Gd_2}, b^{Gd_2})}, f)$, où nous observons que nos résultats admettent des représentations suffisamment bonnes même pour une petite dimension de l'échantillon $n = 50$. Il est bien de noter que, notre méthode de sélection automatique calcule plus rapidement les résultats, également pour une taille d'échantillon plus grande, $n \in 100, 200, 250$.

Dans la suite, nous comparons cette procédure globale à la méthode la locale. Pour ce but, nous considérons un échantillon de 150 observations de couples (X, Y) , dans deux sous-ensembles (100 observations pour l'échantillon d'apprentissage et 50 observations pour l'échantillon de test) et nous construisons l'estimateur de noyau de la densité conditionnelle pour chaque courbe dans l'échantillon d'apprentissage sur les 100 points équidistants de y_j pour $j = 1, \dots, 100$ dans l'intervalle $[\min\{Y_i, i = 1, \dots, n\} \times 0.9, \max\{Y_i, i = 1, \dots, n\} \times 1.1]$ à l'aide des deux paramètres de sélection (local et global). Les largeurs de fenêtre locales (a^{LCV}, b^{LCV}) est sélectionnées plus de $H_n(x, y)$ l'ensemble de $(a(x), b(y))$ tels que, pour $a(x)$ (respectivement, pour $b(x)$) que la boule de centre X (respectivement, l'intervalle de centré y) avec le rayon $a(x)$ (respectivement,

avec rayon $b(y)$) contient exactement des voisins de k de x (respectivement, de y). De plus, nous utilisons les fonctions local de poids suivants :

$$W_{1,x}(t) = \begin{cases} 1, & \text{sid}(t,x) < a(x) \\ 0 & \text{sinon.} \end{cases} \quad W_{2,y}(z) = \begin{cases} 1, & \text{si } |z - y| < b(y) \\ 0 & \text{sinon.} \end{cases}$$

Nous choisissons les largeurs de fenêtre globales du même ensemble H_n définies avant et nous employons les mêmes

figure

fonctions W_1 et W_2 . D'ailleurs, nous considérons les mêmes noyaux, les mêmes semi-métriques et les mêmes K et H comme dans la première illustration. Puis, nous examinons l'exactitude de notre densité conditionnelle à l'aide de la somme généralisée de résiduels carrés (RSS) défini par :

$$RSS(local) = \frac{1}{500} \sum_{i=1}^{50} \sum_{j=1}^{100} \left[f((x_i, y_i)) - \hat{f}_{(a^{LCV}, b^{LCV})}(x_i, y_i) \right]^2$$

et

$$RSS(global) = \frac{1}{500} \sum_{i=1}^{50} \sum_{j=1}^{100} \left[f((x_i, y_i)) - \hat{f}_{(a^{GCV}, b^{GCV})}(x_i, y_i) \right]^2$$

Nous avons effectué plusieurs tests (exactement 25 tests) en changeant des observations entre apprentissage et les échantillons test. Sur le schéma 2 nous traçons le boxplot (boite à moustaches) des erreurs RSS donnés dans les deux cas. Il apparaît clairement, que le choix local de largeur de fenêtre est mieux que le choix globale.

figure

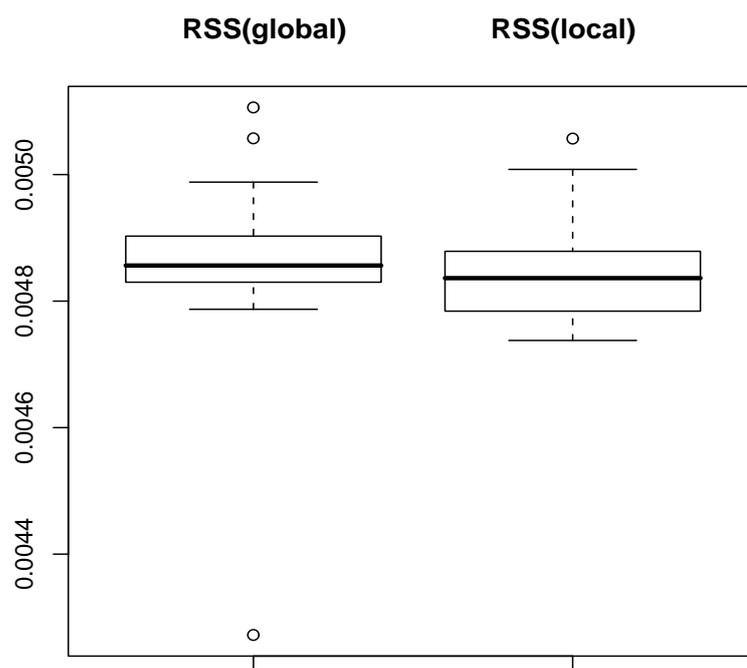


FIGURE 3.2 – Comparison of RSS errors

Bibliographie

- [1] Bashtannyk, D.M. and Hyndman, R.J. (2001). Bandwidth selection for kernel conditional density estimation. *Comput. Statist. Data Anal.*, **36**, Pages 279-298.
- [2] Benhenni, K., Ferraty, F., Rachdi, M. and Vieu, P. (2007). Local smoothing regression with functional data. *Comput. Statist.*, **22**, No. 3, Pages 353–369.
- [3] Bosq, D. (2000). *Linear Processes in Function Spaces : Theory and applications*. Lecture Notes in Statistics, **149**, Springer.
- [4] Dabo-Niang, S. and Laksaci, A. (2007). Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle. *Pub. Inst. Stat. Univ. Paris*, **3**, Pages 27–42.
- [5] Fan, J., Yao, Q. and Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, **83**, Pages 189–206.
- [6] Ferraty, F., Tadj, A., Laksaci, A. and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. of Statist. Plan. and Inf.*, **140**, Pages 335–352 .
- [7] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis. Theory and Practice*. Springer-Verlag.
- [8] Hall, P., Wolk, R.C. and Yao, Q. (1999). Methods for estimating a conditional distribution function. *J. Amer. Statist. Assoc.*, **94**, Pages 154–163.
- [9] Härdle, W. (1991). *Smoothing Techniques with Implementation in S*. Springer, New York.
- [10] Härdle, W., Janssen, P. and Serfling, R. (1991). Strong consistency rates for estimators of conditional functionals. *Ann. Statist.*, **16**, No. 4, Pages 1428–1449.
- [11] Härdle, W. and Marron, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.*, **13**, No. 4, Pages 1465–1481.

-
- [12] Hyndman, R.J. (1995). Highest-density forecast regions for non-linear and non-normal time series models. *J. Forecast.*, **14**, Pages 431–441.
- [13] Hyndman, R.J., Bashtannyk, D.M. and Grunwald, G.K. (1996). Estimating and visualizing conditional densities. *J. Comput. Graph. Statist.*, **5**, Pages 315–336.
- [14] Hyndman, R.J. and Yao, Q. (1998). Nonparametric estimation and symmetry tests for conditional density functions. *Working paper 17/98, Department of Econometrics and Business Statistics, Monash University*.
- [15] Rosenblatt, M. (1969). Conditional probability density and regression estimators. In *Multivariate Analysis II*, Ed. P.R. Krishnaiah. Academic Press, New York and London.
- [16] Roussas, G.G. (1968). On some properties of nonparametric estimates of probability density functions. *Bull. Soc. Math. Greece (N.S.)*, **9**, Pages 29–43.
- [17] Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.*, **9**, Pages 65–78.
- [18] Stone, C.J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.*, **22**, No. 1, Pages 118–184.
- [19] Youndjé, E. (1993). *Estimation non paramétrique de la densité conditionnelle par la méthode du noyau*. PhD Thesis from the Rouen University (in French).
- [20] A. Laksaci, F. Madani and M. Rachdi (2012). Kernel conditional density estimation when the regressor is valued in a semi-metric space. *Communications in Statistics-Theory and Methods*, 2012, vol. 00, issue 0, Pages 00-00.