



UNIVERSITE KASDI MERBAH  
OUARGLA  
Faculté des Mathématiques et des Sciences de la  
Matière

N° d'ordre :  
N° de série :

DEPARTEMENT DE MATHEMATIQUES

MASTER

Spécialité : Mathématiques

Option : Probabilité et Statistique

Par : Ouggad Sara

Thème

**Etude de l'efficacité des estimateurs  
récursifs pour les modèles  
semi-paramétrique.**

Soutenu publiquement le : 31/05/2017

Devant le jury composé de :

Mme .BOUANANE Khadra	M.C.B. Université KASDI Merbah- Ouargla	Président
Mr. CHETTI Djameledine	M.A. Universié KASDI Merbah- Ouargla	Examineur
Mr. BAHEDDI Aissa	Pr. Université KASDI Merbah- Ouargla	Rapporteur

### ملخص

في هذه المذكرة أنا مهتمة بدراسة النموذج شبه معلمي من الشكل  $y = f(x, \theta) + \varepsilon$  من اجل كل  $x \in \mathbb{R}^p$  و  $y \in \mathbb{R}$ .

هدفنا هو دراسة مشاكل تقدير المعلمة  $\theta$  و  $f$  هذا النموذج مع طريقة التراجعية وصياغة مصفوفة لهذه الطريقة

في الجزء الأول، نقترح الطريقة سيد (مقلوب تراجع لشرائح) متواتر، مصفوفة الإهتمام والإتجاه ل  $\theta$  من أجل تقدير المعلم  $\theta$ ، و يقترح طريقة ناداريا واطسون للتقدير متواتر  $f$ .

كلمات مفتاحية: مقدر معلمي، نموذج شبه معلمي، طريقة تراجعية، طريقة الانحدار العكسي بوحدات (SIR)، طريقة واتسن، مرور.

### Abstract

In this memory, I interse have the study of a parametric semi model of the following from

$$y = f(x, \theta) + \varepsilon \text{ for all there } x \in \mathbb{R}^p \text{ and } y \in \mathbb{R} .$$

Our objective of is studied problems of estimate the parameter  $\theta$  and  $f$  of this model with the recursive method and has the matric formulation of this.

In the first part, I propose the method SIR (Slice Invers Regression) recursive ,the matrix of interst and the direction of  $\theta$  to estimate the parameter  $\theta$  ,and to propose the method of Nadarya watson for the recursive estimate of the function link  $f$

key Word: recursive estimate, parametric semi model, method of Slice Invers Regression (SIR), method of Nadarya Watson, the passage.

### Résumé

Dans ce mémoire je m'intéresse a l'étude d'un modèle semi paramétrique de forme suivante  $y = f(x, \theta) + \varepsilon$  pour tout  $x \in \mathbb{R}^p$  et  $y \in \mathbb{R}$  .

Notre objectif est d'étudié des problèmes d'estimation des paramètres  $\theta$  et  $f$  de ce modèle avec la méthode réursive et a la formulation matricielle de ce method

Dans le première partie ,je propose la méthode SIR (Régression inverse par tranches) réursive, la matrice d'intérêt et la direction de  $\theta$  pour estimer le paramètre  $\theta$ , et proposer le méthode de Nadarya Watson pour l'estimation réursive de la fonction de lien  $f$

Mots-clés : estimation réursive, modèle semi-paramétrique, méthode de régression inverse par tranchage (SIR), méthode Nadarya Watson, le passage.

# Table des matières

Dédicaces	i
Remerciement	ii
Notations	vi
Introduction	viii
<b>1 Définition et principes</b>	<b>4</b>
1.1 modèle de régression :	4
1.2 Les modèles paramétriques :	5
1.2.1 Modèle linéaire	7
1.2.2 L'estimation	7
1.2.3 la méthode des moindres carrés	8
1.2.4 le Maximum de vraisemblance :	9
1.3 Les modèles non paramétriques :	10
1.3.1 Modèle non linéaire	12
1.3.2 L'estimation :	12
1.4 Les modèles semi-paramétriques	13
1.4.1 Autre écriture du modèle :	15
1.4.2 La méthode SIR :	16
<b>2 les méthodes récursives</b>	<b>19</b>
2.1 les méthodes récursives :	19

2.1.1	Estimation récursive du paramètre $\theta$ . . . . .	20
2.1.2	Estimation non paramétrique d'une fonction de régression : . . . . .	23
2.1.3	Estimateur de Nadaraya Watson . . . . .	24
<b>3</b>	<b>simulation</b>	<b>26</b>

# Notations

- ▶  $Var$  : la variance .
- ▶  $E$  : *espérance* .
- ▶  $SIR$  : Régression Inverse par tranche.
- ▶  $EDR$  : Effective Dimension Réduction.
- ▶  $\theta$  : paramétré .
- ▶  $f$  : fonction de lien
- ▶  $\xi$  : bruit.
- ▶  $\Sigma$  :matrice varaince covariance.

# Introduction

les modèles de régression sont très utiles pour étudier la liaison entre une variable à expliquer  $y$  et une variable explicative  $x$ .

Les déformation périodiques sont souvent très utiles pour modéliser des phénomènes réels . dans les domaines de la médecine,l'économie,ou encore la météorologie. L'étude de ce type de modèles permet d'avoir une bonne approximation de la réalité.

Dans la littérature statistique, deux grandes classes de modèles sont omniprésentes :les modèles paramétrique et les modèles non paramétrique.

ce deux types de modèles ont leurs avantages et leurs défauts spécifiques.

Afin de conjuguer les avantages des approches paramétriques et non paramétriques, à savoir la capacité d'interprétation des modèles paramétrique et la souplesse des modèles non paramétrique, les modèles semi-paramétriques ont été développés.

Ces modèles dépendent généralement d'un ou de plusieurs paramétrés ainsi que d'une fonction de lien à estimer.

Dans ce mémoire on considère tout d'abord le modèle semi-paramétrique de régression de la forme  $y = f(x, \theta) + \xi$  ou le  $y$  est un réel ( $y \in \mathbb{R}$ ) et le  $x \in \mathbb{R}^p$  ,Notre objectif est d'étudier des problèmes d'estimation de paramètres  $\theta$  et  $f$  de ce modèle récurrente et l'avantage de ce modèle et de prendre en compte l'avantage de ce modèle est de prendre en compte l'arrivée temporelle des informations et d'affiner ainsi au fil du temps les algorithmes d'estimation mis en oeuvre

ce modèle semi-paramétrique permet d'estimer la partie paramétrique  $\theta$  par la méthode SIR(Régression inverse par tranches) introduite par Li(1991) et sans avoir a estimer la fonction de lien  $f$  par Nadarya Watson récursif

En suite la méthode de transition dans la régression par une matrice a partir d'échan-

tillon( $n$ ) à l'échantillon ( $n+1$ ) .

## Problématique

Ce mémoire est consacrée a l'étude de certains modèles semi-paramétriques et proposer des méthodes récursives,Voilà pourquoi nous vous proposons la problématique suivante :  
\_ comment générer un estimateur récursif pour le modèle semi-paramétrique ?

### sous problèmes

- Comment concevoir un estimateur récursif ?
- Comment on définit le générateur d'un estimateur récursif ?
- Est ce que les méthodes récursives d'estimation sont efficaces pour les modèles semi paramétrique ?

### hypothèses

- Prendre en compte l'arrivée temporelle des informations supplémentaire dans la méthode d'estimation du paramétriser  $\theta$ .
- En fixant l'élément de récursivité du problème par une fonction de lien.
- les méthodes récursives d'estimation prend en compte le terme de temps des séries temporelles .

### L'importance de l'étude

l'importance de se travail réside dans la formulation matricielle du modèle et le passage de l'échantillon de taille  $n$  des information à l'échantillon de taille  $n + 1$ .

## Revue de littérature

**1-titre** :Estimation récursive pour les modèle semi-paramétrique

L'auteur :Thi Ngoc Yen

Le résumé :

Dans cette thèse l'auteur s'intéresse au modèle semi-paramétrique de régression ,et à l'étude des problème des estimation des paramétré  $\theta$  et  $f$  ,l'auteur présentait dans sur travail la méthode SIR récursive pour estimé le  $\theta$  et la méthode de Nadarya Watson pour estimé le lien  $f$  et la méthode utilisé pour estimer un modèle semi-paramétrique .

les résultats

L'auteur a proposé une méthode permettant de combiner l'estimation récursive de la fonction de lien  $f$  avec l'estimateur de Nadarya Watson récursif et l'estimation du paramètre  $\theta$  via l'estimateur SIR récursif

**2-titre** :Méthodes récursives en estimation et prévision non paramétrique.

L'auteur :Aboubacar Amiri

le résumé :

Dans cette thèse l'auteur s'intéresse aux méthodes d'estimation non paramétriques a partir d'estimateurs récursifs a noyau ainsi qu'a leurs applications à la prévision .

Résultats

\_ En terme de temps de calcul en utilisant des estimateurs récursifs a noyau de la régression.

\_ Pour des prévisions a long terme on privilégiera les méthode de Box-Jenkins.

**3-titre** :Régression sur variable fonctionnelle :Estimation,tests de structure et Applications

l'auteur :Laurent Delsol

Résumé :dans cette thèse l'auteur s'intéresse plus particulièrement a des modèles de régression dans les quels la variables réponse est réelle tandis que la variable explicative est fonctionnelle ,c'est à dire à valeurs dans un espace de dimension infinie.

Résultats :

\_ L'auteur cite les liés des propriétés asymptotiques de l'estimateur à noyau généralisé au



cas d'une variable explicative fonctionnelle

\_ Toute en supposant que l'échantillon de l'étude de variables du modèle de régression est de nature non paramétrique.

**4-titre** : Modèle linéaire fonctionnel lorsque la variable explicative est bruit

L'auteur : Christophe Crambes

Résumé

Dans cette thèse l'auteur s'intéresse à des modèles linéaires fonctionnels lorsque la variable explicative est un bruit.

Résultats

\_ L'auteur propose un estimateur de  $\alpha$  (fonction inconnue) sur la base des observations \_ Il estime le  $\alpha$  par deux étapes

\_ par la méthode du lissage

\_ par la méthode de régression fonctionnelle sur les composantes principales.

Analyse de la revue de littérature

Toutes les thèses utilisent le modèle de régression et estimer la fonction de lien par la méthode de Nadarya Watson.

Les techniques d'estimation traitent le problème de deux optiques

\_ les espaces de dimension finie

\_ les espaces de dimension infinie

aussi que l'estimation récursive pour les modèles (semi-paramétrique) et l'estimation récursive pour les modèles (non paramétrique)

# Chapitre 1

## Définition et principes

Dans ce chapitre, je présente tout d'abord les modèle de régression dans un cas général, puis on intéresse plus précisément au modèle semi paramétrique de régression dans le cadre du quel on ai travaille et l'estimation récursive pour le modèle semi paramétrique, et présente une synthèse de travaux concernant la méthode de régression inverse par tranchage (SIR). Et on présente des estimateurs non paramétriques d'une densité d'une fonction de régression par Nadarya Watson.

Et je présente la forme matricielle de la modèle  $y = x\theta + \xi$

### 1.1 modèle de régression :

Les modèles de régression sont très utiles pour modéliser la liaison entre une variable à expliquer  $y$  et une variable explicative  $x$  . Dans la littérature statistique, deux grandes classes de modèles de régression sont :

- Les modèles paramétriques
- Les modèles non paramétriques

Ces deux types des modèles sont caractérisés par des avantages et des défauts spécifiques. Afin de conjuguer les avantages des approches paramétriques et non paramétriques, a sa voir la capacité d'interprétation des modèles paramétriques et la souplesse des modèles non paramétrique, on propose l'étude des modèles semi-paramétriques qui ont été développés.

Ces modèles dépendent généralement d'un paramètre de dimension finie, noté  $\hat{\theta}$ , ainsi que d'une fonction de lien  $f$  à estimer.

## 1.2 Les modèles paramétriques :

Pour les modèles de régression paramétrique, la fonction de lien, entre la variable à expliquer ( $y \in \mathbb{R}$ ) et une variable explicative  $x \in \mathbb{R}^p$ , dépend d'un nombre fini de paramètres à estimer <sup>1</sup>.

La forme du modèle paramétriquement :

$$y = f_{\theta}(x) + \xi$$

$f_{\theta}$  appartient à une famille des fonctions paramétrées par  $\theta$  vecteur de paramètres réels,  $\xi$  est une erreur aléatoire.

Dans ce modèle l'objectif est l'estimation du paramètre  $\theta$  et les techniques d'estimations paramétriques (la méthode maximum de vraisemblance, méthode moindres carrés, ?) sont efficaces quand la famille de  $f_{\theta}$  est ponctuellement spécifiée.

la méthode maximum de vraisemblance :

on appelle fonction de vraisemblance pour l'échantillon  $x_1, x_2, \dots, x_n$  la fonction du paramètre  $\theta$

$$L(\theta, x_1, x_2, \dots, x_n) = \begin{cases} \prod P(X = x_i, \theta) & \text{discret} \\ \prod f_x(x_i, \theta) & \text{continue} \end{cases}$$

L'estimation de maximum vraisemblance  $\theta$  est la valeur  $\theta_n$  qui rend maximale la fonction de vraisemblance  $L(\theta, x_1, x_2, \dots, x_n)$ ,  $\hat{\theta}_n$  est la solution du système d'équation

$$\frac{\partial \ln L(\theta, x_1, x_2, \dots, x_n)}{\partial \theta_i} = 0 \quad i = 1, \dots, n$$

La solution de l'équation de vraisemblance n'est pas toujours calculable analytiquement. Pour calculer la variance  $\text{var}(\hat{\theta})$ , la fonction de  $\ln L(x, \theta)$  doit être dérivable d'ordre  $p$ .

---

<sup>1</sup>Estimation de l'écotoxicité de substances chimiques par des méthodes à noyaux

En suite, nous construisons la matrice symétrique d'information de Fisher notée par  $\hat{I}$  :

$$\hat{I} = -E\left[\frac{\partial^2 \ln L(x, \theta)}{\partial \theta_i \partial \theta_j}\right]$$

La matrice inverse  $\hat{I}^{-1}$  représente la matrice de variance covariance, notée  $\hat{\Sigma}$

$$\hat{\Sigma} = \begin{bmatrix} \hat{v}ar(\hat{\theta}_1) & \hat{c}ov(\hat{\theta}_1, \hat{\theta}_2) & \dots & \hat{c}ov(\hat{\theta}_1, \hat{\theta}_p) \\ \hat{c}ov(\hat{\theta}_2, \hat{\theta}_1) & \hat{v}ar(\hat{\theta}_2) & \dots & \hat{c}ov(\hat{\theta}_2, \hat{\theta}_p) \\ \vdots & \vdots & \ddots & \vdots \\ \dots & \dots & \dots & \hat{v}ar(\hat{\theta}_p) \end{bmatrix}$$

Dans ce cas des algorithmes de recherche de maximum (de type Newton) peuvent être utilisés.

La méthode des moindres carrés :

Dans le cas le plus courant, le modèle théorique est une famille de fonctions  $f(x, \theta)$  d'une ou plusieurs variables muettes  $X$ , indexées par un ou plusieurs paramètres  $\theta$  inconnus. La méthode des moindres carrés permettent de sélectionner parmi ces fonctions celle qui reproduisent le mieux les données expérimentales.

On parle dans ce cas d'ajustement par la méthode des moindres carrés. Si les paramètres  $\theta$  ont un sens physique, la procédure d'ajustement donne également une estimation indirecte de la valeur de ces paramètres.

La méthode consiste en une prescription (initialement empirique), qui est la fonction  $f(x, \theta)$  qui décrit ("le mieux") les données est celle qui minimise la somme quadratique des déviations des mesures aux prédictions de  $f(x, \theta)$  si par exemple nous disposons de  $n$  mesures  $(y_i)_{i=1, \dots, n}$ , les paramètres  $\theta$  ("optimaux") au sens de la méthode des moindres carrés sont ceux qui minimisent la quantité :

$$\sum_{i=1}^n (y_i - f(x_i, \theta))^2 = \sum_{i=1}^n r_i^2(\theta)$$

$r_i(\theta)$  : Sont les résidus du modèle.

Avantage spécifiques :

- Avoir une interprétation claire de l'impact de la variable explicative sur la variable expliquée.

### Défauts spécifiques

- le choix d'un bon modèle paramétrique au vu des données n'est pas toujours évident.
- Le modèle paramétrique choisi peut ne pas être en adéquation avec les données (donc parfois "éloigné" de la réalité des données).

## 1.2.1 Modèle linéaire

Le modèle de régression linéaire s'écrit sous la forme matricielle <sup>2</sup>

$$Y = X\theta + \varepsilon \quad (1.1)$$

ou

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{p,1} \\ 1 & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & x_{1,n} & \dots & x_{p,n} \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta_0 \\ \vdots \\ \theta_p \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

est l'écriture est

$$Y = X\theta + \varepsilon \quad (1.2)$$

ou

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{p,1} \\ 1 & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & x_{1,n} & \dots & x_{p,n} \end{pmatrix} \begin{pmatrix} \theta_0 \\ \vdots \\ \theta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

## 1.2.2 L'estimation

Pour estimer cette relation, on utilise un échantillon de l'ensemble des variable, On note  $Y_i$  la  $i$  ème valeur de l'ensemble de l'échantillon de la  $j$  ème variable. la valeur observée

---

<sup>2</sup>Christophe Crambes."Modèle linéaire fonctionnel lorsque la variable explicative est bruitée ".Université Paul Sabatier.mardi 12 juin 2007

de  $y_i$  est alors la somme de deux composantes : l'une déterministe (les  $x_{i,j}$ ,  $\forall j$ ) et l'autre aléatoire  $\varepsilon_i$ .

$$y_i = \theta_1 x_{i,1} + \theta_2 x_{i,2} + \dots + \theta_n x_{i,n} + \varepsilon_i$$

l'objectif est d'estimer les paramètres inconnus du modèle, d'utiliser les données pour étudier la validité de certaines propositions théoriques, on s'appuie dans cette démarche sur un bien connu :

- les estimations du maximum de vraisemblance
- les estimations du moindre carrés

### 1.2.3 la méthode des moindres carrés

Les paramètres de la relation  $y = X\theta + \xi$  sont à estimer par la méthode des moindres carrés ordinaires (*MCO*) et qui forme une méthode simple et très utilisée, même si elle n'est pas toujours la meilleure. Dans ce qui suit, on cherche  $\hat{\theta}$ , un estimateur de  $\theta$ , la vraie valeur de cet estimateur doit vérifier un certain nombre de propriétés que l'on détaillera par la suite. Dans ce qui suit, on appellera  $\hat{\xi}$  les erreurs produites par le modèle estimé. la méthode (*MCO*) se propose de minimiser l'erreur quadratique produite par le modèle. Le programme résolu pour les (*MCO*) est alors le suivant :

$$\min(Y - X\hat{\theta})^2$$

la résolution est simple <sup>3</sup>. Il suffit de dériver l'expression à minimiser par rapport à  $\theta$  et de chercher la valeur de  $\hat{\theta}$  :

$$2X^t(Y - X\hat{\theta}) = 0$$

$$\Leftrightarrow X^t Y = X^t X \hat{\theta}$$

$$\text{et que } \hat{\theta} = (X^t X)^{-1} X^t Y \quad , \text{ avec } (X^t X)^{-1} = \frac{1}{\det(X^t X)} \text{comt}(X^t X)^t$$

donc l'estimateur des paramètres du modèle par la méthode (*MCO*) est :

$$\hat{\theta} = (X^t X)^{-1} X^t Y$$

---

<sup>3</sup>Méthode des moindres carrés".site web

### 1.2.4 le Maximum de vraisemblance :

la fonction de densité de probabilité d'une variable aléatoire  $y$  conditionnellement à un ensemble de paramètre  $\theta$  est noté  $f(y, \theta)$ .

Cette fonction identifie le processus générant les données qui son l'échantillon des données et, en même temps, fournit une description mathématique des données que le processus génère. La densité jointe des  $n$  observations indépendantes et distribuées de façon identique (*i.i.d*) et ce processus est le produite des densités individuelles :

$$f(y_1, \dots, y_n \setminus \theta) = \prod_{i=1}^n f(y_i \setminus \theta) = L(\theta, y)$$

Cette densité conjointe est la fonction de vraisemblance ,définie comme une fonction du vecteur de paramètre inconnus ( $\theta$ ),ou  $y$  indique les données observées (qui ne sont donc pas une inconnue).

$$f(y) = \frac{1}{(2\pi\delta^2)^{\frac{1}{2}}} \exp\left(-\frac{(y - x\theta)^2}{2\delta^2}\right)$$

$$\begin{aligned} L(y, \theta) &= \prod_{i=1}^n f(y_i \setminus \theta) \\ &= \frac{1}{(2\pi\delta^2)^{\frac{n}{2}}} \exp\left(-\sum \frac{(y - x\theta)^2}{2\delta^2}\right) \end{aligned}$$

par conséquent :

$$\max L(y, \theta) = \min(y - x\theta)^2 = \hat{\theta}$$

il est généralement plus simple de travailler avec le logarithme de la fonction de vraisemblance .

$$\ln l(y, \theta) = \sum_{i=1}^n \ln f(y_i, \theta)$$

$$\ln L(y, \theta) = -n\theta + \ln \theta \sum_{i=1}^n y_i - \sum_{i=1}^n \ln(y_i)$$

d'ou :

$$\frac{\partial \ln L(y \setminus \theta)}{\partial \theta} = -n + \frac{1}{\theta} \sum_{i=1}^n y_i = 0$$

\*

$$\text{et avec } \hat{\theta}_{EMV} = \bar{y}_n$$

la méthode du maximum de vraisemblance propose de déterminer  $\hat{\theta}$  de façon à ce que la log-vraisemblance soit maximale, il est nécessaire de vérifier que cet estimateur est réalisable.

**Estimateur de  $\delta^2$  :**

$$\hat{\delta}^2 = \frac{1}{n - (p + 1)} (y - x\hat{\theta})^2$$

$\hat{\delta}^2$  et  $\hat{\theta}$  sont indépendants

**Estimateur de  $\hat{\xi}$**

Pour tout  $i \in 1, \dots, n$   $\hat{\xi}_i = y_i - \hat{y}_i$

ou  $\hat{y}_i = \hat{\theta}_0 + \hat{\theta}_1 x_{1,i} + \dots + \hat{\theta}_p x_{p,i}$

**Coefficients de détermination**

<sup>4</sup> ou a :

$$\hat{R}^2 = 1 - \frac{\|\hat{y} - y\|^2}{\|\bar{y}1_n - y\|^2}$$

ou  $\hat{y} = x\hat{\theta}$   $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ , ( $1_n$  désigne le vecteur colonne à  $n$  composants égaux à 1)

ce  $R^2$  est un coefficient réel toujours compris entre 0 et 1 .

### 1.3 Les modèles non paramétriques :

Les modèles de régression non paramétriques apparaissent comme une alternative qui offre la flexibilité désirée dans la modélisation (car aucune hypothèse paramétrique n'est imposée).<sup>5</sup>

---

<sup>4</sup>christophe chesneau ("modèle régression ")

<sup>5</sup>Gaëlle Chagny. Estimation non paramétrique d'une fonction de régression avec des bases déformées. MAP5.2010



La variable à expliquer  $y$  est maintenant reliée à la variable explicative  $x$  par le biais d'une fonction de lien inconnue que l'on doit estimer :

$$y = f(x) + \xi$$

Dans ce modèle l'objectif est l'estimation de la fonction de lien et les techniques d'estimations non paramétriques (les méthodes des estimateurs à noyau , méthode des splines de lissage,..

les méthodes des estimateurs à noyau :

les méthodes des estimateurs à noyau sont les méthode non-paramétrique d'estimation de la densité de probabilité d'une variable aléatoire. Elles se basent sur un échantillon d'une population statistique et permet d'estimer la densité en tout point du support. En ce sens, cette méthode généralise astucieusement la méthode d'estimation par un histogramme.<sup>6</sup> si  $x_1, x_2, \dots, x_n \sim f$  est un échantillon i.i.d (indépendant identiquement est distribuée) d'une variable aléatoire ,alors l'estimateur non-paramétrique par la méthode du noyau de la densité est :

$$\hat{f}_\theta(x) = \frac{1}{n\theta} \sum_{i=1}^n k\left(\frac{x - x_i}{\theta}\right)$$

Ou  $k$  est un noyau et  $\theta$  un paramètre nommé fenêtre, qui régit le degré de lissage de l'estimation,  $k$  est choisi comme la densité d'une fonction gaussienne standard (espérance nulle et variance unitaire)

$$k(x) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{1}{2}x^2}$$

la méthode noyau consiste à assurer la continuité : pour cela, on remplace l'intervall centré en  $x$  et de largeur  $\theta$  par une loi gaussienne centrée en  $x$ . Plus une observation est proche du point de support de  $x$  plus la courbe de cloche lui donne une valeur numérique importante.

Avantages spécifique :

ils offrent davantage de flexibilité (aucune hypothèses paramétrique n'est imposée dans ce modèle, seules des hypothèses de régularité sur  $f$  sont imposées).

---

<sup>6</sup>Estimation par noyau".site web

Défauts spécifique :

- L'interprétation de la fonction de lien  $f$  n'est pas toujours évidente
- Ils faut estimer la fonction de lien  $f$  le plus souvent au moyen de procédure de calculs intensifs en particulier en ce qui concerne la recherche des paramètres de lissage, ce qui est lourd en temps de calcul.

### 1.3.1 Modèle non linéaire

Le modèle de régression non linéaire s'écrit sous la forme matricielle

$$y = f(x_1, \dots, x_p, \theta_0, \dots, \theta_q) + \xi$$

ou :

$f$  : est une fonction supposée connue

$\theta_0, \dots, \theta_q$  sont  $q + 1$  coefficients réels inconnus

$\xi$  : le résiduel et indépendante de  $x_1, \dots, x_p$

### 1.3.2 L'estimation :

Pour estimer cet modèle

$$y = f(x_{1,i}, \dots, x_{p,i}, \theta_0, \dots, \theta_q) + \xi_i$$

On doit expliquer une variable quantitative  $y$  à partir de  $p$  variables  $x_1, \dots, x_p$ .

$$y = f(X_1, \dots, X_p, \theta_0, \theta_1, \dots, \theta_q) + \xi$$

pour tout

$$i \in \{1, \dots, n\}$$

$$(x_{1,i}, x_{2,i}, \dots, x_{p,i})$$

est une réalisation du vecteur aléatoire  $(x_1, x_2, \dots, x_p)$ , sachant que  $(x_1, x_2, \dots, x_p) = (x_{1,i}, x_{2,i}, \dots, x_{p,i})$

,  $y_i$  est une réalisation de

$$y_i = f(x_{1,i}, \dots, x_{p,i}, \theta_0, \dots, \theta_q) + \xi$$

on appelle valeur moyenne de  $y$  lorsque  $(X_1, \dots, X_p) = (x_1, \dots, x_p) = x$  le réel inconnu

$$y_x = E(y / \{(x_1, \dots, x_p) = x\}) = f(x_1, \dots, x_p, \theta_0, \dots, \theta_q) = \theta_0 + \theta_1 g_1(x_1) + \dots + \theta_p g_p(x_p)$$

$$y_n = \sum_{i=1}^n y_{n-1} + \xi_n$$

les données se présentent généralement sous la forme d'un tableau :

groupe	$x_1$	...	$x_p$	Effectif	$N$
$g_1$	$x_{1,1}$	...	$x_{p,1}$	$n_1$	$\sum_{i=1}^{n_1} y_i$
$g_2$	$x_{1,2}$	...	$x_{p,2}$	$n_2$	$\sum_{i=n_1+1}^{n_1+n_2} y_i$
...	...	...	...	...	...
$g_q$	$x_{1,q}$	...	$x_{p,q}$	$n_q$	$\sum_{i=n_{q-1}+1}^{n_{q-1}+n_q} y_i$

A partir de ce tableau on déduit la fonction de vraisemblance ,pour tout  $g \in \{1, \dots, q\}$

$(x_{1,g}, \dots, x_{p,g})$  est une réalisation du vecteur aléatoire réel  $(x_1, \dots, x_p)$ , sachant que  $(x_1, \dots, x_p) = (x_{1,g}, \dots, x_{p,g}) = x_g$

$\sum_{i=n_{g-1}+1}^{n_{g-1}+n_g} (y_i)$  est une réalisation de  $N_g \sim (n_g, p(x_g))$

$$p(x_g) = P(\{y = 1\} \{ (x_1, \dots, x_p) = x_g \})$$

cette loi est utilisée dans tous les outils théoriques , notamment pour définir la vraisemblance associée à  $(N_1, \dots, N_q)$  (et non  $(y_1, \dots, y_n)$ ).

$$L(\theta, y) = \prod_{g=1}^q \binom{n_g}{y_g} p(x_g)^{y_g} (1 - p(x_g))^{n_g - y_g} \quad y = (y_1, \dots, y_q) \in \prod_{g=1}^q \{1, \dots, n_g\}$$

## 1.4 Les modèles semi-paramétriques

Les modèles semi-paramétriques ont alors été développés pour conjuguer les avantages des approches paramétriques et non paramétriques, à savoir la capacité d'interprétation des modèles paramétriques et la souplesse des modèles non paramétriques. Dans les modèles semi-paramétrique, la variable à expliquer  $y$  dépend généralement de  $x$  par le biais d'un nombre fini de paramètre euclidiens  $\theta_1, \theta_2, \dots, \theta_p$  et d'un paramètre fonctionnel  $f$  .<sup>7</sup>

<sup>7</sup>Thi Mong Ngoc ."estimation récursive pour les modèles semi-paramétrique ".Statistiques [math.ST].Université Sciences et Technologies-Bordeaux I.2010

Soit le modèle semi-paramétrique suivant :

$$y = f(x\theta) + \xi$$

ou

$\theta \in \mathbb{R}^p$  : est une paramètre inconnu

le bruit  $\xi$  est indépendant de  $x$  et il n'y a aucune hypothèse sur la loi de  $\xi$

La fonction  $f$  est inconnue.

Dans ce modèle semi-paramétrique l'objectif est l'estimation du paramètre  $\theta$  et la fonction de lien  $f$ . Les techniques d'estimations semi-paramétriques (la méthode SIR (Sliced Inverse Régression) pour estimé le paramètre , ensuite la fonction de lien  $f$  peut être estimée via une méthode non paramétrique ) .

Description du modèle :

soit le modèle semi-paramétrique de régression proposé par Li(1991) :

$$y = f(\theta_1^t x, \dots, \theta_k^t x, \xi)$$

ou :

- $y$  : une variable unidimensionnelle a expliquer
- $x$  : une variable aléatoire explicative de dimension  $p$  avec  $E(x) = \mu$  et une matrice de covariance  $V(x) = \Sigma$  définie positive
- $\theta_1, \dots, \theta_k$  : sont des vecteurs de paramétrés réels de dimension  $p$ , avec  $k < p$ .
- $\xi$  : un terme d'erreur aléatoire indépendant de  $x$ .
- $f$  : le paramétré fonctionnel à valeur dans  $\mathbb{R}$ .

Comme la valeur de  $k < p$ , alors ce modèle est donc aussi un modèle permettant une réduction de dimension de l'espace des variable explicatives et on appelle indices des terme  $(\theta_k^t x)$   $K = 1, \dots, k$ ). Ainsi pour  $k > 1$ , on parlera de modèle multi-indices, et lorsque  $k = 1$ , de modèle à un seul indice  $\theta^t x$ .

Condition fondamentale :

la valeur explicative  $x$  suit une distribution de probabilité non dégénérée telle que

$\forall b \in \mathbb{R}^p$ , l'espérance conditionnelle  $E[b'x/\theta'_1x, \dots, \theta'_kx]$  est linéaire en  $\theta'_1x, \dots, \theta'_kx$  **Définition :**

On appelle  $E$  l'espace (EDR) (Effective Dimension Réduction). De plus, toute combinaison linéaire des  $(\theta'_p)$  appartenant à  $E$ , il est appelée direction (EDR)

l'objectif majeur que l'on se fixe dans ce modèle est l'estimation de la partie paramétrique, c'est à dire l'estimation d'une base de l'espace EDRE (Effective Dimension Réduction estimée). L'estimation de la partie fonctionnelle  $f$  peut ensuite être obtenue avec les méthodes non paramétriques usuelles.

### 1.4.1 Autre écriture du modèle :

On va voir par la suite qu'il sera pratique de considérer la version standardisée<sup>8</sup>  $z$  de  $x$  :

$$z = \Sigma^{-1/2}(x - \mu)$$

où :  $\mu$  représente la moyenne de  $x$  et  $\Sigma$  la matrice variance covariance. On peut alors réécrire le modèle  $y = f(\theta'_1x, \dots, \theta'_kx, \xi)$  sous la forme :

$$y = g(\eta_1^t z, \dots, \eta_k^t z, \xi)$$

où :  $\eta_k = \Sigma^{1/2}\theta_k, k = 1, \dots, k$  notons  $n = [\eta_1, \dots, \eta_k]$  et  $E^s = \text{vect}(n)$

$E^s$  : La direction EDR (Effective Dimension Réduction) standardisée

Remarque :

la fonction  $g$  est identique à la fonction  $f$  à une translation près de ses  $k$  premiers arguments. en effet

$$g(\eta_1^t z, \dots, \eta_k^t z, \xi) = g(\eta_1^t \Sigma^{-1/2}(x - \mu), \dots, \eta_k^t \Sigma^{-1/2}(x - \mu), \xi) = g(\theta_1^t x - \theta_1^t \mu, \dots, \theta_k^t x, \xi)$$

Principe :

---

<sup>8</sup>Aboubacar Amiri. "thèse méthode récursive en estimation et prévision non paramétrique". 42èmes journées de statistique. 2010 Marseille, France

Dans ce modèle, on a vu que le paramètre  $\theta$  n'est pas totalement identifiable seul la direction de  $\theta$  l'est. On parle de direction EDR (comme "Effective Dimension Réduction") Nous proposons une approche récursive de la méthode SIR pour l'estimation de la direction  $\theta$  dans la cadre de ce modèle semi-paramétrique sans avoir à estimer le paramètre fonctionnel  $f$  ni à spécifier la loi de l'erreur  $\xi$ .

Nous proposons dans un premier temps une forme récursive pour l'estimation de la matrice d'intérêt.

Le modèle  $y = f(x\theta) + \xi$  fait partie de la classe des modèles semi-paramétrique .le but de cette gamme de modèle est d'offrir un compromis entre le modèle paramétrique et le modèle non paramétrique. En effet, le modèle paramétrique et le modèle non paramétrique présentent tous deux des inconvénients, et le modèle de régression est trop réducteur et ne permet pas de prendre en compte toutes les interactions.

### 1.4.2 La méthode SIR :

La méthode SIR est une méthode de régression semi-paramétrique reposant sur un argument géométrique.<sup>9</sup>

**SIR** : Slice Invers Régression (régression inverse par tranches)

le concept **Sliced** :discrétisation (ou "tranchage")de  $y$

- qui va permettre de simplifier l'estimation de moments intervenant dans les propriétés géométriques.
- il ne modifie pas la partie paramétrique du modèle.

le concept **Inverse** : sert l'utilisation des propriétés géométriques des moments "inverse" de  $x$  sachant  $y$ . d'ou le calcul de  $E(x|y)$  ,  $V(x|y)$  .

Le nom de la méthode (SIR), fait référence à deux aspects : ‘

---

<sup>9</sup>Saracco Jérôme.Larramédy I réme.Aragon Yves."La régression inverse par tranche ou méthode SIR présentation générale ".Université Toulouse I

- une propriété géométrique essentielle de la courbe de régression inverse faisant intervenir divers moments du couple  $(x, y)$
- une discrétisation (slicing) de la variable dépendante qui ne modifie pas la partie paramétrique du modèle mais qui simplifie considérablement l'estimation des moments intervenant dans la propriété géométrique.

Mais nous verrons d'une part que l'on peut combiner plusieurs discrétisation et d'autre part que l'on peut estimer directement les moments. Toutes les méthodes d'estimation que nous envisageons sont des méthodes de substitution de moments empiriques aux moments théoriques.

Remarque :Le paramètre  $\theta$  n'est pas totalement identifiable, seule la direction  $\theta$  est à identifier d'où la direction EDR (Effective Dimension Réduction) est un compromis.

Application de la méthode SIR :

Bernard-Michel et al (2008) et al (2009) ont proposé de régulariser la méthode *SIR* (Régression inverse par tranches) en introduisant un a priori gaussien. Lorsque la dimension  $p$  est grande, la matrice de variance covariance  $\hat{\Sigma}$  des variable explicatives  $x$  est généralement mal conditionnée rendant alors délicate l'application de la méthode *SIR* (car elle est basée sur l'inversion de la matrice  $\hat{\Sigma}$ ). Dans ce cadre ,Bernard-Michel et al proposent une méthode de régularisation de *SIR* en introduisant un a priori Gaussien .cette régularisation a pour effet l'amélioration des conditions de la matrice de variance?covariance des variable explicatives .

H

Remarque :onconsidreunetransformationmonotone

TdeYT(y) = f\*( $\theta_1^t x, \dots, \theta_k^t x, \xi$ )oulafonctionf\* est la composée des fonctions  $T$  et  $f$  ,ainsi  $T(y)$  obéit à un modèle de type  $y = f(\theta_1^t x, \dots, \theta_k^t x, \xi)$

les vecteur propre  $v_1, \dots, v_k$  associés aux  $K$  valeurs propres non nulles  $\lambda_1, \dots, \lambda_k$  de la matrice  $\Sigma^{-1}\Gamma_T$  sont des directions *EDR*. ces vecteurs étant  $\sum$  -orthonormes ,il ferme-

ment une base de l'espace  $EDRE$  (Effective Dimension Réduction Estimé).

en note  $V = [v_1, \dots, v_k]$  matrice de dimension  $p * k$ , on a alors  $\text{vect}(V) = E$ .

La notation "vect" est la transformation de la matrice  $V$  en un vecteur. soit  $M = [m_1, \dots, m_p]$  une matrice de dimension  $q * p$  ou les  $m_k, k=1, \dots, k$ , sont des vecteurs colonnes de dimension  $q$ . Alors  $\text{vec}(M)$  est la transformation de  $M$  en un vecteur de dimension  $qp$  de la manière suivante :

$$\text{vec}(M) = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_p \end{bmatrix}$$

Diviser le support des  $y_i$  en  $H$  tranches  $s_1, \dots, s_H$  et calculer  $p_h$  la proportion empirique des  $y_i$  tombant dans la tranche  $s_h$

$$\hat{p}_h = \frac{1}{n} \sum_{i=1}^n I_{[y_i \in s_h]} \quad h = 1, \dots, H,$$



# Chapitre 2

## les méthodes récursives

### 2.1 les méthodes récursives :

Soit le modèle suivant :<sup>1</sup>

$$y = f(\theta x, \xi)$$

$\theta$  : est un vecteur de paramètre inconnu de  $\mathbb{R}^p$

$\xi$  : est le terme d'erreur aléatoire indépendant de  $x$  , Aucune hypothèse n'est faite sur la distribution de  $\xi$ .

$f$  : La fonction de lien est un paramètre fonctionnel à valeur dans  $\mathbb{R}$ , inconnu et arbitraire. L'idée est d'utiliser les calculs d'estimations sur la base de données initiales et de les remettre à jour en tenant compte uniquement des nouvelles données arrivant à la base de l'échantillon.

Le gain en termes de temps de calcul peut être très intéressant.

Nous proposons une approche récursive de la méthode (Régression inverse par tranches)SIR pour l'estimateur de la direction  $\theta$  dans le cadre de ce modèle semi-paramétrique sans avoir à estimer le paramètre fonctionnel  $f$  ni à spécifier la loi de l'erreur  $\xi$ .

L'avantage des méthodes récursives :

- Prendre en compte l'arrivée temporelle des informations et affiner ainsi au fil du temps les algorithmes d'estimation mis en ouvre.

---

<sup>1</sup>NFA031CNAM201

- le gain en termes de temps de calcul peut être très intéressant et les applications d'une telle approche sont nombreuses.
- Il n'est pas nécessaire de relancer tous les calculs d'estimation des paramètres du modèle à chaque fois que la base de données est complétée par de nouvelles observations.

Remarque :

Les méthodes récursives d'estimation n'ont jamais été développées dans le cadre de ce modèle semi-paramétrique.

### 2.1.1 Estimation récursive du paramètre $\theta$

Nous rappelons tout d'abord brièvement la méthode SIR. Ensuite, nous proposons un estimateur récursif  $\theta_n$  ( $n$  le nombre de l'échantillon) de la direction de  $\theta$ .

**Méthode SIR :**

la matrice  $\Gamma = Vect[E(x|T(y))]$  est dégénérée dans toute direction  $\sum$  est orthogonal à  $\theta$ . Ainsi le vecteur propre  $\theta^\sim$  associé à la valeur propre  $\lambda$  non nulle de  $\sum^{-1} \Gamma$  est colinéaire à  $\theta$ . ce vecteur propre  $\theta^\sim$  est donc une direction *EDR* de  $\theta$ .<sup>2</sup>

Étude du cas  $H=2$  permettant d'obtenir une expression analytique de  $\theta^\sim$  et de son estimateur  $\theta_n^\sim$

je m'intéresse au cas particulier où l'on ne considère que deux tranches notées  $s_1$  et  $s_2$  lorsque  $H = 2$ , la matrice  $\Gamma$  s'écrit :

$$\Gamma = P_1 z_1 z_1^t + P_2 z_2 z_2^t$$

Avec  $z_h = m_h - \mu$  (pour  $h=1,2$ ) et  $\mu = E(x)$ , alors calculer la valeur propre non nulle  $\lambda$  de  $\sum^{-1} \Gamma$  et de déduire le vecteur propre  $\theta^\sim$  correspondant (colinéaire à  $\theta$ ) s'écrit de la

---

<sup>2</sup>Thi Mong Ngoc ."estimation récursive pour les modèles semi-paramétrique ".Statistiques [math.ST].Université Sciences et Technologies-Bordeaux I.2010

forme :

$$\lambda = \frac{p_1}{p_2} z_1^{t-1} z_1 \quad \text{et} \quad \theta^{\sim} = \Sigma^{-1}(z_1 - z_2)$$

avec

$$\bar{x}_N = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\Sigma_N = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}_N)(x_i - \bar{x}_N)^t$$

$$\Gamma = \sum_{i=1}^H p_h (m_h - \mu)(m_h - \mu)^t$$

$$P_h = P(y_h) = \frac{N_h}{N}$$

$$m_h = E(x/y \in s_h) = \frac{1}{N_h} \sum x_i$$

**La matrice d'intérêt  $\Sigma^{-1}\Gamma$  :**

Lorsque l'on s'intéresse a la forme non récursive des estimateurs, nous disposons d'un échantillon d'observation  $(x_i; y_i), i=1, \dots, n$  des variables aléatoires  $(x, y)$  i.i.d (indépendantes et identiquement distribuées) du modèle  $y = f(\theta x, \xi)$ . Pour la forme récursive des estimateurs, on scinde cet échantillon en deux parties le sous échantillon  $(x_i, y_i), i=1, \dots, n-1$  et l'observation  $(x_n, y_n)$ , et la matrice d'intérêt s'écrit sous.

$$\Sigma_n^{-1}\Gamma_n = \Sigma_{n-1}^{-1}\Gamma_{n-1} + \Sigma_{n-1}^{-1}C_{h^*,n} - \frac{1}{n + \rho_n} \Sigma_{n-1}^{-1} \Phi_n \Phi_n^t \Sigma_{n-1}^{-1} [\Gamma_{n-1}, C_{h^*,n}]$$

$$h^* = 1 \text{ ou } 2$$

En posant :

$$C_{h^*,n} = \frac{-1}{n} \sum_{h=1}^H p_{h,n-1} (z_{h,n-1} \Phi_n^t + \Phi_n z^t z_{h,n-1}^t) + \frac{1}{n^2} \Phi_n \Phi_n^t + p_{h^*,n-1} A_{h^*,n} + \frac{1}{n-1} B_{h^*,n} B_{h^*,n}^t$$

ou :

$$\bar{x}_N = \frac{1}{N} \sum_{i=1}^N x_i = \frac{N-1}{N} \bar{x}_{N-1} + \frac{x_N}{N} = \bar{x}_{N-1} + \frac{1}{N} \Phi_N$$

$$\Phi_n = X_n - \bar{X}_{n-1}$$

$$\rho_n = \Phi_n^t \Sigma_{n-1}^{-1} \Phi_n$$

$$\begin{aligned} \Sigma_N &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}_N)(x_i - \bar{x}_N)^t \\ \Sigma_N &= \frac{N-1}{N} \Sigma_{N-1} + \frac{N-1}{N^2} (x_i - \bar{x}_N)(x_i - \bar{x}_N)^t \end{aligned}$$

$$\begin{aligned}\Sigma_N &= \frac{N-1}{N}\Sigma_{N-1} + \frac{N+1}{N^2}\Phi_N\Phi_N^t\Sigma_N^{-1} \\ \Sigma_N &= \frac{N}{N-1}\Sigma_{N-1}^{-1} - \frac{N}{(N-1)(n+\rho_n)}\Sigma_{N-1}^{-1}\Phi_N\Phi_N^t\Sigma_{n-1}^{-1}. \\ P_{h,n} &= \begin{cases} \frac{n-1}{n}Ph^*, n-1 + \frac{1}{n} & \text{si } h = h^* \\ \frac{n-1}{n}P_{h,n-1} & \text{sinon} \end{cases}\end{aligned}$$

$$\Gamma_N = \frac{N-1}{N}\Gamma_{N-1} + \frac{N-1}{N^2}\sum_{h=1}^H P_{h,N-1}(z_{h,N-1}\Phi_N^t + \Phi_N z_{h,N-1}^t) + \frac{N-1}{N^3}\Phi_N\Phi_N^t + \frac{N-1}{N}P_{h^*,N-1}A_{h^*,N} + \frac{1}{N}B_{h^*,N}B_{h^*,N}^t$$

$$z_{h,N} = m_{h,N} - \bar{x}_N \quad z_{h,N} = \begin{cases} z_{h^*,N-1} - \frac{1}{N}\Phi_N + \frac{1}{N_{h^*,N-1}+1}\Phi_{h^*,N} & \text{si } h = h^* \\ z_{h^*,N-1} - \frac{1}{N}\Phi_N & \text{sinon} \end{cases}$$

avec

$$m_{h,N} = \begin{cases} m_{h^*,N-1} + \frac{1}{N_{h^*,N-1}+1}\Phi_{h^*,N} & \text{si } h = h^* \\ m_{h,N-1} & \text{sinon} \end{cases}$$

$$\begin{aligned}A_{h^*,n} &= \frac{1}{n_{h^*,n-1}+1}(z_{h^*,n-1}\Phi_{h^*,n} + \Phi_{h^*,n}z_{h^*,n-1}^t) \\ &- \frac{1}{n(n_{h^*,n-1}+1)}(\Phi_n\Phi_{h^*,n}\Phi_n^t) + \frac{1}{(n_{h^*,n-1}+1)^2}\Phi_{h^*,n}\Phi_{h^*,n}^t\end{aligned}$$

$$B_{h^*,n} = z_{h^*,n-1} - \frac{1}{n}\Phi_n + \frac{1}{n_{h^*,n-1}+1}\Phi_{h^*,n}$$

La direction de  $\theta$  :

la forme réursive de l'estimateur est :

$$\hat{\theta}_n = \frac{n}{n-1}\hat{\theta}_{n-1} - \frac{n}{(n-1)(n-\rho_n)}\Sigma_{n-1}^{-1}\Phi_n\Phi_{n-1}^t - \frac{(-1)^{h^*}n}{(n_{h^*,n-1}(n-1))}(\Sigma_{n-1}^{-1} - \frac{1}{n+\rho_n}\Phi_n\Phi_n^t\Sigma_{n-1}^{-1})\Phi_{h^*,n}$$

avec

$$\Phi_n = x_n - \bar{x}_{n-1}$$

$$\Phi_{h^*,n} = x_n - m_{h^*,n-1}$$

$$\rho_n = \Phi_n^t\Sigma_{n-1}^{-1}\Phi_n$$

## 2.1.2 Estimation non paramétrique d'une fonction de régression :

On présente brièvement quelques estimateurs non paramétriques d'une fonction de régression<sup>3</sup>

. Estimation non paramétrique :

Les modèles non paramétriques sont des modèles dont la structure n'est pas spécifiée a priori mais déterminée à partir des données. Je m'intéresse à deux types d'estimation non paramétrique qui sont d'une part, l'estimation d'une densité de probabilité et, d'autre part, l'estimation d'une fonction de régression.

Pour l'estimation d'une fonction de régression suppose que l'on dispose d'une suite  $(x_n, y_n)$  des couples aléatoires indépendants liés, via la fonction de lien.

$$y_n = f(x_n) + \xi_n$$

$f$  : la fonction de régression est inconnue

$\xi_n$  : une suite de variables aléatoires indépendantes centrée et de même loi normale de variance  $\delta^2$

Pour proposer un estimateur non paramétrique d'une fonction de régression par :

- L'estimateur de *Nadaraya – Watson*

Régression par la méthode du noyau :

La régression du Noyau est une technique non paramétrique de statistique pour l'attente conditionnelle d'une variable aléatoire. L'objectif est de trouver une relation non linéaire entre une paire de variable aléatoire  $x$  et  $y$

$$E(x|y) = f(x)$$

ou  $f$  est une fonction inconnue.

Noyau :

---

<sup>3</sup>Philippe Fraysse . "thèse estimation récursive dans certains modèle de déformation ". Université Bordeaux1.2013

Un noyau  $K$  est une fonction définie sur et à valeur dans  $\mathbb{R}^+$  qui est intégrable et vérifie pour tout  $x \in \mathbb{R}$  :

$$K(x) = K(-x) \quad \text{et} \quad \int_{-\infty}^{+\infty} K(x)dx = 1$$

L'estimateur de Nadaraya Watson :

Si les couple  $(x_n, y_n)$  nous arrivent séquentielle-ment, il est préférable d'utiliser une version récursive de l'estimateur de *Nadaraya Watson*, qui est défini pour tout  $n \geq 1$  et pour  $x \in \mathbb{R}$  par :

$$\hat{f}_n(x) = \begin{cases} \frac{\sum_{k=1}^n \frac{1}{h_k} k\left(\frac{x_k-x}{h_k}\right) y_k}{\sum_{k=1}^n \frac{1}{h_k} k\left(\frac{x_k-x}{h_k}\right)} & \text{si } \sum_{k=1}^n \frac{1}{h_k} k\left(\frac{x_k-x}{h_k}\right) \neq 0 \\ 0 & \text{sinon} \end{cases}$$

### 2.1.3 Estimateur de Nadaraya Watson

on se limite au cas d'une seule variable explicative  $x_1$ . on a appelle estimateur de *Nadaraya Watson*<sup>4</sup>

$$\hat{f}_n(x) = \frac{\sum_{i=1}^n k_h(x - x_{1,i}) y_i}{\sum_{i=1}^n k_h(x - x_{1,i})}$$

ou  $k_h(u) = \left(\frac{1}{h}\right) k_{\frac{u}{h}}$

$h$  : fenêtre

L'idée de la construction de  $\hat{f}_n$  est la suivante en notant  $g(x, y)$  la densité de  $(x_1; y)$  et  $g(x)$  celle de  $x_1$  on a

$$f(x) = E(y/x_1 = x) = \int yg(y/x)dy = \frac{(\int yg(y, x)dy)}{g(x)} = \frac{\int yg(y, x)dy}{\int g(y, x)dy}$$

un estimateur de  $\int yg(x)dy$  est

$$\frac{1}{n} \sum_{i=1}^n k_h(x - x_{1,i}) y_i$$

#### Convergence en probabilité :

Pour tout  $x \in \mathbb{R}$  par :

$$\hat{f}_n(x) = \frac{1}{n} \sum_{k=1}^n \frac{1}{h_k} \left( \frac{x_k - x}{h_k} \right)$$

---

<sup>4</sup>Christophe Crambés. "Modèle linéaire fonctionnel lorsque la variable explicative est bruitée ". Université Paul Sabatier. mardi 12 juin 2007

On suppose que  $f$  soit dérivable à dérivée bornée et que la fonction de lissage est  $h_k$  :

$$\text{si } \lim_{n \rightarrow +\infty} \hat{f}_n(x) = f(x) \quad P.s$$

# Chapitre 3

## simulation

On considère un échantillon de taille  $n$  et on utilise la méthode de simulation pour vérifier l'efficacité de la méthode récursive dans un modèle semi-paramétrique.

On considère une matrice de taille  $n=6, n=7, n=8, n=9$  et  $n=10$  comme exemple d'étude et comparée entre la méthode des moindres carrés  $\hat{\theta}$  et la méthode de récursive de direction de  $\tilde{\theta}$ .

$$\hat{\theta} = (X^t X)^{-1} X^t Y$$

$$\hat{\theta}_n = \frac{n}{n-1} \hat{\theta}_{n-1} - \frac{n}{(n-1)(n-\rho_n)} \Sigma_{n-1}^{-1} \Phi_n \Phi_{n-1}^t - \frac{(-1)^{h^*} n}{(n_{h^*,n-1}(n-1))} (\Sigma_{n-1}^{-1} - \frac{1}{n+\rho_n} \Phi_n \Phi_n^t \Sigma_{n-1}^{-1}) \Phi_{h^*,n}$$



```

> x1<-matrix(data=c(2,1,2,4,7,3,1,3,3,6,2,1,3,5,4,2,1,2,6,1,7,1,3,4,4,1,1,5,4,7),6,5)
> x1
      [,1] [,2] [,3] [,4] [,5]
[1,]    2    1    3    6    4
[2,]    1    3    5    1    1
[3,]    2    3    4    7    1
[4,]    4    6    2    1    5
[5,]    7    2    1    3    4
[6,]    3    1    2    4    7
> y1<-c(1,3,4,2,7,3)
> y1
[1] 1 3 4 2 7 3
> a1<-t(x1)
> a1
      [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    2    1    2    4    7    3
[2,]    1    3    3    6    2    1
[3,]    3    5    4    2    1    2
[4,]    6    1    7    1    3    4
[5,]    4    1    1    5    4    7
> b1<-a1%*%y1
> b1
      [,1]
[1,]    79
[2,]
[3,]    51
[4,]    72
[5,]    70
> c1<-a1%*%x1
> c1
      [,1] [,2] [,3] [,4] [,5]
[1,]    83    52    40    64    80
[2,]    52    60    46    46    55
[3,]    40    46    59    64    49
[4,]    64    46    64   112    77
[5,]    80    55    49    77   108
> d1<-solve(c1)
> d1
      [,1] [,2] [,3] [,4] [,5]
[1,]  0.05994456 -0.036118516  0.0284793801 -0.017512600 -0.0264450500
[2,] -0.03611852  0.080238581 -0.0639726571  0.027429287 -0.0046392184
[3,]  0.02847938 -0.063972657  0.0965136618 -0.044994474 -0.0002263816
[4,] -0.01751260  0.027429287 -0.0449944739  0.039290005 -0.0085944809
[5,] -0.02644505 -0.004639218 -0.0002263816 -0.008594481  0.0374410068
> o1<-d1%*%b1
> o1
      [,1]
[1,]  1.23396317
[2,] -0.37363726
[3,]  0.65401344
[4,] -0.05205325
[5,] -0.33523669
> x2<-matrix(data=c(2,1,2,4,7,3,1,1,3,3,6,2,1,3,3,5,4,2,1,2,4,6,1,7,1,3,4,2,4,1,1,5,4,7,1),7,5)
)
> x2
      [,1] [,2] [,3] [,4] [,5]
[1,]    2    1    3    6    4
[2,]    1    3    5    1    1
[3,]    2    3    4    7    1
[4,]    4    6    2    1    5
[5,]    7    2    1    3    4
[6,]    3    1    2    4    7
[7,]    1    3    4    2    1
> y2<-c(1,3,4,2,7,3,5)
> y2
[1] 1 3 4 2 7 3 5
> a2<-t(x2)
> a2
      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]    2    1    2    4    7    3    1
[2,]    1    3    3    6    2    1    3
[3,]    3    5    4    2    1    2    4
[4,]    6    1    7    1    3    4    2
[5,]    4    1    1    5    4    7    1

```

```

> b2<-a2%*%y2
> b2
      [,1]
[1,] 84
[2,] 66
[3,] 71
[4,] 82
[5,] 75
> c2<-a2%*%x2
> c2
      [,1] [,2] [,3] [,4] [,5]
[1,] 84 55 44 66 81
[2,] 55 69 58 52 58
[3,] 44 58 75 72 53
[4,] 66 52 72 116 79
[5,] 81 58 53 79 109
> d2<-solve(c2)
> d2
      [,1] [,2] [,3] [,4] [,5]
[1,] 0.05993307 -0.036115461 0.028102764 -0.017383980 -0.026385269
[2,] -0.03611546 0.080237768 -0.063872440 0.027395062 -0.004655126
[3,] 0.02810276 -0.063872440 0.084159635 -0.040775363 0.001734594
[4,] -0.01738398 0.027395062 -0.040775363 0.037849107 -0.009264187
[5,] -0.02638527 -0.004655126 0.001734594 -0.009264187 0.037129738
> o2<-d2%*%b2
> o2
      [,1]
[1,] 1.2416725
[2,] -0.3756887
[3,] 0.9068999
[4,] -0.1384183
[5,] -0.3753778
> x3<-matrix(data=c(2,1,2,4,7,3,1,2,1,3,3,6,2,1,3,4,3,5,4,2,1,2,4,5,6,1,7,1,3,4,2,1,4,1,1,5,4,
7,1,2),8,5)
> x3
      [,1] [,2] [,3] [,4] [,5]
[1,] 2 1 3 6 4
[2,] 1 3 5 1 1
[3,] 2 3 4 7 1
[4,] 4 6 2 1 5
[5,] 7 2 1 3 4
[6,] 3 1 2 4 7
[7,] 1 3 4 2 1
[8,] 2 4 5 1 2
> y3<-c(1,3,4,2,7,3,5,2)
> y3
[1] 1 3 4 2 7 3 5 2
> a3<-t(x3)
> a3
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,] 2 1 2 4 7 3 1 2
[2,] 1 3 3 6 2 1 3 4
[3,] 3 5 4 2 1 2 4 5
[4,] 6 1 7 1 3 4 2 1
[5,] 4 1 1 5 4 7 1 2
> b3<-a3%*%y3
> b3
      [,1]
[1,] 88
[2,] 74
[3,] 81
[4,] 84
[5,] 79
> c3<-a3%*%x3
> c3
      [,1] [,2] [,3] [,4] [,5]
[1,] 88 63 54 68 85
[2,] 63 85 78 56 66
[3,] 54 78 100 77 63
[4,] 68 56 77 117 81
[5,] 85 66 63 81 113
> d3<-solve(c3)
> d3
      [,1] [,2] [,3] [,4] [,5]

```

```

[1,] 0.05869906 -0.034698255 0.023135600 -0.014424737 -0.026446674
[2,] -0.03469825 0.078610176 -0.058167897 0.023996518 -0.004584605
[3,] 0.02313560 -0.058167897 0.064165794 -0.028863813 0.001487426
[4,] -0.01442474 0.023996518 -0.028863813 0.030752670 -0.009116934
[5,] -0.02644667 -0.004584605 0.001487426 -0.009116934 0.037126682
> o3<-d3%*%b3
> o3
      [,1]
[1,] 1.17086464
[2,] -0.29436938
[3,] 0.62188401
[4,] 0.03138307
[5,] -0.37890121
> x4<-matrix(data=c(2,1,2,4,7,3,1,2,1,1,3,3,6,2,1,3,4,3,3,5,4,2,1,2,4,5,7,6,1,7,1,3,4,2,1,3,4,
1,1,5,4,7,1,2,4),9,5)
> x4
      [,1] [,2] [,3] [,4] [,5]
[1,] 2 1 3 6 4
[2,] 1 3 5 1 1
[3,] 2 3 4 7 1
[4,] 4 6 2 1 5
[5,] 7 2 1 3 4
[6,] 3 1 2 4 7
[7,] 1 3 4 2 1
[8,] 2 4 5 1 2
[9,] 1 3 7 3 4
> y4<-c(1,3,4,2,7,3,5,2,1)
> y4
[1] 1 3 4 2 7 3 5 2 1
> a4<-t(x4)
> a4
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,] 2 1 2 4 7 3 1 2 1
[2,] 1 3 3 6 2 1 3 4 3
[3,] 3 5 4 2 1 2 4 5 7
[4,] 6 1 7 1 3 4 2 1 3
[5,] 4 1 1 5 4 7 1 2 4
> b4<-a4%*%y4
> b4
      [,1]
[1,] 89
[2,] 77
[3,] 88
[4,] 87
[5,] 83
> c4<-a4%*%x4
> c4
      [,1] [,2] [,3] [,4] [,5]
[1,] 89 66 61 71 89
[2,] 66 94 99 65 78
[3,] 61 99 149 98 91
[4,] 71 65 98 126 93
[5,] 89 78 91 93 129
> d4<-solve(c4)
> d4
      [,1] [,2] [,3] [,4] [,5]
[1,] 0.05820289 -0.037024268 0.026450051 -0.015778470 -0.025052100
[2,] -0.03702427 0.067706033 -0.042630044 0.017650340 0.001953036
[3,] 0.02645005 -0.042630044 0.042025137 -0.019820830 -0.007828383
[4,] -0.01577847 0.017650340 -0.019820830 0.027059214 -0.005312047
[5,] -0.02505210 0.001953036 -0.007828383 -0.005312047 0.033207002
> o4<-d4%*%b4
> o4
      [,1]
[1,] 1.2047415
[2,] -0.1355576
[3,] 0.3955852
[4,] 0.1238110
[5,] -0.4741177
> x5<-matrix(data=c(2,1,2,4,7,3,1,2,1,2,1,3,3,6,2,1,3,4,3,3,3,5,4,2,1,2,4,5,7,1,6,1,7,1,3,4,2,
1,3,5,4,1,1,5,4,7,1,2,4,2),10,5)
> x5
      [,1] [,2] [,3] [,4] [,5]
[1,] 2 1 3 6 4

```

```

[2,] 1 3 5 1 1
[3,] 2 3 4 7 1
[4,] 4 6 2 1 5
[5,] 7 2 1 3 4
[6,] 3 1 2 4 7
[7,] 1 3 4 2 1
[8,] 2 4 5 1 2
[9,] 1 3 7 3 4
[10,] 2 3 1 5 2
> y5<-c(1,3,4,2,7,3,5,2,1,6)
> y5
[1] 1 3 4 2 7 3 5 2 1 6
> a5<-t(x5)
> a5
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,] 2 1 2 4 7 3 1 2 1 2
[2,] 1 3 3 6 2 1 3 4 3 3
[3,] 3 5 4 2 1 2 4 5 7 1
[4,] 6 1 7 1 3 4 2 1 3 5
[5,] 4 1 1 5 4 7 1 2 4 2
> b5<-a5%*%y5
> b5
      [,1]
[1,] 101
[2,] 95
[3,] 94
[4,] 117
[5,] 95
> c5<-a5%*%x5
> c5
      [,1] [,2] [,3] [,4] [,5]
[1,] 93 72 63 81 93
[2,] 72 103 102 80 84
[3,] 63 102 150 103 93
[4,] 81 80 103 151 103
[5,] 93 84 93 103 133
> d5<-solve(c5)
> d5
      [,1] [,2] [,3] [,4] [,5]
[1,] 0.052953865 -0.027380953 0.018473634 -0.008961891 -0.025711900
[2,] -0.027380953 0.049989672 -0.027976049 0.005127160 0.003165197
[3,] 0.018473634 -0.027976049 0.029904162 -0.009462348 -0.008831016
[4,] -0.008961891 0.005127160 -0.009462348 0.018206943 -0.004455206
[5,] -0.025711900 0.003165197 -0.008831016 -0.004455206 0.033124066
> o5<-d5%*%b5
> o5
      [,1]
[1,] 0.99249979
[2,] 0.25436547
[3,] 0.07306243
[4,] 0.39943630
[5,] -0.50079643

```

Après le calcul des deux méthodes par la matrice pour l'échantillon de taille (n=6,7,8,9 et 10) on résume dans les tableaux en coordination . avec le  $\theta_{moy} = \sum_{i=6}^{10} \frac{\theta_i}{5}$

P \ N	N=6	N=7	N=8	N=9	N=10	$\theta_{i_{moy}}$
$\theta_1$	1.2339	1.2339	1.1708	1.2047	0.9924	0.367148
$\theta_2$	-0.3736	-0.375	-0.2943	-0.1355	0.2543	-0.18482
$\theta_3$	0.6540	0.906	0.6218	0.3955	0.0730	0.53006
$\theta_4$	-0.0520	-0.1384	0.0313	0.1238	0.3994	0.07282
$\theta_5$	-0.3352	-0.3713	-0.3789	-0.4741	-0.5007	0.41204

$\tilde{\theta}_1$	$\tilde{\theta}_2$	$\tilde{\theta}_3$	$\tilde{\theta}_4$	$\tilde{\theta}_5$	$\tilde{\theta}_{moy}$
1.1685894	1.1432135	1.0998997	0.9514861	0.7939727	1.0314323
-0.3355761	-0.2946205	-0.2382450	-0.0205609	-0.0971246	-0.1972254
0.4932551	0.8138125	0.5892525	0.2547155	0.3701785	0.5042428
-0.04319247	-0.10460251	-0.0095663	-0.2964732	-0.2938081	-0.1495285
0.1915465	0.1685710	0.0944899	0.1505803	0.1371147	0.1484605

En suite minimiser l'erreur quadratique de  $(\hat{\theta} - \tilde{\theta})$  est résumé dans les tableaux suivante :  
avec  $E(\theta_i - \theta_{moy})^2 = Var(\theta)$

$E(\hat{\theta}_1 - \theta_{moy})^2$	$E(\hat{\theta}_2 - \theta_{moy})^2$	$E(\hat{\theta}_3 - \theta_{moy})^2$	$E(\hat{\theta}_4 - \theta_{moy})^2$	$E(\hat{\theta}_5 - \theta_{moy})^2$
0.76475025	0.75134224	0.645933690	0.70157376	0.3910001
0.03640464	0.03621409	0.011990250	0.00243049	0.6069968
0.01536112	0.14133088	0.008416228	0.01810639	0.2089038
0.01557504	0.04460544	0.001722250	0.00260100	0.1066676
0.55845729	0.61371556	0.625681000	0.78535044	0.8332038

$E(\tilde{\theta}_1 - \theta_{moy})^2$	$E(\tilde{\theta}_2 - \theta_{moy})^2$	$E(\tilde{\theta}_3 - \theta_{moy})^2$	$E(\tilde{\theta}_4 - \theta_{moy})^2$	$E(\tilde{\theta}_5 - \theta_{moy})^2$
0.018812076	0.012495041	0.004687788	6.391392e-03	0.0563870521
0.019140911	0.009485802	0.001682606	3.121035e-02	0.0100201742
0.000120730	0.095833387	0.007226646	6.226388e-02	0.0179732419
0.011307355	0.002018346	0.019589422	2.159274e-02	0.0208165984
0.001856405	0.000404433	0.002912824	4.493637e-06	0.0001287267

En suite on a calculé  $\sum_{i=6}^{10} E(\theta_i - \theta_{moy})^2$  est a donné les deux vecteurs suivante :

$$\sum_{i=6}^{10} E(\hat{\theta}_i - \hat{\theta}_{moy})^2 = \begin{pmatrix} 0.65092001 \\ 0.13880726 \\ \tilde{0.07842369} \\ 0.03423426 \\ 0.68328163 \end{pmatrix}, \quad \sum_{i=6}^{10} E(\tilde{\theta}_i - \tilde{\theta}_{moy})^2 = \begin{pmatrix} 0.019754670 \\ 0.0114307969 \\ 0.036683578 \\ 0.015064892 \\ 0.001061376 \end{pmatrix}$$

Et en dénouement en comparé entre les deux méthode

# Résultat

Après les calculs en comparant les deux méthodes (moindres carrés, la direction de  $\theta$ ) on remarque que l'erreur quadratique de la méthode de direction de  $\theta$  est beaucoup moins grande à partir de la méthode des moindres carrés

$$\text{et : } \lim_{n \rightarrow \infty} \tilde{\theta} = 0$$

donc la méthode de direction de  $\theta$  (sans biais).



# Conclusion

Dans ce mémoire on étudie l'estimateur récursif du modèle semi-paramétrique. Après une présentation théorique on a étudié le calcul par la méthode moindres carrés dans la régression en considérant l'échantillon de taille  $n$  et en passant à l'échantillon de taille  $(n + 1)$ ,  $n=6, 7, 8, 9$  et  $10$ .

De plus on a utilisé la méthode de récursion de direction de  $\theta$  pour  $n=6, 7, 8, 9$  et  $10$ , et on a calculé la variance pour les deux approches. Afin de montrer l'efficacité des estimateurs récursifs.

# Bibliographie

- [1] NFA031CNAM2012
- [2] Thi Mong Ngoc ."estimation récursive pour les modèles semi-paramétrique".Statistiques [math.ST].Université Sciences et Technologies-Bordeaux I.2010
- [3] christope chesneau."Modèle de régression" .Université de Caen
- [4] Aboubacar Amiri."thèse méthode récursive en estimation et prévision non paramétrique".42èmes journées de statistique.2010 Marseille,France
- [5] Laurent Delsol."Regression sur vaucture et Applications" [math].Université Paul Sabatier -Toulouse.2008
- [6] Carole Toque."L'indentification de modèle factoriels de séries Temporelles".Ecole Doctorale d'informatique télécommunications et électronique de Paris.13.9.2006
- [7] Plilippe Fraysse."these estimation récursive dans certains modèle de déformation".Université Bordeaux1.2013
- [8] Gaelle Chagny."Estimation non paramétrique d'une fonction de régression avec des bases déformées".MAP5.2010
- [9] Thi Mong Ngoc NGUYEN.Bernard BERCU et Jéome SARACCO."Estimateur récursif de la fonction de lien dans un modèle semi-paramétrique" .9ème Colloque "jeurnés Probabilistes et statisticiens .Mai2010
- [10] Christophe Crambes".Modèle linéaire fonctionnel lorsque la variable explicative est bruitée".Université Paul Sabatier.mardi 12 juin 2007

- [11] Saracco Jérôme.Larramédy I réme.Aragon Yves."La régression inverse par tranche ou méthode SIR présentation générale".Université Toulouse I
- [12] " Méthode des moindres carres".site web
- [13] "Estimation par noyau".site web .