

Remerciement

A l'échéance de ce mémoire de fin d'étude. Je dois en premier lieu l'énorme remerciement à notre dieu le Miséricordieux, aussi que mes très chers parents pour leurs patience et le grand sacrifice pour l'aboutissement de mes études

Aussi mes remerciements les plus sincères à mon encadreur **Dr. Fatima Benziadi** qui m'a fait l'honneur d'être mon encadreur . Je la remercie profondément pour son encouragement continue et aussi d'être toujours la pour m'écouter, m'aider et me guider à retrouver le bon chemin par sa sagesse et ses précieux conseils. ainsi que son soutien moral et sa preuve de compréhension, ce qui m'a donné la force et le courage d'accomplir ce mémoire.

Je voudrais également remercier ma cousine **Djellouli Zohra** qui a toujours été là pour moi. Son soutien inconditionnel et son encouragements a été d'une grande aide.

Aussi au jury à leur tête le président et les examinateurs pour avoir honorer de leurs présences à ma soutenance.

Sans oublier tous mes enseignants durant les années des études, mes collègues aussi que toutes les personnes qui ont contribués de près ou de loin pour mener à bien mon travail.

Merci à tous

Lexique

Abréviations

SIR	Sliced Inverse Regression
i.i.d	indépendantes et identiquement distribuées.
p.s.	presque sûrement.
MCO	Moindres Carrés Ordinaires
SCR	somme des carrés des résidus
CNC	Computer Numerical Control

Notations

$\mathbb{1}_A$	Indicatrice sur A.
0_n	la matrice nulle de dimension $n \times n$.
I_p	la matrice identité de dimension $p \times p$
X'	Transposé de X
X^{-1}	Inverse de X
$rg(A)$	le rang de la matrice A
$A \otimes B$	le produit de Kronecker entre A et B.
$\xrightarrow{\mathcal{L}}$	la convergence en loi.
$\arg \min$	l'argument du minimum
$Pr(y \in D)$	la proportion de y contenant dans D

Introduction

Au fil des premières années de 1800, Legendre et Gauss ont façonné une forme de raisonnement et approche qui a ensuite été nommée « Régression ». C'était un moment crucial pour Statistiques et le premier et le plus célèbre différent prioritaire sur la découverte dans ce domaine[35], dont l'objectif est d'étudier les relations entre une variable à expliquer y et une variable explicative x (unidimensionnelle ou multidimensionnelle). cette étude permet de modéliser des phénomènes réels dans les domaines de l'économie, la biostatistique ou encore les sciences de l'environnement. Dans la littérature statistique, deux grandes classes de modèles de régression sont omniprésentes : les modèles paramétriques et les modèles non paramétriques. Ces deux types de modèles sont caractérisés par des avantages et des défauts spécifiques. Afin de 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, les modèles semi-paramétriques ont été développés. Dans ce mémoire, nous nous focalisons au modèle semi-paramétrique de régression de la forme $y = r(\theta'x, \varepsilon)$, lorsque la variable à expliquer y est à valeurs dans \mathbb{R} et la covariable x appartient à \mathbb{R}^p . La méthode d'estimation de la partie paramétrique du ce modèle, s'appelle la méthode SIR, pour « Sliced Inverse Regression » (que l'on peut traduire par « Régression inverse par tranches ». Cette méthode a été introduite par Duan et Li (1991)[27]. Elle ne nécessite aucune spécification en ce qui concerne la distribution de l'erreur ε et la fonction de lien r , repose sur une propriété géométrique de la courbe de régression inverse qui basée sur une condition cruciale de linéarité de la distribution de la covariable.

Notre objectif dans ce travail est d'estimer le paramètre θ de ce modèle par des méthodes récursives. pour ce faire nous allons proposer une approche récursive de la méthode SIR. L'avantage majeur de méthodes récursives en général est leur rapidité

par rapport aux méthodes non récursives. De plus elles permettent une mise à jour rapide des estimateurs lorsque les données sont observées de manière séquentielle sans être obligé de stocker en mémoire toutes les observations passées.

Ce travail est scindée en quatre chapitres ;

Dans **le premier chapitre** , nous donnons une présentation générale de différents modèles de régression, puis nous décrivons le modèle de régression semi-paramétrique dans le cadre duquel nous allons travailler. Nous énonçons enfin une condition fondamentale sur la distribution de la variable explicative, condition nécessaire dans la théorie de SIR.

Dans **le deuxième chapitre**, nous présentons une synthèse de travaux concernant la méthode SIR. Nous présentons tout d'abord cette approche dans le cas univarié ($y \in \mathbb{R}$), puis nous mentionnons quelques méthodes dans le cas multivarié ($y \in \mathbb{R}^q$), Enfin, nous donnons certaines ses applications.

Dans **le troisième chapitre**, Nous décrivons l'approche récursive de la méthode SIR. Nous nous intéresserons au cas particulier où l'on ne considère que deux tranches, nous donnons dans un premier temps l'expression analytique de l'estimateur (non récursive) de la partie paramétrique du modèle, puis nous proposons une forme récursive pour ce estimateur. Nous montrons enfin des propriétés asymptotiques pour l'estimateur proposé.

Dans **le dernier chapitre**, nous illustrons sur des simulations avec logiciel R, le bon comportement numérique de cet estimateur récursif.

Table des matières

Introduction général	7
1 Modèles de Régression	11
1.1 Modèles paramétriques	11
1.1.1 Modèle d'étude	12
1.1.2 Méthode des moindres carrés ordinaires	14
1.1.3 Quelques propriétés de l'estimateur $\hat{\theta}$	15
1.1.4 Méthode du maximum de vraisemblance	17
1.1.5 Quelques propriétés de $\tilde{\theta}$ et $\tilde{\sigma}_\varepsilon^2$	19
1.1.6 Avantages et défauts spécifiques	20
1.2 Modèles non paramétriques	20
1.2.1 Lissage par noyau	21
1.2.2 L'estimation par projection	23
1.2.3 Avantages et défauts spécifiques	24
1.3 Modèles semi-paramétriques	25
1.3.1 Modèles semi-paramétriques de référence	26
1.3.2 Condition fondamentale	27
2 SIR classique	29
2.1 SIR univariée	29
2.1.1 Philosophie et théorie de la méthode SIR	30
2.1.2 Résultats asymptotiques de SIR	35
2.2 Méthodes SIR multivariées	36
2.3 Mise en œuvre de la méthode SIR	38

3	SIR réursive	41
3.1	Introduction	41
3.2	Procédure d'estimation	42
3.2.1	L'expression analytique de l'estimateur $\tilde{\theta}_n$	42
3.2.2	Estimateur récursif de la direction de θ lorsque $H = 2$	43
3.3	Résultats asymptotiques	47
4	Quelques résultats de simulation	49
4.1	Etude de la qualité des estimateurs pour l'approche SIR réursive	51
4.1.1	Etude sur quelques échantillons	51
4.1.2	Etude sur $\mathcal{B} = 500$ réplifications d'échantillons	52
4.2	Comparaison des temps de calculs entre les approches SIR réursive et SIR classique (basées sur $H = 2$ tranches)	54
4.3	Perspective	55
	bibliographie	57

Chapitre 1

Modèles de Régression

Les modèles de régression fréquent sont avantageusement mis en évidence pour fin de modélisation de liens entre une variable à expliquer y et une variable explicative x . Ils sont appliqués à de nombreux domaines tels que l'économie, la biostatistique ou encore les sciences de l'environnement. Dans la littérature statistique, deux grandes classes de modèles de régression sont prévalente : **modèles paramétriques et les modèles non paramétriques**.

Dans ce chapitre, nous étudions ces types, en indiquant l'objectif, les avantages, les défauts et les méthodes d'estimation de chaque type, puis nous suggérons un autre type compromis entre eux qui nous permet de surmonter ses défauts.

1.1 Modèles paramétriques

Un modèle paramétrique est une classe particulière des modèles statistiques. Plus précisément, est une famille de distributions de probabilité qui dépend d'un nombre fini de paramètres. Il est de la forme :

$$y = r_{\theta}(x) + \varepsilon \tag{1.1}$$

où r_{θ} est la fonction de lien, entre la variable à expliquer $y \in \mathbb{R}$ et une variable explicative $x \in \mathbb{R}^p$, appartient à une famille de fonctions paramétrées par θ , vecteur de paramètres réels, et ε est un terme d'erreur aléatoire.

Dans un tel modèle, l'objectif est l'estimation du paramètre θ . Les techniques d'estimations paramétriques (maximum de vraisemblance, moindres carrés ordinaires,..)

sont efficaces quand la famille de r_θ est correctement spécifiée.

1.1.1 Modèle d'étude

Dans cette étude, nous nous focalisons sur le modèle linéaire multiple :

$$y = \theta_0 + \sum_{j=1}^p \theta_j x_j + \varepsilon \quad (1.2)$$

- y est la variable à expliquer (à valeur dans \mathbb{R}) ;
- x_1, \dots, x_p sont les variables explicatives (à valeurs dans \mathbb{R}) ;
- ε est le terme d'erreur aléatoire du modèle ;
- $\theta_0, \theta_1, \dots, \theta_p$ sont les paramètres à estimer.

Pour n observations, on peut écrire le modèle (1.2) sous la forme :

$$y_i = \theta_0 + \sum_{j=1}^p \theta_j x_{ij} + \varepsilon_i \quad \text{pour } i \in \{1, \dots, n\} \quad (1.3)$$

- ε_i est une variable aléatoire, non observée ;
- x_{ij} est observée et non aléatoire ;
- y_i est observée et aléatoire.

Pour pouvoir dire des choses pertinentes sur ce modèle, il faut néanmoins imposer des hypothèses les concernant. Voici celles que nous ferons dans un premier temps :

L'hypothèse (A1)

$$\mathbb{E}[\varepsilon_i] = 0, \forall i = 1, \dots, n,$$

ou de manière équivalente :

$$\mathbb{E}[y_i] = \theta_0 + \sum_{j=1}^p \theta_j x_{ij} \quad \forall i = 1, \dots, n,$$

L'hypothèse (A2)

$$\mathbb{V}[\varepsilon_i] = \sigma_\varepsilon^2 \quad \forall i = 1, \dots, n,$$

ou de manière équivalente :

$$\mathbb{V}[y_i] = \sigma_\varepsilon^2 \quad \forall i = 1, \dots, n,$$

L'hypothèse (A3)

$$\text{Cov}(\varepsilon_i, \varepsilon_{i'}) = 0 \quad \forall i \neq i',$$

ou de manière équivalente :

$$\text{Cov}(y_i, y_{i'}) = 0 \quad \forall i \neq i'.$$

Commentaires : les hypothèses (A1), (A2) et (A3) indiquent que les erreurs ε_i sont centrées, de même variance (homoscédasticité) et non corrélées.

Notons que le modèle (1.3) peut encore s'écrire de façon matricielle :

$$Y = X\theta + \varepsilon \tag{1.4}$$

où

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{1_1} & \dots & x_{1_p} \\ 1 & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & x_{n_1} & \dots & x_{n_p} \end{pmatrix}, \theta = \begin{pmatrix} \theta_0 \\ \vdots \\ \vdots \\ \theta_p \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

- Y désigne le vecteur à expliquer de taille n ;
- X la matrice explicative de taille $n \times (p + 1)$;
- ε le vecteur d'erreurs de taille n .

Donc on a

$$\begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{1_1} & \dots & x_{1_p} \\ 1 & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & x_{n_1} & \dots & x_{n_p} \end{pmatrix} \begin{pmatrix} \theta_0 \\ \vdots \\ \vdots \\ \theta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Les hypothèses peuvent alors s'écrire sous forme matricielle :

L'hypothèse (A1')

$$\mathbb{E}[\varepsilon] = 0_n,$$

ou de manière équivalente :

$$\mathbb{E}[Y] = X\theta \in \mathbb{R}^n$$

L'hypothèse ($\mathcal{A}2'$)

$$\mathbb{V}[\varepsilon] = \sigma_\varepsilon^2 I_n,$$

ou de manière équivalente :

$$\mathbb{V}[Y] = \sigma_\varepsilon^2 I_n.$$

En plus de ces hypothèses, on suppose que $n > (p + 1)$ et $rg(X) = p + 1$. On a donc plus d'observations que de variables et il n'existe pas de liaison linéaire entre les variables explicatives x_j c'est à dire pas de multicolinéarité.

Remarque

En plus de θ , le paramètre σ_ε^2 est aussi inconnu et doit être estimé.

Pour estimer les paramètres de ce modèle en utilisant :

- Méthode des moindres carrés ordinaires
- Méthode du maximum de vraisemblance

1.1.2 Méthode des moindres carrés ordinaires

Méthode des moindres carrés ordinaires est une méthode simple et très utilisée, indépendamment élaborée par **Legendre** [26] (approche algébrique) et **Gauss** [26] (approche probabiliste) au début du 19^e siècle.

Dans un tel modèle, on applique cette méthode pour trouver l'estimateur de θ que l'on note $\hat{\theta}$. En cherchant $\theta \in \mathbb{R}^{p+1}$ qui minimise la quantité :

$$S(\theta) = (Y - X\theta)^2$$

Le problème est :

$$\min_{\theta \in \mathbb{R}^{p+1}} S(\theta)$$

Avec

$$\begin{aligned} S(\theta) &= \sum_{i=1}^n [y_i - (\theta_0 + \sum_{j=1}^p \theta_j x_{ij})]^2 \\ &= (Y - X\theta)'(Y - X\theta) \\ &= Y'Y - 2Y'X\theta + \theta'X'X\theta \end{aligned}$$

Le minimum est atteint pour

$$\frac{\partial S(\theta)}{\partial \theta} = 0$$

On en déduit après quelques manipulations :

$$\hat{\theta} = (X'X)^{-1}X'Y, \quad (1.5)$$

Vocabulaire :

- $\hat{Y} = X\hat{\theta}$ est appelé **la valeur prédite**.
- $\hat{\varepsilon} = Y - \hat{Y}$ est appelé **le résidu**

Avant de poursuivre, notons que le calcul de l'estimateur des moindres carrés ordinaires $\hat{\theta}$ est purement déterministe : il ne fait en rien appel aux hypothèses $(\mathcal{A}1')$ et $(\mathcal{A}2')$ sur le modèle. Celles-ci vont en fait servir dans la suite à expliciter les propriétés statistiques de cet estimateur.

1.1.3 Quelques propriétés de l'estimateur $\hat{\theta}$

Proposition 1.1.3.1. *Sous les hypothèses $(\mathcal{A}1')$ et $(\mathcal{A}2')$, on a :*

1. $\mathbb{E}[\hat{\theta}] = \theta$
2. $\mathbb{V}(\hat{\theta}) = \sigma_{\varepsilon}^2(X'X)^{-1}$.

Preuve :

1). Nous simplifions l'expression de $\hat{\theta}$ pour faciliter le calcul de l'espérance.

$$\begin{aligned} \hat{\theta} &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'(X\theta + \varepsilon) \\ &= \theta + (X'X)^{-1}X'\varepsilon \end{aligned}$$

Prenons en espérance,

$$\begin{aligned} \mathbb{E}[\hat{\theta}] &= \mathbb{E}[\theta + (X'X)^{-1}X'\varepsilon] \\ &= \theta + \mathbb{E}[(X'X)^{-1}X'\varepsilon] \\ &= \theta + (X'X)^{-1}X'\mathbb{E}[\varepsilon] \\ &= \theta \end{aligned}$$

($\mathbb{E}[\varepsilon] = 0$ D'après l'hypothèse ($\mathcal{A}2'$)).

Commentaire : L'estimateur $\hat{\theta}$ est **sans biais**

2). Nous Calculons la variance de $\hat{\theta}$,

$$\begin{aligned}\mathbb{V}(\hat{\theta}) &= \mathbb{E}[(\hat{\theta} - \theta)(\hat{\theta} - \theta)'] \\ &= \mathbb{E}[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}] \\ &= (X'X)^{-1}X'\mathbb{E}[\varepsilon\varepsilon']X(X'X)^{-1}\end{aligned}$$

D'un autre coté, on a :

$$\mathbb{E}[\varepsilon\varepsilon'] = \begin{pmatrix} \mathbb{E}[\varepsilon_1^2] & \mathbb{E}[\varepsilon_1\varepsilon_2] & \dots & \mathbb{E}[\varepsilon_1\varepsilon_n] \\ & \mathbb{E}[\varepsilon_2^2] & & \\ & & \ddots & \\ & & & \mathbb{E}[\varepsilon_n^2] \end{pmatrix}$$

Puisque :

- $\mathbb{E}[\varepsilon_i^2] = \sigma_\varepsilon^2$,
- $\mathbb{E}[\varepsilon_i\varepsilon_j] = 0$,

(D'après les hypothèses (A2)et(A3)).

On trouve :

$$\mathbb{E}[\varepsilon\varepsilon'] = \sigma_\varepsilon^2 I_n,$$

et en déduire que :

$$\mathbb{V}(\hat{\theta}) = \sigma_\varepsilon^2 (X'X)^{-1},$$

Commentaire :

Il est aussi de **variance minimale** parmi tous les estimateurs linéaires sans biais par rapport à Y, d'après la propriété de Gauss-Markov.

Propriété de Gauss-Markov :

L'estimateur des MCO est optimal parmi tous les estimateurs linéaires sans biais (il est de variance minimale).

Remarque :

La méthode des moindres carrés ordinaires ne fournit pas un estimateur de σ_ε^2 . Néanmoins, on peut en donner un estimateur sans biais grâce aux résidus.

Estimateur sans biais de σ_ε^2

L'estimateur est défini par :

$$s^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-p-1} = \frac{\sum_{i=1}^n (\hat{\varepsilon}_i)^2}{n-p-1} = \frac{SCR}{n-p-1}$$

1.1.4 Méthode du maximum de vraisemblance

Additivement de la méthode des moindres carrés ordinaires, **Ronald Aylmer Fisher** [17] en 1922 a développé une méthode permettant aussi d'estimer les paramètres de notre modèle d'étude (1.2) : θ et σ_ε^2 . Il l'a nommé **la méthode du maximum de vraisemblance**.

Le principe de cette méthode est de construire une fonction appelée fonction de vraisemblance (construite à partir de la fonction de densité de y_i pour $i \in \{1, \dots, n\}$) et de maximiser son logarithme par rapport θ et σ_ε^2 .

Exemple : $y_i \sim \mathcal{N}(\theta_0 + \sum_{j=1}^p \theta_j x_{ij}, \sigma_\varepsilon^2) \quad \forall i = (1 \dots n)$.

Densité de probabilité de y_i

$$f_{\theta, \sigma_\varepsilon^2}(y_i) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp \left[-\frac{1}{2\sigma_\varepsilon^2} \left(y_i - \left(\theta_0 + \sum_{j=1}^p \theta_j x_{ij} \right) \right)^2 \right] \quad (1.6)$$

Fonction de vraisemblance

C'est la loi conjointe de (y_1, y_2, \dots, y_n) . Si les « y_i » sont indépendantes, cette fonction s'écrit :

$$\begin{aligned} L(\theta_0, \dots, \theta_p, \sigma_\varepsilon^2) &= f_{\theta, \sigma_\varepsilon^2}(y_1, \dots, y_n) \\ &= \prod_{i=1}^n f_{\theta, \sigma_\varepsilon^2}(y_i) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \right)^n \exp \left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^n \left(y_i - \theta_0 - \sum_{j=1}^p \theta_j x_{ij} \right)^2 \right] \\ &= \left(\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \right)^n \exp \left[-\frac{1}{2\sigma_\varepsilon^2} (Y - \theta X)'(Y - \theta X) \right] \end{aligned} \quad (1.7)$$

Fonction log-vraisemblance

La transformation logarithmique de la fonction de vraisemblance ci-dessus donné la fonction dite **log-vraisemblance** qui servira de base à l'estimation des paramètres $\{\theta, \sigma^2\}$. La fonction de log-vraisemblance s'écrit :

$$\ln L(\theta_0, \dots, \theta_p, \sigma_\varepsilon^2) = \ln \left[\left(\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \right)^n \right] + \ln \left(\exp \left[-\frac{1}{2\sigma_\varepsilon^2} (Y - \theta X)' (Y - \theta X) \right] \right) \quad (1.8)$$

En effet : $\ln(\mathbf{uv}) = \ln u + \ln v$; $\ln \exp(a) = a$; $\ln \mathbf{x}^2 = 2 \ln x$; $\mathbf{1}/\sqrt{2\pi\sigma_\varepsilon^2} = (2\pi\sigma_\varepsilon^2)^{-1/2}$,

$$\Rightarrow \ln \left[\left(\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \right)^n \right] = \ln \left[(2\pi\sigma_\varepsilon^2)^{-\frac{n}{2}} \right] = -\frac{n}{2} [\ln \pi + \ln(2\sigma_\varepsilon^2)] = -\frac{n}{2} \ln \pi - \frac{n}{2} \ln(2\sigma_\varepsilon^2),$$

$$\Rightarrow \ln \left(\exp \left[-\frac{1}{2\sigma_\varepsilon^2} (\mathbf{Y} - \theta \mathbf{X})' (\mathbf{Y} - \theta \mathbf{X}) \right] \right) = -\frac{1}{2\sigma_\varepsilon^2} (\mathbf{Y} - \theta \mathbf{X})' (\mathbf{Y} - \theta \mathbf{X}).$$

Avec : $(\mathbf{Y} - \theta \mathbf{X})' (\mathbf{Y} - \theta \mathbf{X}) = Y'Y - 2X'Y\theta + \theta'X'X\theta$.

Etant donné ces expressions, la fonction log-vraisemblance (1.8) se réécrit finalement :

$$\begin{aligned} \ln L(\theta_0, \dots, \theta_p, \sigma_\varepsilon^2) &= -\frac{n}{2} \ln \pi - \frac{n}{2} \ln(2\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2} (\mathbf{Y} - \theta \mathbf{X})' (\mathbf{Y} - \theta \mathbf{X}) \\ &= -\frac{n}{2} \ln \pi - \frac{n}{2} \ln(2\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2} (Y'Y - 2X'Y\theta + \theta'X'X\theta) \end{aligned} \quad (1.9)$$

Cette expressions est celle retenue pour l'estimation des paramètres.

L'estimation des paramètres θ et σ^2

Pour estimer les paramètres inconnus θ et σ^2 par la méthode du maximum de vraisemblance, la démarche va consister à maximiser la fonction log-vraisemblance (1.10), ce qui revient à annuler ses dérivées premières par rapport aux arguments θ et σ^2 comme suit :

$$\Rightarrow \frac{\partial \ln L}{\partial \theta} = -\frac{1}{2\sigma_\varepsilon^2} (-2X'Y + 2X'X\theta) = 0, \quad (1.10)$$

$$\Rightarrow \frac{\partial \ln L}{\partial \sigma_\varepsilon^2} = \frac{\partial}{\partial \sigma_\varepsilon^2} \left(-\frac{n}{2} \ln(2\sigma_\varepsilon^2) \right) + \frac{\partial}{\partial \sigma_\varepsilon^2} \left[-\frac{1}{2\sigma_\varepsilon^2} (Y'Y - 2X'Y\theta + \theta'X'X\theta) \right] = 0, \quad (1.11)$$

$$\Rightarrow \frac{\partial}{\partial \sigma_\varepsilon^2} \left(-\frac{n}{2} \ln(2\sigma_\varepsilon^2) \right) = -\frac{n}{2\sigma_\varepsilon^2}, \quad (1.12)$$

$$\begin{aligned}
\Rightarrow \frac{\partial}{\partial \sigma_\varepsilon^2} \left[-\frac{1}{2\sigma_\varepsilon^2} \left(Y'Y - 2X'Y\theta + \theta'X'X\theta \right) \right] &= -\frac{1}{2} * (-1) * (\sigma_\varepsilon^2)^{1-1} (Y - \theta X)'(Y - \theta X) \\
&= \frac{1}{2\sigma_\varepsilon^4} (Y - \theta X)'(Y - \theta X) \\
&= \frac{\varepsilon' \varepsilon}{2\sigma_\varepsilon^4}
\end{aligned} \tag{1.13}$$

D'après (1.12) et (1.13), on peut écrire l'expression (1.11) comme suit :

$$\frac{\partial \ln L}{\partial \sigma_\varepsilon^2} = -\frac{n}{2\sigma_\varepsilon^2} + \frac{\varepsilon' \varepsilon}{2\sigma_\varepsilon^4} = \frac{-n\sigma_\varepsilon^2 + \varepsilon' \varepsilon}{2\sigma_\varepsilon^4} = 0. \tag{1.14}$$

Nous pouvons déduire à partir des expressions (1.10) et (1.14), l'estimateur du maximum de vraisemblance des coefficients « $\tilde{\theta}$ » et celui de la variance de l'erreur « $\tilde{\sigma}_\varepsilon^2$ » :

$$\tilde{\theta} = (X'X)^{-1}X'Y$$

$$\tilde{\sigma}_\varepsilon^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n}$$

1.1.5 Quelques propriétés de $\tilde{\theta}$ et $\tilde{\sigma}_\varepsilon^2$

Pour $\tilde{\theta}$:

- l'estimateur de maximum de vraisemblance du paramètre θ et l'estimateur de MCO sont égaux : $\tilde{\theta} = \hat{\theta}$.

Alors

- $\mathbb{E}(\tilde{\theta}) = \theta \dots$ (sans biais),
- $\mathbb{V}(\tilde{\theta}) = \sigma_\varepsilon^2 (X'X)^{-1} \dots$ (il est de variance minimale).

Pour $\tilde{\sigma}_\varepsilon^2$:

- $\mathbb{E}(\tilde{\sigma}_\varepsilon^2) = \frac{1}{n} \mathbb{E}(\hat{\varepsilon}'\hat{\varepsilon}) = \frac{1}{n} (n - p - 1) \sigma_\varepsilon^2 = \sigma_\varepsilon^2 - \frac{(p + 1)}{n} \sigma_\varepsilon^2$

$$\Rightarrow \lim_{n \rightarrow \infty} \mathbb{E}(\tilde{\sigma}_\varepsilon^2) = \sigma_\varepsilon^2. \tag{1.15}$$

Alors $\tilde{\sigma}_\varepsilon^2$ est **asymptotiquement sans biais**

- $\mathbb{V}(\tilde{\sigma}_\varepsilon^2) = \frac{1}{n^2} \mathbb{V}(\hat{\varepsilon}' \hat{\varepsilon})$

Puisque $\mathbb{V}(\hat{\varepsilon}' \hat{\varepsilon}) > 0$

$$\Rightarrow \lim_{n \rightarrow \infty} \mathbb{V}(\tilde{\sigma}_\varepsilon^2) = \lim_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{V}(\hat{\varepsilon}' \hat{\varepsilon}) = 0$$

Alors « $\tilde{\sigma}_\varepsilon^2$ » est **converge**.

1.1.6 Avantages et défauts spécifiques

Ces modèles fournissent une interprétation claire de l'influence de la variable explicative sur la variable à expliquer.

Cependant, le choix d'un bon modèle paramétrique au vu des données peut s'avérer difficile et ainsi conduire dans le cas d'un mauvais choix à des conclusions erronées.

Afin de surmonter ce problème de sélection du bon modèle paramétrique, les modèles non paramétriques ont été proposés.

1.2 Modèles non paramétriques

Les modèles de régression non paramétrique apparaissent comme une alternative qui offre une flexibilité dans la modélisation du fait qu'aucune hypothèse paramétrique n'est imposée dans le modèle et que seules des hypothèses de régularité de la fonction de lien sont imposées

Supposons que le comportement d'une variable aléatoire y soit lié à celui d'une autre variable aléatoire x . Il est classique d'essayer tout d'abord d'exprimer y linéairement en fonction de x . C'est le problème bien connu de la régression linéaire. Il serait peut-être plus judicieux de tenter d'exprimer y à travers une fonction « r » non nécessairement linéaire de x , c'est-à-dire de trouver une expression de la forme :

$$y = r(x) + \varepsilon, \tag{1.16}$$

où le terme ε correspond à une erreur ou un bruit de modélisation aléatoire, nous rappelons aussi que la fonction r est inconnu. Dans ce qui suit, nous allons présenter quelques approches pour estimer cette fonction.

1.2.1 Lissage par noyau

Une classe de méthode de régression non-paramétrique consiste à effectuer au voisinage de chaque point une régression locale, celle-ci pouvant être linéaire (Nadaraya-Watson) ou polynomiale (polynômes locaux). La localisation est assurée par une fonction de poids K , appelée noyau, qui vérifie : $K \geq 0$, $\int K = 1$ et $K(-u) = K(u)$.

Estimateur de Nadaraya-Watson

Supposons que l'on dispose de n couples indépendants $(x_1, y_1), \dots, (x_n, y_n)$ et de même loi que (x, y) et nous présentons maintenant le premier et le célèbre estimateur à noyau de la fonction de régression introduit par Nadaraya (1964)[13] et Watson (1964)[13].

Définition 1.2.1.1. (*Estimateur de Nadaraya-Watson*) *L'estimateur de Nadaraya-Watson de la régression de y sur x est donné par :*

$$\hat{r}_n(x) = \begin{cases} \frac{\sum_{i=1}^n K\left(\frac{x-x_i}{h_n}\right) y_i}{\sum_{i=1}^n K\left(\frac{x-x_i}{h_n}\right)} & \text{si } \sum_{i=1}^n K\left(\frac{x-x_i}{h_n}\right) \neq 0 \\ \frac{1}{n} \sum_{i=1}^n y_i & \text{sinon} \end{cases} \quad (1.17)$$

h_n s'appelle **la fenêtre** désigne une suite de nombres réels strictement positifs vérifiant $h_n \rightarrow 0$ et $nh_n \rightarrow 0$ lorsque $n \rightarrow \infty$.

Polynômes locaux

L'estimation de la fonction de régression par la méthode des polynômes locaux est fondée sur une simple généralisation de l'estimateur précédent.

Définition 1.2.1.2. (*estimateur par polynômes locaux*) Soit $K : \mathbb{R} \rightarrow \mathbb{R}$ un noyau, $h_n > 0$ paramètre de lissage et $l \geq 0$ un entier.

Le vecteur

$$\hat{\theta}_n(x) = \arg \min_{\theta \in \mathbb{R}^{l+1}} \sum_{i=1}^n \left[Y_i - \theta' U \left(\frac{X_i - x}{h_n} \right) \right]^2 K \left(\frac{X_i - x}{h_n} \right) \quad (1.18)$$

où

$$U(u) = \left(1, u, \frac{u^2}{2}, \dots, \frac{u^l}{l!} \right)'$$

est appelé estimateur localement polynomial d'ordre l de

$$\theta(n) = (r(x), r^{(1)}(x)h_n, r^{(2)}(x)h_n^2, \dots, r^{(l)}(x)h_n^l)'$$

La statistique $\hat{r}_n(x) = U'(x)\hat{\theta}_n(x)$ est appelé estimateur localement polynomial d'ordre l de $r(x)$.

Posons à présent $B_{n,x} = \frac{1}{nh_n} \sum_{i=1}^n U \left(\frac{X_i - x}{h_n} \right) U' \left(\frac{X_i - x}{h_n} \right) K \left(\frac{X_i - x}{h_n} \right)$.

Si $B_{n,x}$ est définie positive alors l'estimateur localement polynomial $\hat{r}_n(x)$ de $r(x)$ est un estimateur linéaire :

$$\hat{r}_n(x) = \sum_{i=1}^n Y_i W_{n,i}(x) \quad (1.19)$$

où $W_{n,i}(x) = \frac{1}{nh_n} U'(0) B_{n,x}^{-1} U \left(\frac{X_i - x}{h_n} \right) K \left(\frac{X_i - x}{h_n} \right)$

Quelques exemples des noyaux

Les noyaux les plus couramment utilisés en pratique sont :

- Noyau uniforme(Rosenblatt)

$$K(x) = \frac{1}{2} \mathbf{1}_{[-1,1]}(x).$$

- Noyau triangulaire

$$K(x) = (1 - |x|) \mathbf{1}_{[-1,1]}(x).$$

- Noyau parabolique(d'epanechnikov)

$$K(x) = \frac{3}{4} (1 - x^2) \mathbf{1}_{[-1,1]}(x).$$

- Noyau gaussien

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \quad \forall x \in \mathbb{R}.$$

1.2.2 L'estimation par projection

Soit $(x_i, y_i) \forall i = 1, \dots, n$ un échantillon i.i.d et soit $(\Phi_j)_{j \geq 1}$ une base **orthonormée** de $\mathbb{L}^2[I]$ ($I \subseteq \mathbb{R}$: c'est l'espace des toutes fonctions de carré intégrable), on se donne $d \geq 1$ et posant

$$S_d = \langle \Phi_1, \Phi_2, \dots, \Phi_d \rangle$$

S_d : l'espace engendré par la base $\Phi_1, \Phi_2, \dots, \Phi_d$.

Supposons que $r \in \mathbb{L}^2[I]$ ($I \subseteq \mathbb{R}$), on note r_d la projection **orthogonal** de r sur S_d , donc on peut l'écrire :

$$r_d(x) = \sum_{j=1}^d \theta_j \Phi_j(x), \quad (1.20)$$

où

$$\begin{aligned} \theta_j &= \langle r(x), \Phi_j \rangle \\ &= \int_I r(x) \Phi_j dx. \end{aligned}$$

Alors

$$\hat{r}_d(x) = \sum_{j=1}^d \hat{\theta}_j \Phi_j(x). \quad (1.21)$$

L'estimateur de θ_j

Le meilleur estimateur de θ_j est donné par :

$$\hat{\theta}_j = \frac{1}{n} \sum_{i=1}^n y_i \Phi_j(x_i). \quad (1.22)$$

On réécrit (1.21) comme suit :

$$\begin{aligned}
 \hat{r}_d(x) &= \sum_{j=1}^d \hat{\theta}_j \Phi_j(x) \\
 &= \frac{1}{n} \sum_{j=1}^d \sum_{i=1}^n y_i \Phi_j(x_i) \Phi_j(x) \\
 &= \sum_{i=1}^n W_{n,i}(x_i) y_i
 \end{aligned} \tag{1.23}$$

Avec :

- $W_{n,i} = \frac{1}{n} \sum_{j=1}^d \Phi_j(x_i) \Phi_j(x),$
- $\sum_{i=1}^n W_{n,i}(x_i) = 1.$

Enfin, ces estimateurs par projection s'étendent facilement au cadre multidimensionnel, cependant comme pour la méthode de lissage par noyaux, la vitesse de convergence se dégrade très rapidement avec la dimension.

1.2.3 Avantages et défauts spécifiques

Avantages spécifiques : Ils caractérisent par une grande souplesse de modélisation. ce qui permet d'élargir le champ des situations modélisables.

Défauts spécifique

- Il faut estimer la fonction de lien 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.
- L'interprétation de la fonction de lien n'est pas toujours évidente.
- Les données vont être de plus en plus dispersées au fur et à mesure que la dimension des covariables x augmente. Ce problème est connu sous le terme « fléau (ou malédiction) de la dimension ».

Pour surmonter ces problèmes, des modèles de régression semi-paramétrique ont été développés.

1.3 Modèles semi-paramétriques

Les modèles semi-paramétriques de régression sont un compromis entre la régression paramétrique et l'approche entièrement non paramétrique, ont été évolués pour conjuguer les avantages des approches paramétriques (la capacité d'interprétation des modèles paramétriques) et non paramétriques (la souplesse des modèles non paramétriques). Nous considérons le modèle de régression semi-paramétrique suivant (introduit par Li (1991)[27]) :

$$y = r(\theta_1'x, \dots, \theta_K'x, \varepsilon), \quad (1.24)$$

La variable à expliquer y est reliée à la variable explicative p -dimensionnelle x par un vecteur explicatif de dimension réduite : $(x\theta_1, \dots, x\theta_K)$, $K < p$. Aucune hypothèse n'est faite ni sur le paramètre fonctionnel r , ni sur la distribution de l'erreur ε .

Nous nous intéressons à la partie paramétrique du modèle, c'est-à-dire que l'on désire connaître l'espace engendré par les K vecteurs de paramètres inconnus $\theta_1, \dots, \theta_K$. Cet espace est appelé l'espace EDR (comme « Effective Dimension Reduction »). Remarquons que les vecteurs $\{(\theta_k), k = 1, \dots, K\}$ ne sont pas identifiables individuellement (car le paramètre fonctionnel r est inconnu), ainsi seul l'espace EDR est globalement identifiable.

On appelle direction EDR toute combinaison linéaire des (θ_k) appartenant à l'espace EDR. Duan et Li (1991)[27] ont présenté une méthode permettant d'estimer une base de cet espace EDR sans avoir à estimer le paramètre fonctionnel r ni à spécifier la loi de l'erreur ε . Ils ont appelé cette méthode "Sliced Inverse Regression" (SIR), que l'on peut traduire en français par "régression inverse par tranches". Une fois qu'on a réduit la dimension de p à K en estimant cet espace EDR, il ne reste qu'à utiliser une méthode non-paramétrique (Ex : **Nadaraya-Watson**) afin d'estimer le paramètre fonctionnel r .

Dans ce qui suit, nous donnons une description détaillée des modèles semi-paramétriques de référence (cas univarié et cas multivarié). Nous énonçons ensuite une condition fondamentale sur la distribution de la variable explicative, condition nécessaire dans la théorie de SIR.

1.3.1 Modèles semi-paramétriques de référence

Cas où la variable y est unidimensionnelle :

Le modèle de référence sur lequel on travaille est le modèle de régression semi-paramétrique suivant :

$$y = r(\theta'_1 x, \dots, \theta'_K x, \varepsilon), \quad (1.25)$$

où

- x est une variable aléatoire explicative de dimension p ayant une espérance $\mathbb{E}[x] = \mu$, une matrice de covariance $\mathbb{V}(x) = \Sigma$ définie positive, et vérifiant une condition fondamentale précisée ci-après ;
- $\theta_1, \dots, \theta_K$ sont des vecteurs de paramètres réels de dimension p , inconnus et linéairement indépendants ;
- ε est un terme d'erreur aléatoire indépendant de x , aucune hypothèse n'est faite sur la distribution de ε ,
- r est le paramètre fonctionnel à valeur dans \mathbb{R} , inconnu et arbitraire.

Cas où la variable y est multidimensionnelle :

Une version multivarié du modèle semi-paramétrique de référence (1.25) peut parfois être écrit de la manière spéciale suivante :

$$\begin{aligned} y^1 &= r_1(\theta'_1 x, \dots, \theta'_K x, \varepsilon^1) \\ &\vdots \\ y^q &= r_q(\theta'_1 x, \dots, \theta'_K x, \varepsilon^q) \end{aligned} \quad (1.26)$$

où

- x est une variable aléatoire explicative de dimension p ayant une espérance $\mathbb{E}[x] = \mu$, une matrice de covariance $\mathbb{V}(x) = \Sigma$ définie positive, et vérifiant une condition fondamentale précisée ci-après ;
- $\theta_1, \dots, \theta_K$ sont des vecteurs paramètres inconnus de dimension p , linéairement indépendants ;
- les termes ε^j ($\forall j = 1, \dots, q$) sont des erreurs aléatoires indépendantes de x ;
- r_j , $j = 1, \dots, q$ sont des paramètres fonctionnels à valeur dans \mathbb{R} , inconnus et arbitraires.

1.3.2 Condition fondamentale

Soit $x = (x_1, \dots, x_p)'$, on dit que x est vérifie **la condition fondamentale** (où la condition de linéarité (LC)) :

- Si x suit une distribution de probabilité non dégénérée telle que, $\forall b \in \mathbb{R}^p$, l'espérance conditionnelle $\mathbb{E}[b'x | \theta'_1 x, \dots, \theta'_K x]$ est linéaire en $\theta'_1 x, \dots, \theta'_K x$, c'est-à-dire qu'il existe des constantes c_0, \dots, c_K telles que :

$$\mathbb{E}[b'x | \theta'_1 x, \dots, \theta'_K x] = c_0 + c_1 \theta'_1 x + \dots + c_K \theta'_K x, \quad (1.27)$$

- En particulier, la condition est vérifiée lorsque la loi de x est **elliptique**. C'est-à-dire que la densité f_x de x est de la forme :

$$f_x(t|m, \Phi) = \frac{1}{\sqrt{\det \Phi}} h((t - m)' \Phi^{-1} (t - m)). \quad (1.28)$$

Chapitre 2

SIR classique

La régression inverse en tranches (Sliced Inverse Regression (SIR)) est une méthode de régression semi-paramétrique reposant sur un argument géométrique. Elle a été introduite par Duan et Li (1991)[27] permettant d'estimer la partie paramétrique du modèle semi-paramétrique « θ » sans avoir à estimer la partie fonctionnelle « r » ni à spécifier la loi de l'erreur ε . Au contraire des autres méthodes de régression semi-paramétrique, elle ne requiert que des temps de calculs informatiques très courts, de plus c'est un outil très utile pour la réduction de dimension dans le domaine des statistiques multivariés. Dans ce chapitre, nous présentons tout d'abord cette méthode SIR dans le cas univarié, ensuite nous mentionnons quelques méthodes dans le cas multivarié. Nous terminons ce chapitre en donnant certaines ses applications .

2.1 SIR univariée

Avant de présenter plus en détail la méthode SIR, de brèves réflexions d'une part sur le nom de la méthode, d'autre part sur le principe d'estimation qu'elle met en œuvre nous paraissent nécessaire. Le nom de la méthode, SIR, fait référence à deux aspects : (1) une propriété géométrique essentielle de la courbe de régression inverse faisant intervenir divers moments du couple (y, x) ; (2) 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.

2.1.1 Philosophie et théorie de la méthode SIR

L'idée la plus importante de ces méthodes est de considérer la régression « inverse » (c'est à dire d'inverser les rôles de y et de x) et de trouver une connexion entre l'espace EDR et les « courbes » des deux premiers moments inverses $\mathbb{E}[x|y]$ et $\mathbb{V}(x|y)$. Ainsi, l'approche SIR repose exclusivement sur une propriété de la courbe de régression inverse $\mathbb{E}[x|y]$. L'utilisation de cette courbe représente un certain nombre d'avantages comme la simplicité de l'idée sous-jacente ainsi que celle de l'implémentation informatique. En effet, d'un point de vue purement calculatoire, **le fléau de la dimension** disparaît vu que p régressions unidimensionnelles sont maintenant nécessaires à l'estimation de $\mathbb{E}[x|y]$ au lieu d'une unique régression p -dimensionnelle pour estimer $\mathbb{E}[y|x]$.

Nous étudions maintenant la propriété des courbes de régression inverse dans le cadre du modèle de référence (1.25) et sous la condition fondamentale, propriété permettant de caractériser l'espace EDR E . Nous introduirons ensuite la procédure d'estimation de ce espace.

Les courbes de régression inverse

Considérons tout d'abord une transformation monotone T de y et x . Nous appelons :

- courbe de régression inverse la fonction de \mathbb{R} dans \mathbb{R}^p définie par :

$$y \longrightarrow \mathbb{E}[x|T(y)].$$

- courbe de régression inverse centrée la fonction de \mathbb{R} dans \mathbb{R}^p définie par :

$$y \longrightarrow \mathbb{E}[x|T(y)] - \mathbb{E}[x].$$

Dans le cadre du modèle (1.25) et sous la condition fondamentale (1.27), Li (1991)[27] a montré la propriété géométrique suivante : la courbe de régression inverse centrée $\mathbb{E}(x|T(y)) - \mathbb{E}(x)$ lorsque y varie, est contenue dans le sous-espace linéaire de \mathbb{R}^p engendré par les vecteurs $\Sigma\theta_1, \dots, \Sigma\theta_K$. Une conséquence directe est que la matrice de covariance $M_T = \mathbb{V}(\mathbb{E}(x|T(y)))$ est dégénérée dans toute direction Σ -orthogonale aux $\theta_k, k = 1, \dots, K$. Il s'en suit que les vecteurs propres associés aux K valeurs propres non nulles de la matrice $\Sigma^{-1}M_T$ sont des directions EDR, c'est-à-dire appartiennent à l'espace EDR.

Remarque. Il sera pratique dans la description de l'approche SIR de considérer la version standardisée z de x :

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

On peut alors réécrire le modèle (1.25) sous la forme :

$$y = f(\eta'_1 z, \dots, \eta'_K z, \varepsilon), \quad (2.1)$$

où $\eta_k = \Sigma^{1/2}\theta_k$, $k = 1, \dots, K$, sont des directions EDR standardisées. L'espace engendré par ces directions est alors appelé espace EDR standardisé et noté E^s .

De manière analogue à ce qui précède, pour ce modèle et sous la condition fondamentale (1.27), la courbe de régression inverse standardisée $\mathbb{E}[z|T(y)]$ lorsque y varie appartient à l'espace EDR standardisé E^s . La matrice de covariance de la courbe de régression inverse standardisée $M_T^s = \mathbb{V}(\mathbb{E}(z|T(y)))$ est dégénérée dans toute direction orthogonale aux η_k . Ainsi Les vecteurs propres associés aux K valeurs propres non nulles de M_T^s , dénotés v_1, \dots, v_K , sont des directions EDR standardisées.

L'espace EDR peut être déduit de cet espace EDR standardisé : il est engendré par les vecteurs $\Sigma^{-1/2}v_k$, $k = 1, \dots, K$.

Conséquence : L'espace EDR E (ou E^s) est indépendant de la fonction monotone T , seule la base de cet espace est modifiée. Or, choisir pour T une discrétisation de y conduit à une écriture de M_T (ou M_T^s) particulièrement simple. Ceci est la base de la méthode d'estimation SIR.

Algorithme général des méthodes SIR

Considérons un échantillon $\{(x_i, y_i), i = 1, \dots, n\}$ de variables aléatoires i.i.d issues du modèle (1.25), Nous voulons estimer une base de l'espace EDR E . Pour ce faire, nous devons à estimer la matrice M_T (ou M_T^s) mais d'abord, il est nécessaire de choisir une transformation T pour laquelle cette matrice sera facilement estimable

Le choix particulier de T : Li (1991)[27] a proposé de prendre pour la transformation T une discrétisation de y en H tranches distinctes :

$$s_1, \dots, s_h, \dots, s_H.$$

On appelle cette transformation "le découpage en tranches".

Dans ce cadre, les espérances conditionnelles $\mathbb{E}[z|T(y)]$ s'écrivent sous la forme :

$$m_h^s = \mathbb{E}[z|y \in s_h], \quad h = 1, \dots, H.$$

En notant $p_h = Pr(y \in s_h)$, la matrice M^s devient

$$M_T^s = \sum_{h=1}^H p_h m_h^s m_h^{s'}.$$

L'algorithme est représenté par les étapes suivants :

✓ **Étape 1** : Standardisation des x_i

$$z_i = \hat{\Sigma}_n^{-1/2}(x_i - \bar{x}_n), \quad i = 1, \dots, n,$$

où

- \bar{x}_n est la moyenne empirique de l'échantillon des x_i : $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$;
- $\hat{\Sigma}_n$ est la matrice de covariance empirique : $\hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)(x_i - \bar{x}_n)'$.

✓ **Étape 2** : Découpage du support des y_i en H tranches s_1, \dots, s_H ,

✓ **Étape 3** : Pour chacune des tranches :

- Calcul de la proportion empirique des y_i appartenant à la tranche :

$$p_{h,n} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[y_i \in s_h]}, \quad h = 1, \dots, H.$$

- Calcul de la moyenne empirique des z_i :

$$m_{h,n}^s = \frac{1}{np_{h,n}} \sum_{i=1}^n z_i \mathbb{1}_{[y_i \in s_h]}, \quad h = 1, \dots, H.$$

✓ **Étape 4** : Construction d'une estimation de la matrice de covariance des moyennes M^s :

$$\hat{M}_T^s = \sum_{h=1}^H p_{h,n} m_{h,n}^s m_{h,n}^{s'}.$$

✓ **Étape 5** : Décomposition spectrale de \hat{M}_T^s .

Notons alors, \hat{v}_k^s pour $k = 1, \dots, K$, les K vecteurs propres associés aux K plus grandes valeurs propres de \hat{M}_T^s

✓ **Étape 6** : Construction des K directions EDR estimées :

$$\hat{v}_k = \hat{\Sigma}^{-1/2} \hat{v}_k^s, k = 1, \dots, K.$$

Choix du nombre H de tranches.

Pour ce qui concerne le choix du nombre de tranches, les deux cas extrêmes qui consisteraient à prendre $H = 1$ ou $H = n$ (c'est-à-dire une seule tranche ou une seule observation par tranche) n'ont pas d'intérêt. Dans les deux cas, aucune information supplémentaire sur la réponse y n'est apportée aux variables explicatives soit parce qu'on ne fait aucune moyenne ($H = 1$) soit parce qu'il n'y a qu'une moyenne ($H = n$). Entre ces 2 cas extrêmes, il reste une grande liberté de choix pour H . Il convient cependant de noter qu'il est indispensable de prendre un nombre des tranches supérieur à K , afin de retrouver toute la direction EDR, et aussi pour ne pas faire une réduction de dimension artificielle.

D'un point de vue théorique, les méthodes SIR ne sont pas sensibles au choix du tranchage, il suffit de prendre un H permettant d'avoir suffisamment d'individus par tranche (par exemple : 10 tranches pour 100 individus). Cependant d'un point de vue pratique, l'estimation des directions EDR peut être sensible au nombre et à la position des tranches choisies lorsque la taille de l'échantillon est relativement petite (inférieure à 50 individus).

Choix des H tranches

Après avoir discuté le problème du choix du nombre H de tranches, on désire maintenant de construire ces H tranches. En pratique il ya naturellement deux possibilités : fixer la largeur des tranche ou fixer le nombre d'observations par tranche.

Il apparaît préférable de prendre des tranches d'effectifs proches, ce qui permet de s'assurer d'avoir des tranches avec suffisamment d'individus. Une technique pour construire un tel tranchage à partir d'un échantillon est la suivante. Notons \hat{r}_j ($j = 1, \dots, H - 1$) le (j/H) ième quantile empirique des y_i .

Les tranches sont alors définies par :

$$s_1 = \{y_i | y_i \leq \hat{r}_1\}, \dots, s_h = \{y_i | \hat{r}_{h-1} < y_i \leq \hat{r}_h\}, \dots, s_H = \{y_i | y_i \geq \hat{r}_{H-1}\}$$

Discussion sur le paramètre inconnu K

Dans ce qui précède, on a considéré que la dimension K de réduction du modèle était connue, malheureusement ce n'est pas le cas dans les applications réelles. Dans cette situation, Li (1991)[27] a proposé un test d'hypothèse sur la nullité des dernières

$(p - K)$ valeurs propres de la matrice $\Sigma^{-1}M_T$. Il était intéressé au cas particulier où la distribution de x est normale. Cette approche a été étendue pour les distributions elliptiques par Schott (1994)[34] et Barríos et Vallila (2007)[3]. Néanmoins, l'usage de ces tests souvent pose des problèmes, notamment en ce qui concerne le niveau global de ceux-ci. En d'autres termes, l'identification de la vraie dimension K n'est plus évidente lorsque la taille de l'échantillon est faible. Une approche différente repose sur la qualité de l'estimation des espaces de réduction effective de dimension. À cet égard, Ferré (1998)[15] introduit un critère conduit à un choix de la dimension K à partir de la quantité :

$$R(k) = \frac{1}{k} \mathbb{E}[tr(\hat{\pi}_k \pi_k)], \quad (2.2)$$

où π_k (resp. $\hat{\pi}_k$) représente le projecteur propre sur le sous-espace engendré par les k vecteurs propres associés aux k plus grandes valeurs propres de $\Sigma^{-1}M_T$ (resp. $\hat{\Sigma}^{-1}\hat{M}_T$). Cette fonction permet de mesurer les écarts respectifs entre les sous-espaces de l'espace EDR et leurs estimateurs. Mais, pour des raisons pratiques, on ne peut pas utiliser le critère directement sous cette forme, puisque π_k est à priori inconnu. On a alors recours à un développement limité de l'expression $\mathbb{E}[tr(\hat{\pi}_k \pi_k)]$. Donc pour x suit une loi elliptique, on a :

$$R(k) = 1 - \frac{1}{Nk} \left\{ \sum_{i=1}^k \sum_{j=k+1}^K \frac{\lambda_i + \lambda_j + (\kappa - 1)\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2} - (p - K) \sum_{i=1}^k \frac{1}{\lambda_i} \right\} + \mathcal{O}_p(n^{-3/2}), \quad (2.3)$$

où λ_i représente la valeur propre non nulle de la matrice $\Sigma^{-1}M_T$, pour $i = 1, \dots, K$ et κ est le paramètre de kurtosis de la loi de x .

En outre, l'estimateur asymptotiquement sans biais $\hat{R}(k)$ de $R(k)$ défini sous la forme :

$$\hat{R}(k) = 1 - \frac{1}{Nk} \left\{ \sum_{i=1}^k \sum_{j=k+1}^K \frac{\hat{\lambda}_i + \hat{\lambda}_j + (\hat{\kappa} - 1)\hat{\lambda}_i \hat{\lambda}_j}{(\hat{\lambda}_i - \hat{\lambda}_j)^2} - (p - K) \sum_{i=1}^k \frac{1}{\hat{\lambda}_i} \right\}, \quad (2.4)$$

où $\hat{\lambda}_i$ représente la valeur propre non nulle de la matrice $\hat{\Sigma}^{-1}\hat{M}_T$, pour $i = 1, \dots, K$. Clairement, pour $k = 1, \dots, p$, $\hat{R}(k)$ prend ses valeurs dans $[0, 1]$, $\hat{R}(p)$ étant toujours égal à 1. Alors, on retient la dimension qui correspond à un "maximum local" (le plus proche de 1) pour $\hat{R}(k)$, $k = 1, \dots, p-1$. Cette approche est plus efficace que les tests d'hypothèses, notamment pour des tailles d'échantillons faibles.



Remarque. Cette approche SIR (SIR_I) souffre cependant d'un problème pathologique connu : elle est en effet « aveugle » dans le cas où le modèle présente une « dépendance symétrique ». Dans ce cas, la méthode SIR peut ne pas trouver toutes les directions EDR. Pour résoudre ce problème pathologique et pour retrouver toutes les directions EDR, une extension naturelle est de considérer des moments conditionnels de x sachant y d'ordre supérieur. Dans ce but, la méthode SIR_{II} utilise des propriétés de la « courbe » de variance conditionnelle $\mathbb{V}(X|Y)$. Afin d'utiliser les informations provenant des deux premiers moments conditionnels, Li (1991)[27] propose de combiner les deux approches $SIR-I$ et SIR_{II} au travers la méthode SIR_α . En outre, Cook et Weisberg (1991)[10] proposent la méthode $SAVE$ (Sliced Average Variance Estimate) qui repose sur l'utilisation de moment conditionnel d'ordre 2.

2.1.2 Résultats asymptotiques de SIR

Nous donnons ci-dessous les différents résultats asymptotiques obtenus pour la méthode SIR classique :

Convergence en probabilité

Dans le cadre du modèle de référence, Li (1991)[27] a donné le théorème suivant :

Théorème 2.1.2.1. *Sous l'hypothèse(1.27) en supposant que les données sont échantillonnées de manière indépendante à partir du modèle (1.25), on a : Les directions EDR estimées $\hat{v}_1, \dots, \hat{v}_K$ convergent à la vitesse $\mathcal{O}_p(n^{-1/2})$ vers des vraies directions EDR*

Normalité asymptotique

Avant de donner les lois asymptotiques des éléments propres caractérisant l'espace EDR estimé, Saracco (1997)[32] a posé quelques hypothèses :

- ($\mathcal{H}1$) Les observations (x_i, y_i) , $i = 1, \dots, n$, sont échantillonnées de manière indépendante à partir du modèle (1.25)
- ($\mathcal{H}2$) Le support de y est divisé en H tranches fixés $s_1, \dots, s_h, \dots, s_H$ telles que : $p_h > 0$, ($= Pr(y \in s_h)$).

- ($\mathcal{H}3$) La matrice de covariance Σ est définie positive.
- ($\mathcal{H}4$) La dimension K est connu et les K plus grandes valeurs propres de $\Sigma^{-1}M$, sont non nulles et satisfaire : $\lambda_1 > \dots > \lambda_K > 0$.

En notant $\omega = \{\lambda_1, \dots, \lambda_K\}$, la matrice $P = \sum_{\lambda_k \in \omega} P_{\lambda_k}$ (où $P_{\lambda_k} = v_k v_k' \Sigma$) est le projecteur propre Σ -orthogonal sur E .

On définit $\hat{\omega} = \{\hat{\lambda}_1, \dots, \hat{\lambda}_K\}$ l'ensemble des K plus grandes valeurs propres de $\hat{\Sigma}^{-1}\hat{M}_T$, et on suppose que $\hat{\lambda}_1 > \dots > \hat{\lambda}_K > \hat{\lambda}_{K+1}$. Le projecteur propre sur l'espace EDR estimé \hat{E} est $\hat{P} = \sum_{\hat{\lambda}_k \in \hat{\omega}} \hat{P}_{\hat{\lambda}_k}$, où $\hat{P}_{\hat{\lambda}_k} = \hat{v}_k \hat{v}_k' \hat{\Sigma}$.

Saracco (1997)[32] a donné deux théorèmes de normalités asymptotiques, Le théorème 1 donne la distribution asymptotique de $\Sigma^{-1}M_T$. À partir de cette distribution limite, les distributions asymptotiques des éléments propres décrivant l'espace EDR estimé sont dérivées dans le théorème 2. Ces éléments propres sont le projecteur propre sur l'espace EDR estimé, les directions EDR estimées et leurs valeurs propres correspondantes.

Théorème 1 : Sous les hypothèses ($\mathcal{H}1$), ($\mathcal{H}2$) et ($\mathcal{H}3$), on a :

$$\sqrt{n} \text{vec}(\hat{\Sigma}^{-1}\hat{M}_T - \Sigma^{-1}M_T) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathcal{C}).$$

Théorème 2 : Sous les hypothèses ($\mathcal{H}1$), ($\mathcal{H}2$), ($\mathcal{H}3$) et ($\mathcal{H}4$), on a :

$$(i) \quad \sqrt{n} \text{vec}(\hat{P} - P) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathcal{C}_P).$$

$$(ii) \quad \sqrt{n} (\hat{v}_k - v_k) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathcal{C}_N).$$

$$(iii) \quad \sqrt{n} (\hat{\lambda}_k - \lambda_k) \xrightarrow{\mathcal{L}} \mathcal{N}(0, [v_k' \otimes v_k' \Sigma] \mathcal{C} [v_k \otimes \Sigma v_k]).$$

-Les expressions des matrices de variance asymptotique indiquées dans les théorèmes précédents ($\mathcal{C}, \mathcal{C}_P, \mathcal{C}_N$) sont données par Saracco (voir [[32]]).

2.2 Méthodes SIR multivariées

Dans le cadre du modèle (1.26) et parmi les différentes œuvres présentées dans ce type, nous nous intéressons aux méthodes proposés par Li, Aragon, Shedden et

Thomas-Agnan (2003)[28]. Ces diverses méthodes basées sur des philosophies légèrement différentes. Dans cette section, nous présentons succinctement les aspects théoriques sous-jacents aux méthodes.

i) Méthode "Directe" : "Complete Slicing" et "Marginal Slicing".

Dans la méthode **Complete Slicing**, la procédure SIR est directement appliquée sur y ($y \in \mathbb{R}^q$). Pour établir des tranches de tailles presque égales, l'approche récursive suivante est utilisée. La première composante de y est découpée en tranches. Puis, chaque tranche est séparément découpée en tranches encore selon la prochaine composante de y , et ainsi de suite. Le problème majeur qui apparaît ici est que le nombre de tranches augmente très rapidement avec la dimension de y . Par conséquent, une façon de contourner ce problème est de limiter le tranchage aux composants qui se sont révélés importantes dans certains sens. Ce thème-là correspond à la méthode **Marginal Slicing** qui consiste à appliquer SIR sur une transformation de y . Le choix de cette transformation est très vaste et dépend en fait du contexte d'intérêt. Par exemple, il peut être la moyenne ou la médiane des y^j . Il peut également s'agir les premières composantes de l'analyse en composantes principales des y^j .

ii) Méthodes "agrégante", "Pooled Marginal Slicing".

L'idée principale de ces méthodes est de considérer les q méthodes SIR univariées de chaque composante y^j de y sur z (basé sur un tranchage T_j) et de combiner les matrices M^s correspondantes (notées M_j^s) dans la mise en commun suivant :

$$M^P = \sum_{j=1}^q w_j M_j^s$$

où les w_j sont une suite de poids positifs tels que $\sum_{j=1}^q w_j = 1$ et $M_j^s = \mathbb{V}(\mathbb{E}[z|T_j(y)])$.

Ainsi, les K vecteurs propres associés aux K plus grandes valeurs propres (non nulles) de M^P sont des directions EDR standardisées.

Pour la version empirique du Pooled Marginal Slicing, chaque matrice M_j^s est esti-

mée par \hat{M}_j^s , alors l'estimateur de M^P est :

$$\hat{M}^P = \sum_{j=1}^q w_j \hat{M}_j^s$$

- Les directions EDR standardisées estimées sont alors les K vecteurs propres associés aux K plus grandes valeurs propres de \hat{M}^P .

iii) Méthodes de l'"Alternating SIR" : directions EDR et variables les plus prévisibles..

De manière analogue aux travaux de de Hotteling (1935, 1936)[22], Li, Aragon, Shedden et Thomas-Agnan (2003)[28] proposent une théorie concernant les variables les plus prévisibles (Most Predictable) mais contrairement à Hotteling. Leur théorie utilise des prédicteurs non linéaires, et se ramène à une décomposition aux valeurs propres du même type que celle de SIR, il suffit en effet de simplement intervertir les rôles de x et de y .

La méthode **Alternating SIR** repose sur l'observation de cette dualité entre les directions EDR (combinaisons linéaires de x) et les directions des variables les plus prévisibles (combinaison linéaires de y). Au niveau de la mise en œuvre pratique, l'utilisation de cette dualité revient à utiliser un algorithme qui à chaque itération calcule les directions EDR et les directions les plus prévisibles. On arrête les itérations lorsque la ou les directions retenues se stabilisent. La convergence de l'algorithme en un nombre fini d'étapes a été montré par Li, Aragon, Shedden et Thomas-Agnan (2003)[28].

2.3 Mise en œuvre de la méthode SIR

Il convient de noter que lorsque la dimension p est grande, la matrice de variance-covariable $\hat{\Sigma}$ des variables 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 Σ). Dans ce cadre, Bernard-Michel et al (2009)[5] proposent une méthode de régularisation de SIR en introduisant un a priori Gaussien (GRSIR). Cette régularisation a pour effet l'amélioration du conditionnement de la matrice de variance-covariance des variables explicatives. La méthode GRSIR est utilisée pour estimer

les propriétés physiques du sol martien à partir d'images hyperspectrales. D'autres applications de la méthode SIR dans le cadre de problématiques réelles sont disponibles dans la littérature. Nous pouvons citer par exemple une méthode d'estimation semi-paramétrique de courbes de référence pour des propriétés biophysiques de la peau, (voir [[18]]). De plus, on peut mettre en œuvre la méthode SIR dans le domaine thermodynamique par exemple, la modélisation des erreurs thermiques des machines-outils CNC (voir [[37]]).

Chapitre 3

SIR réursive

3.1 Introduction

L'intérêt majeur des approches réursives est qu'elles permettent une mise à jour rapide des estimateurs lorsque les données sont observées de manière séquentielle sans être obligé de stocker en mémoire toutes les observations passées. Dans ce chapitre, nous nous intéressons au modèle de régression semi-paramétrique suivant : lorsque $x \in \mathbb{R}^p$ et $y \in \mathbb{R}$,

$$y = r(\theta'x) + \epsilon \tag{3.1}$$

où l'erreur ϵ est indépendante de x et de loi inconnue et arbitraire, la fonction de lien r est inconnue et arbitraire et le paramètre θ est un vecteur de \mathbb{R}^p , n'est pas totalement identifiable, seule la direction de θ l'est. On parle de direction EDR (comme "effective dimension reduction").

Notre objectif est d'étudier le problème d'estimation du paramètre θ de ce modèle avec des méthodes réursives. À cet égard, nous proposons une approche réursive de la méthode SIR dont l'objectif majeur est l'estimation de la direction du vecteur de paramètre θ du modèle (3.1) sans avoir à estimer la partie fonctionnelle « r » ni à spécifier la loi de l'erreur ϵ . Nous introduisons en ce qui suit, la procédure d'estimation de direction θ par cette approche. Nous donnons ensuite quelques propriétés asymptotiques de l'estimateur proposé.

3.2 Procédure d'estimation

Supposons un échantillon d'observations $(x_i, y_i), i = 1, \dots, N$ des variables aléatoires (x, y) indépendantes et identiquement distribuées issues du modèle (3.1). Nous avons vu précédemment que la matrice $M_T = \mathbb{V}(\mathbb{E}[x|T(y)])$ est dégénérée dans n'importe quelle direction Σ -orthogonale à θ . Ainsi, le vecteur propre $\tilde{\theta}$ associé à la valeur propre non nulle de $\Sigma^{-1}M_T$ (colinéaire à θ) est une direction EDR. On désire a construire un estimateur récursive de la direction de θ . Nous donnons dans un premier temps, l'expression analytique de l'estimateur (non récursif) $\tilde{\theta}_n$ de $\tilde{\theta}$. Puis, nous recherchons une forme récursive pour cet estimateur.

3.2.1 L'expression analytique de l'estimateur $\tilde{\theta}_n$

Nous nous focaliserons au cas particulier où l'on ne considère que deux tranches ($H=2$) notées s_1 et s_2 . La raison essentielle de ce choix est que nous pouvons obtenir facilement une expression analytique de l'estimateur de la direction de θ . Dans ce cas, la Matrice M_T s'écrit sous la forme :

$$M_T = p_1 z_1 z_1' + p_2 z_2 z_2', \quad (3.2)$$

avec $z_h = m_h - \mu$ (pour $h = 1, 2$) où m_h, p_h et μ sont définis au chapitre précédent. On peut alors montrer que la valeur propre non nulle λ de $\Sigma^{-1}M_T$ et le vecteur propre $\tilde{\theta}$ correspondant (colinéaire à θ) s'écrivent :

$$\lambda = \frac{p_1}{p_2} z_1' \Sigma^{-1} z_1 \quad \text{et} \quad \tilde{\theta} = \Sigma^{-1}(z_1 - z_2). \quad (3.3)$$

Preuve. Lorsque $H=2$, Nous avons :

$$p_1 z_1 + p_2 z_2 = p_1 m_1 + p_2 m_2 - (p_1 + p_2)\mu = p_1 m_1 + p_2 m_2 - \mu = \mu - \mu = 0. \quad (3.4)$$

Vérifions que $\tilde{\theta} = \Sigma^{-1}(z_1 - z_2)$ est bien vecteur propre de $\Sigma^{-1}M_T$ et déterminons la valeur propre associée. En utilisant (3.2), on a

$$\begin{aligned} \Sigma^{-1}M_T\tilde{\theta} &= \Sigma^{-1}M_T\Sigma^{-1}(z_1 - z_2) \\ &= q_1 p_1 \Sigma^{-1} z_1 - q_{12} p_1 \Sigma^{-1} z_1 + q_{12} p_2 \Sigma^{-1} z_2 - q_2 p_2 \Sigma^{-1} z_2, \end{aligned}$$

où $q_1 = z_1' \Sigma^{-1} z_1$, $q_{12} = z_1' \Sigma^{-1} z_2$ et $q_2 = z_2' \Sigma^{-1} z_2$.

De plus, en utilisant (3.4), nous avons : $p_1 q_1 = -p_2 q_{12}$ et $p_2 q_2 = -p_1 q_{12}$. Nous en déduisons que :

$$\begin{aligned} \Sigma^{-1} M_T \tilde{\theta} &= p_2 q_{12} \Sigma^{-1} (z_2 - z_1) + p_1 q_{12} \Sigma^{-1} (z_2 - z_1) \\ &= -q_{12} \Sigma^{-1} (z_1 - z_2). \end{aligned}$$

Or en utilisant (3.4) à nouveau, nous avons : $-q_{12} = \frac{p_1}{p_2} z_1' \Sigma^{-1} z_1$. Finalement, en posant $\lambda = \frac{p_1}{p_2} z_1' \Sigma^{-1} z_1$, nous avons : $\Sigma^{-1} M_T \tilde{\theta} = \lambda \tilde{\theta}$. Donc $\tilde{\theta}$ est bien vecteur propre associé à la valeur propre λ de la matrice $\Sigma^{-1} M_T$. ■

Nous pouvons aussi déduire de (3.3) des estimateurs λ_n et $\tilde{\theta}_n$ de λ et $\tilde{\theta}$:

$$\lambda_n = \frac{p_{1,n}}{p_{2,n}} z_{1,n}' \Sigma_n^{-1} z_{1,n} \quad \text{et} \quad \tilde{\theta}_n = \Sigma_n^{-1} (z_{1,n} - z_{2,n}). \quad (3.5)$$

où $z_{h,n} = m_{h,n} - \bar{x}_n$ pour $h = 1, 2$ où \bar{x}_n , Σ_n , $p_{h,n}$ et $m_{h,n}$ sont définis au chapitre précédent.

3.2.2 Estimateur récursif de la direction de θ lorsque $\mathbf{H} = \mathbf{2}$

Considérons que l'échantillon $\{(x_i, y_i), i = 1, \dots, n\}$ est scindé en deux sous-ensembles, le sous-échantillon des premières $(n-1)$ observations $\{(x_i, y_i), i = 1, \dots, n-1\}$ et de la nouvelle observation (x_n, y_n) . Nous donnerons tout d'abord les formes récursives de x_n , Σ_n , Σ_n^{-1} , puis celle de $p_{h,n}$ et $m_{h,n}$. Nous décrirons ensuite la forme récursive de la matrice d'intérêt $\Sigma_n^{-1} M_{T,n}$. Enfin, nous déterminerons les formes récursives de $\tilde{\theta}_n$ de λ_n .

Forme récursive de \bar{x}_n , Σ_n , Σ_n^{-1} , $p_{h,n}$ et $m_{h,n}$

La forme récursive de la moyenne empirique \bar{x}_n des n observations x_1, \dots, x_n est la suivante :

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i = \frac{n-1}{n} \bar{x}_{n-1} + \frac{1}{n} x_n = \bar{x}_{n-1} + \frac{1}{n} \Phi_n \quad (3.6)$$

avec $\Phi_n = x_n - \bar{x}_{n-1}$. Elle fait intervenir, dans le premier terme, la moyenne empirique \bar{x}_{n-1} des $n - 1$ premières observations x_i , ainsi que la n -ème observation x_n dans le second terme. De manière similaire, la forme récursive de la matrice des variances-covariances empiriques Σ_n est donnée par :

$$\begin{aligned}\Sigma_n &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)(x_i - \bar{x}_n)' \\ &= \frac{n-1}{n} \Sigma_{n-1} + \frac{n-1}{n^2} (x_n - \bar{x}_{n-1})(x_n - \bar{x}_{n-1})' \\ &= \frac{n-1}{n} \Sigma_{n-1} + \frac{n-1}{n^2} \Phi_n \Phi_n'.\end{aligned}\quad (3.7)$$

On peut montrer de même, via l'équation de Riccati, que l'inverse de Σ_n a pour forme récursive :

$$\Sigma_n^{-1} = \frac{n}{n-1} \Sigma_{n-1}^{-1} - \frac{n}{(n-1)(n+\rho_n)} \Sigma_{n-1}^{-1} \Phi_n \Phi_n' \Sigma_{n-1}^{-1} \quad (3.8)$$

avec $\rho_n = \Phi_n' \Sigma_{n-1}^{-1} \Phi_n$.

Nous supposons que la nouvelle observation (x_n, y_n) est telle que $y_n \in s_{h^*}$ avec $h^* = 1$ ou 2 . On peut alors donner la forme récursive des estimateurs $p_{h,n}$ et $m_{h,n}$ de p_h et m_h :

$$p_{h,n} = \begin{cases} \frac{n-1}{n} p_{h^*,n-1} + \frac{1}{n} & \text{si } h = h^* \\ p_{h^*,n-1} & \text{sinon} \end{cases} \quad (3.9)$$

$$m_{h,n} = \begin{cases} m_{h^*,n-1} + \frac{1}{n_{h^*,n-1} + 1} \Phi_{h^*,n} & \text{si } h = h^* \\ m_{h,n-1} & \text{sinon} \end{cases} \quad (3.10)$$

où $\Phi_{h^*,n} = x_n - m_{h^*,n-1}$

Nous avons : $z_{h,n} = m_{h,n} - \bar{x}_n$. A partir des relations (3.6) et (3.10), on obtient la forme récursive de $z_{h,n}$

$$z_{h,n} = \begin{cases} z_{h^*,n-1} - \frac{1}{n} \Phi_n + \frac{1}{n_{h^*,n-1} + 1} \Phi_{h^*,n} & \text{si } h = h^* \\ z_{h,n-1} - \frac{1}{n} \Phi_n & \text{sinon} \end{cases} \quad (3.11)$$

Forme récursive de la matrice d'intérêt $\Sigma_n^{-1}M_{T,n}$

Après avoir les formes récursives de \bar{x}_n , $p_{h,n}$, Σ_n^{-1} et $z_{h,n}$, nous décrivons maintenant la forme récursive de la matrice intérêt $\Sigma_n^{-1}M_{T,n}$. En écrivant $M_{T,n}$ sous la forme :

$$M_{T,n} = \sum_{h \neq h^*} p_{h,n} z_{h,n} z'_{h,n} + p_{h^*,n} z_{h^*,n} z'_{h^*,n} \quad (3.12)$$

nous pouvons déduire des relations (3.9) et (3.11) la forme récursive suivante pour $M_{T,n}$:

$$\begin{aligned} M_{T,n} = & \frac{n-1}{n} M_{T,n-1} - \frac{n-1}{n^2} \sum_{h=1}^H p_{h,n-1} (z_{h,n-1} \Phi'_n + \Phi_n z'_{h,n-1}) \\ & + \frac{n-1}{n^3} \Phi_n \Phi'_n + \frac{n-1}{n} p_{h^*,n-1} A_{h^*,n} + \frac{1}{n} B_{h^*,n} B'_{h^*,n} \end{aligned} \quad (3.13)$$

avec

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

et

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

À partir des relations (3.8) et (3.13), on peut obtenir une forme récursive pour $\Sigma_n^{-1}M_{T,n}$:

$$\Sigma_n^{-1}M_{T,n} = \Sigma_{n-1}^{-1}M_{T,n-1} + \Sigma_{n-1}^{-1}C_{h^*,n} - \frac{1}{n + \rho_n} \Sigma_{n-1}^{-1} \Phi_n \Phi'_n \Sigma_{n-1}^{-1} [M_{T,n-1} + C_{h^*,n}] \quad (3.14)$$

en posant

$$\begin{aligned} C_{h^*,n} = & -\frac{1}{n} \sum_{h=1}^H p_{h,n-1} (z_{h,n-1} \Phi'_n + \Phi_n z'_{h,n-1}) \\ & + \frac{1}{n^2} \Phi_n \Phi'_n + p_{h^*,n-1} A_{h^*,n} + \frac{1}{n-1} B_{h^*,n} B'_{h^*,n} \end{aligned} \quad (3.15)$$

Forme récursive de $\tilde{\theta}_n$

D'après (3.8) et (3.11), nous obtenons la forme récursive de l'estimateur $\tilde{\theta}_n$ de $\tilde{\theta}$

$$\begin{aligned} \tilde{\theta}_n = & \frac{n}{n-1} \tilde{\theta}_{n-1} - \frac{n}{(n-1)(n+\rho_n)} \Sigma_{n-1}^{-1} \Phi_n \Phi_n' \tilde{\theta}_{n-1} \\ & - \frac{(-1)^{h^*n}}{(n_{h^*,n-1}+1)(n-1)} \left(\Sigma_{n-1}^{-1} - \frac{1}{n+\rho_n} \Sigma_{n-1}^{-1} \Phi_n \Phi_n' \Sigma_{n-1}^{-1} \right) \Phi_{h^*,n} \end{aligned} \quad (3.16)$$

Forme récursive de λ_n

Nous avons aussi la forme récursive de l'estimateur λ_n de λ :

$$\lambda_n = \frac{p_{1,n-1} + \frac{1}{n-1} \mathbb{I}[h^* = 1]}{p_{2,n-1} + \frac{1}{n-1} \mathbb{I}[h^* = 2]} v_n' \left(\frac{n}{n-1} \Sigma_{n-1}^{-1} - \frac{n}{(n-1)(n+\rho_n)} \Sigma_{n-1}^{-1} \Phi_n \Phi_n' \Sigma_{n-1}^{-1} \right) v_n$$

avec $v_n = z_{1,n-1} - \frac{1}{n} \Phi_n + \frac{1}{n_{1,n-1}+1} \Phi_{1,n} \mathbb{I}[h^* = 1]$

Il est possible d'écrire cet estimateur sous la forme :

$$\lambda_n = \frac{n}{n-1} \lambda_{n-1} + F(x_n, y_n, n, n_{1,n-1}, n_{2,n-1}, p_{1,n-1}, p_{2,n-1}, \Phi_n, \Sigma_{n-1}^{-1}, z_{1,n-1}, \Phi_{1,n}, \Phi_{2,n}).$$

Nous ne donnons pas volontairement l'expression de $F(\cdot)$ qui, malgré quelques simplifications, reste lourde à écrire.

Remarque 3.2.2.1. *En ce qui concerne l'estimation simultanée de la direction de θ et de la fonction de lien r dans le cadre du modèle (3.1), nous allons considérer l'estimateur de Nadaraya-Watson récursif de r (voir [13]) pour estimer r et le combiner à l'estimateur récursif $\tilde{\theta}_n$ de θ ,*

$$\hat{r}_n(z) = \begin{cases} \frac{\sum_{i=1}^n \frac{1}{h_i} K\left(\frac{z - \hat{z}_i}{h_i}\right) y_i}{\sum_{i=1}^n \frac{1}{h_i} K\left(\frac{z - \hat{z}_i}{h_i}\right)} & \text{si } \sum_{i=1}^n \frac{1}{h_i} K\left(\frac{z - \hat{z}_i}{h_i}\right) \neq 0 \\ \frac{1}{n} \sum_{i=1}^n y_i & \text{sinon} \end{cases} \quad (3.17)$$

où K est un noyau, $z = \theta'x$ et $\hat{z}_i = \tilde{\theta}'_i x_i$ et la suite (h_n) correspond au choix de la largeur de fenêtre.

3.3 Résultats asymptotiques

Avant de donner des résultats asymptotiques pour $\tilde{\theta}_n$, nous supposons les deux hypothèses suivantes :

($\mathcal{H}1'$) Les observations $(x_i, y_i), i = 1, \dots, n$, sont échantillonnées de manière indépendante à partir du modèle (3.1)

($\mathcal{H}2'$) Le support de y est partitionné en deux tranches fixes s_1 et s_2 telles que $p_h \neq 0$ pour $h = 1, 2$.

Nous présentons ci-après deux résultats de convergence, le premier précise la vitesse de convergence presque sûre de $\tilde{\theta}_n$ et le second la normalité asymptotique de cet estimateur. Il serait possible d'obtenir de manière similaire des résultats de convergence de λ_n vers λ .

Convergence presque sûre :

Théorème 3.1. *Sous la condition fondamentale (1.27) et les hypothèses ($\mathcal{H}1'$), ($\mathcal{H}2'$), nous avons, pour $n \rightarrow \infty$:*

$$\|\tilde{\theta}_n - \tilde{\theta}\| = \mathcal{O}\left(\sqrt{\frac{\log(\log n)}{n}}\right) \text{ p.s.}$$

Normalité asymptotique :

Théorème 3.2. *Sous la condition fondamentale (1.27) et les hypothèses ($\mathcal{H}1'$), ($\mathcal{H}2'$), nous avons, pour $n \rightarrow \infty$:*

$$\sqrt{n}(\tilde{\theta}_n - \tilde{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma^{-1} \mathcal{C}_{v_k} \Sigma^{-1}), \quad (3.18)$$

où l'expression de la matrice \mathcal{C}_{v_k} est donnée dans [32]

Chapitre 4

Quelques résultats de simulation

Dans ce chapitre, nous effectuons des études de simulation avec logiciel R pour illustrer le comportement numérique de l'estimateur récursif $\tilde{\theta}_n$ que nous avons proposé. Théoriquement nous avons montré que $\tilde{\theta}_n \xrightarrow{p.s.} \tilde{\theta}$ où $\tilde{\theta}$ est colinéaire à θ . Afin d'évaluer la qualité de l'estimation, nous calculons le cosinus carré de l'angle entre $\tilde{\theta}_n$ et θ défini par :

$$\cos^2(\tilde{\theta}_n, \theta) = \frac{(\langle \tilde{\theta}_n, \theta \rangle)^2}{\|\tilde{\theta}_n\| \times \|\theta\|}$$

où $\|\cdot\|$ désigne la norme associée au produit scalaire usuel $\langle \cdot, \cdot \rangle$. Plus le $\cos^2(\tilde{\theta}_n, \theta)$ est proche de 1, meilleure est la qualité de l'estimation.

Dans les simulations présentées ici, nous avons considéré le modèle de régression suivant :

$$y = (x'\theta)^3 + \varepsilon \tag{4.1}$$

où x suit la loi multinormale $\mathcal{N}_p(0, I_p)$, $\theta = (1, -1, 0, \dots, 0) \in \mathbb{R}^p$ et le terme d'erreur ε , indépendant de x , suit la loi normale $\mathcal{N}(0, 1)$. Ce modèle (4.1) est un modèle homoscédastique ne présentant pas de difficultés particulières pour l'estimation de la direction de θ . Il est considéré parmi les modèles qui sont classiquement utilisés dans les simulations décrites dans la littérature sur la méthode SIR. Nous présentons à la FIGURE 4.1 un nuage de $n=300$ points $\{(x'_i\theta, y_i), i = 1, \dots, n\}$ simulés à partir de modèle(4.1) pour $p=20$.

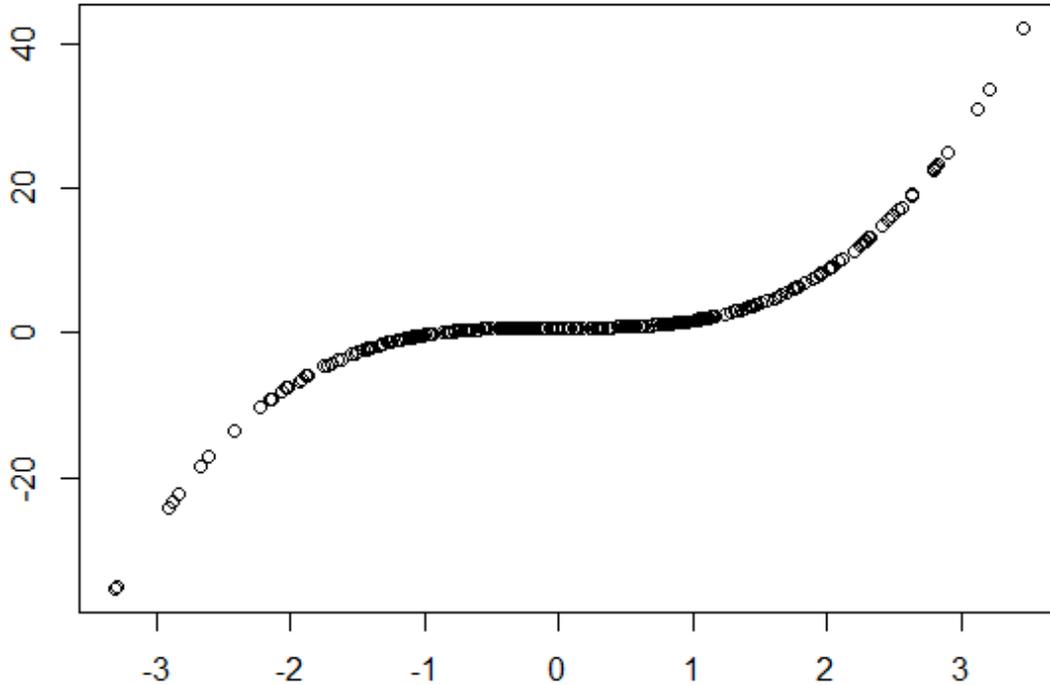


FIGURE 4.1 – Nuage des points $\{(x'_i\theta, y_i), i = 1, \dots, 300\}$ simulés à partir de (4.1) pour $p = 20$

Nous avons simulé, à partir de ce modèle (4.1), des échantillons particuliers et de nombreuses répliques d'échantillons (On prend $\mathcal{B} = 500$ échantillons), de taille 1000 avec successivement $p = 5, 10, 20$ et 40 . Pour chaque échantillon simulé, nous avons calculé pour $n = n_0$ jusqu'à $n = 1000$, l'estimateur récursif $\tilde{\theta}_n$ ainsi que la qualité correspondante, à savoir $\cos^2(\tilde{\theta}_n, \theta)$, la valeur n_0 étant égale à $p + 2$.

Le seuil permettant d'affecter une nouvelle observation (x_n, y_n) dans la première tranche (lorsque $y_n \leq \text{seuil}$) ou dans la deuxième tranche (si $y_n > \text{seuil}$) a été fixé à la valeur médiane des n_0 premières observations utilisées lors de la phase d'initialisation, ce qui permet automatiquement d'avoir des observations dans chacune des deux tranches. Nous décrivons dans un premier temps les résultats de notre simulation, en étudiant la qualité des estimateurs sur les échantillons particuliers puis sur $\mathcal{B} = 500$ échantillons. Nous donnons ensuite des résultats de comparaison des temps de calculs entre les approches SIR récursive et non récursive.

4.1 Etude de la qualité des estimateurs pour l'approche SIR récursive

4.1.1 Etude sur quelques échantillons

À la FIGURE 4.2, nous considérons des échantillons particuliers simulés à partir du modèle (4.1) avec la dimension $p=5$. Sur cette figure, nous avons représenté l'évolution de la mesure de qualité de l'estimateur de $\tilde{\theta}_n$ en fonction de n .

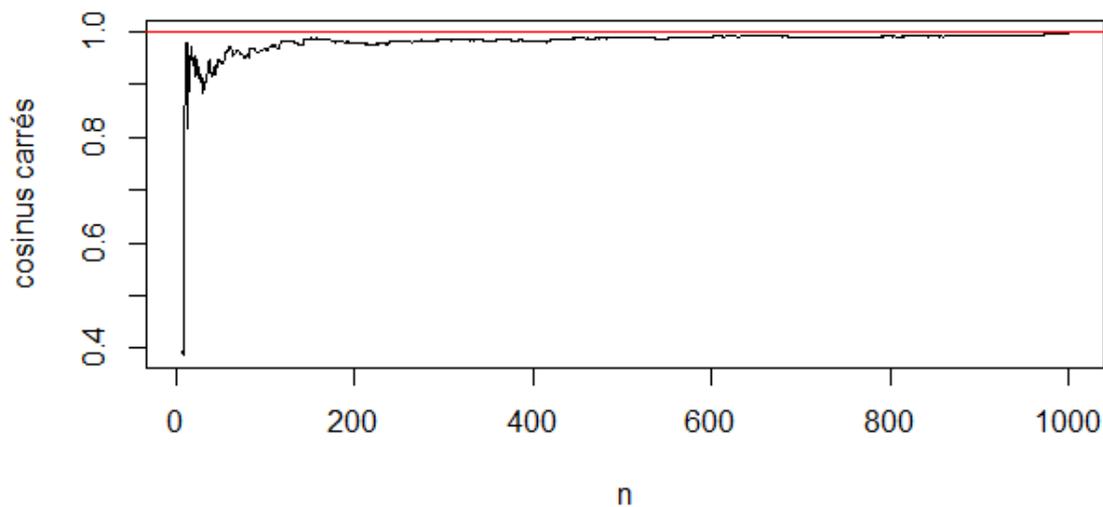


FIGURE 4.2 – Evolution, en fonction de n , de la qualité de l'estimateur $\tilde{\theta}_n$ pour un échantillon issu du modèle (4.1) avec $p = 5$

Nous observons clairement que plus le nombre d'observations est important, plus les valeurs des cosinus carrés $\cos^2(\tilde{\theta}_n, \theta)$ sont proche de 1 (plus la qualité d'estimation est bonne). D'un autre part, il semble que la dimension p de la covariable a aussi naturellement une certaine influence sur la qualité d'estimation de la direction de θ . Dans ce qui suit nous allons illustrer ce point.

Influence de la dimension p de la covariable x sur la qualité de l'estimation.

Nous avons présenté à la FIGURE 4.3, l'évolution, en fonction de n , de la qualité de l'estimateur $\tilde{\theta}_n$ pour un échantillon issu du modèle (4.1) avec différentes dimensions p ($= 10, 20, 40$ et 50).

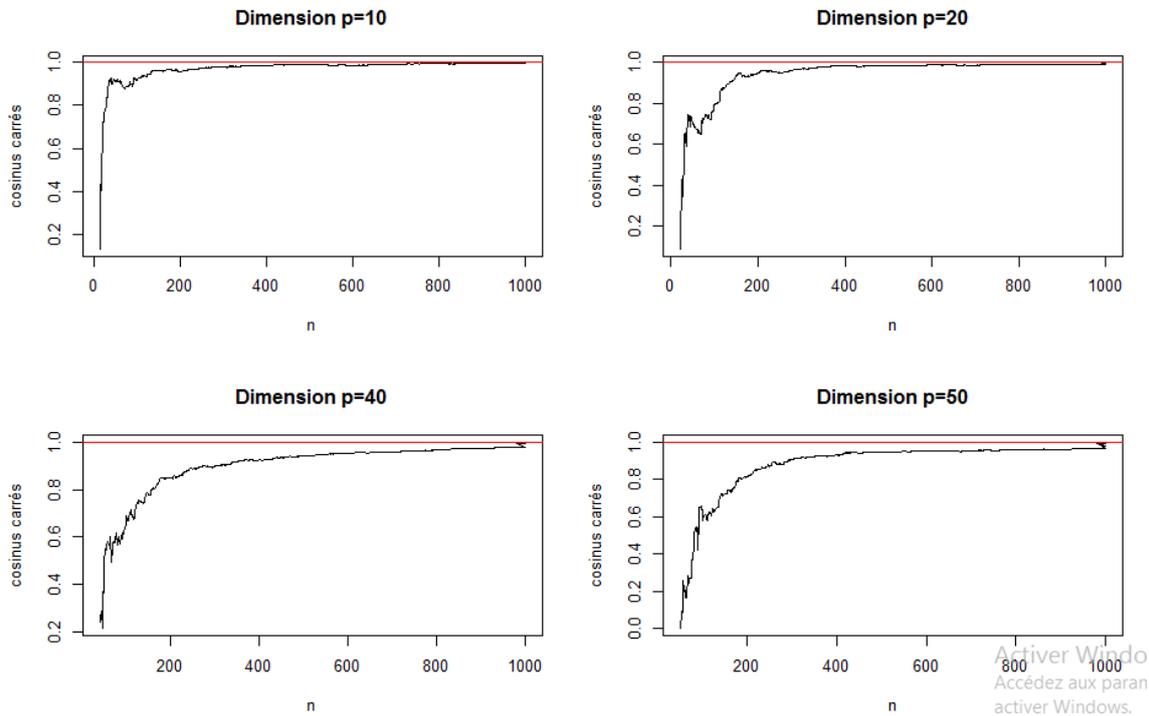


FIGURE 4.3 – Evolution, en fonction de n , de la qualité de l'estimateur $\tilde{\theta}_n$ pour un échantillon issu du modèle (4.1) avec différentes dimensions p ($= 10, 20, 40$ et 50).

Nous voyons clairement que plus la dimension p de la covariable est importante, plus la convergence de $\cos^2(\tilde{\theta}_n, \theta)$ vers 1 est lente. Notons aussi que pour obtenir une bonne estimation de direction de θ lorsque on a des échantillons de taille faible, il suffit de prendre une dimension p de x plus faible.

Dans ce qui suit, nous étendrons cette étude sur $\mathcal{B} = 500$ répliques d'échantillons pour valider nos conclusions.

4.1.2 Etude sur $\mathcal{B} = 500$ répliques d'échantillons

Nous avons représenté à la FIGURE 4.4, les boxplots, pour différentes valeurs de n ($=25, 50, 75, 100, 150, 200, 300, 500, 750, 1000$), des mesures de qualité de $\tilde{\theta}_n$ calculées sur les $\mathcal{B} = 500$ échantillons issus du modèle (4.1) avec $p = 5$. Nous remarquons le même type de phénomène que ceux décrits dans la FIGURE 4.2, plus les valeurs du $\cos^2(\tilde{\theta}_n, \theta)$ se rapprochent de 1 très rapidement.

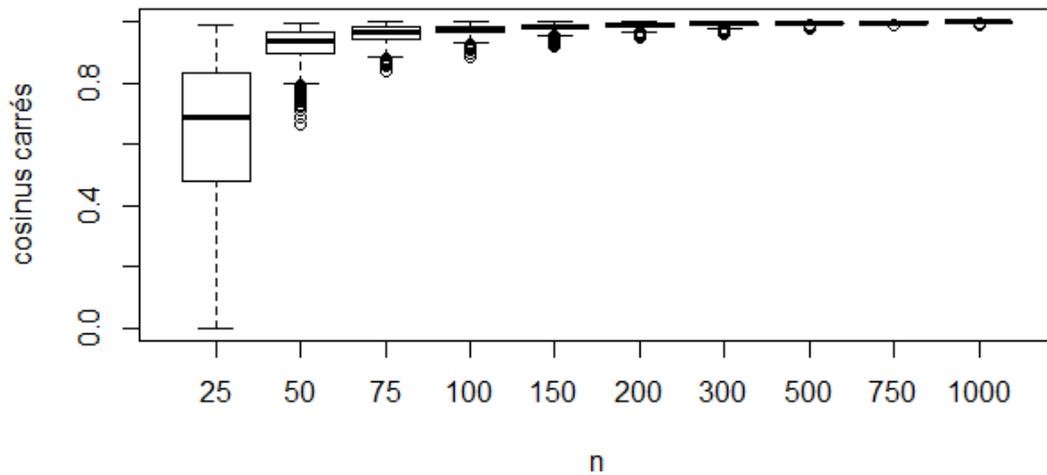


FIGURE 4.4 – Boxplots, pour différentes valeurs de n , des mesures de qualité de $\tilde{\theta}_n$ calculées sur $\mathcal{B} = 500$ échantillons issus du modèle (4.1) avec $p = 5$.

Influence de la dimension p de la covariable x sur la qualité de l'estimation.

Nous avons représenté à la FIGURE 4.5, pour deux tailles d'échantillon ($n = 500$ et 1000), les boxplots, en fonction de la dimension p de x , des mesures de qualité de $\tilde{\theta}_n$ calculées sur $\mathcal{B} = 500$ échantillons issus du modèle (4.1).

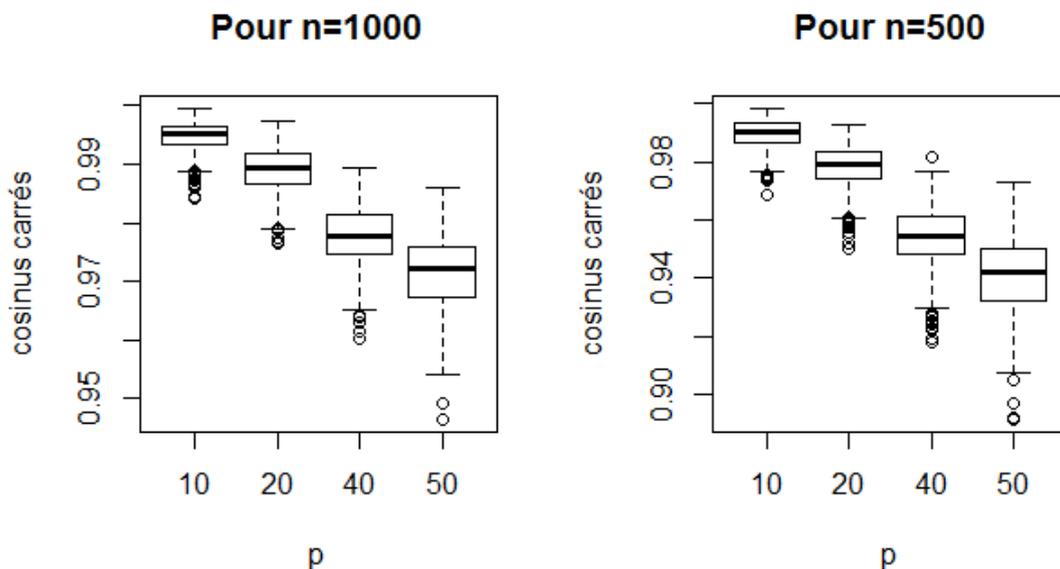


FIGURE 4.5 – Boxplots, en fonction de la dimension p , des mesures de qualité de $\tilde{\theta}_n$ calculées sur $\mathcal{B} = 500$ échantillons issus du modèle (4.1) pour deux tailles d'échantillons $N=500$ ou 1000 .

Remarquons que plus la dimension de la covariable p augmente, plus les valeurs des cosinus carrés $\cos^2(\tilde{\theta}_n, \theta)$ ont tendance à être plus petites.

4.2 Comparaison des temps de calculs entre les approches SIR récursive et SIR classique (basées sur $H = 2$ tranches)

Dans cette partie, nous nous focalisons seulement sur les temps de calcul des estimateurs SIR classique et SIR récursif (basées sur $H = 2$ tranches). Pour chaque méthode, nous mesurons en secondes le temps de calcul nécessaire pour calculer les estimateurs correspondants $\tilde{\theta}_N$ de la direction de θ pour N allant de n_0 à n .

Dans le tableau ci-dessous, nous donnons les valeurs des moyennes et des écarts-types calculés sur $\mathcal{B} = 500$ répliques d'échantillons issus du modèle (4.1) avec différentes valeurs de p et n .

		p=5	p=20	p=40
n=300	SIR classique	1.00(0.019)	1.94(0.053)	3.97(0.011)
	SIR récursif	0.11(0.001)	0.13(0.002)	0.17(0.005)
n=600	SIR classique	1.98(0.014)	4.5(0.093)	10.12(0.097)
	SIR récursif	0.21(0.002)	0.26(0.002)	0.38(0.013)
n=900	SIR classique	3.29(0.021)	7.49 (0.015)	16.67(0.027)
	SIR récursif	0.33(0.004)	0.4(0.004)	0.52(0.007)

TABLE 4.1 – Temps de calculs en secondes des estimateurs $\tilde{\theta}_N$ de la direction de θ (pour N allant de n_0 à n) par la méthode SIR classique et par l'approche SIR récursive : moyennes et écarts-types entre parenthèses calculés sur $\mathcal{B} = 500$ répliques d'échantillons avec différentes valeurs de p et n

À partir de ce tableau, nous constatons que pour calculer les estimations $\tilde{\theta}_{n_0}, \tilde{\theta}_{n_0+1}, \dots, \tilde{\theta}_n$, l'approche SIR récursive proposée est 10 à 30 fois plus rapide (selon les valeurs de p) que la méthode de SIR classique. Notons aussi que lorsque la taille d'échantillon augmente, le temps de calcul semble croître proportionnellement pour la méthode SIR récursive et plus que proportionnellement pour l'approche SIR classique.

Conclusion

Dans ce mémoire, Nous avons présenté tout d'abord les différents modèles de régression, nous avons focalisé au modèle semi-paramétrique de régression de la forme $y = f(\theta'x, \varepsilon)$ où $x \in \mathbb{R}^p$ et $y \in \mathbb{R}$, Notre objectif dans ce travail était d'estimer le paramètre θ de ce modèle par des méthodes récursives. Pour ce fait, nous développons une approche basée sur une méthode introduite par Li (1991)[27], appelée Sliced Inverse Regression (SIR). Puis nous avons proposé une forme récursive de cette approche. Nous avons enfin illustré sur des simulations, le bon comportement numérique de l'approche récursive proposée.

4.3 Perspective

Nous avons présenté dans la section 2.2 les méthodes SIR multivariée. Il serait intéressant de développer des approches récursives pour ces méthodes. Nous pouvons aussi faire de même pour les méthodes SIR_{II} , SIR_{α} et $SAVE$.

Bibliographie

- [1] Amiri, A. (2010). Estimateurs fonctionnels récurrents et leurs applications à la prévision. *Thèse de doctorat, université d'Avignon et des Pays de Vaucluse*.
- [2] Barreda, L., Gannoun, A. and Saracco, J. (2007). Some extensions of multivariate sliced inverse regression. *Journal of Statistical Computation and Simulation*, **77**(1-2), 1-17.
- [3] Barrios, M, P., and Velilla, S. (2007). A bootstrap method for assessing the dimension of a general regression problem, *Statistics and probability letters*, **77**, 247-255.
- [4] Bellman, R. (1957). *Dynamic Programming*, Princeton University Press, Princeton, NJ.
- [5] Bernard-Michel, C., Douté, S., Fauvel, M., Gardes, L. and Girard, S. (2009). Retrieval of Mars surface physical properties from OMEGA hyperspectral images using Regularized Sliced Inverse Regression. *Journal of Geophysical Research - Planets*, **114**, E06005.
- [6] Bernard-Michel, C., Gardes, L. and Girard, S. (2009). Gaussien Regularized Sliced Inverse Regression. *Statistics and Computing*, **19**, 85-98.
- [7] Bourbonnais, R. (2015). *Économétrie*, 9ème ed. Dunod.
- [8] Chesneau, C. (2017). *Modèles de régression*. Master. France. cel-01248297v3
- [9] Chiancone, A. (2006). Réduction de dimension via Sliced Inverse Regression : Idées et nouvelles propositions. *Thèse de doctorat, université de Grenoble*.
- [10] Cook, R. D., and Weisberg, S. (1991). Sliced Inverse Regression for Dimension Reduction : Comment. *Journal of the American Statistical Association*, **86**(414), 328-332.

- [11] Cornillon, P. A. et Løber, E. M. (2010). *Régression avec R*. Université de Rennes 2. Springer.
- [12] Coudret, R. , Girard, S. and Saracco, J. (2014). A new sliced inverse regression method for multivariate response. *Computational Statistics and Data Analysis*, Elsevier, **77**, 285-299.
- [13] Durrieu, G., Nguyen, T.M.N. et Sow, M. (2009). Comparaison d'estimateurs de régression non paramétriques : application en valvométrie. *41èmes Journées de Statistique, SFdS, Bordeaux*. inria-00386716f
- [14] Ferré, L. (1997). Choix de dimension en régression inverse par tranches (SIR). *C. R. Acad. Sci. Paris*, **322**, série I.
- [15] Ferré, L. (1998). Determining the dimension in Sliced Inverse Regression and related methods. *Journal of American Statistical Association*, **93**, 132- 140.
- [16] Ferré, L., Yao, A.F. (1999). Un critère de choix de la dimension dans la méthode SIR_{II} *Revue de statistique appliquée*, **47**(4), 33-46.
- [17] FOULLEY, J. L., DELMAS, C. et Céline et ROBERT-GRANIÉ, C. (2002). Méthodes du maximum de vraisemblance en modèle linéaire mixte. *Journal de la société française de statistique*, **143**(1-2), 5-52.
- [18] Gannoun, A., Girard, S., Guinot, C., Saracco, J. (2002). Trois méthodes non paramétriques pour l'estimation de courbes de référence-application à l'analyse de propriétés biophysiques de la peau. *Revue de statistique appliquée*, **50** (1), 65-89.
- [19] Gannoun, A., Guinot, C. et Saracco, J. (2002). Méthode de régression semi-paramétrique de types « Slicing » ou « Pooled Slicing » : Mise en œuvre sous le logiciel SAS® sous forme de macro-commandes et application sur des jeux de données. *Revue de MODULAD*, **29**, 1-34.
- [20] Gannoun, A., Saracco, J. (2003a). An asymptotic theory for SIR_{α} method. *Statistica Sinica*, **13**, 297-310.
- [21] Gannoun, A., Saracco, J. (2003b). Two Cross Validation Criteria for SIR_{α} and $PSIR_{\alpha}$ methods in view of prediction. *Computational statistics*, **18**, 585-603.
- [22] Hotelling, H.(1936). The Most predictable criterion. *Journal of Educational Psychology*, **26**(2), 139-142.

- [23] Hsing, T. and Carroll, R. J. (1992). An asymptotic theory for Sliced Inverse regression. *The Annals of Statistics*, **20**, 1040-1061.
- [24] Kuentz, V. (2009). Contributions à la réduction de dimension. *Thèse de doctorat, université de Bordeaux I*.
- [25] Kuma, J. K.(2019). Estimation par la méthode du Maximum de Vraisemblance : Éléments de Théorie et pratiques sur Logiciel. Licence. Congo-Kinshasa. cel-02189969ff
- [26] Legendre, A. M. (1805). *Nouvelles méthodes pour la détermination des orbites des comètes*, Firmin-Didot, Paris.
- [27] Li, K-C. (1991). Sliced Inverse Regression for Dimension Reduction, *Journal of the American Statistical Association*, **86**, 316-342.
- [28] Li, K.C., Aragon, Y., Shedden, K. and Thomas-Agnon, C. (2003). Dimension Reduction for Multivariate Response Data. *Journal of the American Statistical Association*, **98**(461), 99 -109.
- [29] Liquet, B. et Saracco, J. (2007). Pooled marginal slicing approach via SIR_α with discrete covariables. *Computational Statistics*, **4**, 599-617.
- [30] Liquet, B. et Saracco, J. (2008). Application of the bootstrap approach to the choice of dimension and the α parameter in the SIR_α method. *Communications in Statistics - Simulation and Computation*, **37** (6), 1198-1218.
- [31] Nguyen, T.M.N., Saracco, J. (2010). Estimation récursive en régression inverse par tranches (sliced inverse regression). *Journal de la Société Française de Statistique*, **151**(2), 19-46.
- [32] Saracco, J. (1997). An asymptotic theory for Sliced Inverse Regression. *Communications in Statistics - Theory and methods*, **26**, 2141-2717.
- [33] Saracco, J. (2005). Asymptotics for pooled marginal slicing estimator based on SIR_α approach. *Journal of Multivariate Analysis*, **96**, 117-135.
- [34] Schott, J. R. (1994). Determining the dimensionality in sliced inverse regression, *Journal of the American Statistical Association*, **89**, 141-148.
- [35] Stigler, S. M. (1981). Gauss and the invention of least squares, *The Annals of Statistics*, **9**, 465-474.

- [36] Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*, Springer-Verlag New York Inc...
- [37] Zhang, T., Ye, W. and Shan, Y. (2016). Application of sliced inverse regression with fuzzy clustering for thermal error modeling of CNC machine tool. *The International Journal of Advanced Manufacturing Technology*, **85**, 2761-2771
- [38] Zhu, L. X. and Ng, K. W. (1995). Asymptotics of sliced inverse regression. *The Annals of Statistics*, **24**, 1053-1068.