



**UNIVERSITE A / MIRA DE BEJAIA.**



**FACULTE DES SCIENCES EXACTES**

**DEPARTEMENT DE MATHEMATIQUES**

Mémoire de fin d'études

**En vue de l'obtention d'un master en mathématiques**

**Option :**

**Statistiques et Analyse Décisionnelle**

**Thème**

**Prévision non paramétrique via la fonction  
de régression**

**Présenté par :**

**M<sup>elle</sup>. MOUHOU Macissilia**

**Soutenu le 18/06/2015**

**Devant le jury composé de :**

**Présidente M<sup>me</sup> : K.TIMERIDJINE**

**M.C.B**

**Université A/Mira, Bejaïa**

**Examineur M<sup>r</sup> : M.BOURAINE**

**M.A.A**

**Université A/Mira, Bejaïa**

**Rapporteur M<sup>me</sup> : H. TABTI**

**M.A.A**

**Université A/Mira, Bejaïa**

**Université de Bejaïa, Juin 2015**



**Promotion : 2014 - 2015**



REPUBLIQUE ALGERIENNE DEMOCRATIQUE et POPULAIRE.  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique.

UNIVERSITE ABDERRAHMANE MIRA de BEJAIA

Faculté des Sciences Exacte  
Département de Mathématiques

**Memoire de fin d'etude**

Option  
Statistiques et Analyse décisionnelle

Thème

**Prévision non paramétrique via la fonction de  
régression**

Présentées par :

Mouhou Macissilia

# Remerciements

Au terme de ce travail, je tiens à remercier :

Tout d'abord, le Dieu le tout-puissant qui m'a donné la force et le savoir afin d'accomplir ce travail.

Je remercie *M<sup>me</sup>TABTI*, ma promotrice, qui m'a fait bénéficier de son savoir, de ses compétences scientifiques et de sa passion pour la recherche. Je vous remercie également madame de m'avoir appris à aller jusqu'au bout de mes idées.

Je tien aussi à remercier également les membres de jury pour avoir accepté d'évaluer mon travail *M<sup>me</sup>TIMERIDJINE* et *M.BOURAINE*.

Mes remerciements vont à tous les membres de ma famille à qui je dois beaucoup, sans leurs aides, ce travail n'aurait pu voir le jour.

Merci à tous ceux qui m'ont aidé sans ménager ni leurs temps, ni leurs encouragements, ni leurs savoirs.

Et enfin, merci à tous le personnel de **RAMDY** pour leurs accueils, encouragements, conseils et pour le temps qu'ils m'ont consacré durant la période de mon stage.

## Dédicaces

A mes chères parents qui m'ont toujours soutenue et je les remercie d'autant que je ne remercie personne, à leur aide, à leur orientations et leurs conseils durant mes études et dans ma vie.

A mes chers frères : Nassim et Hocine.

A mes chères sœurs : Nouara, Lyly et Hania.

A mes chères copines : Fouza et Hassina sans oublier leurs familles.

A mes amies : Souad, Assia, Lila, Tina , Souhila

A toute la promotion de Master 2 SAD (2014/2015)

*Macissilia.M*

# Table des matières

Liste des tableaux	6
Liste des figures	7
Liste des abréviations	8
Introduction générale	9
<b>1 Estimation non paramétrique par la méthode de noyau</b>	<b>11</b>
1.1 Introduction	11
1.2 Critères d'erreurs et définitions	12
1.2.1 Quelques définitions	12
1.2.2 Les critères d'erreur	13
1.3 Estimation par la méthode de noyau	14
1.3.1 Généralité	14
1.4 Choix de noyau $K$	16
1.5 Propriétés de l'estimateur à noyau	18
1.6 Choix du paramètre de lissage	20
1.6.1 Rule of Thumb	22
1.6.2 La validation croisée	23
<b>2 Régression non paramétrique dans le cas réel</b>	<b>24</b>
2.1 Introduction	24
2.2 Quelques modèles de régression	25
2.2.1 Le cadre général	25
2.2.2 Modèle pour variable réelle	27
2.3 Régression avec lissage par moyenne mobile (MA)	28
2.3.1 Etude de la convergence en probabilité	29
2.3.2 Etude de la convergence en distribution et intervalles de confiance	31
2.4 Régression par opérateur à noyau	32
2.4.1 Estimateur de Nadaraya par la fonction de densité conditionnelle	32
2.4.2 Propriétés de l'estimateur à noyau	35
2.4.3 Consistance	35

2.4.4	Absence de biais asymptotique . . . . .	37
2.4.5	Etude de la convergence en probabilité . . . . .	38
2.4.6	Etude de la convergence en loi et intervalles de confiance . . . . .	38
2.5	Le choix de paramètre lissage $h$ . . . . .	41
2.6	Validation croisée . . . . .	41
<b>3</b>	<b>Régression non paramétrique cas des séries chronologiques</b>	<b>42</b>
3.1	Introduction . . . . .	42
3.2	Le prédicteur à noyau . . . . .	44
3.3	Le modèle de dépendance . . . . .	44
3.3.1	Le mélange fort . . . . .	44
3.3.2	Une inégalité exponentielle . . . . .	45
3.3.3	Une inégalité de covariance . . . . .	46
3.4	Régression sous mélange fort . . . . .	46
3.4.1	Quelques résultats préliminaires . . . . .	47
3.4.2	Convergence presque complète ponctuelle . . . . .	48
3.4.3	Convergence presque complète uniforme . . . . .	50
3.4.4	Convergence en moyenne quadratique . . . . .	51
3.5	Propriétés de convergence presque complète du prédicteur à noyau . . . . .	53
3.5.1	Convergence en moyenne quadratique du prédicteur à noyau . . . . .	54
<b>4</b>	<b>Prévision non paramétrique des séries chronologiques</b>	<b>55</b>
4.1	Introduction . . . . .	55
4.2	Prédicteur à noyau . . . . .	57
4.3	Interprétation de prédicteur . . . . .	58
4.3.1	Processus stationnaire . . . . .	58
4.3.2	Processus non stationnaire . . . . .	59
4.4	Similarité . . . . .	60
4.5	Choix de paramètre de troncation $r$ . . . . .	61
4.6	Intervalle de prévision . . . . .	63
4.7	Comparaison . . . . .	64
4.7.1	Critère EMRO . . . . .	64
4.7.2	Critère EMRP . . . . .	64
<b>5</b>	<b>Application</b>	<b>65</b>
5.1	Introduction . . . . .	65
5.2	Prévision paramétrique . . . . .	65
5.2.1	La méthodologie de Box et Jenkins . . . . .	65
5.2.2	Identification de modèle . . . . .	66
5.2.3	L'estimation du modèle . . . . .	69
5.2.4	Validation de modèle . . . . .	70
5.2.5	Choix de modèle . . . . .	72
5.2.6	Prévision . . . . .	73

5.3	Prévision non paramétrique . . . . .	74
5.3.1	Présentation . . . . .	74
5.3.2	Démarche de construction de prédicteur . . . . .	75
5.4	Prévision des séries . . . . .	76
5.4.1	Prévision non paramétrique d'un ARIMA (11,1,7) . . . . .	76
5.4.2	Prévision non paramétrique d'un ARIMA (10,1,7) . . . . .	78
<b>Conclusion générale</b>		<b>81</b>
<b>Bibliographie</b>		<b>83</b>

# Liste des tableaux

1.1	Les noyaux usuels . . . . .	17
2.1	Les noyaux usuels avec la valeur de $b_K$ . . . . .	39
5.1	Tableau d'estimation de modèle ARIMA (11,1,7) . . . . .	70
5.2	Les prévisions paramétrique pour cinq mois (année 2015) pour le modèle ARIMA(11,1,7) . . . . .	73
5.3	Les prévisions paramétrique pour cinq mois (année 2015) pour le modèle ARMA(10,1,7) . . . . .	74
5.4	Tableau des résultats pour le modèle ARIMA(11,1,7) . . . . .	77
5.5	Tableau de critère EMRO pour le ARIMA(11,1,7) . . . . .	77
5.6	Tableau des résultats pour le modèle ARIMA(10,1,7) . . . . .	79
5.7	Tableau de critère EMRO pour le ARIMA(10,1,7) . . . . .	79

# Table des figures

1.1	Les courbes des noyaux . . . . .	17
5.1	La série chronologique des pots de yaourt (2010-2014) . . . . .	67
5.2	La série chronologique différenciée des pots de yaourt (2010-2014) . . . . .	67
5.3	Le graphe de la fonction de l'autocorrelation partielle . . . . .	68
5.4	Le graphe de la fonction de l'autocorrelation . . . . .	68
5.5	Schéma des résidus du modèle ARIMA(10,1,7) . . . . .	71
5.6	Schéma des résidus du modèle ARIMA(11,1,7) . . . . .	72
5.7	Prévision non paramétrique de modèle ARIMA(11,1,7) . . . . .	76
5.8	Prévision non paramétrique de modèle ARIMA(10,1,7) . . . . .	78

# Liste des abréviations

$\mathbb{R}$ :	Ensemble des réels.
$X$ :	Variable aléatoire.
$\mathbf{E}(X)$ :	Espérance de la variable aléatoire $X$ .
$\mathbf{V}(X)$ :	La variance de la variable aléatoire $X$ .
$\mathbf{I}$ :	Fonction indicatrice d'un ensemble.
<i>i.i.d</i>	Indépendantes et identiquement distribuées.
<i>p.co</i>	Converge presque complète.
<i>MSE</i>	Erreur quadratique moyenne (Mean square error).
<i>MISE</i>	Erreur quadratique moyenne intégrée(Mean Integrated square error).
<i>ISE</i>	Erreur quadratique intégrée.
<i>Card</i>	Cardinal.
<i>ARMA</i>	Autoregressif moyenne mobile.
<i>ARIMA</i>	Autoregressif moyenne mobile intégrable.
<i>PACF</i>	La fonction d'autocorrelation partielle
<i>ACF</i>	La fonction d'autocorrelation.

# Introduction générale

La prévision utilisant des séries temporelles est à l'origine du développement de la Statistique Mathématique des processus. En effet l'étude d'une série observée comporte les étapes suivantes : choix d'un modèle, estimation des paramètres du modèle, validation du modèle et prévision.

Bien entendu ces étapes ne se succèdent pas dans un ordre immuable. Ainsi l'estimation de la tendance et de la saisonnalité d'une série intervient souvent avant le choix du modèle.

En statistique paramétrique on choisit un modèle comportant un nombre fini de paramètres inconnus alors que la statistique non paramétrique, préfère considérer des modèles plus vastes tel qu'elle s'intéresse à l'estimation, à partir d'un nombre fini d'observations, d'une fonction inconnue. Cela explique l'apparition des méthodes non paramétriques d'estimation et de prévision.

Les méthodes non paramétriques ont connu un essor important depuis les travaux de Bosq (1979), Collomb (1980) et Robinson (1983). Parmi ces méthodes non paramétriques, la méthode de régression non paramétrique.

Historiquement, le principe des régressions non paramétriques remonte au 19<sup>e</sup> siècle selon Cleveland and Loader (1995), toute fois les premiers travaux modernes sur ce sujet datent des années 50. La première application que nous verrons relève de l'estimation de fonctions de densité par des méthodes d'opérateur à noyau (kernel) avec les travaux fondateurs de Rosenblatt (1956) et de Parzen (1962). Ces premiers travaux ont été étendus à la notion de régression kernel, imparfaitement traduit en français par le terme de régression avec lissage par opérateur à noyau. Dans ce domaine, on identifie deux papiers fondateurs publiés la même année : Nadaraya (1964) et Watson (1964).

Cette méthode non paramétrique (régression non paramétrique) consiste à déterminer la façon dont l'espérance d'une variable dépendante dépend d'un ensemble de variables explicatives. Alors on déduit que cette espérance conditionnelle sachant le passé de ces

variables a le rôle d'un prédicteur. Notre question est :

Comment appliquer ce prédicteur au cas d'une série chronologique et quelles sont les différentes étapes à suivre ?

Le présent travail portera sur les prévisions via la fonction de régression non paramétrique univariée.

Dans le premier chapitre, on s'intéresse au cadre classique d'estimation non paramétrique d'une fonction de régression à partir de régresseurs réels en rappelant les différentes propriétés statistiques d'estimation non paramétrique par la méthode de noyau.

Par ailleurs dans le deuxième chapitre, nous avons accordé une place importante à un domaine d'actualité de la statistique non paramétrique qui est celui de la prévision de série chronologique. Il s'agit d'un domaine sur lequel les premiers résultats conséquents furent établis au début des années 80 (Collomb, 1983 et 1984, et Robinson 1983), et qui a connu depuis des développements continus (citons par exemple les ouvrages généraux de Cyorfi et al 1989, Yoshihara, 1994, hardle et al 1997 et celui de Bosq 1996). Pour ce chapitre nous avons opté pour une démarche différentes de celle des deux précédents en ce sens que nous avons cherché à établir nos résultats sous une formulation des plus récentes. Ainsi le chapitre contient essentiellement des résultats asymptotique qui bien qu'étant de nature classique (Bosq 1996).

Après avoir développé les techniques utilisées dans le cadre de régression non paramétrique en s'intéressant dans les deux cas à l'estimateur par noyau.

Le dernier chapitre s'insère pleinement au cadre d'application des prévisions non paramétriques en utilisant le logiciel R, passant par l'application d'estimation non paramétrique et la méthodologie de Box et Jenkins en estimant le meilleur modèle de prévision anisi déduire les prévisions pour ce modèle. On termine par l'interprétation des résultats en comparant des résultats paramétrique à ceux de la méthode non paramétrique.

# Chapitre 1

## Estimation non paramétrique par la méthode de noyau

### 1.1 Introduction

La théorie de l'estimation est l'une des préoccupations des statisticiens, on la divise en estimation paramétrique et estimation non paramétrique. L'estimation qui a reçu un intérêt croissant tant sur le plan théorique que pratique est celle d'estimation non paramétrique.

Pour cette approche plusieurs travaux sont marqués pour ça citant Rosenblatt (1956) et Parzen (1962) puis de Nadaraya-Watson (1964), citant plusieurs méthodes parmi ces méthodes :

- Estimation non paramétrique par la méthode d'histogramme.
- Estimation non paramétrique par la méthode de noyau.
- Estimation non paramétrique par les fonctions orthogonales.
- Estimation non paramétrique par les fonctions spline.

Au *XVII<sup>e</sup>* John Graunt a introduit l'estimateur de la densité non paramétrique par la méthode d'histogramme. Cette méthode consiste à estimer  $f$  en un point  $x$  par la proportion des variables aléatoires  $(x_1, \dots, x_n)$  qui se trouvent dans un intervalle de longueur un paramètre de lissage  $h_n$  et qui contient  $x$ . Elle est donc basée sur le choix d'un point d'origine et d'une partition donnée par :  $I_j = [a_j, a_{j+1}[$  tel que :  $j = \overline{1, k-1}$ .

Pour  $x_i \in [a_j, a_{j+1}[$ , l'estimateur de  $f$  de type histogramme est donné par :

$$\hat{f}_h(x) = \frac{\text{card}\{x_i \leq a_{j+1}\} - \text{card}\{x_i \leq a_j\}}{nh} \quad (1.1)$$

$$= \frac{n_j}{nh} \quad (1.2)$$

Les propriétés asymptotiques de cet estimateur ont été détaillées dans le livre de Bosq (1979) et Lecoutre (1987) et le livre de Simonoff (1996), l'étude asymptotique de l'erreur quadratique moyenne de  $f_n$  et l'erreur quadratique moyenne intégrée de  $f_n$  ont été établies par Lecoutre (1982), en (1974) Geoffrey a étudié la convergence uniforme et presque complète de cet estimateur.

L'estimateur par l'histogramme a de bonnes propriétés statistiques, mais le désavantage qui gêne est la discontinuité qui ne peut pas s'adapter au cas où  $f$ , la densité à estimer, vérifie certaines hypothèses de régularité.

Afin de résoudre ce problème, l'estimateur de Parzen-Rosenblatt a été introduit, il généralise intuitivement la méthode d'estimation par histogramme, et il est très utilisé en estimation non paramétrique.

Au sujet d'estimation non paramétrique on s'intéresse dans cette partie à la méthode de noyau (méthode de Parzen-Rosenblatt), les critères d'erreurs utilisés, la définition de l'estimateur et les propriétés statistiques font l'objectif de cette partie.

## 1.2 Critères d'erreurs et définitions

Avant de construire des estimateurs de la densité de probabilité, pour mesurer les performances théoriques et identifier le meilleur, il est nécessaire de spécifier un critère d'erreur.

Nous considérons la densité de probabilité  $f$  et son estimateur  $\hat{f}$ .

### 1.2.1 Quelques définitions

#### Définition 1.2.1

On dit qu'un estimateur  $\hat{f}$  de  $f$  est sans biais si :

$$\mathbf{E}(\hat{f}(x)) = f$$

#### Définition 1.2.2

On dit qu'un estimateur  $\hat{f}$  de  $f$  est asymptotiquement sans biais si :

$$\lim_{n \rightarrow \infty} \mathbf{E}(\hat{f}(x)) = f$$

**Définition 1.2.3**

Un estimateur  $\hat{f}$  de  $f$  dit asymptotiquement uniformément sans biais si :

$$\lim_{n \rightarrow \infty} \sup_x |\mathbf{E}[\hat{f}(x) - f(x)]| = 0$$

**1.2.2 Les critères d'erreur****1) L'erreur quadratique intégrée ISE**

$$\begin{aligned} \text{ISE}(\hat{f}(x)) &= \int [f(x) - \hat{f}(x)]^2 dx \\ &= \int [f(x)^2 - 2f(x)\hat{f}(x) + \hat{f}^2(x)] dx \end{aligned}$$

**2) L'erreur quadratique moyenne MSE**

$$\begin{aligned} \text{MSE}(\hat{f}(x)) &= \mathbf{E}(f(x) - \hat{f}(x))^2 \\ &= \mathbf{E}(f^2(x)) - 2\mathbf{E}(f(x)\hat{f}(x)) + \mathbf{E}(\hat{f}^2(x)) \\ &= \mathbf{E}f^2(x) - 2\mathbf{E}f(x)\hat{f}(x) + \mathbf{E}\hat{f}^2(x) + [\mathbf{E}\hat{f}(x)]^2 - [\mathbf{E}f(x)]^2 \\ &= f^2(x) - 2f(x)\mathbf{E}(\hat{f}(x)) + \mathbf{E}(\hat{f}^2(x)) + [\mathbf{E}(\hat{f}(x))]^2 - [\mathbf{E}f(x)]^2 \\ &= [\mathbf{E}(\hat{f}(x)) - f(x)]^2 + \mathbf{E}(\hat{f}^2(x)) - [\mathbf{E}(\hat{f}(x))]^2 \end{aligned}$$

Le MSE est donné par :

$$\text{MSE}(\hat{f}(x)) = [\mathbf{Biais}(\hat{f}(x))]^2 + \mathbf{V}(\hat{f}(x)) \quad (1.3)$$

### 3) L'erreur quadratique moyenne intégrée MISE

$$\begin{aligned}
 \text{MISE}(\hat{f}(x)) &= \int \text{MSE}(\hat{f}(x)) dx \\
 &= \int \mathbf{E}(f(x) - \hat{f}(x))^2 dx \\
 &= \int [(\mathbf{Biais}(\hat{f}(x)))^2 + \mathbf{V}(\hat{f}(x))] dx
 \end{aligned}$$

## 1.3 Estimation par la méthode de noyau

### 1.3.1 Généralité

Parzen (62) a étudié les propriétés fondamentales de l'estimateur à noyau. Il s'agit de l'estimateur le plus populaire. Il est adapté aux variables aléatoires continues. L'estimateur à noyau de la densité en un point  $x$  est construit en comptant le nombre d'observation dans un intervalle autour de  $x$ . Plus formellement :

$$\hat{f}_h(x) = \frac{1}{2nh} \text{Card}\{x_i \in [x - h, x + h]\}$$

Où de façon équivalente :

$$\hat{f}_h(x) = \frac{1}{2nh} \sum_{i=1}^n I_{\left[\frac{x-x_i}{h}\right] \leq 1}(x)$$

Plus généralement, on adapte les notations suivantes :

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

Avec :

$$\hat{f}_h(x) = \frac{1}{2nh} \sum_{i=1}^n I_{\left|\frac{x-x_i}{h}\right| \leq 1}(x)$$

**Définition 1.3.1**

Soit  $(x_1, \dots, x_n)$  un échantillon de loi  $f(x)$  sur  $\mathbb{R}$ . La fonction de répartition :

$$F(x) = \int_{-\infty}^x f(t) dt$$

On appelle la fonction de répartition  $F_n$  de  $F$  au point  $x$ ,  $F_n : \mathbb{R} \rightarrow [0,1]$ . Définie pour tout  $x \in \mathbb{R}$  par :

$$F_n = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(x_i)_{]-\infty, x]} \quad (1.4)$$

La fonction de répartition empirique  $F_n$  est un estimateur simple de  $F$ .

$$n F_n(x) = \sum_{i=1}^n \mathbf{I}(x_i)_{]-\infty, x]} \rightarrow B(n, F_n)$$

$B$  est une binomial avec :

$$\begin{aligned} E(F_n(x)) &= F(x) \\ V(F_n(x)) &= \frac{1}{n} [1 - F(x)] F(x) \end{aligned}$$

Apartir de la définition d'une probabilité et en utilisant l'équation (1.4) la densité  $f$  peut s'écrire :

$$\hat{f}_h(x) = \lim_{h \rightarrow 0} \frac{F_n(x+h) - F_n(x-h)}{2h} \quad (1.5)$$

$$= \frac{1}{2nh} \sum_{i=1}^n I(x_i)_{]x-h, x+h]} \quad (1.6)$$

En posant :

$$\varpi(u) = \begin{cases} \frac{1}{2} & \text{si } -1 < u < 1 \\ 0 & \text{si non} \end{cases}$$

On peut réécrire (1.6) sous la forme :

$$\widehat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n \varpi\left(\frac{x - x_i}{h}\right) \quad (1.7)$$

Nous venons de définir l'estimateur à noyau dit de Rosenblatt (uniforme). Parzen a étudié une classe générale d'estimateurs. Remplaçons la fonction  $\varpi$  par une fonction noyau définie sur  $\mathbb{R}$  appelée  $K$  (Karnel) satisfaisant la condition  $\int_{-\infty}^{\infty} K(u)du = 1$ .

Généralement,  $K$  est une densité de probabilité. Par analogie avec la définition de l'estimateur de Rosenblatt l'estimateur à noyau de Parzen est :

$$\widehat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (1.8)$$

Où  $h$  est paramètre de lissage.

### Définition 1.3.2

Si  $x_1, \dots, x_n$  est un échantillon iid d'une variable aléatoire de densité  $f$  continue, alors l'estimateur non paramétrique de la densité par la méthode du noyau définit comme (1.8)

## 1.4 Choix de noyau $K$

Selon la définition (1.8) toute fonction  $K$  peut servir comme noyau d'estimation d'une densité  $f$ . D'après [13] les noyaux utilisés dans l'estimation de la densité de probabilité sont donnés dans le tableau suivant :

Noyau	K(u)
Uniforme	$\frac{1}{2},  u  \leq 1$
Triangulaire	$(1 -  u ),  u  \leq 1$
Gaussien	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$
Rectangulaire	$\frac{1}{2} I_{ u  \leq 1}$
Biweight	$\frac{15}{16} (1 - u^2)^2,  u  \leq 1$
Triweight	$\frac{35}{32} (1 - u^2)^3,  u  \leq 1$
Epanechnikov	$\frac{3}{4\sqrt{5}} (1 - \frac{u^2}{5}),  u  \leq \sqrt{5}$
Cosine	$\frac{\pi}{4} \cos \frac{u\pi}{2},  u  \leq 1$

TAB. 1.1 – Les noyaux usuels

Les noyaux les plus couramment utilisés en pratique sont :

- Le noyau rectangulaire.
- Le noyau triangulaire.
- Le noyau d'epanechnikov.
- Le noyau gaussien.

D'après [1] les courbes de ces noyaux sont présentées ci-dessous :

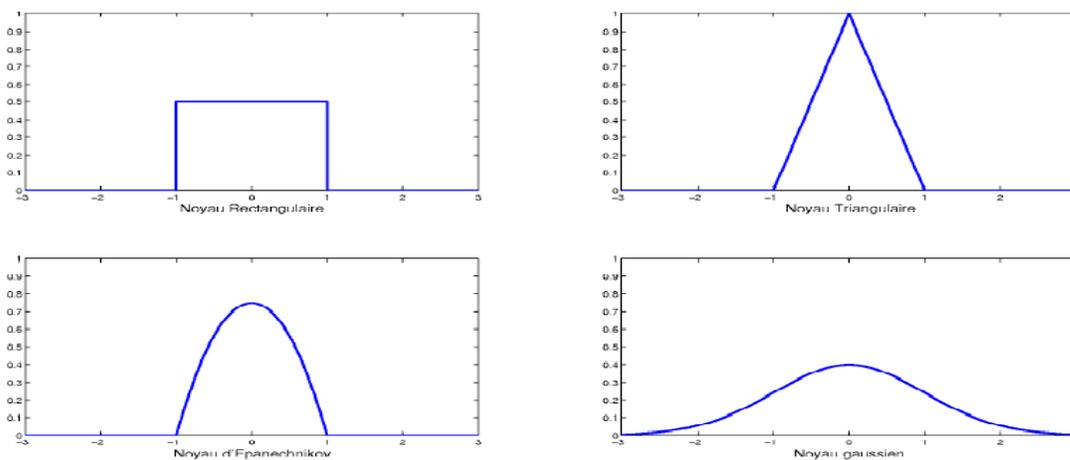


FIG. 1.1 – Les courbes des noyaux

**Remarque**

Lorsqu'on définit un estimateur à noyau, on a non seulement le choix de la fenêtre  $h > 0$  mais aussi celui du noyau  $K$ . Il y a un certain nombre de conditions qui sont considérées comme usuelles pour les noyaux et qui permettent d'analyser le risque de l'estimateur à noyau qui en résulte.

**Les conditions de K**

$K$  vérifie les conditions suivantes :

- $\int K(y)dy = 1$
- $\int yK(y)dy = 0$
- $\int y^2K(y)dy = \sigma_K^2$

Après avoir donné l'idée générale sur l'estimateur et le choix de noyau, les propriétés statistiques détaillent l'estimation.

**1.5 Propriétés de l'estimateur à noyau****• Espérance**

L'espérance mathématique de  $\widehat{f}_h(x)$  est :

$$\begin{aligned} \mathbf{E}(\widehat{f}_h(x)) &= \mathbf{E}\left(\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)\right) \\ &= \frac{1}{nh} \sum_{i=1}^n \mathbf{E}\left(K\left(\frac{x-X}{h}\right)\right) \\ &= \frac{1}{h} \mathbf{E}\left(K\left(\frac{x-X}{h}\right)\right) \\ &= \frac{1}{nh} \int \left(K\left(\frac{x-u}{h}\right)\right) f(u) du \end{aligned}$$

En posant  $y = \frac{x-u}{h} \Rightarrow dy = \frac{-du}{h}$

$$\mathbf{E}(\widehat{f}_h(x)) = \int_{-\infty}^{\infty} K(y) f(x-hy) dy$$

- Biais

Le biais de  $\widehat{f}_h(x)$  est :

$$\begin{aligned} \text{Biais}(\widehat{f}_h(x)) &= \mathbf{E}(\widehat{f}_h(x)) - f(x) \\ &= \int_{-\infty}^{\infty} K(y)f(x - hy)dy - f(x) \end{aligned}$$

- La variance

Le variance de  $\widehat{f}_h(x)$  est :

$$\begin{aligned} V(\widehat{f}_h(x)) &= \mathbf{V} \sum_{i=1}^n \frac{1}{nh} K\left(\frac{x - x_i}{h}\right) \\ &= \frac{1}{n^2 h^2} \sum_{i=1}^n \mathbf{V} K\left(\frac{x - x_i}{h}\right) \\ &= \frac{1}{n^2 h^2} \sum_{i=1}^n [\mathbf{E}(K\left(\frac{x - x_i}{h}\right))^2] - \frac{1}{n^2 h^2} \sum_{i=1}^n (\mathbf{E}(K\left(\frac{x - x_i}{h}\right)))^2 \\ &= \frac{1}{nh^2} \int [K\left(\frac{x - y}{h}\right)]^2 f(y) dy - \frac{1}{n} \left(\frac{1}{h}\right) \int K\left(\frac{x - y}{h}\right) f(y) dy)^2 \end{aligned}$$

Avec un changement de variable,  $y = \frac{x-u}{h}$ , on obtient :

$$V(\widehat{f}_h(x)) = \frac{1}{h} \int (K(y))^2 f(x - yh) d(y) - \frac{1}{h} \left( \int (K(y)) f(x - yh) d(y) \right)^2$$

En faisant le développement de Taylor à l'ordre 2 au point  $h = 0$  de  $f(x - yh)$ . On obtient :

$$f(x-yh) = f(x) - \frac{hy}{1} f'(x) + \frac{h^2 y^2}{2!} f''(x) + 0(h^2)$$

$$\begin{aligned} \mathbf{E}(\widehat{f}_h(x)) &= \int K(y) [f(x) - hyf'(x) + \frac{h^2 y^2}{2!} f''(x)] dy + 0(h^2) \\ &= f(x) \int K(y) dy - hf'(x) \int yK(y) dy + \frac{h^2}{2} f''(x) \int y^2 K(y) dy + 0(h^2) \end{aligned}$$

Si le noyau  $K(y)$  est une fonction symétrique par rapport à 0 c'est à dire :

$$\begin{aligned} \int yK(y)dy &= 0 \\ \int y^2K(y)dy &< \infty \end{aligned}$$

Alors les expressions finales sont données par :

$$\mathbf{E}(\widehat{f}_h(x)) = f(x) + \frac{h^2}{2}f''(x)\mu_2(K) + o(h^2)$$

Où :

$$\mu_2(K) = \int y^2K(y)dy$$

$$\begin{aligned} \text{Biais}(\widehat{f}_h(x)) &= \mathbf{E}(f_h(x)) - f(x) \\ &= \frac{h^2}{2}f''(x)\mu_2(K) \end{aligned}$$

$$\mathbf{V}(\widehat{f}_h(x)) = \frac{f(x)}{nh} \int K^2(y)dy - \frac{f'(x)}{n} \int y^2K(y)dy - \frac{1}{n}(f(x) + \text{biais}f_h(x))^2$$

## 1.6 Choix du paramètre de lissage

La décision d'un choix optimal pour le paramètre de lissage suppose une spécification d'un critère d'erreur qui puisse être optimisé. On cherche à minimiser l'erreur quadratique moyenne intégrée. D'après [13]

$$\begin{aligned} \text{MISE} &= \int \mathbf{E}|\widehat{f}_h(x) - f(x)|^2 dx \\ &= \frac{h^4}{4}\sigma_K^4 R(f'') + \frac{R(K)}{nh} + o(h^5 + \frac{1}{h}) \end{aligned}$$

Et l'erreur quadratique moyenne intégrée asymptotique AMISE.

$$AMISE = MISE - O(h^5 + \frac{1}{h})$$

Tel que :

$$AMISE = \frac{h^4}{4}\sigma_K^4 R(f'') + \frac{R(K)}{nh} \text{ avec } R(q) = \int q^2(x)dx \quad [13]$$

### Remarque

On peut remarquer que le premier terme du nombre droit est un terme biais, alors que le second est un terme de variance. On constate que dans le MISE le terme de Biais est la fonction croissante en h alors que le terme de la variance est une fonction décroissante en h. On peut calculer le paramètre de lissage  $h$ ,  $h^*$  qui minimise l'erreur quadratique moyenne intégrée asymptotique.

$$\frac{d}{dh} AMISE = h^3 \sigma_K^4 R(f'') - \frac{R(K)}{nh^2} = 0$$

$$\begin{aligned} h^3 \sigma_K^4 R(f'') &= \frac{R(K)}{nh^2} \\ nh^5 \sigma_K^4 R(f'') &= R(K) \\ h^5 &= \frac{R(K)}{\sigma_K^4 R(f'')} \\ h &= \left[ \frac{R(K)}{\sigma_K^4 R(f'')} \right]^{\frac{1}{5}} \end{aligned}$$

$$\text{Alors } h^* = \left[ \frac{R(K)}{\sigma_K^4 R(f'')} \right]^{\frac{1}{5}}$$

On obtient :

$$\frac{d}{dh} (AMISE) = 3h^2 \sigma_K^4 R(f'') + \frac{2}{nh^3} R(K) > 0$$

On peut écrire  $h^*$  :

$$h^* = \left[ \frac{R(K)}{\sigma_K^4 R(f'')} \right]^{\frac{1}{5}} \quad (1.9)$$

$$= \psi(K) \rho(f) n^{-\frac{1}{5}} \quad (1.10)$$

La valeur optimale d'AMISE :

$$\begin{aligned} \text{AMISE}^* &= \text{AMISE}(h^*) \\ &= \frac{5}{4} [\sigma_K R(K)^4 R(f'')]^{\frac{1}{5}} n^{-\frac{1}{5}} \end{aligned}$$

Il existe plusieurs méthodes pour choisir le paramètre de lissage, parmi ces méthodes la méthode de Rule of Thunb et la méthode de validation croisée.

### 1.6.1 Rule of Thunb

On a montré que si on choisit le paramètre de lissage de telle sorte que le MISE, soit minimum, alors le paramètre lissage optimale  $h^*$  est donné par (1.10) . Dans cette formule il s'agit d'assigner une valeur au terme  $R(f'')$  pour obtenir une estimation de  $h^*$ . Si on choisit comme état de loi normale de moyenne 0 et variance  $\sigma^2 = 1$ . On a alors [13] :

$$R(f'') = \int (f''(x))^2 dx \quad (1.11)$$

$$= \frac{3}{8} \sqrt{K} \sigma^{-5} \quad (1.12)$$

De plus si on utilise un noyau gaussien, alors la valeur pour  $h^*$  notée dans ce cas par  $h_{rot}$  est obtenu en substituant ce noyau et la valeur obtenue dans (1.10) dans la formule (1.12)

$$\begin{aligned} h_{rot} &= ((4\pi)^{\frac{-1}{10}} [\frac{3}{8} \pi^{\frac{-1}{2}} \sigma]) n^{-\frac{1}{5}} \\ &= \frac{4^{\frac{1}{5}}}{3} \sigma n^{-\frac{1}{5}} \\ &= 1.06 \sigma n^{-\frac{1}{5}} \end{aligned}$$

### 1.6.2 La validation croisée

Pour un estimateur à noyau et de fenêtre  $h$ , la sélection par validation croisée est une approche classique. Il s'agit à minimiser par rapport à  $h$  la vraisemblance pour l'échantillon  $(x_i), i=\overline{1, n}$ , défini par :

$$\text{LCV}(h) = \prod_{i=1}^n f_{h,i}(x_i)$$

Tel que :

$$f_{h,i}(x_i) = \frac{1}{h(n-1)} \sum_{1 \leq j \leq n, i \neq j} K\left(\frac{x_i - x_j}{h}\right)$$

Est l'estimateur à noyau basé sur les  $(n-1)$  observations différentes de  $(x_i)$ , la vraisemblance est alors [13] :

$$\text{LCV}(h) = \prod_{i=1}^n \frac{1}{h(n-1)} \sum_{1 \leq j \leq n, i \neq j} K\left(\frac{x_i - x_j}{h}\right)$$

# Chapitre 2

## Régression non paramétrique dans le cas réel

### 2.1 Introduction

Lorsqu'on souhaite décrire l'influence d'une variable  $Y$  qui est liée à une autre variable  $X$ , on cherche à exprimer  $Y$  en fonction de  $X$ , il est bien connu de la régression.

On sait que le but d'un modèle de régression consiste à déterminer la façon dont l'espérance d'une variable dépendante  $Y$  dépend d'un ensemble de variables explicatives  $X$ .

Supposant que  $X \in \mathbb{R}$ , le problème consiste à déterminer donc pour chaque réalisation  $x$  de la variable  $X$  la valeur de la fonction  $r(x)$  dite fonction de lien, on obtient :

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

Pour caractériser cette fonction de lien, la première approche consiste à utiliser un modèle de régression paramétrique. On suppose que cette fonction peut s'écrire comme une fonction explicite des valeurs de  $X$ . Cette fonction peut être linéaire, logarithmique, non-linéaire etc.

Dans l'approche paramétrique, la fonction de lien a deux propriétés :

- Elle est de forme explicite.
- Elle peut s'écrire en fonction d'un nombre réduit de paramètres.

Dans notre cas on veut estimer la relation entre le niveau moyen de la variable  $Y$  et toutes les valeurs réalisées de  $X$ , tel qu'on ne postule aucune forme spécifique sur la fonction de lien.

Par définition est une approche non paramétrique, contrairement à l'approche paramétrique la fonction de lien :

- N'a pas une forme explicite.
- Ne peut pas s'écrire en fonction d'un nombre réduit de paramètres.

Afin de déterminer cette fonction de lien Nadaraya et Watson (1964) ont proposé une méthode de régression non paramétrique dite la régression de Kernal, cette méthode repose en fait sur des méthodes de lissage. Pour comprendre le principe de cette régression non paramétrique, nous commencerons par exposer les différents modèles de régression, ensuite on passe à la régression avec lissage par moyenne mobile. Une fois que l'on aura démontré un certain nombre de résultats dans ce cas simple, nous nous contenterons d'énoncer plusieurs résultats dans le cas de la régression par noyau.

## 2.2 Quelques modèles de régression

### 2.2.1 Le cadre général

Dans le cadre général la fonction la plus connue est celle de la régression de  $Y$  en  $X$ , définit par :

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

#### Définition 2.2.1

Supposons que le comportement d'une variable aléatoire  $Y$  soit liée à une autre variable  $X$ , on veut exprimer  $Y$  par une fonction  $r$ , donnée l'expression sous la forme :

$$Y = r(X) + \epsilon \quad (2.1)$$

Où  $\epsilon$  est un bruit blanc. La fonction  $r$  peut être estimée par plusieurs méthodes :

- 1) Le régressogramme.
- 2) Le régressogramme mobile.
- 3) La méthode du noyau.

### 1) Régressogramme

A partir des  $n$  observations  $(x_i, y_i)$  on va estimer  $r(t)$  en fabriquant les classes  $C_i$  où se situent les  $x_j$ , pour la classe  $C_i$  où se trouve  $t$  on effectue la moyenne  $y_j$  correspondants aux  $x_j$  de cette classe  $C_i$ . En notant  $k$  le nombre de points  $x_j$  de cette classe  $C_i$ , pour tous  $t$  de  $C_i$ , d'après [5] on estime  $r(t)$  par :

$$\hat{r}_n(t) = \frac{1}{k} \sum_{j=1}^k y_j, k = \sum_{j=1}^k I_{C_j}(x_j)$$

D'où :

$$\hat{r}_n(t) = \frac{\sum_{j=1}^n I_{C_j}(x_j) y_j}{\sum_{j=1}^n I_{C_j}(x_j)}$$

### 2) Régressogramme mobile

La première chose sert à centrer la classe en le point  $t$  où l'on estime la régression, d'après [5] ce régressogramme s'écrit :

$$\hat{r}_n(t) = \frac{\sum_{j=1}^n I_{[t-h(n), t+h(n)]}(x_j) y_j}{\sum_{j=1}^n I_{[t-h(n), t+h(n)]}(x_j)}$$

Où par :

$$\hat{r}_n(t) = \frac{\sum_{j=1}^n I_{[-1, 1]}[x_j - t/h(n)] y_j}{\sum_{j=1}^n I_{[-1, 1]}[x_j - t/h(n)]}$$

### 3) Régressogramme par la méthode de noyau

On dispose d'un échantillon  $(x_i, y_i)_{1 \leq i \leq T}$  et on cherche à identifier la fonction  $r$  telle que :

$$Y_t = r(x_t) + \epsilon_t; \quad 1 \leq t \leq T$$

Watson (1964) et Nadaraya (1964) ont proposé, indépendamment et simultanément, l'estimateur

$$\hat{r}(x) = \begin{cases} \frac{\sum y_i K((x-x_i)/h_T)}{\sum K((x-x_i)/h_T)}; & \text{si } \sum K((x-x_i)/h_T) \neq 0 \\ 0; & \text{sinon} \end{cases}$$

plus de détail voir [5]

## 2.2.2 Modèle pour variable réelle

Nous adapterons la démarche qui consiste à caractériser un modèle de régression par une hypothèse du type :

$$r \in \mathcal{C} \quad (2.2)$$

Où  $\mathcal{C}$  est une classe de fonction, avec ou sans restriction supplémentaire sur la loi de  $Y, X$  ou  $\epsilon$ , [10].

### Définition 2.2.2

Nous disons que le modèle défini par (2.1),(2.2) est un modèle paramétrique pour variable réelle lorsque la classe  $\mathcal{C}$  est indexable par un nombre fini de paramètres réels. Par opposition nous parlerons de modèle non paramétrique pour variable réelle lorsque  $\mathcal{C}$  n'est indexable par un nombre fini de paramètres réels.[10]

De manière naturelle, le premier exemple qui vient l'esprit, est le modèle linéaire qui, que l'on fasse une hypothèse de loi sur  $\epsilon$

$$Y = aX + b + \epsilon; \quad a, b \in \mathbb{R}, \epsilon \sim N(0, \sigma^2) \quad (2.3)$$

Où non :

$$Y = aX + b + \epsilon; \quad (a, b) \in \mathbb{R}^2 \quad (2.4)$$

Est clairement de nature paramétrique au vu de la définition précédente.

Un exemple classique de modèle non paramétrique est celui qui consiste à poser :

$$\mathcal{C} = \text{fonctions } k \text{ fois continument dérivables} \quad (2.5)$$

## 2.3 Régression avec lissage par moyenne mobile (MA)

On suppose le modèle suivant :

$$y_i = r(x_i) + \epsilon_i; \quad i = \overline{1, n}$$

Où  $\epsilon_i$  est un bruit blanc avec  $E(\epsilon_i) = 0$  et  $E(\epsilon_i^2) = \sigma_i^2$ . On suppose que la fonction  $r(\cdot)$  est inconnue et l'on se propose d'estimer cette fonction par une méthode de lissage par moyenne mobile (MA). L'idée consiste tout simplement à appliquer une moyenne mobile aux valeurs de  $Y$  pour obtenir un estimateur de la fonction de lien.

### Définition 2.3.1

[11] L'estimateur de la fonction de lien par moyenne mobile s'écrit sous la forme suivante :

$$\hat{r}(x_i) = (\bar{Y}_j)_{x_j \in V_{k, x_i}} \quad (2.6)$$

Où  $V_{k, x_i}$  désigne un voisinage de  $x_i$  défini par les  $k$  individus ayant les valeurs de  $X$  les plus proches de  $x_i$ .

On peut donner une autre définition de cette fonction de lien.

### Définition 2.3.2

[11] Supposons que les observations  $x_i$  sont ordonnées de façon croissante  $x_1 \leq x_2 \dots \leq x_n$  et que  $k$  est un entier impair, alors :

$$\hat{r}(x_i) = \frac{1}{k} \sum_{j=\underline{i}}^{\bar{i}} Y_j \quad (2.7)$$

Où l'on a  $\underline{i} = i - (k - 1)/2$  et  $\bar{i} = i + (k - 1)/2$

De façon générale, l'estimateur MA de la fonction de lien peut s'écrire sous la forme :

$$\hat{r}(x_i) = \frac{1}{k} \sum_{j=\underline{i}}^{\bar{i}} r(x_j) + \frac{1}{k} \sum_{j=\underline{i}}^{\bar{i}} \epsilon_j \quad (2.8)$$

Etudions la convergence de cet estimateur. Pour cela, nous étudierons successivement :

1. La convergence en probabilité de  $\hat{r}(x_i)$
2. La convergence en loi de  $\hat{r}(x_i)$  afin d'en déduire des intervalles de confiance sur  $r(x_i)$

### 2.3.1 Etude de la convergence en probabilité

Considérons l'écriture suivante :

$$\widehat{r}(x_i) = \frac{1}{k} \sum_{j=\underline{i}}^{\bar{i}} r(x_j) + \frac{1}{k} \sum_{j=\underline{i}}^{\bar{i}} \epsilon_j$$

Utilisons une décomposition en séries de Taylor à l'ordre 2 de la fonction  $\widehat{r}(x_j)$  autour du point de référence  $\widehat{r}(x_i)$ . Il vient,  $\forall j = 1, \dots, n$ .

$$r(x_j) = r(x_i) + r'(x_i)(x_j - x_i) + \frac{r''(x_i)}{2}(x_j - x_i)^2 + o(x_j - x_i)^2 \tag{2.9}$$

$$\simeq r(x_i) + r'(x_i)(x_j - x_i) + \frac{r''(x_i)}{2}(x_j - x_i)^2 \tag{2.10}$$

On obtient alors :

$$\begin{aligned} r(x_i) &\simeq \frac{1}{k} \sum_{j=\underline{i}}^{\bar{i}} [r(x_i) + r'(x_i)(x_j - x_i) + r''(x_i)(x_j - x_i)^2] + \frac{1}{k} \sum_{j=\underline{i}}^{\bar{i}} \epsilon_j \\ &\simeq r(x_i) + \frac{r'(x_i)}{k} \sum_{j=\underline{i}}^{\bar{i}} (x_j - x_i) + \frac{r''(x_i)}{2k} \sum_{j=\underline{i}}^{\bar{i}} (x_j - x_i)^2 + \frac{1}{k} \sum_{j=\underline{i}}^{\bar{i}} \epsilon_j \end{aligned}$$

Or, on sait que si les  $k$  valeurs  $x_j$  sont choisies de façon symétrique par rapport à la valeur pivotale  $x_i$  on a :

$$\sum_{j=\underline{i}}^{\bar{i}} (x_j - x_i) = 0 \tag{2.11}$$

On en déduit donc finalement que :

$$\widehat{r}(x_i) \simeq r(x_i) + r''(x_i) \frac{1}{24} \left(\frac{k}{n}\right)^2 + \frac{1}{k} \sum_{j=\underline{i}}^{\bar{i}} \epsilon_j \tag{2.12}$$

Le dernier terme est une somme de  $k$  termes indépendants et indentiquement distribués dès lors, de variance finie  $\sigma_\epsilon^2/k$  ; par conséquent on obtient le résultat suivant.

**Résultat 1**

L'estimateur MA de la fonction de lien  $r(x_i)$ ;  $\forall i = 1, \dots, n$ , noté  $\hat{r}(x_i)$ ; est tel que :

$$\hat{r}(x_i) = r(x_i) + o\left(\frac{k}{n}\right)^2 + op\left(\frac{1}{k^{1/2}}\right) \quad (2.13)$$

Par conséquent, l'erreur quadratique moyenne vérifie :

$$E[\hat{r}(x_i) - r(x_i)]^2 = o\left(\frac{k}{n}\right)^4 + op\left(\frac{1}{k}\right) \quad (2.14)$$

On sait dès lors que le biais de l'estimateur est défini par :

$$\hat{r}(x_i) - r(x_i) \simeq r''(x_i) \frac{1}{24} \left(\frac{k}{n}\right)^2 + \frac{1}{k} \sum_{j=i}^{\bar{i}} \epsilon_j \quad (2.15)$$

Et que la variance de  $\hat{r}(x_i)$  est approximativement égale à :

$$\text{Var}[\hat{r}(x_i)] = \frac{\sigma_\epsilon^2}{k} \quad (2.16)$$

On en tire la conséquence suivante :

$$\lim_{\frac{k}{n} \rightarrow 0} E[\hat{r}(x_i) - r(x_i)] = 0 \quad (2.17)$$

$$\lim_{k \rightarrow 0} \text{Var}[\hat{r}(x_i) - r(x_i)] = 0 \quad (2.18)$$

De ces deux propriétés, on déduit immédiatement que :

**Résultat 2**

L'estimateur MA de la fonction de lien  $\hat{r}(x_i)$ ;  $i = 1, \dots, n$ , noté  $\hat{r}(x_i)$ ; n'est convergent que si conjointement  $k/n \rightarrow 0$  et  $k \rightarrow 0$

$$\hat{r}(x_i) \rightarrow r(x_i); \quad \text{en probabilité} \quad (2.19)$$

### 2.3.2 Etude de la convergence en distribution et intervalles de confiance

[11] Considérons l'écriture (2.11); si le nombre de points de la MA, c'est à dire  $k$ ; augmente avec  $n$ , alors par un théorème central limite, on montre que le terme de droite est asymptotiquement distribué selon une loi normale de moyenne nulle et de variance finie telle que :

$$\sqrt{k}\left(\frac{1}{k}\sum_{j=i}^{\bar{i}}\epsilon_j\right) \rightarrow N(0,\sigma_\epsilon^2) \quad (2.20)$$

Par conséquent, on en déduit que la quantité :

$$\sqrt{k}[\hat{r}(x_i) - r(x_i) - r''(x_i)\frac{1}{24}\left(\frac{k}{n}\right)^2] \quad (2.21)$$

Converge asymptotiquement vers une loi normale.

$$\sqrt{k}[\hat{r}(x_i) - r(x_i) - r''(x_i)\frac{1}{24}\left(\frac{k}{n}\right)^2] \rightarrow N(0,\sigma_\epsilon^2) \quad (2.22)$$

Supposons que la taille de la fenêtre vérifie la propriété suivante :

$$k = k(n) \quad (2.23)$$

$$= n^\alpha \quad (2.24)$$

Pour  $\alpha = 4/5$ , c'est à dire  $k = n^{4/5}$ , on déduit l'intervalle de confiance pour  $r(x_i)$  par cette définition :

#### 2.3.2 Définition

[11] Si la fenêtre  $k$  est telle que  $\lim kn^{4/5}$ ; un intervalle de confiance au seuil  $\alpha$  sur la valeur de  $r(x_i)$  pour tous les points  $x_1, \dots, x_n$  est donné par :

$$IC_\alpha = \left[\hat{r}(x_i) - C_{1-\alpha/2}\frac{\sigma_\epsilon}{\sqrt{k}}; \hat{r}(x_i) + C_{1-\alpha/2}\frac{\sigma_\epsilon}{\sqrt{k}}\right] \quad (2.25)$$

Où  $C_{1-\alpha/2}$  désigne le fractile de la loi  $N(0; 1)$ .

Nous allons à présent énoncer directement les résultats concernant la régression kernel qui pour l'essentiel ressemblent, dans l'esprit, à ceux que nous venons de démontrer dans le cas de la régression MA.

## 2.4 Régression par opérateur à noyau

L'estimateur de noyau appelé aussi l'estimateur de Nadaraya Watson, défini de la manière suivante :

$$\hat{r}_{NW}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)}; \quad \forall x \in \mathbb{N} \quad (2.26)$$

Dans cette définition  $K$  est une fonction de  $\mathbb{R}$  et  $h$  est un paramètre réel strictement positif. On peut avoir plusieurs types de noyaux sont définis dans le Tab 1.1.

### 2.4.1 Estimateur de Nadaraya par la fonction de densité conditionnelle

Soient  $(X_1, Y_1); \dots; (X_n, Y_n)$  des couples aléatoires indépendants et identiquement distribués (i.i.d.) à valeurs dans  $\mathbb{R}^2$ . Le couple de variable aléatoire  $(X; Y)$  est supposé admettre une densité jointe sur  $\mathbb{R}^2$  notée  $f_{X,Y}(\cdot, \cdot)$  et nous désignons par  $f_X(\cdot)$  la densité marginale (par rapport à la mesure de Lebesgue sur  $\mathbb{R}$ ), d'après [2] donnée par :

$$f_X(x) = g(x) = \int f(x, y) dy$$

La densité conditionnelle de  $Y$  en  $X = x$  est :

$$f_{Y/X}(y/x) = \frac{f(x, y)}{g(x)}; \quad sig(x) \neq 0$$

L'espérance conditionnelle ou la fonction de régression de  $Y$  en  $X = x$  s'écrit :

$$\begin{aligned} r(x) &= E(Y/X = x) \\ &= \int y f_{Y/X}(y/x) dy \\ &= \int \frac{y f(x, y)}{g(x)} dy \end{aligned}$$

Soit  $J(x, y)$  une fonction de densité sur  $\mathbb{R}^2$ ; on pose :

$$K(x) = \int J(x, y) dy$$

Soit  $h_n \mapsto 0$  quand  $n \mapsto \infty$ ; on a alors :

$$f_n(x, y) = \frac{1}{nh_n^2} \sum_{j=1}^n J\left[\left(\frac{x - X_j}{h_n}\right), \left(\frac{y - Y_j}{h_n}\right)\right]$$

Est un estimateur de  $f(x; y)$ .

On a aussi :

$$\begin{aligned} g_n(x) &= \frac{1}{h_n} \sum_{j=1}^n K\left[\left(\frac{x - X_j}{h_n}\right)\right] \\ &= \int f_n(x, y) dy \end{aligned}$$

Est un estimateur de  $g(x)$ . Un estimateur naturel de  $r(x)$  est donné par :

$$\begin{aligned} \hat{r}_n(x) &= \int \frac{y f_n(x, y)}{g_n(x)} dy \\ &= \int f_n(x, y) dy \end{aligned}$$

En posant  $z = \frac{y - Y_j}{h_n}$ , il vient :

$$\hat{r}_n(x) = h_n \sum_{j=1}^n m\left(\frac{x - X_j}{h_n}\right) / K\left(\frac{x - X_j}{h_n}\right) + \sum_{j=1}^n Y_j K\left(\frac{x - X_j}{h_n}\right) / K\left(\frac{x - X_j}{h_n}\right)$$

Où

$$m(x) = \int y J(x, y) dy$$

En choisissant

$$J(x, y) = K(x)L(y)$$

Avec

$$\int yL(y)dy = 0$$

Nous obtenons

$$m(x) = 0$$

D'où l'expression de l'estimateur de la régression :

$$\hat{r}_{NW}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)}; \quad \forall x \in \mathbb{N}$$

Appelé estimateur à noyau,  $K$  étant le noyau et  $h_n$  la fenêtre. Plus généralement, des estimateurs de  $r(x)$  peuvent s'écrire :

$$\hat{r}_{NW}(x) = \sum_{i=1}^n W_{n,i}(x)Y_i$$

Où  $W_{n,i}(x)$  sont des poids dépendant de  $x$  et de  $(X_1, \dots, X_n)$ . La valeur de  $W_{n,i}(x)$  dépend du type d'estimateur considéré, nous citons ci-dessous des exemples connus :

### Estimateur des plus proches voisins

Soit  $k_n$  un paramètre de l'estimation, on pose :

$$W_{n,i}(x) = \begin{cases} \frac{1}{k_n} & \text{si } X_i \text{ est parmi le voisin le plus proche de } k_n \text{ de } x \text{ dans } (X_1, \dots, X_n); \\ 0 & \text{si non} \end{cases}$$

Pour plus des détails (voir Devroye et al., 1994) [2].

### Estimateur à partitions

On utilise une partition  $\pi_n = \{A_{n,j} : j\}$  de  $\mathbb{R}^2$  et on pose :

$$W_{n,i}(x) = \sum_j \frac{I_{A_{n,j}}(X_i)}{\sum_{k=1}^n I_{A_{n,j}}(X_k)} I_{A_{n,j}}(x)$$

Pour plus des détails (voir Devroye et Györ., 1983) [2].

## 2.4.2 Propriétés de l'estimateur à noyau

Les propriétés de consistance et d'absence de biais sont reprises de Nadaraya (1989).

Posons sur la fonction noyau  $K : \mathbb{R} \rightarrow \mathbb{R}$  les hypothèses suivantes :

- $K_1$ )  $K$  est bornée ;  $\sup_{\mu \in \mathbb{R}} |K(\mu)| \leq M < \infty$
- $K_2$ )  $\lim_{|\mu| \rightarrow \infty} |K(\mu)| = 0$
- $K_3$ )  $\int_{\mathbb{R}} |K(\mu)| du < \infty$
- $K_4$ )  $\int_{\mathbb{R}} K(\mu) du = 1$
- $K_5$ )  $\forall \mu \in \mathbb{R}, K(\mu) = K(-\mu)$
- $K_6$ )  $\int_{-\infty}^{+\infty} \mu^2 K(\mu) du < \infty$

Rappelons que :

$$\hat{r}_n(x) = \frac{\hat{g}_n(x)}{\hat{f}_n(x)}$$

Où

$$\hat{g}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h_n}\right)$$

Et

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

Est l'estimateur à noyau de la densité  $f$  de  $X$ .

## 2.4.3 Consistance

### Théorème 2.4.3

Supposons que le noyau  $K$  vérifie les hypothèses  $K_1 - K_6$ , que  $E(Y^2) < \infty$  et que  $f_x(x)$  est strictement positive. Si  $h_n \rightarrow 0$ ;  $nh_n \rightarrow +\infty$  (quand  $n \rightarrow \infty$ ); alors  $\hat{r}_n(x)$  est un estimateur consistant de  $r(x)$ . Voir [2]

Preuve 2.4.3 [2]

Nous déduisons du théorème de Bochner que, lorsque  $h_n \rightarrow 0$  :

$$\begin{aligned} E[\widehat{f}_n(x)] &= E\left[\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)\right] \\ &= \frac{1}{h_n} E\left[K\left(\frac{x - X_i}{h_n}\right)\right] \\ &= \frac{1}{h_n} \int_{\mathbb{R}} K\left(\frac{x - t}{h_n}\right) f_X(t) dt \rightarrow f_X(x) \int_{\mathbb{R}} K(t) dt = f_X(x) \end{aligned}$$

Donc  $\widehat{f}_n(x)$  est un estimateur asymptotiquement sans biais. D'autre part, comme les  $X_i$  sont indépendantes et identiquement distribuées, il vient que :

$$\begin{aligned} \text{Var}[\widehat{f}_n(x)] &= \text{Var}\left[\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)\right] \\ &\leq \frac{1}{n} E\left[\frac{1}{nh_n} K\left(\frac{x - X_i}{h_n}\right)\right]^2 \\ &= \frac{1}{nh_n} \int_{-\infty}^{+\infty} \frac{1}{h_n} K^2\left(\frac{x - t}{h_n}\right) f_X(t) dt \end{aligned}$$

D'après le théorème de Bochner :

$$\int_{-\infty}^{+\infty} \frac{1}{h_n} K^2\left(\frac{x - t}{h_n}\right) f_X(t) dt \rightarrow f_X(x) \int_{\mathbb{R}} K^2(t) dt < \infty$$

quand  $n \rightarrow \infty$

donc :

$$\text{Var}(\widehat{f}_n(x)) \rightarrow 0$$

quand  $nh_n \rightarrow \infty$

Puisque  $\widehat{f}_n(x)$  est un estimateur consistant de  $f_X(x)$  ; il suffit donc de montrer que  $g_n(x)$  est un estimateur consistant de :

$$g_n(x) = \int_{-\infty}^{+\infty} y f(x, y) dy$$

Nous avons :

$$\begin{aligned}
 E[\widehat{g}_n(x)] &= E\left[\frac{1}{nh_n} \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h_n}\right)\right] \\
 &\leq \frac{1}{h_n} E\left[Y K\left(\frac{x - X_i}{h_n}\right)\right] \\
 &= \frac{1}{h_n} \int_{\mathbb{R}} E(Y|X = x) K\left(\frac{x - t}{h_n}\right) f_X(t) dt \\
 &= \frac{1}{h_n} \int_{\mathbb{R}} r(t) K\left(\frac{x - t}{h_n}\right) f_X(t) dt \\
 &= \frac{1}{h_n} \int_{\mathbb{R}} K\left(\frac{x - t}{h_n}\right) r(t) f_X(t) dt \\
 &= \frac{1}{h_n} \int_{\mathbb{R}} K\left(\frac{x - t}{h_n}\right) g(t) f_X(t) dt \rightarrow g(x)
 \end{aligned}$$

Par le théorème de Bochner :

$$\text{Var}(\widehat{g}_n(x)) = E[\widehat{g}_n(x)]^2 - E^2[\widehat{g}_n(x)] \sim \frac{1}{h_n} m(x) \int K^2(t) dt \rightarrow 0 \text{ quand } n \rightarrow \infty$$

Donc :

$$\lim_{n \rightarrow \infty} \text{Var}(\widehat{g}_n(x)) = 0$$

Ceci implique que :

$$\widehat{g}_n(x) \rightarrow g(x) \quad \text{en probabilité}$$

D'où :

$$\widehat{r}_n(x) = \frac{\widehat{g}_n(x)}{\widehat{f}_n(x)} \rightarrow \frac{g(x)}{f(x)} \quad \text{en probabilité}$$

#### 2.4.4 Absence de biais asymptotique

##### Théorème 2.4.4

Sous les conditions  $K_1 - K_6$  et si  $f_x(x)$  est strictement positive, il vient :

a) Lorsque  $Y$  est bornée p.s.,  $h_n \rightarrow 0$  et  $nh_n \rightarrow \infty$  (quand  $n \rightarrow \infty$ ) alors :

$$E(\widehat{r}_n(x)) = E[\widehat{g}_n(x)]/E[\widehat{f}_n(x)] + o((nh_n)^{-1}) \quad (2.27)$$

b) Lorsque  $E[Y^2] < \infty$ ,  $h_n \rightarrow 0$  et  $nh_n^2 \rightarrow \infty$  (quand  $n \rightarrow \infty$ ) alors :

$$E(\widehat{r}_n(x)) = E[\widehat{g}_n(x)]/E[\widehat{f}_n(x)] + o((n^{\frac{1}{2}}h_n)^{-1}) \quad (2.28)$$

a), b) et le théorème de Bochner impliquent que  $\widehat{r}_n(x)$  est un estimateur asymptotiquement sans biais de  $r(x)$ .

### 2.4.5 Etude de la convergence en probabilité

On admet le résultat suivant :

#### Proposition 2.4.5

L'estimateur à noyau de Nadaraya-Watson est convergent. Si les variables  $X$  sont distribuées selon une loi de probabilité de densité  $p(x)$  ; le numérateur converge vers  $r(x_0)p(x_0)$  et le dénominateur converge vers  $p(x_0)$  :

$$\widehat{r}(x_i) \rightarrow r(x_i); \quad \text{en probabilité}$$

Voir [11]

### 2.4.6 Etude de la convergence en loi et intervalles de confiance

Pour construire un IC sur  $r(x_i)$ , on admet le résultat suivant :

#### Proposition 2.4.6

L'estimateur à noyau de Nadaraya-Watson vérifie :

$$\sqrt{h}\sqrt{n}\{\widehat{r}(x_0) - r(x_0) - \frac{1}{2}a_K h^2[r''(x_0) + 2r'(x_0)\frac{p'(x_0)}{p(x_0)}]\} \rightarrow N(0, \frac{b_K \sigma_\epsilon^2}{p(x_0)})$$

Où  $p(\cdot)$  désigne la densité de  $x$  et :

$$a_K = \int u^2 K(u) du; b_K = \int K(u)^2 du$$

Voir Wand et Jones (1995) pour les valeurs de  $a_K$  et  $b_K$  pour de nombreux kernels. D'après [11] on admettra en particulier que :

Noyau	$b_K$
Uniforme	1/2
Triangulaire	2/3
Gaussien	1/2* $\pi$
Biweight	5/7
Triweight	350/429
Espanechnikov	3/5

TAB. 2.1 – Les noyaux usuels avec la valeur de  $b_K$

Des simplifications peuvent être apportées si la valeur de  $h$  décroît avec  $n$  plus rapidement que  $h = n^{-1/5}$ . Dans ce cas, le terme de biais disparaît et donc on obtient le résultat suivant dans le cas d'un kernel uniforme  $b_K = \frac{1}{2}$  :

$$\sqrt{h}\sqrt{n}[\hat{r}(x_0) - r(x_0)] \rightarrow N(0, \frac{\sigma_\epsilon^2}{2p(x_0)})$$

On peut donc en déduire la manière de construire des IC sur les valeurs de  $r(x_i)$ .

**Proposition**

Sous l'hypothèse que  $h$  vérifie  $hn^{-1/5} \rightarrow 0$ , l'écart type de l'estimateur à noyau de Nadaraya-Watson  $\hat{r}(x_0)$  vérifie :

$$\sqrt{h}\sqrt{n}[\hat{r}(x_0) - r(x_0)] \rightarrow N(0, \frac{b_K \sigma_\epsilon^2}{p(x_0)}) \quad [11]$$

où  $p(\cdot)$  désigne la densité de  $x$ . Un intervalle de confiance sur  $\hat{r}(x_0)$  au seuil de  $\alpha\%$  est donc défini par :

$$\text{IC}_\alpha = [\hat{r}(x_0) - C_{1-\alpha/2} s_{\hat{r}(x_0)}; \hat{r}(x_0) + C_{1-\alpha/2} s_{\hat{r}(x_0)}] \quad (2.29)$$

Où  $C_{1-\alpha/2}$  désigne le fractile de la loi  $N(0; 1)$  et où :

$$s_{\hat{r}(x_0)} = \sqrt{\frac{b_K \sigma_\epsilon^2}{\hat{p}(x_0)}} \quad (2.30)$$

Avec :

$$\hat{p}(x_0) = K\left(\frac{x_i - x_0}{h}\right) \quad (2.31)$$

La procédure pour obtenir les IC est donc la suivante :

1. On choisit  $h$  vérifie  $hn^{1/5} \rightarrow 0$  et une fonction kernel  $K(u)$ ; d'où l'on déduit  $b_K$
2. On construit l'estimateur à noyau de Nadaraya-Watson  $\hat{r}(x_0)$ . On recommence pour toutes les valeurs  $x_1, \dots, x_n$ .
3. On calcule l'estimateur de la variance des résidus

$$\sigma_\epsilon^2 = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{r}(x_i)]^2$$

4. On estime la valeur de  $p(x_0)$ , densité de  $X$  au point  $x_0$  (cf. section sur l'estimation des densités) par :

$$\hat{p}(x_0) = K\left(\frac{x_i - x_0}{h}\right)$$

5. On calcule l'intervalle sur  $\hat{f}(x_0)$  au seuil  $\alpha$  défini par :

$$\text{IC}_\alpha = \left[ \hat{r}(x_0) - C_{1-\alpha/2} \sqrt{\frac{b_K \sigma_\epsilon^2}{\hat{p}(x_0)}}; \hat{r}(x_0) + C_{1-\alpha/2} \sqrt{\frac{b_K \sigma_\epsilon^2}{\hat{p}(x_0)}} \right] \quad (2.32)$$

## 2.5 Le choix de paramètre lissage $h$

Pour choisir le paramètre de lissage  $h$ , on choisit la méthode de AMISE, pour plus de détail voir [11]

## 2.6 Validation croisée

D'après [11], Cross Validation Function est définie par la quantité :

$$CV(h) = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{r}(x, h)]^2 \quad (2.33)$$

# Chapitre 3

## Régression non paramétrique cas des séries chronologiques

### 3.1 Introduction

Après avoir vu la régression non paramétrique dans le cas réel, on se pose la question :

Comment estimer la fonction de régression et ces différentes propriétés de convergence dans le cas des séries chronologiques ?

Considérons une série temporelle  $\{(Z_i), i = 1, \dots, n\}$ , où chaque  $Z_i$  est une variable aléatoire réelle. Nous nous en tiendrons à l'estimation de la fonction d'autorégression d'ordre 1, que nous noterons plus simplement :

$$r(t) = E(Z_{i+1}|Z_i = t), \quad t \in \mathbb{R}, \quad i = \overline{1, n} \quad (3.1)$$

En posant :

$$Y_i = Z_{i+1}; X_i = Z_i \quad (3.2)$$

On se trouve en fait en face d'un problème d'estimation de la fonction de régression d'une variable réelle  $X$ , puisque l'on a de manière évidente que :

$$r(t) = E(Y_i|X_i = t), \quad t \in \mathbb{R} \quad (3.3)$$

Il est donc naturel de chercher à utiliser les méthodes de régression étudiées au partie précédente. Dans cette démarche, la seule difficulté consiste donc à adapter les techniques

précédentes au cadre de l'estimation d'une fonction de régression à partir d'échantillon non indépendant. Nous supposons que tous les couples  $(X_i, Y_i)$  définis par (3.2) sont identiquement distribués, chacun ayant même loi qu'un couple  $(X, Y)$

Pour reprendre les notations de chapitre précédent, on notera  $f$  la densité marginale de  $Z$ , densité supposée exister. Le cadre sous lequel nous envisagerons ce problème d'estimation non paramétrique, en ce sens que la définition (2.2.2) est vérifiée pour le couple  $(X, Y)$  défini ci-dessus. Plus précisément notre modèle ne fera que des hypothèses de régularité sur les fonctions  $r$  et  $f$ . Lorsque nous nous intéressons à l'estimation de  $r$  en un point  $x$  fixé de  $\mathbb{R}$  tel que

$$f(x) > 0 \quad (3.4)$$

Notre hypothèse non paramétrique consistera à supposer que pour  $k \in \mathbb{N}$  on a que :

$$r \text{ et } f \text{ sont } k \text{ fois continûment dérivable autour de } x \quad (3.5)$$

Lorsque nous nous intéresserons à une estimation globale de  $r$  sur un compact  $S$  tel qu'il existe  $\theta > 0$  pour lequel

$$\inf_{t \in S} f(t) > \theta > 0 \quad (3.6)$$

Notre hypothèse non paramétrique consistera à supposer que pour  $k \in \mathbb{N}$  on a que

$$r \text{ et } f \text{ sont } k \text{ fois continûment dérivables autour de } S \quad (3.7)$$

Le fait que l'estimation de la fonction de d'autorégression  $r$  puisse être vu comme un problème de régression particulier (3.2) et le fait que l'estimateur de Nadarya Watson combinent leur bonne propriétés mathématiques on s'intéresse dans cette partie à l'estimateur de Nadaraya et les différent type de convergence pour l'estimation de la fonction de régression pour les séries chronologique.

## 3.2 Le prédicteur à noyau

Le prédicteur à noyau (Nadaraya Watson) est défini par :

$$\hat{r}(x) = \frac{\sum_{i=1}^{n-s} Z_{i+s} K\left(\frac{x-Z_i}{h}\right)}{\sum_{i=1}^{n-s} K\left(\frac{x-Z_i}{h}\right)}; \forall x \in \mathbb{R} \quad (3.8)$$

Dans cette définition les paramètres  $h$  et  $K$  sont définis comme dans le cadre usuel de régression, c'est à dire que  $K$  est une fonction de pondération (non nécessairement positive) et que  $h = h(n)$  est un réel strictement positif.

## 3.3 Le modèle de dépendance

Pour obtenir des estimateurs convergents, il est nécessaire de faire des hypothèses d'indépendance asymptotique sur le processus observé. Les hypothèses de mélangeance sont les plus utilisées. Nous allons énoncer les principales définitions.

### 3.3.1 Le mélange fort

On définit le mélange fort par :

#### Définition 3.3.1

Soit  $\{\Delta_i, i \in \mathbb{Z}\}$  une famille de variables aléatoires à valeurs dans un même espace probabilisable  $(E, \mathcal{E})$ . Pour tout couples  $(i, j)$  dans  $\mathbb{Z} \cup \{-\infty, +\infty\}$ , on note  $\sigma_i^j$  la tribu engendrée par  $\{\Delta_k, i < k < j\}$ . On appelle coefficients de mélange fort, les réels :

$$\alpha(n) = \sup_{\{k \in \mathbb{Z}, A \in \sigma_{-\infty}^k, B \in \sigma_{n+k}^{+\infty}\}} |P(A \cap B) - P(A)P(B)| \quad (3.9)$$

Voir [10].

#### Définition 3.3.2

On dit qu'une famille  $\{\Delta_i, i \in \mathbb{Z}\}$  de variables aléatoires à valeurs dans un même espace probabilisable  $(E, \mathcal{E})$  est fortement mélangeante, ou  $\alpha$  mélangeante, si l'on a

$$\lim_{n \rightarrow \infty} \alpha(n) = 0 \quad (3.10)$$

Voir [10].

### Définition 3.3.3

On dit qu'une famille  $\{\Delta_i, i \in \mathbb{Z}\}$  de variables aléatoires à valeurs dans un même espace probabilisable  $(E, \mathcal{E})$  est algébriquement  $\alpha$  mélangeante, s'il existe deux constantes  $c, a \in \mathbb{R}^{*+}$  tel que les coefficients de mélangeante verifants :

$$\alpha(n) \leq cn^{-a} \quad (3.11)$$

Voir [10].

### 3.3.2 Une inégalité exponentielle

Les inégalités de type exponentiel relatives à des sommes de v.a permettent d'obtenir des convergences presque sûres d'estimateurs.

#### Lemme 3.3.1

##### Inégalité de type Fuk-Nagaev[10]

Soit  $\{\Delta_i, i \in \mathbb{N}\}$  est une famille de variables aléatoires à valeurs dans  $\mathbb{R}$  qui vérifient la condition de mélange fort (3.10), avec des coefficients à décroissance algébrique tel que définis en (3.11) on pose :

$$s_n^2 = \sum_{i=1}^n \sum_{j=1}^n |cov(\Delta_i, \Delta_j)| \quad (3.12)$$

Si  $\|\Delta_i\|_\infty < \infty, \forall i$ , alors on a pour tout  $\epsilon > 0$  et pour tout  $r > 1$

$$P[|\sum_{k=1}^n \Delta_k| > 4\epsilon] \leq 4(1 + \frac{\epsilon^2}{rs_n^2})^{-r/2} + 2ncr^{-1}(\frac{2r}{\epsilon})^{a+1} \quad (3.13)$$

### 3.3.3 Une inégalité de covariance

L'utilisation de l'inégalité exponentielle précédente va nécessiter le calcul d'expression de type  $s_n^2$ , et donc nécessiter l'utilisation l'inégalité de covariance pour variables mélangées, on donne

#### Lemme 3.3.2

Soit  $\{\Delta_i, i \in \mathbb{Z}\}$  est une famille de variables aléatoires à valeurs dans  $\mathbb{R}$  qui vérifient la condition de mélange fort (2.10), et tel que  $\|\Delta_i\|_\infty < \infty, \forall i$ . On a pour tout  $i \neq j$  :

$$| \text{cov}(\Delta_i, \Delta_j) | \leq 4\|\Delta_i\|_\infty\|\Delta_j\|_\infty\alpha_{|i-j|} \quad (3.14)$$

Voir [10].

## 3.4 Régression sous mélange fort

Dans le cas des séries chronologiques la différence viendra du fait que nous ferons l'hypothèse de mélangeance suivante sur les couples  $(X_i, Y_i)$  :

$$\text{La suite } \{(X_i, Y_i), i = 1, \dots, n\} \quad \text{est } \alpha \text{ mélangeante} \quad (3.15)$$

Et nous supposons que pour tout  $i \neq j$

$$(X_i, Y_i) \quad \text{admet une densité notée } f_{ij} \quad (3.16)$$

L'estimateur que nous étudierons est l'estimateur usuel de Nadaraya Watson déjà étudié dans le premier cas :

$$\hat{r}_{NW}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)}$$

Les résultats que nous allons présenter peuvent être vus comme des extensions au cadre mélangeant de ceux du chapitre précédent. Nous établissons maintenant des propriétés de convergence presque complète et moyenne quadratique, en donnant chaque fois sous deux modèles différents (modèle de continuité et modèle dérivabilité) un résultat ponctuel et un résultat uniforme.

### 3.4.1 Quelques résultats préliminaires

Pour étendre des résultats de convergence du cadre iid à un cadre de dépendance, la difficulté se traduit le plus souvent par la nécessité d'avoir à inclure des termes de type de covariance dans les calculs habituels de variance. Vu la forme des estimateurs que nous allons étudier  $\widehat{r}_{NW}$ , nous allons avoir à appliquer des inégalités du type de celles présentées ci dessus à des variables aléatoires de la forme :

$$\Delta_i = Y_i^l K\left(\frac{x - X_i}{h}\right) - EY_i^l K\left(\frac{x - X_i}{h}\right), l = 0 \text{ ou } l = 1 \tag{3.17}$$

C'est la raison pour laquelle nous avons décidé, en guise de préliminaire, de donner un résultat général concernant la quantité

$$S_n^{2*} = \sum_{i=1}^n \sum_{j \neq i} |cov(\Delta_i, \Delta_j)| \tag{3.18}$$

Les hypothèses que nous nécessiterons sont voisines de celles utilisées au chapitre précédent dans le cadre iid. Nous introduirons les conditions suivantes :

$$K \text{ est integrable, borné et à support compact} \tag{3.19}$$

$$|Y| < M < \infty \tag{3.20}$$

Et nous aurons parfois besoin de notion de noyau d'ordre  $k$  dont nous rappelons qu'elle s'écrit :

$$\int t^j K(t) dt = 0, \forall j = 1, \dots, k-1 \text{ et } 0 < \left| \int t^j K(t) dt \right| < \infty \quad (3.21)$$

Bien que nous utilisions des hypothèses et arguments différents, la démonstration de la proposition suivante s'inspire des calculs de Bosq (1996.p43).

**Proposition 3.4.1**

Sous les conditions (3.4),(3.5),(3.15),(3.16), (3.19) et (3.20) on a :

$$s_n^{2*} = o(nh) + O(n^2\alpha((h \log n)^{-1})) \quad [10] \quad (3.22)$$

Voir Preuve de cette proposition dans [10]

**Proposition 3.4.2**

Si  $S$  est un compact tel que (3.6) et (3.7) soient vérifiées, et si la condition de la proposition (3.4.1) sont vérifiées, alors on a

$$\sup_{x \in S} s_n^{2*} = o(nh) + O(n^2\alpha((h \log n)^{-1})) \quad [10] \quad (3.23)$$

Voir preuve de ces propositions dans [10]

**3.4.2 Convergence presque complète ponctuelle**

Nous allons tout d'abord donner des propositions générales qui concernent l'estimateur

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

de  $f$ , et l'estimateur

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

de  $g = rf$ . A partir de ces propositions générales nous obtiendrons quasi directement les propriétés de convergence presque complète qui nous intéressent.

**Proposition 3.4.3**

Supposons que les conditions de la proposition (3.4.1) soient vérifiées. Supposons que la condition de décroissance algébrique (3.11) soit satisfaite pour une valeur de  $\alpha$  vérifiant avec la fenêtre  $h$  les hypothèses suivantes :

$$\exists \theta > 0, \exists c_1 > 0, \exists c_2 > 0, c_2 n^{(\frac{3-\alpha}{\alpha+1} + \theta)} \leq h \leq c_1 n^{(\frac{1}{\alpha-1} - \theta)} \quad [10] \quad (3.24)$$

Alors il existe  $v > 0$  et  $\epsilon > 0$  tel que l'on ait :

$$P[| E\widehat{g}(x) - \widehat{g}(x) | > \epsilon \sqrt{\frac{\log n}{nh}}] = O(n^{-1-v}) \quad [10] \quad (3.25)$$

Et

$$P[| Ef(x) - \widehat{f}(x) | > \epsilon \sqrt{\frac{\log n}{nh}}] = O(n^{-1-v}) \quad [10] \quad (3.26)$$

**Théorème 3.4.1**

**Vitesse de convergence presque complète ponctuelle sous condition de dérivabilité.**

Considérons le modèle de dérivabilité défini par (3.4) et (3.5) avec  $k > 0$  et supposons que les conditions (3.11), (3.15), (3.16), (3.19) soient réalisées, on a :

$$\widehat{r}_{NW}(x) - r(x) = O\left(\left(\frac{n}{\log n}\right)^{\frac{-k}{2k+1}}\right), p.co \quad [10] \quad (3.27)$$

Voir [10]

### Théorème 3.4.2

**Vitesse de convergence presque complète ponctuelle sous condition de continuité.**

Considérons le modèle de continuité défini par (3.4) et (3.5) avec  $k > 0$  et supposons que les conditions (3.11),(3.15),(3.16),(3.19), (3.24) soient réalisées, on a :

$$|\widehat{r}_{NW}(x) - r(x)| \rightarrow 0, p.co \quad [10] \quad (3.28)$$

Voir [10].

### 3.4.3 Convergence presque complète uniforme

L'obtention de résultats analogue uniforme sur un compact  $S$  est subordonnée à une hypothèse supplémentaire sur le noyau  $K$  :

$$\exists \beta > 0, \exists C < \infty, \forall x \in S, \forall y \in S, |K(x) - K(y)| \leq C |x - y|^\beta \quad [10] \quad (3.29)$$

Nous aurons besoin de renforcer la condition (3.24) de la façon suivante :

$$\exists \theta > 0, \exists c_1 > 0, \exists c_2 > 0, c_2 n^{\left(\frac{(3-\alpha)\beta}{\beta(\alpha+1)+2\beta+1} + \theta\right)} \leq h \leq c_1 n^{\left(\frac{1}{\alpha-1} - \theta\right)} \quad (3.30)$$

### Remarque

Pour fixer les idées considérons le cas particulier où  $\beta = 1$ , et un modèle de dérivabilité (i.e un modèle avec  $k > 0$ ). La condition (3.30) est satisfaite dès que la largeur de fenêtre est de la forme habituelle :

$$h = C\left(\frac{\log n}{n}\right)^{\frac{1}{2k+1}} \quad [10] \quad (3.31)$$

Et quand le taux de mélange vérifie

$$\alpha > \max\{2k + 2; 4 + 4/k\} \quad [10] \quad (3.32)$$

### Théorème 3.4.3

#### Vitesse de convergence presque complète uniforme sous condition de dérivabilité

Considérons le modèle de dérivabilité défini par (3.6) et (3.7) avec  $k > 0$  et supposons que les conditions (3.11),(3.15),(3.16),(3.30),soient réalisées. On a :

$$\sup_{x \in S} | \hat{r}_{NW}(x) - r(x) | = O\left(\left(\frac{n}{\log n}\right)^{\frac{-k}{2k+1}}\right), p.co \quad [10] \quad (3.33)$$

Voir [10].

### Théorème 3.4.4

#### convergence uniforme presque complète sous condition de continuité

Considérons le modèle de continuité défini par (3.6) et (3.7) avec  $k = 0$  et supposons que les conditions (3.11),(3.15),(3.16),(3.30), soient réalisées. On a :

$$\sup_{x \in S} | \hat{r}_{NW}(x) - r(x) | \rightarrow 0, p.co \quad [10] \quad (3.34)$$

Voir [10].

### 3.4.4 Convergence en moyenne quadratique

Nous allons énoncer maintenant des résultats de convergence en moyenne quadratique, sous des modèles de dérivabilité ou des modèles de continuité , dans chaque nous donnerons une version ponctuelle et une version intégrée sur un compact

**Théorème 3.4.4****Convergence en moyenne quadratique ponctuelle sous condition de dérivabilité**

Supposons que les conditions du théorème (3.4.1) soient vérifiées. Supposons que la structure de dépendance soit celle définie par les conditions (3.11),(3.15),(3.16) avec

$$\alpha > 6 \quad \text{et} \quad h = Cn^{-1/5} \quad [10]$$

Alors on a :

$$E[\widehat{r}_{NW}(x) - r(x)]^2 = O(n^{-4/5}) \quad (3.35)$$

Voir [10].

**Théorème 3.4.6****Erreur quadratique moyenne intégrée sous condition de dérivabilité**

Supposons que les conditions du théorème (3.4.1) soient vérifiées. Supposons que la structure de dépendance soit celle définie par les conditions (3.11),(3.15),(3.16) avec

$$\alpha > 6 \quad \text{et} \quad h = Cn^{-1/5}$$

Alors on a :

$$EQMI(\widehat{r}_{NW}) = O(n^{-4/5}) \quad (3.36)$$

Voir [10].

**Théorème 3.4.7****Convergence en moyenne quadratique ponctuelle sous condition de continuité**

Supposons que les conditions du théorème (3.4.1) soient vérifiées. Supposons que la structure de dépendance soit celle définie par les conditions (3.11),(3.15),(3.16) avec

$$h^{(\alpha-1)} = On^{-(1+\theta)}$$

Alors on a :

$$E[\hat{r}_{NW}(x) - r(x)]^2 \rightarrow 0 \quad (3.37)$$

Voir [10].

### **Théorème 3.4.8**

**Erreur quadratique moyenne intégrée sous condition de continuité** Supposons que les conditions du théorème (2.4.1) soient vérifiées. Supposons que la structure de dépendance soit celle définie par les conditions (2.15),(2.16) avec

$$h^{(\alpha-1)} = On^{-(1+\theta)}$$

Alors on a :

$$EQMI(\hat{r}_{NW}) \rightarrow 0 \quad (3.38)$$

Voir [10].

## **3.5 Propriétés de convergence presque complète du prédicteur à noyau**

- Supposons que les conditions de théorème (3.4.1) soient réunis pour les variable  $(X_i, Y_i)$  définis par (3.2). Alors on a

$$\hat{r}(x) - r(x) = o\left(\frac{n}{\log n}\right)^{\frac{-k}{2k+1}}; \quad p.co \quad (3.39)$$

- Supposons que les conditions de théorème (3.4.2) soient réunis pour les variable  $(X_i, Y_i)$  définis par (3.2). Alors on a

$$\hat{r}(x) - r(x) \rightarrow 0; \quad p.co \quad (3.40)$$

• Supposons que les conditions de théorème (3.4.3) soient réunis pour les variable  $(X_i, Y_i)$  définis par (3.2). Alors on a

$$\sup_{x \in S} |\hat{r}(x) - r(x)| = o\left(\frac{n}{\log n}\right)^{\frac{-k}{2k+1}}; \quad p.co \quad (3.41)$$

• Supposons que les conditions de théorème (3.4.4) soient réunis pour les variable  $(X_i, Y_i)$  définis par (3.2). Alors on a :

$$\sup_{x \in S} |\hat{r}(x) - r(x)| \rightarrow 0; \quad p.co \quad (3.42)$$

### 3.5.1 Convergence en moyenne quadratique du prédicteur à noyau

• Supposons que les conditions de théorème (3.4.5) soient réunis pour les variable  $(X_i, Y_i)$  définis par (3.2). Alors on a

$$E |\hat{r}(x) - r(x)| = o(n^{-4/5}) \quad (3.43)$$

• Supposons que les conditions de théorème (3.4.6) soient réunis pour les variable  $(X_i, Y_i)$  définis par (3.2). Alors on a

$$EQMI(\hat{r}) = o(n^{-4/5}) \quad (3.44)$$

• Supposons que les conditions de théorème (3.4.7) soient réunis pour les variable  $(X_i, Y_i)$  définis par (3.2). Alors on a

$$E |\hat{r}(x) - r(x)| \rightarrow 0 \quad (3.45)$$

• Supposons que les conditions de théorème (3.4.8) soient réunis pour les variable  $(X_i, Y_i)$  définis par (3.2). Alors on a

$$EQMI(\hat{r}) \rightarrow 0 \quad (3.46)$$

# Chapitre 4

## Prévision non paramétrique des séries chronologiques

### 4.1 Introduction

Soient  $X_1, \dots, X_T$  une série chronologique, on voudrait prédire à partir des valeurs observées  $X_{T+h}$  où  $h$  est l'horizon de prévision ( $h \in \mathbb{N}^*$ ). Le prédicteur de  $X_{T+h}$  est basé sur  $X_1, \dots, X_T$  est donné par

$$\mathbf{E}(X_{T+h} | X_T, \dots, X_1) \tag{4.1}$$

Cette espérance conditionnel est impossible à l'estimer si on fait pas l'hypothèses supplémentaires sur le processus  $(X_t)_{t \in \mathbb{Z}}$ .

L'expression (4.1) peut bien entendu s'estimer par une régression de  $X_{T+h}$  sur les variables  $X_1, \dots, X_T$ . Tout ce qui a été vu précédemment sur la régression non paramétrique peut donc s'appliquer en terme d'autorégression non paramétrique sur le processus  $(X_t)_{t \in \mathbb{Z}}$ . La seule différence notable est qu'on perd l'indépendance qu'on avait dans le cas de la régression. Si on suppose à priori que le processus est  $k$ -Markovien, alors :

$$\mathbf{E}(X_{T+h} | X_T, \dots, X_1) = \mathbf{E}(X_{T+h} | X_T, \dots, X_{T-k+1})$$

On cherche alors à régresser  $X_{T+h}$  sur son propre passé proche. Une telle démarche permet souvent d'obtenir une prévision de  $X_{T+h}$  même si le processus n'est pas Markovien. Plusieurs méthodes non paramétriques applicables à la prévision, on cite la méthode de

la médiane conditionnelle, la méthode de mode conditionnel, pour ces deux méthode on donne la définition de chaque méthode :

### a) La méthode de la médiane conditionnelle

soit :

$$F_{X_T}^{X_1, \dots, X_{T-1}} = P(X_T \leq \lambda | X_1 = x_1, \dots, X_{T-1} = x_{T-1}) \quad [4]$$

La fonction de répartition conditionnelle de  $X_T$ , elle admet une densité si cette loi conditionnelle est absolument continue par rapport à la mesure, la médiane conditionnelle consiste à considérer la médiane :

$$m(x_1, \dots, x_{T-1}) = \inf \left\{ y \in \mathbb{R} \mid F_{X_T}^{X_1, \dots, X_{T-1}}(x_1, \dots, x_{T-1}; \lambda) \geq \frac{1}{2} \right\}$$

On utilise comme prédicteur :

$$\widehat{X}_T = \widehat{m}(x_1, \dots, x_{T-1})$$

Où  $\widehat{m}$  est un estimateur de la médiane conditionnelle  $m(x_1, \dots, x_{T-1})$ , qui nécessite de manière primilaire, une estimation de la fonction de répartition conditionnelle  $F_{X_T}^{X_1, \dots, X_{T-1}}(x_1, \dots, x_{T-1}; \lambda)$ .

### b) La méthode de mode conditionnel

Soit  $f_{X_T}^{X_1, \dots, X_{T-1}}(x_1, \dots, x_{T-1})$  la fonction de densité, le mode conditionnel consiste à considérer le mode :

$$\theta(x_1, \dots, x_{T-1}) = \arg_{\lambda \in \mathcal{S}} \sup f_{X_T}^{X_1, \dots, X_{T-1}}(x_1, \dots, x_{T-1}; \lambda) \quad [4]$$

Où est un sous ensemble de  $\mathbb{R}$ , on définit le prédicteur par :

$$\widehat{X}_T = \widehat{\theta}(x_1, \dots, x_{T-1})$$

Dans la suite de ce chapitre on s'intéresse au prédicteur par la régression non paramétrique à noyau, la définition et interprétation de ce prédicteur dans le cas d'un processus

stationnaire et dans le cas d'un processus non stationnaire est l'étude de cette partie pour passer aux différents étapes de prévision non paramétrique des séries chronologiques nous définirons le choix de tracateur et la similarité.

## 4.2 Prédicteur à noyau

Soit  $X_{n,(r)}$  le  $r$  uple de variable aléatoire suivant :

$$X_{n,(r)} = (X_n, X_{n-1}, \dots, X_{n-r+1}) \quad [12] \quad (4.2)$$

$X = (X_n)_{n \in \mathbb{N}}$  désigne le processus, pour l'horizon de prévision fixe noté  $h$ , la fonction d'autorégression  $f$  notée par :

$$f(x) = \mathbf{E}(X_{n+h} | X_{n,(r)} = x) \text{ pour } x \in \mathbb{R}^r \quad (4.3)$$

L'estimateur à noyau de la fonction  $r$  est défini par :

$$f_n(x) = \frac{\sum_{t=r, \dots, n-h} \mathbf{K}((x - X_{t,(r)})/h_n) X_{t+h}}{\sum_{t=r, \dots, n-h} \mathbf{K}((x - X_{t,(r)})/h_n)} \quad (4.4)$$

Où  $K$  est un noyau positif,  $h = (h_n)_n$  est une suite réelle positive décroissante. Le prédicteur à noyau s'en déduit :

$$X_{n+h|n} = f_n(X_{n,(r)}) \quad (4.5)$$

L'estimateur de (3.4) peut s'écrire aussi lorsque  $n$  et  $h$  fixés sous la forme suivante :

$$f_n(x) = \sum_{t=r, \dots, n-h} w_{n,t} X_{t+h} \quad (4.6)$$

Où définit par :

$$w_{n,t} = \frac{K\left(\frac{x - X_{t,(r)}}{h_n}\right)}{\sum_{m=r, \dots, n-h} K\left(\frac{x - X_{m,(r)}}{h_n}\right)} \quad [12] \quad (4.7)$$

C'est à dire  $X_{n+h|n}$  s'obtient comme le barycentre des  $(X_{t+h})_t$  affectés des poids  $(w_{n,t})$ , si  $K$  un noyau positif choisi de sorte que  $K(x)$  décroisse lorsque  $\|x\|$  croit, alors  $w_{n,t}$  admet l'interprétation suivante :

- $X_{n,(r)}$  est le bloc témoin auquel les autres blocs  $(X_{n,(r)})$  sont comparés.
- $w_{n,t}$  eut alors s'interpréter comme un indice de similarité du bloc  $X_{t,r}$  au bloc témoin.

La prévision  $X_{n+h|n}$  s'obtient donc comme le barycentre des futurs d'horizon  $h$  de tous les blocs de longueur  $r$  du passé affectés des coefficients de similarité  $w_{n,t}$  de ces blocs de bloc témoin.

En outre, on peut remarquer dans (4.7) que les numérateurs de ces poids  $(w_{n,t})_t$  sont indépendantes de l'horizon de prévision  $h$ . Les mêmes poids, calculés une fois pour toutes  $n$ , étant donné, sont utilisés pour calculer les prévisions à un horizon quelconque, après une troncature dans le temps est une normalisation convenable. A des fin d'interprétation de la prévision, il convient d'examiner les instants  $t$  aux quels correspondent les grandes valeurs des poids.

### 4.3 Interprétation de prédicteur

Soit  $X(t)_{t \in \mathbb{Z}}$  un processus strictement stationnaire de carré intégrable, le prédicteur naturel de  $X_{T+h}$  ( $h \in \mathbb{N}^*$ ), appelé l'horizon de prévision, donné par :

$$\mathbf{E}(X_{T+h} | X_T, \dots, X_1)$$

Dans le cas particulier où le processus est  $k$  markovien, on a alors :

$$\mathbf{E}(X_{T+h} | X_T, \dots, X_1) = \mathbf{E}(X_{T+h} | X_T, \dots, X_{T-k+1}) \quad (4.8)$$

Sous des hypothèses adaptées précédemment, les quantités  $\mathbf{E}(X_{T+h} | X_T, \dots, X_1)$  et la prévision à l'horizon 1 se calcule en faisant la moyenne des  $X_{t+1}$ , pour  $t$  variant, avec les poids  $\hat{w}_{t,T,r}$ , même chose pour la prévision à l'horizon 2, en faisant la moyenne des  $X_{t+2}$  avec même pondération de  $\hat{w}_{t,T,r}$

#### 4.3.1 Processus stationnaire

Soit  $X(t)_{t \in \mathbb{Z}}$  un processus strictement stationnaire de carré intégrable,

Il existe un paramètre  $r$  coïncide avec  $k$  si le processus est  $k$  markovien, s'appelle le paramètre de troncature peut s'interpréter comme l'ordre de processus markovien approchant asymptotiquement le processus  $(X_t)$ .

### 4.3.2 Processus non stationnaire

Soit  $X(t)_{t \in \mathbb{Z}}$  un processus non stationnaire est défini par :

$$X_T = T_t + \eta_t; t \in \mathbb{Z} \quad (4.9)$$

Tel que  $T_t$  est la tendance et  $\eta_t$  le processus stationnaire, on estime la tendance  $T_t$  par :

$$\hat{\eta}_t = X_T - \hat{T}_t; t \in \mathbb{Z} \quad (4.10)$$

Dans ce cas on peut déduire les prédictions de processus initial  $(X_t)_{t \in \mathbb{Z}}$  si la tendance n'est plus éliminer.

Alors d'après ce que on a vu précédement, dans l'approche non paramétrique consiste asymptotiquement à approcher l'espérance conditionnel  $\mathbb{E}(X_{T+h}|X_T, \dots, X_1)$  et  $(X_{T+h}|X_T, \dots, X_{T-r+1})$ .

L'étude de ce problème à été menée par Bosq (1991 et 1996).

Supposons  $(X_t)$  un processus réel markovien, et que nous voulions prévoir  $X_{T+1}$  à partir de  $X_1, \dots, X_T$ .

Notons la densité de  $(\eta_1, \eta_2)$ ,  $f$  la densité de  $\eta_1$  et  $m$  la régression de  $\eta_2$  sur  $\eta_1$   $\varphi = r f$

Faisons sur la tendance  $(T_t)$  l'hypothèse suivantes [4] :

$H_1$   $(T_t)$  est bornée, il existe  $\bar{f}, \bar{\varphi}, \delta$  positifs tel que :

$$T^\delta \sup_{x \in \mathbb{R}} \left| \frac{1}{T} \sum_{i=1}^T f(x - T_t) - \bar{f}(x) \right| \rightarrow 0$$

Et

$$T^\delta \sup_{x \in \mathbb{R}} \left| \frac{1}{T} \sum_{i=1}^T \int y g(x - T_t, y - T_t) dy - \bar{\varphi}(x) \right| \rightarrow 0$$

Si  $f$  et  $\varphi$  sont bornées et si  $(T_t)$  est périodique de période  $s$  alors  $H_1$  vérifiée avec :

$$\bar{f}(\lambda) = \frac{1}{s} \sum_{t=1}^s f(\lambda - T_t)$$

Et

$$\bar{\varphi}(\lambda) = \frac{1}{s} \sum_{t=1}^s \int y g(x - T_t, y - T_t) dy$$

Si  $f$  lipschisienne et si  $T_t \rightarrow \alpha$  (respectivement si  $(T_t)$  est bornée et si  $\frac{1}{T} \sum_{t=1}^T |T_t| \rightarrow 0$  alors  $\bar{f}(\lambda) = f(\lambda - \alpha)$  (respectivement  $\bar{f} = f$ ).

De plus si  $\varphi$  lipschisienne, si  $f$  est bornée, si  $(T_t)$  est bornée, si  $\frac{1}{T} \sum_{t=1}^T |T_t| \rightarrow 0$ , alors  $\bar{\varphi} = \varphi$

La présence de point aberrant dans la série ne remet pas en cause la condition de convergence de :  $\frac{1}{T} \sum_{t=1}^T |T_t| \rightarrow 0$

On peut alors définir une pseudo-régression par :

$$\bar{r}(x) = \begin{cases} \frac{\bar{\varphi}(x)}{\bar{f}(x)} & \text{si } \bar{f}(x) > 0 \\ E(\eta_1) + \limsup \frac{1}{T} \sum_{t=1}^T T_t & \text{si } \bar{f}(x) = 0 \end{cases}$$

$\bar{r}(x)$  est une approximation au  $E(X_{T+1}|X_T = x)$ , l'estimateur à noyau :

$$r_T(X_T) = \frac{\sum_{t=1}^{T-1} X_{T+1} K\left(\frac{X_T - X_t}{\delta_T}\right)}{\sum_{t=1}^{T-1} K\left(\frac{X_T - X_t}{\delta_T}\right)}$$

Le choix de la largeur de fenêtre  $\delta$  égal à :

$$\delta(T) = \frac{1}{T^{\frac{1}{4+r}}} \hat{\sigma}(T) \quad (4.11)$$

Où

$$\hat{\sigma}(T) = \frac{1}{T} \sum_{t=1}^T (X_t - \bar{X}_t)^2 \quad (4.12)$$

## 4.4 Similarité

Le prédicteur se décompose donc, d'après ce qui précède, en deux phases :

- Rechercher des périodes semblables aux  $r$  dernière observation : on utilise des poids.
- calculer le barycentre des futures de ces blocs affectés des poids.

La première phase mérite d'être isolée et permet l'étude des périodes semblables à deux fins : description de la chronique et extraction des connaissances issue de la description (classe de périodes, décroissance des poids,..) pour aider au calibrage des paramètres de prédicteur.

Dans cette phase afin de prendre en compte l'instationnarité de la chronique, deux indices de similarité sont considérés. Ils sont basés sur (3.6) et sont définis comme suit, ayant fixé un instant  $n$

### Similarité entre blocs bruts

Définit par :

$$\text{sim}_1(x, X_{t,(r)}) = \mathbf{K}\left(\frac{x - X_{t,(r)}}{h}\right) \quad [4] \quad (4.13)$$

### Similarité entre blocs centrés

Définit par :

$$\text{sim}_1(x, X_{t,(r)}) = \mathbf{K}\left(\frac{c(x) - c(X_{t,(r)})}{h}\right) \quad [4] \quad (4.14)$$

Où  $c(y) = y - m(y)$  tel que  $m(y)$  est la moyenne de bloc  $y$ .

Dans ce cadre, pour construire la prévision, on remplacera  $X_{t+h}$  par  $Z_{t+h}$  avec :

$$Z_{t+h} = X_{t+h} + (m(X_{n,(r)}) - m(X_{t,(r)})) \quad (4.15)$$

Et dans  $w_{n,t} \text{sim}_1$  par  $\text{sim}_2$

## 4.5 Choix de paramètre de troncation $r$

Pour un  $r$  fixé, tel que  $r$  est inconnu et doit être aussi estimé. On appellera  $\hat{r}$  un tel estimateur respectant  $E(X_{t+h}|X_T, \dots, X_{T-\hat{r}+1})$  soit plus proche de  $E(X_{t+h}|X_T, \dots, X_1)$ .

On propose un prédicteur  $X_{T+h}$ , pour chaque entier  $t$  avec  $r + h - 1 \leq t \leq T - h$  sous la forme :

$$X_{T+h}^{(r)} = \sum_{j=r}^{t-h} \hat{w}_{j,T}^{(r)} X_{j+h} \quad [4] \quad (4.16)$$

Ce qui donne une erreur de prévision estimée :

$$e_t^h = |X_{j+h}^{(r)} - X_{j+h}| \quad (4.17)$$

Il semble alors très classique d'utiliser un critère d'erreur quadratique pour estimer le  $r$  adéquat, c'est à dire minimiser :

$$E_T^{(r)} = \frac{1}{T - 2h - r} \sum_{r+h-1}^{T-h} (e_t^h)^2 \quad (4.18)$$

Quans  $h$  est grand, on utilise aussi des prédicteurs sous la forme :

$$\widehat{X}_{t+h}^{r,p} = \frac{1}{p} \sum_{j=t-h-p+1}^{t-h} w_{j,t}^{r,p} X_{j+h} \quad (4.19)$$

Où :

$$w_{j,t}^{r,p} = \frac{K \frac{X_j^{(r)} - X_t^{(r)}}{\delta(T)}}{\sum_{j=t-h-p+1}^{t-h} K \left( \frac{X_j^{(r)} - X_t^{(r)}}{\delta(T)} \right)} \quad (4.20)$$

Le prédicteur  $\widehat{X}_{t+h}^{r,p}$  donne une estimation à  $X_{t+h}$  à partir des observations  $X_{j+h}$  pour  $j = t - h, \dots, t - h - (p - 1)$ . Pour  $p$  est fixé, pour  $t$  variant, on obtient  $\widetilde{T}$  prévision  $\widehat{X}_{t+h}^{r,p}$  avec la meme précision, ou  $\widetilde{T}$  verifie :

$$t - h - (p - 1) \geq r$$

et

$$t \leq T - h$$

c'est à dire :

$$\widetilde{T} + p + r = T - 2h + 1 \quad (4.21)$$

On choisi  $\widehat{r}$  la valeurde  $r$  minimisant :

$$E_{T,h}^{r,p} = \frac{1}{\widetilde{T}} \sum_{j=r+h+p-1}^{T-h} (\widehat{X}_{t+h}^{r,p} - X_{t+h})^2 \quad (4.22)$$

Pour déterminer  $\widehat{r}$  il faut choisir  $p$ , d'après (4.21) on peut choisir  $p = \widetilde{T}$  [4].

## 4.6 Intervalle de prévision

On va réutiliser les erreurs  $e_t^h$  pour  $t = r + h - 1, \dots, T - h$ . On définit  $\hat{q}^h$  le quantile empirique à 95/ basé pour les  $e_t^h$ , pour cela notons  $N_t^h$  le nombre de  $e_t^h$  supérieur à  $e_t^h$  et  $t_0$  un indice tel que  $N_{t_0}^h$  soit le plus grand entier vérifiant :

$$(T-2h-r+1)^{-1}N_{t_0}^h \leq \alpha \quad [4]$$

On a alors :

$$\hat{q}^h = e_{t_0}^h$$

Ce qui nous donne un estimé de l'intervalle de prévision (à 95) pour  $X_{T+h}$  :

$$[X_{T+h} - \hat{q}^h; X_{T+h} + \hat{q}^h]$$

Puis que  $(X_t)$  Stationnaire, et que la loi de  $(X_t)$  est continue,  $\hat{q}_\alpha^h$  est un estimateur raisonnable de la vraie valeur  $q_{T,\alpha}^h$  définit par :

$$P(|X_{T+h} - \hat{X}_{T+h}| \leq q_{T,\alpha}^h) = \alpha$$

Cette procédure fournit alors un intervalle de prévision à  $\alpha\%$  (souvent pris à 95%) pour  $X_{T+h}$  il y a 95 chances sur 100 que la valeur prévue de  $X_{T+h}$  appartienne à l'intervalle :

$$[X_{T+h} - \hat{q}^h; X_{T+h} + \hat{q}^h]$$

On pourra remarquer que cette manière de fournir un intervalle de prévision ne fait aucun hypothèse sur la forme des lois sous jacentes (Box et Jenkins font l'hypothèse de normalité pour fournir un intervalle de prévision). De plus il existe des points aberrants, ils se retrouveront dans les  $(1 - \alpha\%)$  restant lorsqu'on ordonne les erreurs par ordre croissant, n'interviendront nullement dans le calcul de l'intervalle de révision, ce qui donne un argument supplémentaire quand à la robustesse de la méthode.

## 4.7 Comparaison

Pour comparer les résultats (prévision) de notre série chronologique par la méthode paramétrique Box et Jenkins à celle méthode non paramétrique, on utilise deux critères de comparaison sont :

### 4.7.1 Critère EMRO

Le critère EMRO nommer par l'erreur moyenne relative observée définit par :

$$\text{EMRO} = \frac{1}{h} \sum_{i=T-h+1}^T \frac{|X_i - \hat{X}_i|}{|X_i|} \quad [4] \quad (4.23)$$

Qui mesure l'écart relatif entre les prévisions et les observations. Si on designe par  $\hat{q}_i$  le quantile estimé du quantile théorique  $q_i$  à 95% définit par :

$$P(|X_i - \hat{X}_i| < q_i) = 0.95 \quad (4.24)$$

### 4.7.2 Critère EMRP

Le critère EMRP nommer par l'erreur moyenne relative de prévision définit par :

$$\text{EMRP} = \frac{1}{h} \sum_{i=T-h+1}^T \frac{\hat{q}_i}{|\hat{X}_i|} \quad [4] \quad (4.25)$$

Qui mesure la qualité relative des prévisions.

# Chapitre 5

## Application

### 5.1 Introduction

Nous présentons, dans ce chapitre, le travail de simulation effectué pour accorder notre théorie à une application, pour cela on prend une série chronologique représente la production des pots de yaourt produites mensuellement pendant 5 ans, de janvier 2010 jusqu'à décembre 2014, au niveau de la SARL RAMEDY.

Dans la première partie de ce chapitre, on applique la méthode paramétrique de prévision (méthodologie de Box et Jenkins), ensuite dans la deuxième partie, on applique la méthode non paramétrique de prévision. On termine par une interprétation et une comparaison des résultats obtenus des deux méthodes de la prévision. Ces deux applications sont faites avec le logiciel R.

### 5.2 Prévision paramétrique

Afin de prédire les valeurs futures pour notre série, on applique l'une des méthodes de prévision paramétrique, celle de Box et Jenkins

#### 5.2.1 La méthodologie de Box et Jenkins

La méthodologie de Box et Jenkins vise à formuler un modèle permettant de représenter une chronique avec comme finalité de prévoir des valeurs futures. De ce fait, l'objectif de cette méthodologie est de modéliser une série temporelle en fonction de ses valeurs passées pour déterminer le processus ARMA

Cette méthodologie suggère une procédure à quatre étapes essentielles :

- Identification du modèle.
- Estimation du modèle.
- Validation du modèle.
- Prévision.

### 5.2.2 Identification de modèle

L'identification est le processus dont le quel on reconnait la classe d'appartenance d'un modèle. Elle repose sur le principe suivant : les données dans une série chronologique, sont liées entre elles. La forme et la force de cette interdépendance caractéristiques sont indiquées statistiquement par : l'autocorrélation que nous abrègerons par ACF (autocorrelation function) et l'autocorrélation partielle ou PACF (partial autocorrelation function). L'étude du comportement de ces deux paramètres permet de classer chaque série chronologique dans l'une des trois catégories suivantes :

- Processus moyenne mobile (*MA*).
- Processus autoregressif (*AR*)
- Processus mixte (*ARMA*).
- Processus mixte (*ARIMA*).
- Processus mixte (*SARIMA*).

#### Remarque

Notre serie chronologique n'est pas stationnaire d'après notre test (PP.test tel que :la p.value = 0.02138) alors on fais une différenciation pour notre série.

Après avoir différencier notre serie chronologique et d'après notre test (PP.test tel que :la p.value = 0.01) alors notre série est stationnaire, on déduit que le modèle obtenu est le modèle ARIMA avec un ordre de différenciation ( $d=1$ ).

On passe à l'identification de l'ordre par les deux graphes de la fonction de l'autocorrelation partiele et le graphe de l'autocorrelation.

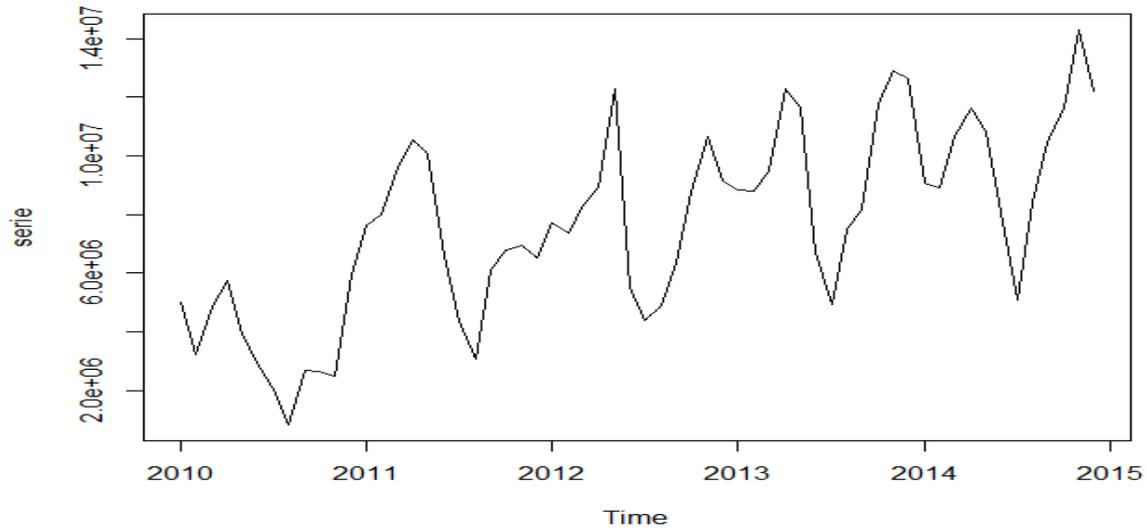


FIG. 5.1 – La série chronologique des pots de yaourt (2010-2014)

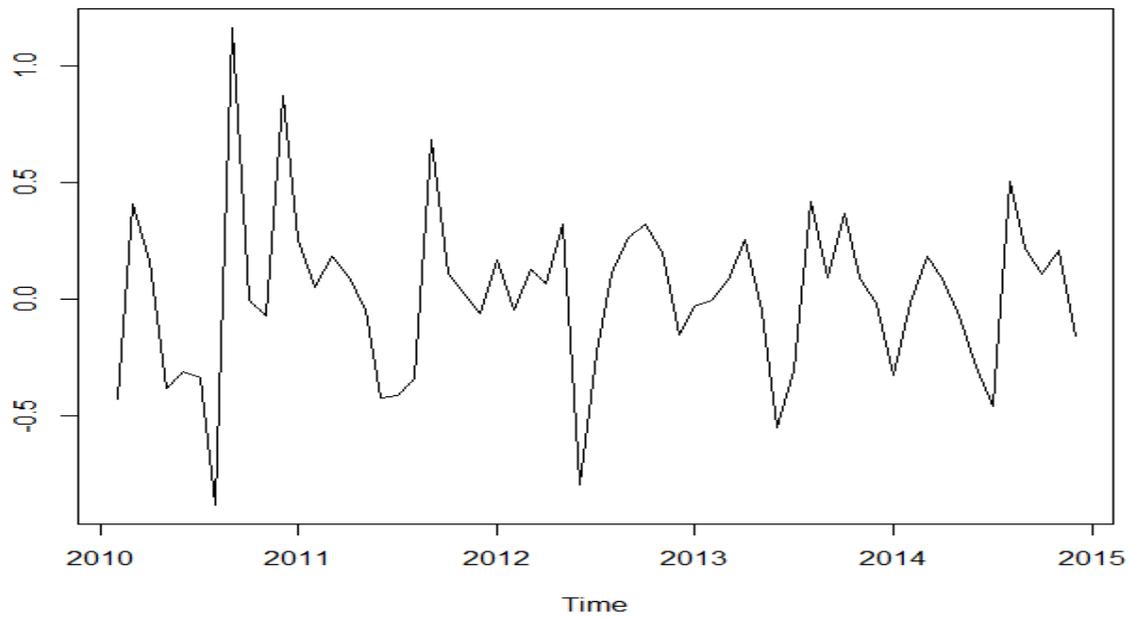


FIG. 5.2 – La série chronologique différenciée des pots de yaourt (2010-2014)

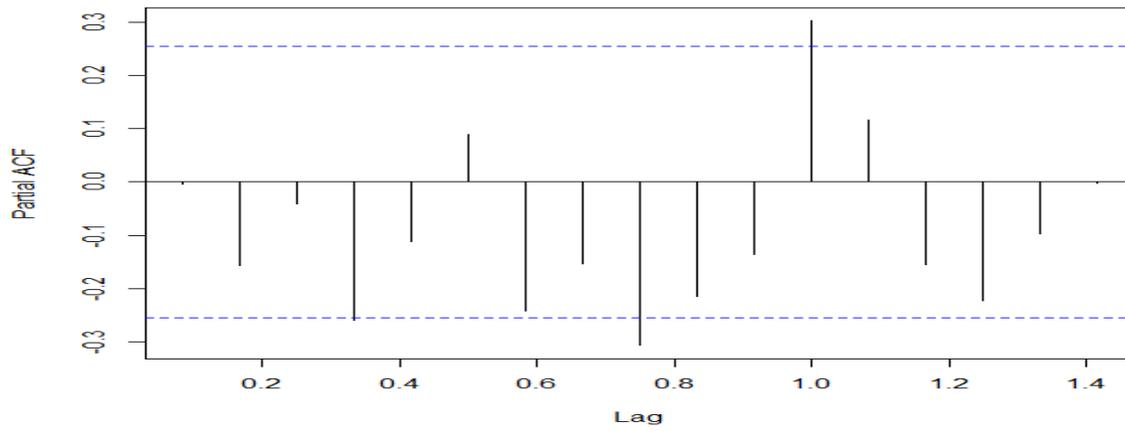


FIG. 5.3 – Le graphe de la fonction de l'autocorrelation partielle

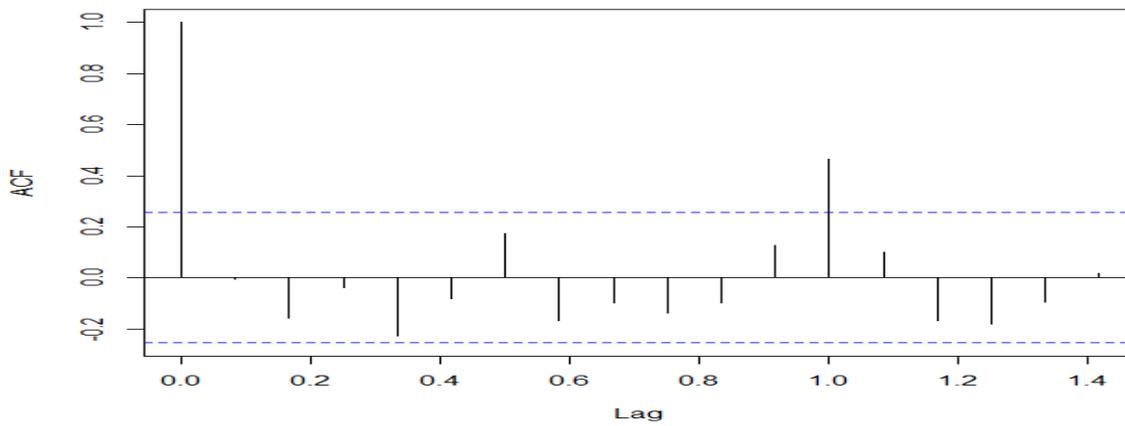


FIG. 5.4 – Le graphe de la fonction de l'autocorrelation

**Remarque**

D'après les graphes de PACF et l'ACF nous allons choisir tous les modèles possibles.

Pour cela, nous allons essayer toutes les combinaisons d'ARIMA(p,d,q) tel que  $d = 1$  on pose  $p \in \{1, \dots, 11\}$ ,  $q \in \{1, \dots, 12\}$ .

**5.2.3 L'estimation du modèle**

L'estimation des paramètres d'un modèle peut être vue comme étant un raffinement de l'analyse faite lors de l'étape d'identification. En effet, une fois les autocorrélations et autocorrélations partielle sont calculées et qu'un modèle de base (*AR, MA, ou ARMA*) a été choisi en suivant quelques principes, l'estimation consiste à calculer les paramètres requis par le modèle en question et à discuter de leur qualité et de leur aptitude à modéliser la série donnée.

Il existe plusieurs méthodes d'estimation de ces paramètres alors pour le test de student, il s'agit de tester si c'est paramètres sont considérés comme significatifs (acceptables) ou sont à rejeter.

On peut aussi procéder à l'estimation par intervalle de confiance. Si les bornes de l'intervalle sont de même signe alors le paramètre ne peut pas prendre la valeur zéro et, dans ce cas, ce paramètre peut être considéré significativement différent de zéro au seuil de signification  $\alpha$  donné. Le résultat obtenu par le test de student et toujours confirmé par l'estimation par intervalle de confiance.

**Remarque**

Les meilleurs modèles obtenus d'après le critère d'AIC sont :

- Le modèle ARMA (11, 1, 7) avec un AIC= 1861,57.
- Le modèle ARMA (10, 1, 7) avec un AIC = 1862.9.

Les paramètres estimés de cette modélisation de modèle ARIMA (11,1,7) et les intervalles de confiance estimés sont contenus dans le tableau suivant :

paramètre	paramètre estimé	intervalle de confiance	AIC
AR(1)	-0.1112	[-0.6;0.3]	1862.9
AR(2)	-1.1134	[-1.4;-0.8]	1862.9
AR(3)	-0.4038	[-0.9;0.1]	1862.9
AR(4)	-0.7219	[-1.2;-0.2]	1862.9
AR(5)	-0.8637	[-1.3;-0.4]	1862.9
AR(6)	-0.3278	[-0.8;0.1]	1862.9
AR(7)	-0.9817	[-1.2;-0.7]	1862.9
AR(8)	-0.3204	[-0.7;0.1]	1862.9
AR(9)	-0.7673	[-1.1;-0.4]	1862.9
AR(10)	-0.5046	[-0.8;-0.2]	1862.9
AR(11)	-0.3782	[-0.8; 0.1]	1862.9
MA(1)	-0.3411	[-0.8;0.1]	1862.9
MA(2)	1.4693	[1.0;1.9]	1862.9
MA(3)	0.0783	[-0.7;0.8]	1862.9
MA(4)	0.6587	[-0.1;1.4]	1862.9
MA(5)	0.7628	[0.1;1.5]	1862.9
MA(6)	-0.1929	[-0.6;0.3]	1862.9
MA(7)	0.6703	[0.3;1.0]	1862.9

TAB. 5.1 – Tableau d'estimation de modèle ARIMA (11,1,7)

0 n'appartient pas aux intervalles de confiance des modèles :AR(2), AR(4), AR(5), AR(7),AR(9), AR(10) ,MA(2), MA(5) et MA(7) d'où l'acceptation de ces modèles . Donc on accepte globalement ARIMA (11,1,7).

#### 5.2.4 Validation de modèle

Pour confirmer la validation du modèle on étudie les résidus de chaque ARIMA(p,d,q)

- Pour le modèle ARIMA (10,1,7) :

Les graphiques des résidus sont comme suit :

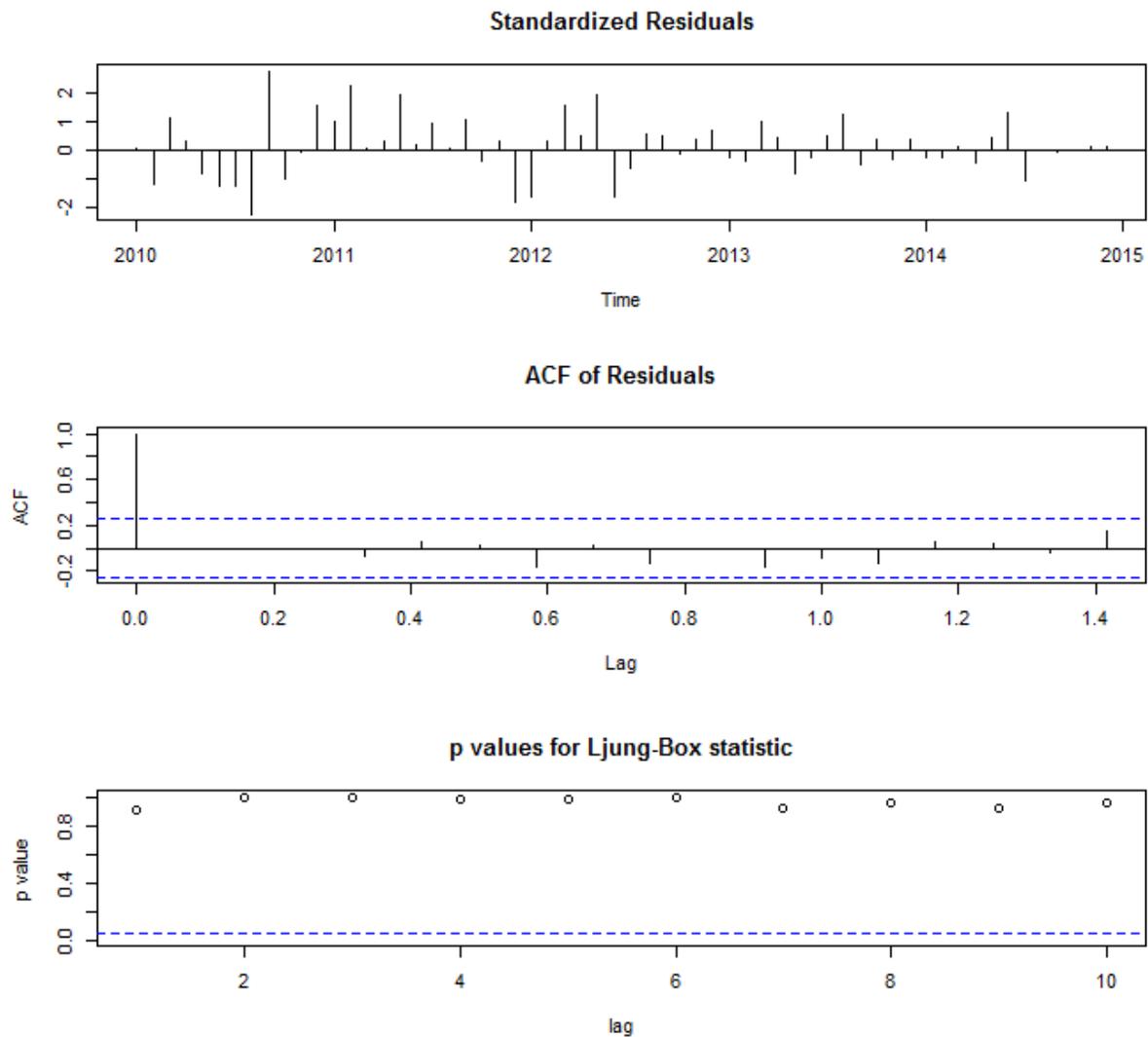


FIG. 5.5 – Schéma des résidus du modèle ARIMA(10,1,7)

D'après le graphique de l'ACF des résidus on constate qu'il n'y a pas de pics significatifs, tous les pics se trouvent à l'intérieur de la bande bleue. Et le test de Box-Pierce confirme que les résidus forment un bruit blanc au seuil de signification  $\alpha = 0,05$  car  $p\text{-value} = 0,9996 > 0,05$ .

- Pour le modèle ARIMA (11,1,7) :

Les graphiques des résidus sont comme suit :

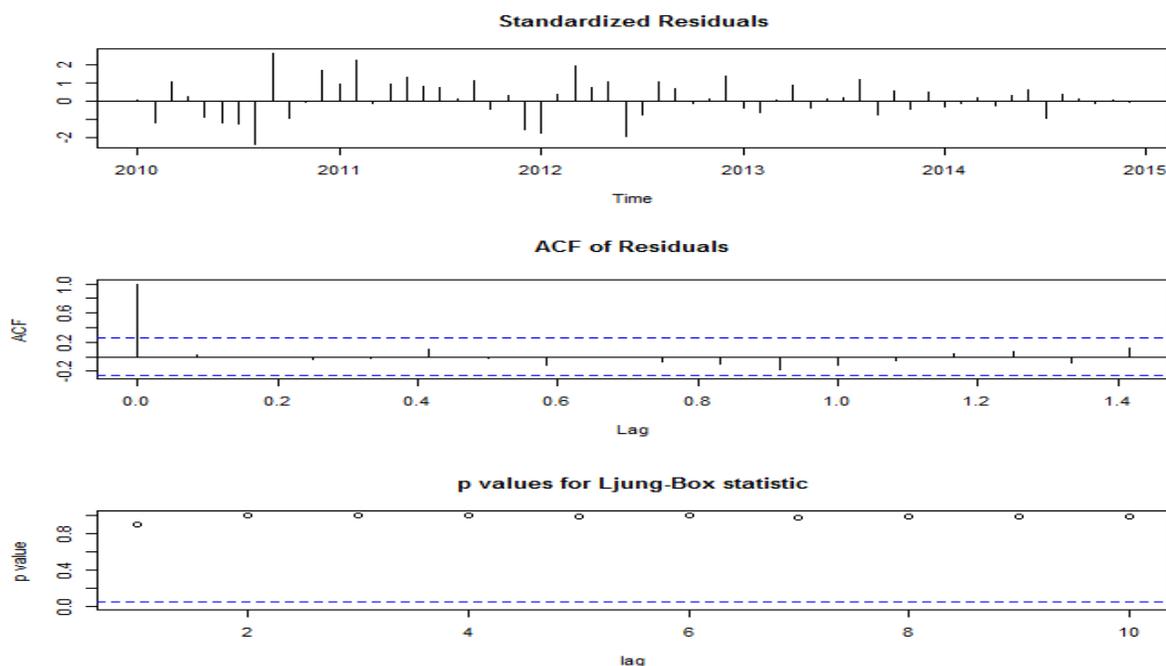


FIG. 5.6 – Schéma des résidus du modèle ARIMA(11,1,7)

D'après le graphique de l'ACF des résidus on constate qu'il n'y a pas de pics significatifs, tous les pics se trouvent à l'intérieur de la bande bleue. Et le test de Box-Pierce confirme que les résidus forment un bruit blanc au seuil de signification  $\alpha = 0,05$  car p-value = 0,9945 > 0,05.

### 5.2.5 Choix de modèle

Nous avons deux modèles qui ajustent correctement le modèle. Le meilleur modèle est celui qui minimise le critère d'AIC. Ainsi, le modèle retenu est un ARIMA(11,1,7) car il a le plus petit AIC = 1861.57

D'où  $(X_t)$  est ajustée par l'équation :

$$\Phi(B)(I - B)^d X_t = \Theta(B)\epsilon_t$$

Pour notre modèle ARIMA(11,1,7) on a :

$$\begin{aligned}\Phi(B)X_t = & -0.1112X_{t-1} - 1.1134X_{t-2} - 0.4038X_{t-3} \\ & -0.7219X_{t-4} - 0.8637X_{t-5} - 0.3278X_{t-6} - 0.9817X_{t-7} \\ & -0.3204X_{t-8} + -0.7673X_{t-9} - 0.5046X_{t-10} - 0.3782X_{t-11}\end{aligned}$$

Et

$$\begin{aligned}\Theta(B)\epsilon_t = & -0.1087\epsilon_{t-1} + 0.8285\epsilon_{t-2} + 1.1693\epsilon_{t-3} + 0.511\epsilon_{t-4} + 0.5554 \\ & \epsilon_{t-5} + 0.9365\epsilon_{t-6} + 0.1194\epsilon_{t-7}\end{aligned}$$

### 5.2.6 Prévision

Après la modélisation de notre série, les futures valeurs de cette série pour cinq mois de l'année 2015 sont regroupées dans ce tableau :

Mois	Futurs valeurs
Janvier	8602814.7
Février	9252182.245
Mars	9737548.988
Avril	10938398.94
Mai	11106154.81

TAB. 5.2 – Les prévisions paramétrique pour cinq mois (année 2015) pour le modèle ARIMA(11,1,7)

#### Remarque

On donne les prévisions de modèle ARIMA(10,1,7) qui est le deuxième modèle meilleur après ARIMA(11,1,7) pour faire une comparaison avec les résultats des prévisions non paramétrique.

Mois	Futurs valeurs
Janvier	8580991.292
Février	9898165.55
Mars	10719124.07
Avril	10407320.3
Mai	10732209.38

TAB. 5.3 – Les prévisions paramétrique pour cinq mois (année 2015) pour le modèle ARMA(10,1,7)

### 5.3 Prédiction non paramétrique

Nous étudions dans cette partie notre série chronologique et nous comparons l'efficacité entre la méthode non paramétrique et la méthode paramétrique (Box et Jenkins).

Pour la série simulé est de modèle ARIMA, nous avons bien retenu l'identification de notre modèle dans la partie précédente. Pour cette partir on cherche à tronquer noter série de manière à avoir une comparaison des prévisions par rapport aux vraies observations ou vrais série.

#### 5.3.1 Présentation

Dans toute la suite,  $K$  le noyau introduit (4.6) est le noyau gaussien :

$$K(y) = c \cdot \exp(-\|y\|^2), c = (2\pi)^{-r/2}$$

Il s'agit ici d'utiliser la notion de similarité pour décrire la chronique on s'intéressera à associer une période témoin en général un jour, semaine,..., dans notre série mensuelle on prend les mois (c'est à dire par mois). Les périodes qui lui ressemblent le plus au sens de l'ordre induit par les similarités par apport au témoin.

Le prédicteur à noyau dépend de :

- choix de noyau (noyau gaussien)
- la taille du bloc  $r$
- la vitesse de fermeture de la fenêtre  $h_n$

On calibre les constantes  $r$  ;  $C_h$  par validation croisée. Plus précisément, ayant observé  $X_1, \dots, X_n$ , et cherchant à prévoir  $X_{n+1}, \dots, X_{n+k_{prev}}$ , on détermine les valeurs de  $r$  ;  $C_h$  qui

minimisent la quantité suivante (erreur quadratique)

$$\sum_{t=n/2}^{n-k_{prev}} \sum_{i=1}^{k_{prev}} (\widehat{X}_{t:1} - X_{t+i})^2$$

### 5.3.2 Démarche de construction de prédicteur

Cette partie présente la structure retenue pour le prédicteur, en l'accompagnant par des justifications préalables.

Les prévisions non paramétriques de la série chronologique des pots de yaourt sont liées à la définition des deux paramètres fondamentaux qui sont :

#### 1) Le paramètre de troncature $r$

Le choix d'un modèle linéaire est assez arbitraire et son approximation par un processus ARMA( $p, q$ ) reste imprécise puisque, en général,  $p$  et  $q$  sont des indices de troncature et non pas des paramètres inconnus. Alors dans cette partie et pour les modèles obtenus on prend  $r$  est l'ordre de chaque  $p$  et  $q$  dans le modèle ARIMA.

#### 2) La largeur de la fenêtre $h_n$

La technique utilisée est inspirée d la validation croisée en régression (cf Vieu 91 et 92, Hardle 90). Le principe de celle ci consiste pour l'horizon de prévision  $h$  fixé en la minimisation d'une mesure de déviation quadratique entre  $f$  et son estimateur  $f_n$ , vu comme une fonction de  $h$ , que nous noterons ici simplement  $f_h$ . On montre alors sous les hypothèses peu restrictives sur le noyau  $K$ . Des hypothèses de régularité sur  $f$  et  $g$  et le mélange de  $X$ , qu'il est asymptotiquement équivalent de minimiser le critère dit de validation croisée suivant :

$$CV(h) = \sum_{i=r, \dots, n-h} \{X_{i+h} - f_{h,-i}(X_{i,(r)})\}^2 / (n-h)$$

Où  $f_{h,-i}(X_{i,(r)})$  est l'estimateur de  $f_h(x)$  calculé l'aide de toutes les données sauf la  $i^{eme}$

On cherche  $h_n$  de la forme suivante :

$$h_n = h.(h(x_1, \dots, x_n)n^{-1/r+4})$$

## 5.4 Prévion des séries

Dans cete partie nous prenons les deux meilleurs modèles estimés dans la partie des prévisions paramétriques, pour chaque modèle on applique les prévisions non paramétrique prenant en compte :

- Le même horizon de prévision, dans ces modèle on apris l'horizon égal à 5.
- La longueur de prévision égal à 5 c'est à dire : prédire 5 valeur (5 mois).
- Le même intervalle de confiance au niveau 0.05.
- On retrouve bien l'ordre de l'ARIMA comme choix optimal de r.

### 5.4.1 Prévion non paramétrique d'un ARIMA (11,1,7)

Pour les prévision de ce modèle on a :

- Les données sont simulées suivant un modèle ARIMA(11,1,7).
- Les paramètres obtenus par validation croisée sont  $r = 11$  et  $C_h = 1$ .

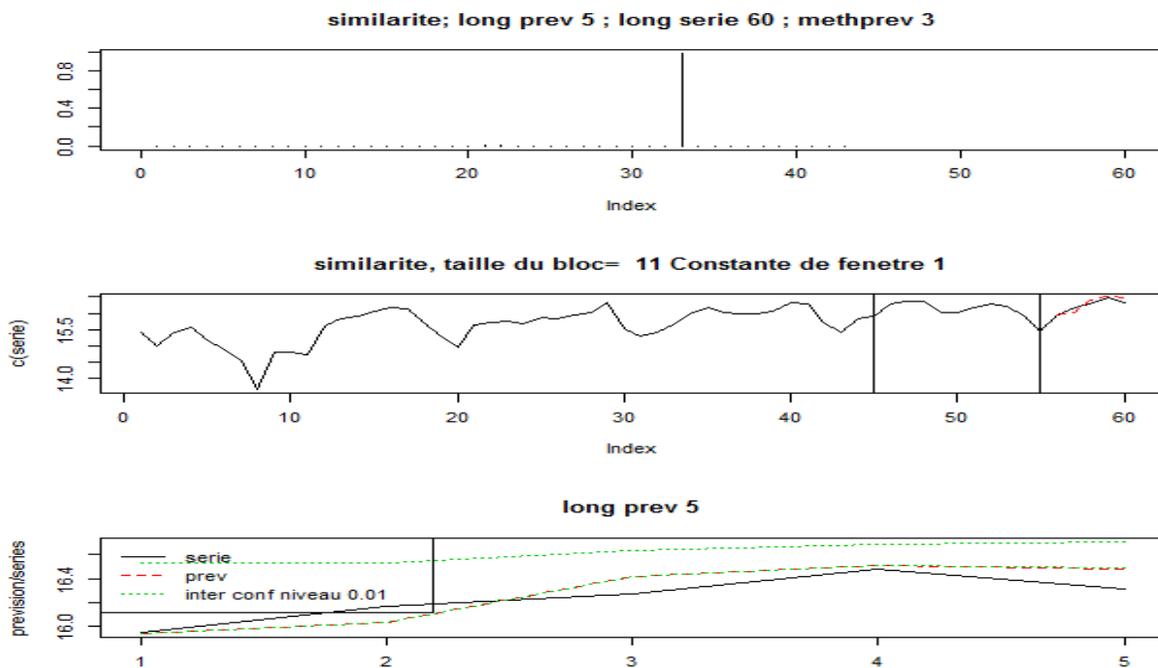


FIG. 5.7 – Prévion non paramétrique de modèle ARIMA(11,1,7)

**Remarque**

Les graphiques représentent : en haut les similarités, celui de milieu la série et la prévision obtenue, en bas représente la région de confiance et la prévision.

D'après ces graphes, nous remarquons que la série tracée en rouge représente des prévisions, elle est à l'intérieur de l'intervalle de confiance représenté en vert, elle est au-dessus de notre série tracée en noire.

Graphiquement nous avons obtenu des bons résultats, les prévisions obtenues sont :

*prevision*

[1] 8414345 9211820 13214480 14442137 14119636

**Comparaison des résultats pour le ARIMA (11,1,7)**

Pour comparer les résultats paramétrique avec les résultats non paramétrique, nous regroupons les données dans le tableau suivant :

Les mois	résultat paramétrique	résultat non paramétrique	données réelles
Janvier	8602814.7	8414345	11743934
Fevrier	9252182.245	9211820	11632942
Mars	9737548.988	13214480	13173237
Avril	10938398.94	14442137	14802878
Mai	11106154.81	14119636	15147469

TAB. 5.4 – Tableau des résultats pour le modèle ARIMA(11,1,7)

**Interprétation des résultats pour le modèle ARIMA(11,1,7)**

D'après ce tableau ci-dessus, nous remarquons que les valeurs prédites en appliquant dans le cas non paramétrique sont très proches des valeurs réelles que cette obtenues par la méthode de Box et Jenkins ,prenons les trois mois : Mars, Avril et Mai, pour plus de précision nous donnons les valeurs de critère EMRO défini dans le tableau suivant :

Le critère	EMRO
Box Jenkins	0.2521586
Non Paramétrique	0.116

TAB. 5.5 – Tableau de critère EMRO pour le ARIMA(11,1,7)

### Interprétation des résultats de EMRO pour modèle ARIMA(11,1,7)

D'après ce tableau ci-dessus, nous remarquons que la valeur de EMRO obtenue dans les deux méthodes est différente  $0.116 < 0.2521586$ . Cela assure la faveur de la méthode non paramétrique, qui est ainsi une méthode alternative très intéressante à la classique méthode de Box et Jenkins.

#### 5.4.2 Prédiction non paramétrique d'un ARIMA (10,1,7)

Pour les prévisions de ce modèle on a :

- Les données sont simulées suivant un modèle ARIMA(10,1,7).
- Les paramètres obtenus par validation croisée sont  $r = 10$  et  $C_h = 1$ .

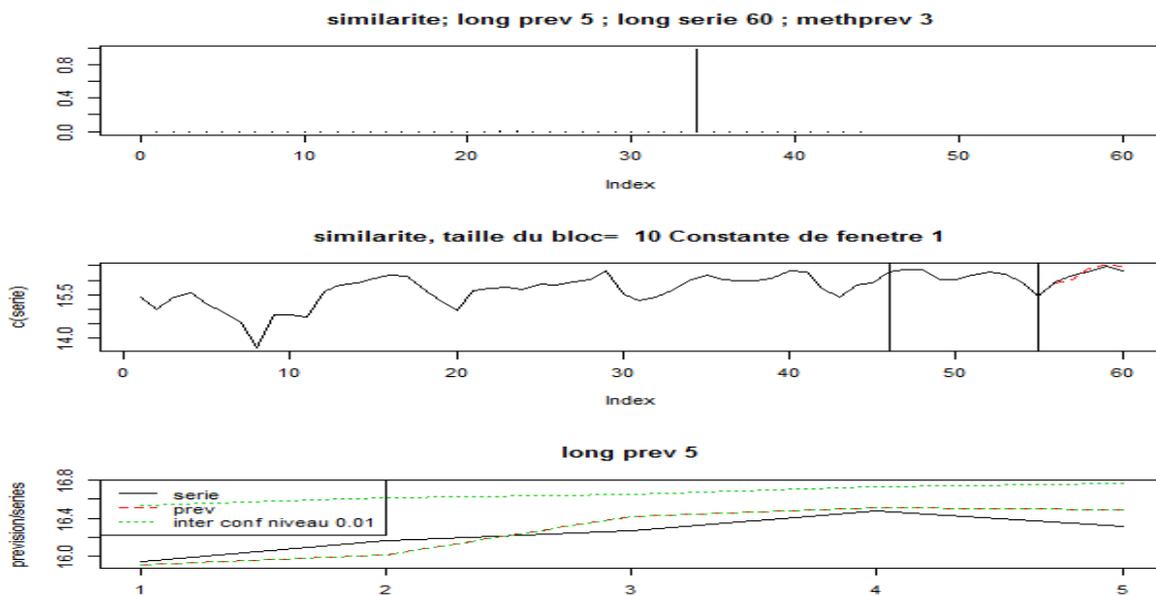


FIG. 5.8 – Prédiction non paramétrique de modèle ARIMA(10,1,7)

#### Remarque

Les graphiques représentent en haut les similarités, celui de milieu la série et la prédiction obtenue, en bas la prédiction et la région de confiance.

D'après ces graphes, on remarque que la série tracée en rouge représente des prévisions, est à l'intérieur de l'intervalle de confiance tracé en vert, et elle est au-dessus de notre série

tracée en noire.

Graphiquement nous avons obtenu des bons résultats, les prévisions obtenues sont :

*prévision*

[1] 8252244 9089689 13260346 14539993 14203261

### Comparaison des résultats pour le ARIMA (10,1,7)

Pour comparer les résultats paramétrique avec les résultats non paramétrique, on regroupe les données dans le tableau suivant :

Les mois	résultat paramétrique	résultat non paramétrique	données réelles
Janvier	8580991.292	8252244	11743934
Fevrier	9898165.55	9089689	11632942
Mars	10719124.07	13260346	13173237
Avril	10407320.39	14539993	14802878
Mai	10732209.38	14203261	15147469

TAB. 5.6 – Tableau des résultats pour le modèle ARIMA(10,1,7)

### Interprétation des résultats pour le modèle ARIMA(10,1,7)

D'après ce tableau ci-dessus, nous remarquons les valeurs prédite que les valeurs prédite en appliquant dans le cas non paramétrique sont très proches des valeurs réelles que cette obtenues par la méthode de Box et Jenkins ,prenons les trois mois : Mars, Avril et Mai, pour plus de précision nous donnons les valeurs de critère EMRO définit dans le tableau suivant :

Le critère	EMRO
Box Jenkins	0.2386343
Non Paramétrique	0.124

TAB. 5.7 – Tableau de critère EMRO pour le ARIMA(10,1,7)

### Interprétation des résultats de EMRO pour modèle ARIMA(10,1,7)

D'après ce tableau au-dessus, nous remarquons que la valeur de EMRO obtenue dans les deux méthodes est différente  $0.124 < 0.2386343$  . Cela assure la faveur de la méthode

non paramétrique, qui est ainsi une méthode alternative très intéressante à la classique méthode de Box et Jenkins.

# Conclusion générale

Le travail présenté dans ce mémoire traite les prévisions non paramétriques des séries chronologiques en utilisant la régression non paramétrique à noyau qui donne le rôle de prédicteur . Les prévisions non paramétriques que nous présentons dans ce mémoire, elles donnent leurs preuves par rapport à la prévision paramétrique, celle de Box Jenkins, sont simples à appliquer, robustes et souvent supérieures aux méthodes paramétriques.

Dans la première partie de ce mémoire, nous avons exposé la méthode non paramétrique d'estimation de la densité de probabilité, à savoir l'estimation par la méthode de noyau avec les différentes propriétés statistique et le choix de noyau.

Dans la deuxième partie, nous avons présenté la méthode d'estimation de la fonction de régression pour le cas des variables réelles, nous avons défini la méthode non paramétrique par moyenne mobile (MA) et la méthode non paramétrique à noyau.

La troisième partie penche sur la régression non paramétrique des séries chronologiques, les hypothèses de mélange et les différents types de convergence présentent durant toute la présentation de cette partie.

Dans le quatrième chapitre, on s'intéresse au prévision non paramétrique des séries chronologiques via la fonction de régression, cette partie est motivée par les différentes étapes à suivre pour prédire les valeurs futures choisissant le noyau, l'horizon de prévision et la longueur de prévision passant par le paramètre de troncation.

Enfin, afin d'accorder la théorie à une application, on a pris une série chronologique représente des pots de yaourt produisent mensuellement pendant 5 ans pour la SARL

RAMEDY, on a appliqué deux méthodes de prévision : la première méthode est la méthode paramétrique ( Box et Jenkins ), la deuxième méthode est la méthode non paramétrique. Nous avons conclu des résultats pour ces deux méthodes, en comparant avec les données réalisées, plus de précision avec le critère d'erreur moyenne relative observée.

# Bibliographie

- [1] Arnak, S. *Statistique avancée : méthodes non paramétriques*. Ecole Centrale de Paris.
- [2] BEGHRICHE, H. *Sur L'estimation de la Fonction de Régression*. Mémoire de magistère en Mathématiques. Université de Mentouri-Constantine.
- [3] BOUTAHAR, M. *Séries temporelle : estimation paramétrique et non paramétrique avec le logiciel R*. univsité-nantes. 4 Décembre 2007.
- [4] Carbon, M. *Approche non paramétrique en régression*. Chapitre 9.
- [5] Carbon, M., Franco, C. *Estimation non paramétrique de la densité et de la régression-Prévision non paramétrique*. Laboratoire de Probabilités et Statistique.
- [6] CHIKHI, M., TERRAZA, M. *Un essai de prévision non paramétrique de l'action France telecom*. Décembre 2002
- [7] D.Bosq. *Nonparametric Statistics for Stochastic Processes*. Université Pierre et Marie Curie . Institut de Statistique. 1996
- [8] D.Bosq. *La statistique non paramétrique des processus*. Univers Politecn. TORINO Vol.45. année(1987)
- [9] Ferraty, F., Vieu, V. *Nonparametric Functional Data Analysis. Theory and Practice*. Springer Science+Business Media. 2006
- [10] Ferraty, F., Vieu, V. *Statistique fonctionnelle : Modèles-Non paramétrique de régression*. Notes de cours de DEA (2002-2003).
- [11] Hurlin, C. *Régressions Non Paramétriques Univariées*. Mémoire de master econometrie et statistique appliquée. Université d'Orléans.(2007-2008)
- [12] POGGI, J.M. *Prévision non paramétrique de la consommation électrique*. Revue de statistique appliqué, tome 42, n°4. p83-98 RO03, 2011.

- 
- [13] SAADI, N. *Estimation non paramétrique de la densité de probabilité par des fonctions orthogonales*. Mémoire de magister. University de Bejaia 2008.
- [14] Zougab, N. *Méthode de Monte Carlo dans l'estimation non paramétrique de la densité de probabilité par noyau associé*. Mémoire de magister. Université de Bejaia 2014.

## Résumé

Après un aperçu sur l'estimation non paramétrique de la densité et la régression, nous décrivons les différentes propriétés statistiques et différentes lois de convergence prenant en compte le choix de notre noyau pour les deux cas connus : cas des variables réelles, cas des processus continus (série chronologique).

Nous détaillons et interprétons une méthode de prévision, dite prévision non paramétrique. Nous montrons les différents aspects techniques ainsi pratiques, et nous comparons les données reçues à celle de la méthodologie de Box et Jenkins.

**Mots clés :** Estimation, densité de probabilité, noyau, régression non paramétrique, estimateur à noyau, prévision non paramétrique.

## Abstract

After an overview of nonparametric density estimation and regression, we describe the different statistical properties and different laws taking convergence into account the choice of our kernel for the two known cases : the case of real variables, case of continuous process (time series).

We describe and interpret a forecasting method, called nonparametric forecast. We show the different technical aspects and practice, and we compare the data received to that of the Box-Jenkins methodology.

**Key words :** Estimation, estimation of probability density, Kernel estimate, nonparametric regression, nonparametric prediction.