



HAL
open science

Estimation récursive pour les modèles semi-paramétriques

Thi Mong Ngoc Nguyen

► **To cite this version:**

Thi Mong Ngoc Nguyen. Estimation récursive pour les modèles semi-paramétriques. Statistiques [math.ST]. Université Sciences et Technologies - Bordeaux I, 2010. Français. NNT: . tel-00938607

HAL Id: tel-00938607

<https://theses.hal.science/tel-00938607>

Submitted on 29 Jan 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° d'ordre : 4107



THÈSE

présentée à

L'UNIVERSITÉ BORDEAUX 1



ÉCOLE DOCTORALE DE MATHÉMATIQUES ET INFORMATIQUE

par **Thi Mong Ngoc NGUYEN**

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : Mathématiques Appliquées - Statistique

ESTIMATION RÉCURSIVE POUR DES MODÈLES SEMI-PARAMÉTRIQUES

Soutenue le 26 novembre 2010 à l'Institut de Mathématiques de Bordeaux

Devant la commission d'examen composée de :

M. Bernard BERCU	PROF.	Université Bordeaux I	Directeur de thèse
M. Philippe BERTHET	PROF.	Université Toulouse III	Examinateur
M. Gérard BIAU	PROF.	Université Paris VI	Examinateur
M. Gilles DURRIEU	PROF.	Université de Bretagne-Sud	Examinateur
M. Stéphane GIRARD	PROF.	INRIA Rhône-Alpes	Rapporteur
M. Bruno PORTIER	PROF.	INSA de Rouen	Rapporteur
M. Jérôme SARACCO	PROF.	Institut Polytechnique de Bordeaux	Directeur de thèse

Institut de Mathématiques de Bordeaux
Université Bordeaux 1
352, cours de la Libération -F 33405
TALENCE cedex

École doctorale de Mathématiques
et Informatique de Bordeaux
U.F.R de Mathématiques Informatique
Bât A33, 1^{er} étage, 352, cours de la
Libération -F 33405 TALENCE cedex

à ma famille

Remerciements

Je souhaite tout d'abord exprimer toute ma gratitude à mes directeurs de thèse, Bernard Bercu et Jérôme Saracco, qui ont dirigé l'ensemble de mes travaux de recherche. Je tiens particulièrement à les remercier pour leur disponibilité, leurs remarques pertinentes, leurs précieux conseils et les nombreuses connaissances statistiques qu'ils m'ont permis d'acquérir.

Je remercie également Messieurs Bruno Portier, Professeur à l'INSA de Rouen et Stéphane Girard, Professeur à l'INRIA Rhône-Alpes, pour avoir accepté d'être les rapporteurs de mon travail de thèse ainsi que pour les précieuses remarques qui ont servi à améliorer ce manuscrit.

Je remercie Monsieur Gilles Durrieu, Professeur à l'Université de Bretagne-Sud d'avoir accepté de faire parti de mon jury. J'ai travaillé avec lui dans le cadre d'une application réelle présentée au chapitre 6. Je tiens à le remercier pour son accueil et sa disponibilité lors de mes séjours au sein du laboratoire EPOC.

Mes remerciements s'adressent aussi à Messieurs Philippe Berthet, Professeur à l'Université Toulouse III et Gérard Biau, Professeur à l'Université Paris VI pour avoir accepté de prendre part à l'examen de cette thèse en tant que membres du jury.

Je remercie Mohamedou Sow pour son amitié et sa sympathie, travailler avec lui est un réel plaisir.

D'une manière plus générale, je remercie tous les membres de l'IMB. Plus particulièrement, je pense aux membres de l'équipe Probabilités et Statistique.

Merci à tous mes amis pour leur soutien tout au long de mon séjour en France et pour tous les bons moments passés ensemble. Je tiens particulièrement à remercier Hung et Jonathan pour m'avoir motivé de faire une thèse.

Je tiens à remercier mes parents sans qui cette expérience n'aurait jamais été possible. Merci à mon frère et à mes deux soeurs pour leur soutien et leur appui moral.

Résumé

Dans cette thèse, nous nous intéressons au modèle semi-paramétrique de régression de la forme $y = f(\theta'x, \varepsilon)$, lorsque $x \in \mathbb{R}^p$ et $y \in \mathbb{R}$. Notre objectif est d'étudier des problèmes d'estimation des paramètres θ et f de ce modèle avec des méthodes récursives.

Dans la première partie, l'approche que nous développons est fondée sur une méthode introduite par Li (1991), appelée Sliced Inverse Regression (SIR). Nous proposons des méthodes SIR récursives pour estimer le paramètre θ . Dans le cas particulier où l'on considère le nombre de tranches égal à 2, il est possible d'obtenir une expression analytique de l'estimateur de la direction de θ . Nous proposons une forme récursive pour cet estimateur, ainsi qu'une forme récursive de l'estimateur de la matrice d'intérêt. Ensuite, nous proposons une nouvelle approche appelée "SIRoneslice" (récursive ou non récursive) de la méthode SIR basée sur l'utilisation de l'information contenue dans une seule tranche optimale (qu'il faudra choisir parmi un nombre quelconque de tranches). Nous proposons également un critère "bootstrap naïf" pour le choix du nombre de tranches. Des résultats asymptotiques sont donnés et une étude sur des simulations démontre le bon comportement numérique des approches récursives proposées et l'avantage principal de l'utilisation la version récursive de SIR et de SIRoneslice du point de vue des temps de calcul.

Dans la seconde partie, nous travaillons sur des données de valvométrie mesurées sur des bivalves. Sur ces données, nous comparons le comportement numérique de trois estimateurs non paramétriques de la fonction de régression : celui de Nadaraya-Watson, celui de Nadaraya-Watson récursif et celui de Révész qui est lui aussi récursif.

Dans la dernière partie de cette thèse, nous proposons une méthode permettant de combiner l'estimation récursive de la fonction de lien f par l'estimateur de Nadaraya-Watson récursif et l'estimation du paramètre θ via l'estimateur SIR récursif. Nous établissons une loi des grands nombres ainsi qu'un théorème de limite centrale. Nous illustrons ces résultats théoriques par des simulations montrant le bon comportement numérique de la méthode d'estimation proposée.

Mots-clés : estimation récursive, modèle semi-paramétrique de régression, méthode de régression inverse par tranchage, bootstrap, estimation à noyau, loi des grands nombres, théorème de limite centrale, martingale.

Abstract

In this thesis, we focus on the semi-parametric regression model $y = f(\theta'x, \varepsilon)$, where $x \in \mathbb{R}^p$ and $y \in \mathbb{R}$. Our objective is to study problems of estimation of the parameters θ and f in this model with recursive methods.

In the first part, we develop an approach which is based on a method introduced by Li (1991), called Sliced Inverse Regression (SIR). We propose recursive methods for estimating the parameter θ . In the particular case when the number of slices equal to 2, it is possible to obtain an analytic expression of the estimator of the direction θ . We

propose a recursive form for this estimator, and a recursive estimation of the matrix of interest. Moreover, we propose an estimator of the direction of θ based on the use of only one “optimal” slice chosen among any number of slices. We call this new method SIRoneslice. We also offer a standard “naive bootstrap” for the choice of the number of slices. We establish some asymptotic properties of the estimators and a simulation study illustrates the good numerical behavior of the estimators proposed. The recursive approach (SIR and SIRoneslice) has the advantage to be computationally faster than the non recursive.

In the second part, we work on data measured on bivalves. On these data, we compare the numerical behavior of three nonparametric estimators of the regression function : Nadaraya-Watson, recursive Nadaraya-Watson and Révész which is also recursive.

In the last part of this thesis, we propose a method which combines the recursive estimation of the link function f by the recursive Nadaraya-Watson estimator and recursive estimation of parameter θ via the recursive SIR estimator. We establish a law of large numbers and a central limit theorem. We illustrate these theoretical results by simulations showing the good numerical behavior of the proposed estimation method.

Keywords : recursive estimation, semiparametric regression model, sliced inverse regression (SIR), bootstrap, kernel estimator, law of large numbers, central limit theorem, martingale.

Table des matières

I	Méthode SIR	1
1	SIR classique	5
1.1	Modèles de régression	5
1.1.1	Modèles paramétriques	5
1.1.2	Modèles non paramétriques	6
1.1.3	Modèles semi-paramétriques	6
1.2	Modèle semi-paramétrique de référence	7
1.2.1	Description du modèle de référence	7
1.2.2	Condition fondamentale	9
1.2.3	Etude des courbes de régression inverse	12
1.3	Méthode SIR	15
1.3.1	Algorithme général des méthodes SIR	15
1.3.2	Détermination de la dimension K du modèle	17
1.3.3	L'état actuel des résultats asymptotiques de SIR	19
1.3.3.1	Convergence en probabilité	19
1.3.3.2	Normalité asymptotique	19
1.4	Quelques extensions de la méthode SIR	20
1.4.1	SIR multivariée	20
1.4.2	SIR_{II} , SIR_{α} et $SAVE$	23
1.4.3	Alternatives au tranchage	24
1.4.3.1	Méthode du Pooled-Slicing	24
1.4.3.2	Estimation de l'espérance de la variance conditionnelle	24
1.4.3.3	Régression inverse par lissage	25
1.5	Applications de la méthode SIR	26
2	Approche récursive pour la méthode SIR	27
2.1	Principe	27
2.2	Estimation récursive du paramètre θ	28
2.2.1	Méthode SIR	28
2.2.2	Estimateur récursif de la matrice d'intérêt $\Sigma^{-1}\Gamma$	28
2.2.3	Estimateur récursif de la direction de θ	29
2.3	Résultats asymptotiques	29
2.4	Quelques résultats de simulations	30

3	Approche récursive ou non de SIR basée sur une seule tranche	61
3.1	Principe	61
3.2	Approche SIR basée sur une seule tranche : “SIRoneslice”	62
3.3	Version récursive de SIRoneslice	63
3.4	Etude sur simulations	63
II	Estimation fonctionnelle	95
4	Estimation non paramétrique d’une densité	99
4.1	Histogramme	99
4.2	Estimation par la méthode du noyau	101
4.2.1	Estimateur de Parzen-Rosenblatt	101
4.2.2	Estimateur récursif à noyau	103
5	Estimateurs d’une fonction de régression	105
5.1	Modèle de régression	105
5.2	Régressogramme	106
5.3	Régression par la méthode du noyau	107
5.4	L’estimateur de Nadaraya-Watson	107
5.4.1	Consistance de l’estimateur Nadaraya-Watson	108
5.4.1.1	Calcul de la variance	108
5.4.1.2	Calcul du biais	109
5.4.2	Normalité asymptotique	110
5.5	Nadaraya-Watson récursif et Révész récursif	111
5.5.1	Les estimateurs récursifs	111
5.5.2	Propriétés asymptotiques	111
6	Application en valvométrie	113
6.1	Introduction	113
6.2	Modèle et Estimateurs	114
6.3	Enregistrement et Acquisition des données	115
6.4	Résultats	115
III	Conjonction des deux algorithmes d’estimation	125
7	Résultats asymptotiques	129
7.1	Modèle et Estimateurs	129
7.2	Hypothèses	131
7.3	Théorèmes	131
7.4	Démonstration des théorèmes	132
7.4.1	Démonstration du théorème 7.1 (Loi des Grands Nombres)	132
7.4.1.1	Etude de la vitesse de convergence de $(N_n(y))$	132
7.4.1.2	Etude de la vitesse de convergence de $(P_n(y))$	136
7.4.1.3	Etude de la vitesse de convergence de $(Q_n(y))$	140
7.4.1.4	Obtention de la convergence de $\hat{f}_n(y)$	146

7.4.2	Démonstration du théorème 7.2 (Théorème de Limite Centrale)	146
8	Quelques résultats de simulation	149
8.1	Modèles simulés et procédure d'estimation	149
8.2	Qualité des estimateurs	150
8.3	Illustration de la normalité asymptotique de l'estimateur	154
8.4	Conclusion	159

Introduction

Les modèles de régression sont très utiles pour modéliser la liaison 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 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. Ces modèles dépendent généralement d'un paramètre de dimension finie, noté θ , ainsi que d'une fonction de lien f à estimer.

Dans cette thèse, nous nous intéressons au modèle semi-paramétrique de régression de la forme $y = f(\theta'x, \varepsilon)$ lorsque la variable à expliquer y est à valeurs dans \mathbb{R} et la covariable x appartient à \mathbb{R}^p . Notre objectif est d'étudier des problèmes d'estimation des paramètres θ et f de ce modèle avec des méthodes récursives. Les méthodes récursives n'ont jamais été développées pour l'estimation de ce type de modèle semi-paramétrique. D'une manière générale, l'avantage des méthodes récursives est de prendre en compte l'arrivée temporelle des informations et d'affiner ainsi au fil du temps les algorithmes d'estimation mis en œuvre. Un intérêt majeur de ces méthodes est qu'il n'est pas nécessaire de relancer le calcul de l'estimateur sur la totalité des données à chaque fois que la base de données est complétée par de nouvelles observations. L'idée est ici d'utiliser les estimations calculées sur la base de données initiales et de les remettre à jour en tenant uniquement compte des nouvelles données arrivant dans la base. Le gain en terme de temps de calculs peut être très intéressant et les applications d'une telle approche sont nombreuses.

Plus précisément, dans un premier temps, l'approche que nous avons développée est fondée sur une méthode récente, introduite par Li (1991), appelée Sliced Inverse Regression (SIR) que l'on peut traduire par "Régression inverse par tranches". Cette méthode permet d'estimer la partie paramétrique θ du modèle semi-paramétrique sans avoir à estimer la fonction de lien f . Nous avons proposé des méthodes SIR récursives pour estimer le paramètre θ . Après avoir obtenu une estimation de θ , nous avons ensuite proposée une procédure récursive d'estimation de la fonction de lien f .

Dans la suite de ce mémoire, les travaux réalisés durant ma thèse sont regroupés en trois parties. La première partie concerne la méthode SIR, la seconde partie porte sur l'estimation fonctionnelle. La dernière partie combine des estimateurs récursifs SIR et fonctionnels. Détaillons maintenant plus précisément ces trois parties.

Dans la partie 1, nous donnons dans un premier temps une présentation de l'état actuel

des travaux sur la méthode SIR. Nous présentons ensuite deux articles (un publié et un soumis à l'heure actuelle). Cette partie se divise en trois chapitres.

Dans le chapitre 1 intitulé “Méthode SIR”, nous présentons une synthèse de travaux concernant la méthode de régression inverse par tranches (SIR). Nous présentons tout d'abord différents modèles de régression, puis nous nous intéressons à un modèle de régression semi-paramétrique dans le cadre nous avons travaillé. Ensuite, nous rappelons brièvement la méthode SIR introduite par Li (1991) et Duan et Li (1991). Cette méthode repose sur une propriété de la courbe de régression inverse et elle permet d'estimer la partie paramétrique du modèle semi-paramétrique sans avoir à estimer la partie fonctionnelle ni à spécifier la loi de l'erreur ε . Dans ce chapitre, nous précisons des techniques de choix de la dimension de la partie paramétrique du modèle, ainsi que quelques propriétés asymptotiques de SIR. Nous terminons ce chapitre en mentionnant quelques extensions de la méthode SIR et des applications de cette méthode.

Nous nous intéressons à une approche récursive de la méthode SIR dans le second chapitre intitulé “Approche récursive pour la méthode SIR”. Dans le cas particulier où l'on considère le nombre de tranches égal à 2, il est possible d'obtenir une expression analytique de l'estimateur de la direction du paramètre θ . Nous proposons une forme récursive pour cet estimateur, ainsi qu'une forme récursive de l'estimateur de la matrice d'intérêt. Ensuite, nous donnons des propriétés asymptotiques de cet estimateur (la convergence presque sûre avec la vitesse correspondante ainsi que la normalité asymptotique). Enfin, nous illustrons sur des simulations le bon comportement numérique de l'approche récursive proposée. Les résultats que nous obtenons confirment l'avantage majeur de cette approche récursive : sa rapidité par rapport aux méthodes SIR non récursive et SIR classique.

Dans le chapitre 3 intitulé “Approche récursive ou non de SIR basée sur une seule tranche”, nous nous intéressons à une approche que l'on a appelée SIRoneslice (récursive ou non récursive) de la méthode SIR basée sur l'utilisation de l'information contenue dans une seule tranche optimale (en un certain sens), dans le cas d'un nombre quelconque de tranches. Nous proposons tout d'abord un estimateur de la direction de θ basé sur l'information contenue dans une seule tranche optimale (qu'il faudra choisir dans l'ensemble des tranches). Nous proposons aussi un critère “bootstrap naïf” pour le choix du nombre de tranches. Nous fournissons ensuite la version récursive de cette approche. Nous avons obtenu le résultat de convergence presque sûre avec la vitesse correspondante de notre estimateur. Nous terminons en montrant les bonnes performances numériques de cette méthode SIRoneslice (récursive ou non récursive) sur des simulations, ainsi que l'avantage principal de l'utilisation la version récursive SIR et SIRoneslice du point de vue des temps de calculs.

Dans la seconde partie intitulée “Estimation fonctionnelle”, nous présentons des estimateurs non paramétriques d'une densité ou d'une fonction de lien. Cette partie regroupe également trois chapitres. Nous donnons tout d'abord la construction de l'estimateur à noyau classique (estimateur de Parzen-Rosenblatt) d'une densité de probabilité. Ensuite, nous présentons l'estimateur récursif d'une densité de probabilité introduit par Wolverton-Wagner, puis celui de Deheuvels. Nous précisons également quelques résultats de convergence de ces estimateurs.

Nous poursuivons dans le chapitre suivant en présentant le célèbre estimateur à

noyau de la fonction de régression introduit par Nadaraya-Watson. Nous nous intéressons ensuite à l'estimateur de Nadaraya-Watson récursif ainsi qu'à l'estimateur récursif de Révész. Enfin, nous rappelons les principales propriétés asymptotiques de ces estimateurs.

Le dernier chapitre de cette deuxième partie intitulé "Application en valvométrie" présente les travaux que j'ai réalisés dans le cadre d'une collaboration avec Gilles Durrieu et Mohamedou Sow du laboratoire EPOC (Environnements et Paléoenvironnement OCéaniques). Nous nous plaçons dans le cadre d'une application réelle dont l'objectif est d'étudier, à partir de données de valvométrie mesurées sur des bivalves, les perturbations des rythmes qui peuvent induire une perturbation du milieu (comme une pollution). Sur ces données, nous avons comparé le comportement numérique des trois estimateurs présentés au second chapitre. Les résultats nous montrent qu'au sens du critère de la variance, les estimateurs récursifs sont statistiquement meilleurs que la version non récursive.

Dans la troisième et dernière partie de cette thèse, nous nous intéressons au modèle de régression $Y_{n+1} = f(\theta'X_n) + \varepsilon_{n+1}$. Notre objectif est d'étudier le comportement de l'estimateur de $f(x'\theta)$. Nous proposons tout d'abord l'estimateur récursif combinant l'estimateur de Nadaraya-Watson récursif de f (qui a été présenté dans la seconde partie) à l'estimateur SIR récursif $\hat{\theta}_n$ de θ (qui a été étudié dans la partie 1). Ensuite, nous étudions quelques propriétés asymptotiques de l'estimateur proposé en donnant deux théorèmes (convergence presque sûre avec la vitesse correspondante et normalité asymptotique). Puis, nous détaillons la démonstration de ces deux théorèmes. Enfin, nous illustrons sur des simulations le bon comportement numérique de l'estimateur récursif proposé.

Nous terminons cette thèse par une conclusion générale et quelques perspectives de recherche sur l'ensemble des trois parties abordées.

Première partie
Méthode SIR

Introduction

La régression inverse par tranches (Sliced Inverse Regression (SIR)) est une méthode de régression semi-paramétrique. Au contraire des autres méthodes de régression semi-paramétrique, elle ne requiert que des temps de calculs informatiques très courts. Nous nous intéressons ici aux méthodes d'estimation récursives. L'avantage des méthodes récursives est de prendre en compte l'arrivée temporelle des informations et d'affiner ainsi au fil du temps les algorithmes d'estimation mis en œuvre. Un intérêt majeur de ces méthodes est qu'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. Le gain en terme de temps de calculs peut être très intéressant. Dans cette partie, après une présentation de l'état actuel des travaux sur la méthode SIR, nous présentons deux articles (un publié et un soumis à l'heure actuelle). Cette première partie est ainsi découpée en trois chapitres.

Le chapitre 1 intitulé "SIR classique" propose une synthèse de travaux concernant la méthode SIR. Nous présentons tout d'abord les modèles de régression dans le cas assez général, puis nous nous intéressons à un modèle de régression semi-paramétrique dans le cadre duquel nous avons travaillé. Ensuite, nous rappelons brièvement la méthode SIR introduite par Li (1991) et Duan et Li (1991), cette méthode permet d'estimer la partie paramétrique des modèles semi-paramétriques considérés sans avoir à estimer la partie fonctionnelle ni à spécifier la loi de l'erreur ε . Une possibilité est de s'intéresser à la distribution conditionnelle de y (variable dépendante sachant x variable explicative multidimensionnelle). Les méthodes de réduction de dimension supposent que l'on peut remplacer x par un vecteur de dimension inférieure $(x'\theta_1, \dots, x'\theta_K)$, avec $K < p$, sans perdre d'information sur le lien entre y et x . On cherche alors à estimer une base du sous-espace de réduction de dimension effective (engendré par les vecteurs $\theta_k, k = 1, \dots, K$). L'idée de base des méthodes SIR est d'échanger le rôle de x et y (afin de réduire la dimension du problème) et d'étudier les moments conditionnels de x sachant y . Les méthodes SIR sont basées sur une propriété géométrique de la courbe de régression inverse qui repose sur une condition cruciale de linéarité de la distribution de la covariable. Dans ce chapitre, nous précisons aussi des techniques de choix de la dimension de la partie paramétrique du modèle, ainsi que quelques propriétés asymptotiques de SIR. Enfin, nous mentionnons quelques extensions de la méthode SIR.

Dans le second chapitre intitulé "Approche récursive pour la méthode SIR", nous nous intéressons à une approche récursive de la méthode SIR. Un avantage majeur de l'utilisation de la forme récursive est que les temps de calculs des estimateurs sont beaucoup plus courts que ceux obtenus avec la forme non récursive, en particulier

lorsque la dimension de x est grande.

Nous proposons tout d'abord une approche récursive de la méthode SIR en donnant une forme récursive de l'estimateur de la matrice l'intérêt, ainsi qu'une forme récursive de l'estimateur de la partie paramétrique du modèle semi-paramétrique dans le cas particulier où l'on considère le nombre de tranches égal à 2. Ensuite, la convergence presque sûre avec la vitesse correspondante ainsi que la normalité asymptotique de l'estimateur ont été obtenues. Enfin, nous illustrons sur des simulations le bon comportement numérique de l'approche récursive proposée et nos résultats nous confirment l'avantage majeur de cette approche récursive : sa rapidité par rapport aux méthodes SIR non récursive et SIR classique.

Dans le chapitre 3 intitulé “Approche récursive ou non de SIR basée sur une seule tranche”, nous nous intéressons à une approche (récursive ou non) de la méthode SIR basée sur l'utilisation de l'information contenue dans une seule tranche optimale (en un certain sens), dans le cas d'un nombre quelconque de tranches.

Nous proposons tout d'abord un estimateur de la direction (de la partie paramétrique du modèle semi-paramétrique) basé sur l'information contenue dans une seule tranche optimale (en un certain sens qu'il faudra choisir parmi le nombre quelconque de tranches). Nous appelons cette nouvelle méthode “SIRoneslice”. Nous proposons une critère “bootstrap naïf” pour le choix du nombre de tranches. Nous fournissons ensuite la version récursive de SIRoneslice. Puis, nous donnons un résultat asymptotique de l'approche SIRoneslice. Enfin, nous montrons sur des simulations la bonne performance numérique de la méthode SIRoneslice, ainsi que l'avantage principal de l'utilisation de la version récursive de SIR et de SIRoneslice du point de vue des temps de calculs.

Les simulations ont été faites à l'aide du logiciel R dans cette première partie.

Chapitre 1

SIR classique

Dans ce chapitre, nous présentons une synthèse de travaux concernant la méthode de régression inverse par tranchage (Sliced Inverse Regression en anglais ou SIR). Nous présentons tout d'abord les modèles de régression dans un cadre assez général, puis nous nous intéressons plus précisément au modèle semi-paramétrique de régression dans le cadre duquel nous avons travaillé. Pour ce modèle, nous spécifions aussi la condition cruciale sur la distribution de la variable explicative, condition qui est indispensable au bon fonctionnement théorique de l'approche SIR. Nous rappelons brièvement alors la méthode SIR introduite par Li (1991) et Duan et Li (1991). Cette méthode repose sur une propriété de la courbe de régression inverse et elle permet d'estimer la partie paramétrique du modèle. Nous précisons ensuite des techniques de choix de la dimension de la partie paramétrique du modèle, ainsi que quelques propriétés asymptotiques de SIR. Nous terminons ce chapitre en mentionnant quelques extensions de la méthode SIR.

1.1 Modèles de régression

Les méthodes de régression sont très utiles pour modéliser la liaison entre une variable à expliquer y et une variable explicative x . Elles sont utilisées dans de nombreux domaines des sciences appliquées tels que l'économie, l'agronomie, la biostatistique, la biométrie 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.

1.1.1 Modèles paramétriques

Pour les modèles de régression paramétriques, 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. Ils sont de la forme :

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

où f_{θ} appartient à une famille de fonctions paramétrées par θ , vecteur de paramètres réels, et où 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, ...) sont efficaces quand la famille de f_θ est correctement spécifiée. Ces modèles permettent une interprétation claire de l'impact de la variable explicative sur la variable à expliquer.

Cependant, le choix d'un bon modèle paramétrique au vu des données n'est pas toujours évident. Ainsi le modèle paramétrique choisi peut ne pas être en adéquation avec les données réelles et donc parfois être très "éloigné" de la réalité des données. En conséquence les conclusions en découlant peuvent alors être erronées.

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

1.1.2 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 dans le modèle, seules des hypothèses de régularité de la fonction de lien sont imposées). 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 :

$$y = f(x) + \varepsilon.$$

Le thème commun de la régression fonctionnelle pour estimer $f(x)$ est l'idée d'un lissage local. Sa qualité dépend de la présence de suffisamment de données dans le voisinage de x . Lorsque la variable explicative est unidimensionnelle, de nombreuses techniques de lissage sont disponibles : méthode des noyaux, splines de lissage, ... Mais lorsque la dimension de x devient importante, le nombre d'observations nécessaires pour le lissage local croît de manière exponentielle avec cette dimension. À moins de disposer d'un échantillon de taille gigantesque, ces méthodes d'estimation non paramétrique ne seront plus valides à cause de la faible abondance de données dans les régions d'intérêt. De plus, l'interprétation de la fonction de lien n'est pas toujours évidente.

Pour surmonter ce problème de dimension et d'interprétation de l'effet de la covariable x , des modèles de régression semi-paramétrique ont été développés.

1.1.3 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 de tels modèles, la variable à expliquer y dépend généralement de x par le biais d'un nombre fini de paramètres euclidiens $\theta'_1, \dots, \theta'_K$ et d'un paramètre fonctionnel f . Dans la section suivante, nous présentons le modèle de référence dans le cadre on va travailler dans la suite de ce mémoire.

1.2 Modèle semi-paramétrique de référence et régression inverse

Nous décrivons le modèle de référence et précisons les paramètres d'intérêt que l'on est amené à estimer. Ensuite, nous donnons une condition fondamentale sur la distribution de la variable explicative. Puis, nous étudions une propriété géométrique de la courbe de régression inverse dans ce cadre.

1.2.1 Description du modèle de référence

Nous nous intéressons au modèle semi-paramétrique de régression proposé par Li (1991) :

$$y = f(\theta'_1 x, \dots, \theta'_K x, \varepsilon), \quad (1.2.1)$$

où :

- (i) y est une variable unidimensionnelle à expliquer ;
- (ii) 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 ;
- (iii) $\theta_1, \dots, \theta_K$ sont des vecteurs de paramètres réels de dimension p , inconnus et linéairement indépendants, avec $K < p$;
- (iv) ε est un terme d'erreur aléatoire indépendant de x , aucune hypothèse n'est faite sur la distribution de ε ;
- (v) f est le paramètre fonctionnel à valeur dans \mathbb{R} , inconnu et arbitraire.

La valeur de K étant supposée strictement inférieure à p , ce modèle est donc aussi un modèle permettant une réduction de dimension de l'espace des variables explicatives. Dans un tel modèle, on appelle *indices* les termes $(\theta'_k x)$, $k = 1, \dots, K$. Ainsi pour $K > 1$, on parlera de modèle multi-indices, et lorsque $K = 1$, de modèle à un seul indice $\theta'_1 x$.

En permettant à la fonction f et à l'erreur ε d'être complètement non spécifiées, on n'impose aucune hypothèse structurelle sur la forme du lien qui existe entre y et les K indices $\theta'_1 x, \dots, \theta'_K x$. Le modèle (1.2.1) est en fait équivalent à l'assertion suivante :

Conditionnellement à $\theta'_1 x, \dots, \theta'_K x$, y est indépendant de x .

Ainsi dans le cadre du modèle de référence (1.2.1), l'étude de la relation entre y et x se réduit à l'examen de la structure qui existe entre y et les K indices $\theta'_1 x, \dots, \theta'_K x$.

Cependant, il convient de remarquer que sous les hypothèses du modèle (1.2.1), les paramètres (θ_k) ne sont pas totalement identifiables.

En effet, soit $B = [\theta_1, \dots, \theta_K]$. Pour toute matrice A inversible d'ordre K , on peut réécrire (1.2.1) sous la forme :

$$y = f(B'x, \varepsilon) = f(\{A'\}^{-1}(BA)'x, \varepsilon).$$

Etant donné que la fonction f est arbitraire, le terme (A^{-1}) peut être "absorbé" par f . Ainsi, le modèle (1.2.1) ne permet pas de distinguer les vecteurs colonnes de B de ceux

de la matrice BA . Par conséquent, on ne peut estimer au mieux que le sous-espace engendré par les colonnes de B .

Notons $E = \text{Vect}(B)$ le sous-espace linéaire de \mathbb{R}^p de dimension $K < p$ engendré par les colonnes de la matrice B , c'est à dire par les $\theta_k, k = 1, \dots, K$.

Définition 1.1. *On appelle E l'espace EDR ("effective dimension reduction"). De plus, toute combinaison linéaire des (θ_k) appartient à E , et 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 EDR E . L'estimation de la partie fonctionnelle f peut ensuite être obtenue avec les méthodes non paramétriques usuelles.

Remarque 1.1. Quelques remarques sur le modèle

- (i) Quand $K = 1$, on s'intéresse simplement à l'estimation de la direction du vecteur θ . Duan et Li (1991) ont étudié un tel modèle.
- (ii) Pour que les θ_k soient individuellement identifiables, il est nécessaire d'imposer en plus des contraintes de normalisation, d'orthogonalité et de sens aux θ_k .
- (iii) Cette limitation concernant la non identifiabilité des θ_k n'affecte cependant pas la spécification du modèle dans l'étape suivant la réduction de dimension. En effet, n'importe quelle base de l'espace EDR E peut être ensuite utilisée pour l'étape d'estimation du paramètre fonctionnel f .
- (iv) Dans le modèle de référence (1.2.1), on suppose K , le paramètre de réduction de dimension, connu. Divers travaux ont été effectués sur le choix de la dimension K quand on est confronté à un échantillon (voir la section 1.3.2 pour des détails et références).

Autre écriture du modèle

On va voir par la suite qu'il sera pratique de considérer la version standardisée z de x :

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

On peut alors réécrire le modèle (1.2.1) sous la forme :

$$y = g(\eta'_1 z, \dots, \eta'_K z, \varepsilon), \tag{1.2.2}$$

où $\eta_k = \Sigma^{1/2}\theta_k, k = 1, \dots, K$. Notons $N = [\eta_1, \dots, \eta_K]$ et $E^s = \text{Vect}(N)$.

Définition 1.2. *E^s est appelé l'espace EDR standardisé, et toute direction appartenant à E^s est appelée direction EDR standardisée.*

Remarque 1.2. La fonction g est identique à la fonction f à une translation près de ses K premiers arguments. En effet,

$$\begin{aligned} g(\eta'_1 z, \dots, \eta'_K z, \varepsilon) &= g(\eta'_1 \Sigma^{-1/2}(x - \mu), \dots, \eta'_K \Sigma^{-1/2}(x - \mu), \varepsilon) \\ &= g(\theta'_1 x - \theta'_1 \mu, \dots, \theta'_K x - \theta'_K \mu, \varepsilon). \end{aligned}$$

1.2.2 Condition fondamentale

Le théorème de caractérisation de l'espace EDR (à la base des méthodes SIR) énoncé à la section 1.2.3 nécessite que x vérifie la condition suivante :

(H) *La variable explicative x suit une distribution de probabilité non dégénérée telle que, $\forall b \in \mathbb{R}^p$, l'espérance conditionnelle $\mathbb{E}[b'x \mid \theta'_1 x, \dots, \theta'_K x]$ 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 :*

$$\mathbb{E}[b'x \mid \theta'_1 x, \dots, \theta'_K x] = c_0 + c_1 \theta'_1 x + \dots + c_K \theta'_K x.$$

Il est intéressant de noter que cette condition est vérifiée lorsque la distribution de x est de type elliptique. En effet, si x est elliptique, **(H)** est vraie quels que soient les vecteurs θ_k . On peut donc la remplacer par l'hypothèse plus forte (mais plus simple à imaginer) suivante :

(H') *La variable explicative x suit une distribution elliptique.*

On rappelle que les distributions multinormales, les distributions multidimensionnelles de Laplace, de Student et de Cauchy ainsi que la loi multinormale contaminée dans la variance, sont des exemples de distributions elliptiques.

Petite discussion sur la condition fondamentale

Rappelons tout d'abord que la condition fondamentale **(H)** est vérifiée lorsque la distribution de x est elliptique. Remarquons aussi qu'il est très difficile de vérifier a priori si la condition **(H)** est satisfaite vu qu'elle dépend de directions inconnues. De plus, la plupart des données collectées ne suivent pas une distribution elliptique.

Cette discussion se décompose en deux parties principales : dans un premier temps, on s'intéresse au cas où la variable explicative x est de faible dimension et on verra comment on peut "forcer" les données à être elliptiques si elles ne le sont pas ; puis, dans une seconde partie, on présente un résultat de Hall et Li (1993). Ils montrent par un argument bayésien que, quand la dimension de x est grande, la condition **(H)** est presque toujours approximativement vérifiée. Lorsque la dimension de x est intermédiaire, il est préférable d'utiliser une variante de la méthode SIR, appelée "cluster-based" SIR (voir Kuentz et Saracco (2009)).

Etude du cas où x est de faible dimension

Lorsque la dimension de la variable explicative x n'est pas trop importante et que les données ne suivent pas une distribution elliptique, l'idée est d'essayer de forcer ces données à se comporter comme si tel était le cas.

Si les données n'ont pas encore été collectées, il est possible grâce à un plan d'expérience convenable d'atteindre l'objectif recherché. D'autres méthodes sont disponibles pour obtenir l'ellipticité de l'échantillon sélectionné.

La méthode la plus simple est celle du rééchantillonnage gaussien de Brillinger (1983). L'idée est de simuler tout d'abord un échantillon gaussien de taille n ayant comme espérances et variances celles de l'échantillon des données initiales. Les points ainsi engendrés seront les "attracteurs". Pour chaque attracteur, il faut ensuite sélectionner le point des données initiales le plus proche de cet attracteur. (Certains points des

données seront sélectionnés plusieurs fois, alors que d'autres ne seront pas sélectionnés du tout). Les points sélectionnés sont supposés suivre une distribution normale (ou du moins plus normale que la distribution initiale).

La seconde méthode est due à Cook et Nachtsheim (1994). Elle permet de calculer des poids pour chaque point de l'échantillon afin de disposer de variables explicatives proches de l'ellipticité. Soit $\{x_i, i = 1, \dots, n\}$ l'échantillon initial. L'idée est de construire une mesure discrète de probabilité Q_n appropriée qui met une masse seulement sur les points x_i et telle qu'elle soit proche d'une distribution de probabilité elliptique cible Q_t . Deux questions se posent : celle du choix de la distribution cible Q_t , et celle de la construction d'une mesure discrète Q_n qui est une approximation d'une distribution continue spécifiée Q_t .

En ce qui concerne le choix de Q_t , cette distribution cible est fondée sur l'ellipsoïde de volume minimal (EVM) qui supprime une proportion α (fixée par l'utilisateur) des points de l'échantillon. L'avantage de l'EVM est de faire décroître l'influence des individus extrêmes. Rappelons que toute variable elliptique w d'espérance μ_w et de matrice de covariance Σ_w a une représentation de Cambanis. Ainsi, pour caractériser une distribution elliptique, il est nécessaire de spécifier μ_w , Σ_w et R_w . La distribution cible a alors pour deux premiers paramètres les moments empiriques de l'échantillon "restant". Le choix de la distribution de R_w est ouvert. Cependant, un choix par défaut qui semble donner des résultats convenables en pratique est la distribution empirique Q_{R_w} : il s'agit de la distribution empirique des longueurs des données non éliminées, lorsque l'on a standardisé en prenant pour μ_w et Σ_w les valeurs précisées ci-dessus. Un problème reste cependant ouvert, celui du choix optimal du pourcentage α des observations à éliminer.

En ce qui concerne la construction d'une mesure discrète Q_n qui est une approximation d'une distribution continue spécifiée Q_t , une solution naturelle est l'utilisation des mosaïques de Voronoi (ou cellules de Dirichlet) définies de la manière suivante :

$$d(x_i) = \{x \in \mathbb{R}^p \mid \|x - x_i\| \leq \|x - x_j\|, j \neq i\}, \text{ avec } i = 1, \dots, n^0.$$

où n^0 est le nombre de données restantes après l'étape d'élimination avec l'ellipsoïde de volume minimal. On affecte aux individus éliminés une masse de probabilité nulle, et la masse de probabilité sur les autres points x_i sera :

$$\int_{d(x_i)} dQ_t(x).$$

Le calcul de cette intégrale peut être obtenu par la méthode de Monte-Carlo.

Une fois la distribution Q_n obtenue, on peut appliquer la méthode de régression inverse par tranches en utilisant pour chaque individu le poids fourni par Q_n .

Etude du cas où x est de grande dimension

Lorsque le nombre de variables explicatives est très grand, il devient très difficile de forcer les données à être elliptiques avec des méthodes du type de celles présentées précédemment. Cependant, Hall et Li (1993) ont obtenu un résultat très intéressant. Ils ont étudié les formes des projections des données de grande dimension sur des espaces de dimension plus faible. Supposons x standardisée. Pour une projection $\theta'x$

avec $\|\theta\|=1$, ils se sont intéressés à trouver une direction b telle que la courbe de régression de $b'x$ sur $\theta'x$ soit aussi non linéaire que possible. Et, ils ont montré que, quand la dimension de x est très grande, pour la plupart des directions θ , même la régression la plus non linéaire est encore presque linéaire. Ils ont généralisé ce résultat en remplaçant $\theta'x$ par $B'x$ où $B = [\theta_1, \dots, \theta_K]$, les θ_k étant des vecteurs aléatoires orthonormaux. Ils ont ensuite précisé comment ces résultats permettent d'élargir le champ d'applicabilité de la régression inverse par tranches.

Il apparaît intéressant de détailler relativement rapidement leurs résultats. Soit x un vecteur aléatoire de \mathbb{R}^p tel que $\mathbb{E}[x] = 0$ et $\mathbb{V}(x) = I_p$. Soit une direction unitaire θ de \mathbb{R}^p . Hall et Li s'intéressent à la courbe de régression de $b'x$ sur $\theta'x$ où b est une direction de \mathbb{R}^p , c'est-à-dire $\mathbb{E}[b'x \mid \theta'x]$. Si la distribution de x satisfait la condition

$$\mathbb{E}[b'x \mid \theta, \theta'x = t] \text{ est linéaire en } t, \forall b \in \mathbb{R}^p \quad (*)$$

alors le graphe de $b'x$ sur $\theta'x$ montre une liaison linéaire, et il est de plus possible de montrer que :

$$\mathbb{E}[x \mid \theta, \theta'x = t] = t,$$

ou de façon équivalente que :

$$\|\mathbb{E}[x \mid \theta, \theta'x = t]\|^2 - t^2 = 0.$$

Si la distribution de x ne satisfait pas tout à fait la condition (*), les deux relations précédentes ne sont alors plus vérifiées. Utilisons le terme de gauche de la dernière relation, $\|\mathbb{E}[x \mid \theta, \theta'x = t]\|^2 - t^2$, comme mesure de déviation de la linéarité, et intéressons-nous à la mesure de l'ensemble

$$\{\theta \mid \|\mathbb{E}[x \mid \theta, \theta'x = t]\|^2 - t^2 > \epsilon, \|\theta\|=1\}$$

pour $\epsilon > 0$. Hall et Li montrent que, sous certaines hypothèses techniques pas très contraignantes, la mesure de cet ensemble relativement à la sphère $\{\theta \mid \|\theta\|=1\}$, converge vers 0 quand la dimension de x tend vers l'infini. De manière équivalente, si θ est traité comme un vecteur aléatoire provenant de la sphère unité de \mathbb{R}^p , alors :

$$\|\mathbb{E}[x \mid \theta, \theta'x = t]\|^2 - t^2 \xrightarrow{\mathcal{P}} 0 \quad \text{pour } p \rightarrow \infty.$$

Pour la généralisation au cas multi-indices, en posant $B = [\theta_1, \dots, \theta_K]$ où les θ_k sont K vecteurs orthonormaux aléatoires provenant de la la sphère unité de \mathbb{R}^p , Hall et Li ont obtenu le résultat suivant : pour p tendant vers l'infini,

$$\|\mathbb{E}[x \mid B, B'x = t]\|^2 - t^2 \xrightarrow{\mathcal{P}} 0, \quad \text{avec } t \in \mathbb{R}^K.$$

Ces résultats ont un impact sur le champ d'applicabilité de la méthode SIR. Hall et Li présentent un argument bayésien. Les vecteurs inconnus θ_k du modèle de référence peuvent être vus comme ayant été choisis au hasard mais orthogonaux sur la sphère unité de \mathbb{R}^p . D'où, d'après les résultats ci-dessus, tant que la dimension de x est grande, les chances pour que la condition **(H)** soit vérifiée sont importantes. Ainsi, pour de nombreux ensembles de données, une application "aveugle" des procédures de régression

(SIR ou autre), sans vérifier la condition **(H)**, peut être utile et fournir encore une réponse à peu près correcte pour l'estimation des directions EDR. Cependant un diagnostic de vérification est recommandé après cette utilisation de SIR afin de regarder si, parmi les directions trouvées, il y en a certaines qui s'éloignent sérieusement de l'ellipticité. Et, une fois que l'on a repéré ces directions les plus dangereuses, on peut par exemple appliquer la méthode de Cook et Nachtsheim pour forcer l'ellipticité des données sur ces directions là.

Pour conclure cette discussion, il est intéressant de noter que dans le cas d'un modèle à un seul indice, Duan et Li (1991) ont étudié le comportement de l'estimateur $\hat{\theta}$ de la direction de θ lorsque l'hypothèse d'ellipticité de la distribution de x était violée. Ils ont établi une borne pour le biais qui apparaît lorsque la distribution de x n'est pas elliptique.

1.2.3 Etude des courbes de régression inverse

Nous étudions maintenant une propriété géométrique des courbes de régression inverse dans le cadre du modèle de référence et sous la condition fondamentale. Cette propriété va permettre de caractériser l'EDR E .

Remarquons tout d'abord que si l'on considère une transformation monotone T de y , on a :

$$T(y) = f^*(\theta'_1 x, \dots, \theta'_K x, \varepsilon), \quad (1.2.3)$$

où la fonction f^* est la composée des fonctions T et f , ainsi $T(y)$ obéit à un modèle de type (1.2.1). De même, le modèle (1.2.2) peut alors s'écrire :

$$T(y) = g^*(\eta'_1 z, \dots, \eta'_K z, \varepsilon), \quad (1.2.4)$$

où g^* est la fonction composée de T et de g . Ici, $T(y)$ obéit à un modèle de type (1.2.2).

Définition 1.3. *Etant donnée une telle transformation T , on appelle :*

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

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

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

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

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

$$y \longrightarrow \mathbb{E}[z \mid T(y)].$$

Ces différentes courbes sont dans l'espace \mathbb{R}^p , cependant, Li (1991) a énoncé le théorème de caractérisation de l'espace EDR E (respectivement de l'espace EDR standardisé E^s) à partir des courbes de régression inverse.

Théorème 1.1. Caractérisation de l'espace EDR

*Dans le cadre du modèle (1.2.3) et sous la condition fondamentale **(H)**, la courbe de régression inverse centrée appartient au sous-espace linéaire de dimension K de \mathbb{R}^p engendré par les vecteurs $\Sigma\theta_k$, $k = 1, \dots, K$.*

Théorème 1.2. Caractérisation de l'espace EDR standardisé

Dans le cadre du modèle (1.2.4) et sous la condition fondamentale **(H)**, la courbe de régression inverse standardisée appartient à l'espace EDR standardisé E^s .

Preuve (du théorème 1.2) :

Soit $b \in \mathbb{R}^p$ un vecteur orthogonal à l'espace EDR standardisé, c'est-à-dire : $\forall k = 1, \dots, K, b' \eta_k = 0$. D'après le modèle (1.2.4), on a :

$$\begin{aligned} b' \mathbb{E}[z | T(y)] &= \mathbb{E} [\mathbb{E}[b'z | T(y), \eta'_1 z, \dots, \eta'_K z] | T(y)] \\ &= \mathbb{E} [\mathbb{E}[b'z | \eta'_1 z, \dots, \eta'_K z] | T(y)]. \end{aligned}$$

Montrons que $\mathbb{E}[b'z | \eta'_1 z, \dots, \eta'_K z] = 0$, ou de manière équivalente que

$$\mathbb{E} [(\mathbb{E}[b'z | \eta'_1 z, \dots, \eta'_K z])^2] = 0.$$

En effet, on a :

$$\begin{aligned} \mathbb{E} [(\mathbb{E}[b'z | \eta'_1 z, \dots, \eta'_K z])^2] &= \mathbb{E} [\mathbb{E}[b'z | \eta'_1 z, \dots, \eta'_K z] \mathbb{E}[b'z | \eta'_1 z, \dots, \eta'_K z]] \\ &= \mathbb{E} [\mathbb{E}[b'z \mathbb{E}[b'z | \eta'_1 z, \dots, \eta'_K z] | \eta'_1 z, \dots, \eta'_K z]] \\ &= \mathbb{E}[b'z \mathbb{E}[b'z | \eta'_1 z, \dots, \eta'_K z]]. \end{aligned}$$

Or, d'après **(H)**, $\mathbb{E}[b'z | \eta'_1 z, \dots, \eta'_K z]$ est linéaire en $\eta'_1 z, \dots, \eta'_K z$ pour tout $b \in \mathbb{R}^p$. On en déduit que :

$$\begin{aligned} \mathbb{E} [(\mathbb{E}[b'z | \eta'_1 z, \dots, \eta'_K z])^2] &= b' \mathbb{E} \left[z \sum_{k=1}^K a_k \eta'_k z \right] \\ &= \sum_{k=1}^K a_k b' \mathbb{E}[z z'] \eta_k = 0. \end{aligned}$$

On a donc : $\forall b \in (E^s)^\perp, b' \mathbb{E}[z | T(y)] = 0$.

Ainsi, la courbe, $y \rightarrow \mathbb{E}[z | T(y)]$, appartient à l'espace EDR standardisé, soit :

$$\mathbb{E}[z | T(y)] = \sum_{k=1}^K c_k(y) \eta_k.$$

□

Remarque 1.3. Il est important de noter que seuls les coefficients $c_k(y)$ dans la démonstration ci-dessus dépendent du choix de la fonction T .

Remarque 1.4. Le théorème 1.2 garantit uniquement le fait que la courbe de régression inverse standardisée appartient à l'espace EDR standardisé E^s . Ainsi, cette courbe n'engendre pas forcément E^s . De même, dans le théorème 1.1, la courbe de régression inverse centrée appartient bien au sous-espace linéaire engendré par les vecteurs $\Sigma \theta_k, k = 1, \dots, K$, mais elle n'engendre pas nécessairement ce sous-espace. Il peut ainsi parfois arriver que ces courbes n'engendent que des sous-espaces de dimension strictement inférieure à K .

Conséquence sur les matrices de covariances des courbes de régression inverse

Li a déduit des deux théorèmes de caractérisation, les résultats directs suivants qui sont à la base des méthodes d'estimation. On suppose dorénavant que les courbes de régression inverse, centrée ou standardisée, engendrent les sous-espaces linéaires de dimension K auxquels elles appartiennent. Notons par Γ_T et Γ_T^s les matrices de covariances des courbes de régression inverse centrée et standardisée :

$$\Gamma_T := \mathbb{V}(\mathbb{E}[x | T(y)]) \quad \text{et} \quad \Gamma_T^s := \mathbb{V}(\mathbb{E}[z | T(y)]).$$

Conséquence du théorème 1.2. D'après ce théorème, on a :

$$\begin{aligned} \Gamma_T^s &= \mathbb{V}(\mathbb{E}[z | T(y)]) \\ &= \mathbb{E}[E[z | T(y)]E[z | T(y)]'] - \mathbb{E}[E[z | T(y)]\mathbb{E}[E[z | T(y)]]'] \\ &= \mathbb{E} \left[\left(\sum_{k=1}^K c_k(y)\eta_k \right) \left(\sum_{k=1}^K c_k(y)\eta_k \right)' \right] - \mathbb{E}[z]\mathbb{E}[z]' \\ &= \mathbb{E} \left[\sum_{k=1}^K \sum_{h=1}^K c_k(y)c_h(y)\eta_k\eta_h' \right] \\ &= \sum_{k=1}^K \sum_{h=1}^K \mathbb{E}[c_k(y)c_h(y)] \eta_k\eta_h'. \end{aligned}$$

D'où, pour tout $b \in \mathbb{R}^p$ orthogonal à l'espace EDR standardisé, c'est-à-dire $\forall k = 1, \dots, K, \eta_k' b = 0$, on a : $\Gamma_T^s b = 0$. La matrice Γ_T est donc dégénérée dans toutes directions orthogonales à E^s . Ainsi, les K vecteurs propres, v_1^s, \dots, v_K^s , associés aux K plus grandes valeurs propres de Γ_T^s (c'est-à-dire les K valeurs non nulles de Γ_T^s) appartiennent à l'espace EDR standardisé, et sont donc des directions EDR standardisées.

Pour revenir à l'échelle originale, on utilise la transformation linéaire suivante :

$$v_k = \Sigma^{-1/2} v_k^s, \quad k = 1, \dots, K.$$

Les vecteurs v_k appartiennent à l'espace $EDR E$ et sont Σ -orthonormés.

Conséquence du théorème 1.1. On peut montrer de même que les K vecteurs propres, v_1, \dots, v_K , associés aux K plus grandes valeurs propres de la matrice Σ -symétrique $\Sigma^{-1}\Gamma_T$ sont des directions EDR Σ -orthonormées et engendrent donc bien l'espace $EDR E$.

Comme on l'a vu 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 Γ_T (ou Γ_T^s) particulièrement simple. Ceci est la base de la méthode d'estimation SIR.

Remarque 1.5. La matrice de variance conditionnelle $\mathbb{V}(z | T(y))$ permet aussi de trouver les directions EDR standardisées. Pour cela, il suffit simplement d'observer l'identité de l'analyse de la variance (ANOVA) :

$$\mathbb{V}(z) = \mathbb{E}[\mathbb{V}(z | T(y))] + \mathbb{V}(\mathbb{E}[z | T(y)]),$$

soit $\mathbb{E}[\mathbb{V}(z \mid T(y))] = I_p - \Gamma_T^s$. Ainsi, après une décomposition spectrale de $\mathbb{E}[\mathbb{V}(z \mid T(y))]$, les vecteurs propres associés aux K plus petites valeurs propres sont des directions *EDR* standardisées. Des estimateurs fondés sur l'analyse spectrale d'estimateurs de $\mathbb{E}[\mathbb{V}(z \mid T(y))]$ ont été développés et étudiés par Hsing et Carrol (1992) et Zhu et Ng (1995).

1.3 Le régression inverse par tranche : Méthode SIR

Nous nous intéressons à la version échantillon de cette méthode et nous plaçons dans le cas où $y \in \mathbb{R}$. On décrit d'abord l'algorithme de la méthode, et on énonce ensuite des propriétés asymptotiques.

1.3.1 Algorithme général des méthodes SIR

D'après les résultats théoriques de la section précédente, l'espace *EDR* standardisé E^s est indépendant du choix de la transformation monotone T , seule la base de cet espace est modifiée.

En pratique, on dispose d'un échantillon $\{(x_i, y_i), i = 1, \dots, n\}$ de variables aléatoires i.i.d. provenant du modèle de référence. On désire estimer une base de l'espace *EDR* E . On est alors amené à estimer la matrice Γ_T (ou Γ_T^s).

L'idée est de choisir une fonction T de y telle que cette matrice soit facile à estimer. Ainsi, la régression inverse par tranches prend pour fonction T une discrétisation de y en H tranches distinctes : $s_1, \dots, s_h, \dots, s_H$. On appelle tranchage une telle transformation.

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

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

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

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

L'algorithme de SIR est donc le suivant :

- **Étape 1** : Standardiser les x_i en z_i par la transformation affine

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

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

- **Étape 2** : Diviser le support des y_i en H tranches s_1, \dots, s_H , et calculer \hat{p}_h la proportion empirique des y_i tombant dans la tranche s_h :

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

- **Étape 3** : Calculer les moyennes empiriques par tranche des z_i :

$$\hat{m}_h^s = \frac{1}{n\hat{p}_h} \sum_{i=1}^n z_i \mathbf{1}_{[y_i \in s_h]}, \quad h = 1, \dots, H.$$

- **Étape 4** : Construire une estimation de la matrice de covariance des moyennes par tranches :

$$\hat{\Gamma}_T^s = \sum_{h=1}^H \hat{p}_h \hat{m}_h^s \hat{m}_h^{s'}.$$

Faire la décomposition aux valeurs propres de cette matrice.

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

- **Étape 5** : Calculer les K directions *EDR* estimées :

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

Remarque 1.6. Cet algorithme est très facile à mettre en œuvre au niveau informatique.

Remarque 1.7. Ce choix du nombre de tranches H n'a pas encore été évoqué. Le nombre peut être choisi arbitrairement. 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. En effet, dans le premier cas, les matrices $\hat{\Gamma}_T^s$ et $\hat{\Gamma}_T$ sont égales à la matrice nulle, et dans le second cas, $\hat{\Gamma}_T^s = I_p$ et $\hat{\Gamma}_T = \hat{\Sigma}$. Entre ces 2 choix extrêmes, il reste une grande liberté de choix pour H . Il convient cependant de noter que plus H augmente, plus le nombre de vecteurs propres non triviaux augmente. Ainsi, il apparaît nécessaire de choisir H suffisamment grand pour retrouver toutes les directions *EDR*. En particulier, pour ne pas faire une réduction de dimension artificielle, il est indispensable que le nombre de tranches H soit supérieur à K , la dimension de réduction du modèle. L'arbitraire qui peut présider au choix du nombre et de la position des tranches peut sensiblement dégrader la qualité des estimations en présence d'échantillons de petite taille. Il est donc impératif de considérer des solutions alternatives au tranchage dans ce cadre (voir par exemple Arangon et Saracco, 1997).

En pratique, d'après des observations faites sur des simulations, le choix de H n'a pas trop d'impact sur la qualité des estimations quand on travaille avec des échantillons de taille raisonnable ($n \geq 100$). L'idée est alors de prendre un nombre de tranches suffisant sans être trop important (par exemple 10 tranches pour $100 \leq n \leq 1000$, lorsque l'on prend $K < 10$ en fonction de la taille de l'échantillon (ce qui permet d'avoir suffisamment d'individus par tranches)).

En fait, le choix de H est beaucoup moins crucial que le choix du paramètre de lissage en régression nonparamétrique. En théorie, un mauvais choix pour le paramètre de lissage en estimation fonctionnelle de la densité ou de la régression peut entraîner une vitesse de convergence très faible, alors que pour la méthode SIR, quel que soit le choix de H (pour $K < H \leq [n/2]$), on a toujours une convergence en probabilité des directions *EDR* estimées vers de vraies directions *EDR* à la vitesse $1/\sqrt{n}$ (voir la section 1.3.3).

Remarque 1.8. Le problème du choix du nombre H de tranches ayant été discuté, la question qui se pose maintenant est de savoir comment construire ces H tranches. Deux options se présentent naturellement : ou bien prendre des tranches de même amplitude, ou bien faire des tranches contenant à peu près le même nombre d'individus.

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 :

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

Il est tout de même intéressant de remarquer qu'à l'étape 4, le fait d'avoir des tranches de tailles inégales est pris en compte.

1.3.2 Détermination de la dimension K du modèle

Jusqu'à présent, on a supposé que la dimension K de réduction du modèle était connue. Cependant dans la plupart des applications, cette dimension K est inconnue et doit être par conséquent estimée à partir de l'échantillon dont on dispose. Dans cette partie, différents travaux permettant de déterminer la dimension K sont présentés. La première approche de ce problème est fondée sur les tests d'hypothèses : (Li, 1991) se place dans le cas où x est normalement distribuée, et Schott (1994) dans le cas où la distribution de x est elliptique. La seconde approche (Ferré, 1998, Liquet et Saracco, 2008)) repose sur la qualité de l'estimation des espaces de réduction effective de dimension.

i) Tests d'hypothèses Li (1991) remarque que pour récupérer vraiment avec succès de K dimensions pour la réduction, les K premières valeurs propres de Γ_T^s (ou de manière équivalente de $\Sigma^{-1/2}\Gamma_T\Sigma^{-1/2}$) doivent être significativement différentes de zéro par rapport à l'erreur d'échantillonnage. Il s'intéresse alors à la distribution asymptotique de la moyenne des $(p - K)$ plus petites valeurs propres de $\hat{\Sigma}^{-1/2}\hat{\Gamma}_T\hat{\Sigma}^{-1/2}$, notée $\hat{\lambda}_{(p-K)}$, lorsque la distribution de x est normale. Il montre que $n(p - K)\hat{\lambda}_{(p-K)}$ suit une loi du χ^2 à $(p - K)(H - K - 1)$ degrés de liberté. En utilisant ce résultat, il propose la démarche suivante : débiter avec $k = 1$, si $n(p - K)\hat{\lambda}_{(p-K)}$ est supérieur à la valeur du χ^2 correspondant (au niveau 95% par exemple), alors on peut inférer qu'il y a au moins $(k + 1)$ composantes significatives dans le modèle.

Lorsque la distribution de x est elliptique, les résultats sont plus complexes. Ce cas a été traité par Schott (1994) et Barrios et Vallila (2007).

Cependant, il convient de noter que l'utilisation de ces tests emboîtés pose des problèmes, notamment en ce qui concerne le niveau global de ceux-ci. De plus, du point de vue population, le terme d'erreur ϵ ne joue aucun rôle dans la recherche de l'espace EDR . Mais lorsque la taille de l'échantillon est faible, certaines directions estimées de l'espace EDR peuvent être perturbées par cette erreur aléatoire. Dans de telles situations, l'identification de la vraie dimension K n'est plus évidente. Dans la sous-partie suivante, nous décrivons une approche différente.

ii) Autre approche Des méthodes, initialement proposées par Ferré (1997), permettant le choix du nombre K de combinaisons linéaires à retenir en régression inverse par tranches, reposent sur la maximisation d'un critère mesurant la ressemblance entre deux sous-espaces. Pour ce faire, introduisons, pour chaque valeur de $k = 1, \dots, p$, la quantité :

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

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 $\Sigma^{-1}\Gamma_T$ (resp. $\hat{\Sigma}^{-1}\hat{\Gamma}_T$). Clairement, on voit que plus $R(k)$ est près de 1, meilleure est l'estimation du sous-espace de dimension k . Cependant, cette quantité $R(k)$ est inconnue en pratique. Ferré (1996) propose un estimateur $\hat{R}(k)$ de $R(k)$ à partir de développements limités.

Lorsque x suit une loi elliptique, pour un échantillon de taille n , et quel que soit $k < K$, Ferré montre que :

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

où v_1, \dots, v_K représentent la suite décroissante des valeurs propres non nulles de $\Sigma^{-1}\Gamma_T$ et κ est le paramètre de Kurtosis de la loi elliptique de x . Il est alors aisé d'en déduire un estimateur asymptotiquement sans biais, $\hat{R}(k)$, en posant :

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

où $\hat{v}_1, \dots, \hat{v}_K$ représentent la suite décroissante des valeurs propres de $\hat{\Sigma}^{-1}\hat{\Gamma}_T$. Les quantités intervenant dans l'expression de $\hat{R}(k)$ étant fournies par les résultats de la décomposition spectrale de $\hat{\Sigma}^{-1}\hat{\Gamma}_T$ par l'utilisateur, il est possible de calculer sans difficultés les valeurs de $\hat{R}(k)$ pour un échantillon donné. La dimension K étant inconnue a priori, il est possible de calculer $\hat{R}(k)$ pour $k = 1, \dots, p$.

On peut alors choisir la dimension à retenir comme étant celle qui correspond à un "maximum local" (le plus proche de 1) pour $\hat{R}(k)$, $\hat{R}(p)$ étant toujours égal à 1. Cette démarche est plus parcimonieuse que celle des tests emboîtés et peut s'avérer plus efficace, notamment pour des tailles d'échantillons faibles.

Ferré (1997) propose aussi une autre manière de rechercher la dimension K . Il utilise une méthode de validation croisée. L'idée de base de la validation croisée est de supprimer une (ou plus d'une) observation de l'échantillon, de calculer de valeur de la fonction de perte, et enfin de minimiser cette dernière. Dans la régression inverse par tranches "classique" (Li (1991), Duan et Li (1991)), l'échantillon est découpé en H tranches. Supprimer une observation ne va alors modifier qu'une seule tranche, les autres restant identiques, appliquer la validation croisée n'a ici que peu de sens. Cependant, la version de la régression inverse par tranches proposée par Hsing et Carroll (1992), où ils estiment $\Lambda := \mathbb{E}[\mathbb{V}(x | y)]$ en faisant des tranches (distinctes ou glissantes) d'exactly deux individus, est elle plus intéressante dans ce cadre. En effet, en supprimant deux observations de l'échantillon initial, cela revient à éliminer la totalité d'une tranche. Ainsi, en tenant compte de ces remarques, on peut mettre

en œuvre une technique de validation croisée. Soient π_k (resp. $\hat{\pi}_k$) le projecteur propre associé aux k plus petites valeurs propres de $\Sigma^{-1}\Lambda$ (resp. $\hat{\Sigma}^{-1}\hat{\Lambda}$). Soit la fonction $CV(\cdot)$ définie pour $k = 1, \dots, p$ par :

$$CV(k) = \frac{1}{n_0} \sum_{i=1}^{n_0} \text{Trace}(\hat{\pi}_k \hat{\pi}_k^{(i)}) / k,$$

où $\hat{\pi}_k^{(i)}$ est l'estimation de π_k obtenue en supprimant les $2i - 1$ et $2i$ -ièmes observations de l'échantillon dans les cas des tranches distinctes (avec n_0 égal à la partie entière de $n/2$), ou bien en supprimant les observations i et $i + 1$ (avec n_0 égal à $n - 1$).

La dimension optimale est alors celle qui maximise $CV(k)$ pour k allant de 1 à p (ou de façon équivalent qui minimise la fonction de perte $1 - CV(\cdot)$). Cependant, le temps de calcul de chaque $CV(k)$ est relativement important. Afin de réduire les temps de calculs, Ferré (1998) propose alors une approximation de $CV(k)$.

1.3.3 L'état actuel des résultats asymptotiques de SIR

Rappelons ici les notations suivantes :

Les vecteurs propres, v_1, \dots, v_K (resp. $\hat{v}_1, \dots, \hat{v}_K$) associés aux K valeurs propres non nulles, $\lambda_1, \dots, \lambda_K$ (resp. $\hat{\lambda}_1, \dots, \hat{\lambda}_K$) de la matrice $\Sigma^{-1}\Gamma_T$ (resp. $\hat{\Sigma}^{-1}\hat{\Gamma}_T$) sont des directions *EDR* (resp. *EDR* estimées). Ces vecteurs étant Σ -orthonormés (resp. $\hat{\Sigma}$ -orthonormés), ils forment une base de l'espace *EDR* E (resp. *EDR* \hat{E}). En notant $V := [v_1, \dots, v_K]$ (resp. $\hat{V} := [\hat{v}_1, \dots, \hat{v}_K]$), matrice de dimension $p \times K$, on a alors : $\text{vect}(V) = E$ (resp. $\text{vect}(\hat{V}) = \hat{E}$).

La notation "vect" est la transformation de vectorisation. Soit $M = [m_1, \dots, m_p]$ une matrice de dimensions $q \times p$ où les m_k , $k = 1, \dots, p$, 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 : $\text{vec}(M) = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_p \end{pmatrix}$.

1.3.3.1 Convergence en probabilité

Dans le cadre du modèle de référence, Li (1991) donne le théorème suivant :

Théorème 1.3. *Sous l'hypothèse (H), en supposant que les données sont échantillonnées de manière indépendante à partir du modèle (1.2.1), on a :*

Les directions EDR estimées $\hat{v}_1, \dots, \hat{v}_K$ convergent à la vitesse $\mathcal{O}_p(n^{-1/2})$ vers de vraies directions EDR.

1.3.3.2 Normalité asymptotique

Avant de donner les lois asymptotiques des éléments propres caractérisant l'espace *EDR* estimé, Saracco (1997) pose quelques hypothèses :

(A1) Les observations (x_i, y_i) , $i = 1, \dots, n$, sont échantillonnées de manière indépendante à partir du modèle (1.2.1).

(A2) Le support de y est partitionné en H tranches fixes $s_1, \dots, s_h, \dots, s_H$ telles que : $p_h > 0, (= Pr(y \in s_h))$.

(A3) $\lambda_1 > \dots > \lambda_K > 0$.

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

On définit $\hat{w} = \{\hat{\lambda}_1, \dots, \hat{\lambda}_K\}$ l'ensemble des K plus grandes valeurs propres de $\hat{\Sigma}^{-1} \hat{\Gamma}_T$, 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 $\hat{P} = \sum_{\hat{\lambda}_k \in \hat{w}} \hat{P}_{\hat{\lambda}_k}$, où $\hat{P}_{\hat{\lambda}_k} = \hat{v}_k \hat{v}_k' \hat{\Sigma}$.

Saracco (1997) donne les théorèmes de normalités asymptotiques suivants :

Les expressions des matrices de variance asymptotique indiquées dans les différents théorèmes (Δ_1, Δ_2 et Δ_3) sont données dans Saracco (1997).

i) Distribution asymptotique de la matrice d'intérêt

Théorème 1.4. *Sous les hypothèses (A1) et (A2), on a :*

$$\sqrt{n} \text{vec}(\hat{\Sigma}^{-1} \hat{\Gamma}_T - \Sigma^{-1} \Gamma_T) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Delta_1).$$

ii) Distribution asymptotique du projecteur propre sur \hat{E}

Théorème 1.5. *Sous les hypothèses (A1), (A2) et (A3), on a :*

$$\sqrt{n} \text{vec}(\hat{P} - P) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Delta_2).$$

iii) Distribution asymptotique des directions EDR estimées

Théorème 1.6. *Sous les hypothèses (A1), (A2) et (A3), on a :*

$$\sqrt{n} (\hat{b}_k - b_k) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Delta_3).$$

iv) Distribution asymptotique des estimateurs des valeurs propres non nulles

Théorème 1.7. *Sous les hypothèses (A1), (A2) et (A3), on a :*

$$\sqrt{n} (\hat{\lambda}_k - \lambda_k) \xrightarrow{\mathcal{L}} \mathcal{N}(0, [b_k' \otimes b_k' \Sigma] \Delta_1 [b_k \otimes \Sigma b_k]).$$

1.4 Quelques extensions de la méthode SIR

1.4.1 SIR multivariée

Dans cette partie, on considère des modèles semi-paramétriques dans lesquels la variable dépendante et la variable explicative sont multidimensionnelles : la variable à expliquer y est ici de dimension q , soit $y \in \mathbb{R}^q$.

Dans un premier temps, on décrit rapidement le nouveau modèle de régression semi-paramétrique de référence. On fait ensuite le lien entre la régression inverse et ce modèle de référence. Dans un troisième temps, on présente les différentes méthodes d'estimation qui en découlent. (Pour plus de détails concernant cette partie, voir Li, Aragon, Shedden et Thomas-Agnan, 2003).

Le modèle multivarié de référence

Il s'agit d'un modèle de régression semi-paramétrique de la forme suivante :

$$y_j = f_j(\theta'_1 x, \dots, \theta'_K x, \varepsilon_j), \quad \text{pour } j = 1, \dots, q \quad (1.4.1)$$

où :

- (i) $y = (y_1, \dots, y_j, \dots, y_q)'$ est une variable dépendante de dimension q ;
- (ii) 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 ;
- (iii) $\theta_1, \dots, \theta_K$ sont des vecteurs paramètres inconnus de dimension p , linéairement indépendants ;
- (iv) les termes ε_j sont des erreurs aléatoires indépendantes de x ;
- (v) f_j , $j = 1, \dots, q$ sont les paramètres fonctionnels à valeur dans \mathbb{R} , inconnus et arbitraires.

Tout comme dans le cas du modèle univarié, on s'intéresse à la partie paramétrique de ce modèle. L'objectif des méthodes présentées dans cette partie est l'estimation d'une base de l'espace EDR E . En effet, pour les mêmes raisons que celles décrites dans le cas d'un modèle univarié, si on ne donne aucune contrainte de normalisation et d'orthogonalité aux θ_k , $k = 1 \dots, K$, ces derniers ne sont pas identifiables, et on ne peut au mieux qu'estimer une base de l'espace engendré par les θ_k .

Condition fondamentale

Pour que le courbe de régression inverse possède des propriétés permettant la caractérisation de l'espace EDR E , il est nécessaire de rajouter à ce modèle la même que celle mentionnée condition fondamentale **(H)** ou **(H')** sur la distribution de x dans le cadre du modèle univarié.

Régression inverse et modèle multivarié

Li, Aragon, Shedden et Thomas-Agnan (2003) ont proposé différentes méthodes d'estimation de l'espace EDR . Ces méthodes reposent sur des philosophies légèrement différentes (tout en utilisant toujours la propriété de caractérisation de l'espace EDR par la courbe de régression inverse). On va présenter ici uniquement les aspects sur population sous-jacents aux méthodes.

La première méthode est fondée sur une utilisation directe de la propriété de la fonction de régression inverse. La seconde est une méthode qui agrège diverses informations provenant des composantes de y , et enfin la dernière méthode repose sur l'observation d'une certaine dualité entre les directions EDR et les directions des variables les plus prédictives (que l'on définira).

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

Considérons une fonction mesurable G de \mathbb{R}^q dans \mathbb{R}^d (avec $1 \leq d \leq q$). $G(y)$ satisfait un modèle ayant les mêmes indices $\theta'_1 x, \dots, \theta'_K x$ que le modèle de régression

semi-paramétrique multivarié de référence. Ainsi, de même que dans le cas univarié le théorème de caractérisation de l'espace EDR est vérifié, et ceci quel que soit le choix de la fonction G .

Différents choix intéressants sont possibles pour cette fonction G , et ils sont suggérés en fonction du problème particulier que l'on est amené à traiter. Par exemple, on peut prendre :

- $G(y) = y$;
- $G(y) = y_j$ une composante particulière de y ;
- $G(y) = c_j$ la première composante principale de y ;
- $G(y) = \bar{y}$, moyenne sur les composantes de y ;
- $G(y) = \text{médiane de } y \text{ sur les composantes de } y$.

Il est intéressant de noter que l'on résume ici l'information sur la variable à expliquer (composantes principales, médiane, moyenne, ...) sans tenir compte de la variable explicative x , ce qui n'est pas le cas des deux autres types de méthodes. L'étape de tranchage porte sur $G(y)$ et non pas sur y directement. Les méthodes d'estimation appartenant à cette famille sont les méthodes du Complete Slicing (cas $G(y) = y$) et du Marginal Slicing (cas des autres transformations G).

ii) Méthodes “agrégante” : “Pooled Marginal Slicing”.

Cette méthode repose sur l'observation suivante. Étant données deux fonctions à valeurs réelles G_1 et G_2 de y ainsi que leurs matrices de covariances

$$\Gamma_1^s = \mathbb{V}(\mathbb{E}[z | G_1(y)]) \quad \text{et} \quad \Gamma_2^s = \mathbb{V}(\mathbb{E}[z | G_2(y)]),$$

on peut construire une combinaison convexe Γ^P de Γ_1^s et Γ_2^s . Cette matrice $\Gamma^P = w_1\Gamma_1^s + w_2\Gamma_2^s$ est telle que les K vecteurs propres associés aux K plus grandes valeurs propres (non nulles) de Γ^P sont des directions EDR standardisées.

Ceci se généralise sans aucun problème au cas où l'on considère plus de deux fonctions G_j , et en particulier au cas où l'on travaille avec q fonctions G_j qui sont telles que : pour $j = 1, \dots, q$, $G_j(y) = T_j(y_j)$ est une discrétisation de y_j , la j ème composante de y , en H_j tranches.

La méthode du Pooled Marginal Slicing utilise ces q fonctions. On définit alors la matrice agrégée :

$$\Gamma^P = \sum_{j=1}^q w_j \Gamma_j^s$$

où les w_j sont une suite de poids positifs tels que $\sum_{j=1}^q w_j = 1$, et $\Gamma_j^s = \mathbb{V}(\mathbb{E}[z | T_j(y_j)])$. Ainsi, les K vecteurs propres associés aux K plus grandes valeurs propres (non nulles) de cette matrice sont des directions EDR standardisées.

iii) Méthodes de l’“Alternating SIR” : directions EDR et variables les plus prévisibles.

Li, Aragon, Shedden et Thomas-Agnan (2003) proposent une théorie concernant les variables les plus prévisibles (Most Predictable) en s'inspirant des travaux de Hotteling (1935, 1936). Leur théorie (contrairement à celle d'Hotteling) 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 .

Cette méthode repose donc sur l'observation de cette dualité entre les directions *EDR* et les directions des variables les plus prévisibles. Commençons tout d'abord par introduire les MP-variables, puis nous précisons cette dualité et les conséquences intéressantes que cela entraîne.

Pour toutes combinaisons linéaires $\theta'y$ de y sachant $x = x_0$, la meilleure prédiction non linéaire, au sens de l'erreur quadratique moyenne, est : $\mathbb{E}[\theta'y \mid x = x_0]$.

L'erreur quadratique moyenne de prédiction associée est alors égale à : $\mathbb{V}(\theta'y \mid x = x_0)$.

La MP-variable $\theta'_1 y$ est celle qui minimise le ratio $\frac{\mathbb{E}[\mathbb{V}(\theta'y \mid x)]}{\mathbb{V}(\theta'y)}$.

Cela conduit donc à une décomposition aux valeurs propres identique à celle de *SIR* exceptée le fait que les rôles de x et y ont été échangés. Cette dualité entre les indices *EDR* $\theta'x$ et les indices MP $\theta'y$ sont à la base de la méthode "Alternating SIR". Dans son algorithme, les étapes de recherche des indices *EDR* alternent avec les étapes de recherche des indices MP (avec les tranchages partant respectivement sur les indices $\theta'y$ et $\theta'x$).

Li, Aragon, Shedden et Thomas-Agnan (2003) ont montré qu'en prenant comme étape initiale pour l'algorithme les résultats de l'analyse canonique entre x et y , sa convergence est garantie en un nombre fini d'étapes.

1.4.2 *SIR_{II}*, *SIR _{α}* et *SAVE*

Il est important de remarquer que sous les hypothèses **(H)** ou **(H')**, la méthode *SIR* garantit seulement les calculs des vraies espérances conditionnelles, de la vraie matrice Γ_T et de sa décomposition spectrale. Elle garantit aussi le fait que les vecteurs propres associés à des valeurs propres non nulles sont des directions *EDR*, mais ces vecteurs propres ne vont pas nécessairement engendrer l'espace *EDR*. Un cas pathologique a été identifié par Cook et Weisberg (1991) : il s'agit d'un modèle du type $y = f(\theta'x) + \varepsilon$ où la distribution de $\theta'x$ est symétrique par rapport à $\theta'\mu$ et pour lequel la fonction f est symétrique par rapport à l'axe vertical d'abscisse $\theta'\mu$. Par exemple, pour le modèle suivant $y = (x'\theta)^2 + \varepsilon$ où $x \sim \mathcal{N}_p(0, I_p)$ et $\varepsilon \sim \mathcal{N}_p(0, 1)$, la méthode *SIR* va échouer dans la recherche de la direction de θ .

Pour éviter de tels problèmes, une approche est d'explorer des moments conditionnels de x sachant $T(y)$ d'ordre supérieur. Différentes méthodes ont été proposées sur ce thème.

Cook et Weisberg (1991) se sont intéressés aux moments conditionnels d'ordre 2 dans leur méthode *SAVE* (Sliced Average Variance Estimate). Ils proposent de rechercher les vecteurs propres correspondant aux K plus grandes valeurs propres de la matrice $M_{SAVE} = \sum_{h=1}^H (I_p - \mathbb{V}(z \mid y \in s_h))$. Les K vecteurs propres ainsi obtenus sont des directions *EDR* standardisées.

Li (1991) a lui aussi travaillé dans ce sens et il a proposé la méthode "*SIR_{II}*" qui prend en compte la "courbe" de la covariance conditionnelle : $y \mapsto \mathbb{V}(z \mid T(y))$. Cette méthode se ramène à la décomposition spectrale de la matrice $M_{SIR_{II}} = \mathbb{E}[(\mathbb{V}(z \mid T(y)) - \mathbb{E}[\mathbb{V}(z \mid T(y))])^2]$. Les K vecteurs propres associés aux K plus grandes valeurs propres de $M_{SIR_{II}}$ sont des directions *EDR* standardisées.

Une autre méthode (Li (1991)), notée " SIR_α ", combine les approches SIR_I et SIR_{II} via la décomposition spectrale de la matrice

$$M_{SIR_\alpha} = (1 - \alpha)(M_{SIR_I})^2 + \alpha M_{SIR_{II}}, \quad \text{où } M_{SIR_I} = \mathbb{V}(\mathbb{E}[z \mid T(y)]).$$

Le choix de $\alpha \in [0, 1]$ étant à adapter selon les différentes situations. On peut remarquer que pour $\alpha = 0.5$, que l'on a $M_{SIR_\alpha} = 1/2 M_{SAVE}$.

Gannoun et Saracco (2003) ont proposé deux critères de type validation croisée pour déterminer une valeur optimale de α pour la méthode SIR_α .

Liquet et Saracco (2008) ont présenté une approche de type bootstrap permettant de déterminer simultanément le paramètre α de la méthode SIR_α et la dimension K du modèle.

Barreda et al (2007) ont étendu ces méthodes multivariés en utilisant l'approche SIR_α .

1.4.3 Alternatives au tranchage

L'idée de la méthode SIR était de choisir une transformation T telle que Γ_T soit facile à estimer. Dans la suite, nous présentons quelques méthodes alternatives reposant aussi sur cette philosophie.

1.4.3.1 Méthode du Pooled-Slicing

Dans le cadre de petits échantillons, il apparaît intéressant de rechercher des méthodes alternatives à un découpage en tranches particulier fixé par l'utilisateur. En effet, l'estimateur des directions EDR peut être, dans ce cas, sensible au nombre et à la position des tranches. Dans le but d'améliorer les estimations et de supprimer le choix du tranchage particulier demandé à l'utilisateur, un remède est de combiner les résultats provenant de plusieurs tranchages différents. Ainsi, la méthode Pooled-Slicing proposée par Aragon et Saracco (1997) repose sur l'idée de balayer un certain nombre de tranchages différents et d'agréger ensuite le tout.

Dans la version sur population de cette méthode, on considère D discrétisations de y : T_d pour $d = 1, \dots, D$. Notons chacune de ces discrétisations : $\Gamma_d = \mathbb{V}(\mathbb{E}[z \mid T_d(y)])$.

Soit $\{w_d\}_{d=1, \dots, D}$ une suite de poids positifs. On définit alors la matrice agrégée suivante : $\Gamma^P = \sum_{d=1}^D w_d \Gamma_d$.

Les vecteurs propres associés aux valeurs propres non nulles de la matrice Γ^P sont des directions EDR standardisées.

La convergence en probabilité des directions EDR standardisées estimées vers de vraies directions EDR standardisées à la vitesse $\mathcal{O}_p(n^{-1/2})$ a été obtenue par Aragon et Saracco (1997).

1.4.3.2 Estimation de l'espérance de la variance conditionnelle

Dans le cadre de petits échantillons, il apparaît intéressant de rechercher des méthodes alternatives à un découpage en tranches particulier fixé par l'utilisateur. En effet, Hsing et Carroll (1992) et Zhu et Ng (1995) proposent une estimation de la matrice

Λ , et ils en déduisent l'estimation des directions *EDR* standardisées, définies comme étant les K vecteurs propres associés aux K plus petites valeurs propres de Λ_T .

Les estimateurs de Λ_T présentés par Hsing et Carroll (1992) sont basés sur les statistiques d'ordre et les valeurs concomitantes. Soient $y_{(1)} \leq \dots \leq y_{(N)}$ les statistiques d'ordre de y_i , $i = 1, \dots, N$. Notons $z_{(i)*}$ la valeur concomitante de z associée à $y_{(i)}$ (voir Yang (1977)). Ils sont basés sur des “tranches” distinctes ou glissantes de 2 individus :

$$\hat{\Lambda}_T = \frac{1}{n} \sum_{i=1}^{[n/2]} (z_{(2i)*} - z_{(2i-1)*})(z_{(2i)*} z_{(2i-1)*})' \quad (\text{Tranches distinctes})$$

où $[n/2]$ est la partie entière de $n/2$,

$$\tilde{\Lambda}_T = \frac{1}{2n} \sum_{i=2}^n (z_{(i)*} - z_{(i-1)*})(z_{(i)*} z_{(i-1)*})' \quad (\text{Tranches glissantes}).$$

Ici le choix des tranches est fixé. Par contre, les estimateurs de Zhu et Ng (1995) utilisent des tranches contenant c individus et le choix de c reste à faire (on retombe alors sur un problème de choix de tranche). Hsing et Carroll (1992) ont montré que, sous certaines hypothèses techniques lourdes,

$$n^{\frac{1}{2}}(\hat{\Lambda}_T - \Lambda_T) = \mathcal{O}_p(1) \quad \text{et} \quad n^{\frac{1}{2}}(\tilde{\Lambda}_T - \Lambda_T) = \mathcal{O}_p(1).$$

De même, Zhu et Ng (1995) ont établi la convergence à la vitesse $\mathcal{O}_p(n^{-1/2})$ de leur estimateur de Λ_T .

1.4.3.3 Régression inverse par lissage

Cette méthode alternative ne nécessite aucun découpage en tranches. Elle repose sur l'utilisation de méthodes de régression fonctionnelle (par exemple celle des noyaux ou des splines de lissage) pour estimer la matrice $\Gamma := \mathbb{V}(\mathbb{E}[z | y])$ lorsque y est une variable continue. Aragon et Saracco (1997) proposent d'estimer cette matrice composante par composante.

Soit z^j la j ème composante de z . Estimer la courbe de régression inverse standardisée revient en fait à estimer les p fonctions de régression $y \mapsto \mathbb{E}[z^j | y]$ pour $j = 1, \dots, p$. Notons $v^j(y) := \mathbb{E}[z^j | Y = y]$. Vu que $\mathbb{E}[z] = 0$, on a la relation :

$$\forall j, l = 1, \dots, p, \quad \Gamma_{j,l} := \text{cov}(v^j(y), v^l(y)) = \mathbb{E}[\mathbb{E}[z^j | y] \mathbb{E}[z^l | u]] = \int v^j(y) v^l(y) f(y) dy.$$

Si on remplace dans cette intégrale les fonctions v^j et f par leur estimateur à noyau correspondant, on obtient un estimateur des composantes de Γ de type “intégrale” :

$$\hat{\Gamma}_{j,l} = \int \hat{v}^j(y) \hat{v}^l(y) \hat{f}(y) dy.$$

Alternativement, un estimateur de type “somme” peut être utilisé :

$$\tilde{\Gamma}_{j,l} = \frac{1}{n} \sum_{i=1}^n \hat{v}^j(y_i) \hat{v}^l(y_i) \mathbf{1}_{[\hat{f}(y_i) > b_n]},$$

où b_n est un réel positif dépendant de la taille n de l'échantillon, il est appelé paramètre d'élagage ("trimming parameter").

Sous les hypothèses techniques, Aragon et Saracco (1997) ont montré que : $\hat{\Gamma}$ converge vers Γ en moyenne quadratique et $\tilde{\Gamma}$ converge vers Γ en probabilité.

Zhu et Yu (2007) ont proposé et étudié un estimateur de Γ de type somme (sans paramètres d'élagage) dans lequel les fonctions γ_i^j sont estimées par des splines de lissage.

1.5 Applications de la méthode SIR

Bernard-Michel et al (2008) et (2009a) ont proposé de régulariser la méthode SIR en introduisant un a priori Gaussien. Lorsque la dimension p est grande, la matrice de variance-covariance $\hat{\Sigma}$ 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 $\hat{\Sigma}$). 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 du conditionnement de la matrice de variance-covariance des variables explicatives.

La méthode SIR régularisée est utilisée pour estimer les propriétés physique du sol martien à partir d'images hyperspectrales, voir Bernard-Michel *et al* (2009b).

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 Gannoun *et al* (2002).

Chapitre 2

Approche récursive pour la méthode SIR

L'intérêt des méthodes récursives est de prendre en compte l'arrivée temporelle des informations et d'affiner ainsi au fil du temps les estimations. L'idée est d'utiliser les estimations calculées sur la base de données initiales et de les remettre à jour en tenant uniquement compte des nouvelles données arrivant dans la base. Le gain en terme de temps de calculs peut être très intéressant et les applications d'une telle approche sont nombreuses.

L'approche développée dans ce chapitre a donné lieu à l'article intitulé "Estimation récursive en régression inverse par tranches (sliced inverse regression)" écrit en collaboration avec Jérôme Saracco. Cet article est paru dans le *Journal de la Société Française de Statistique* et est disponible à la fin de ce chapitre.

2.1 Principe

Dans ce chapitre, nous nous intéressons au modèle de régression semi-paramétrique proposé par Duan et Li (1991) et décrit au chapitre 1. Lorsque la variable à expliquer y est à valeurs dans \mathbb{R} et la covariable x appartient à \mathbb{R}^p , ce modèle s'écrit sous la forme :

$$y = f(\theta'x, \varepsilon), \quad (2.1.1)$$

où :

- (i) le paramètre θ est un vecteur inconnu de \mathbb{R}^p ;
- (ii) le bruit ε est un terme d'erreur aléatoire indépendant de x , aucune hypothèse n'est faite sur la distribution de ε ;
- (iii) la fonction de lien f est un paramètre fonctionnel à valeur dans \mathbb{R} , inconnu et arbitraire.

Dans le cadre de ce modèle, on a vu que le paramètre θ n'est pas totalement identifiable, seule la direction de θ l'est. On parle de direction EDR (comme "effective dimension reduction").

Nous proposons une approche récursive de la méthode SIR pour l'estimateur de la direction θ dans le cadre de ce modèle semi-paramétrique sans avoir à estimer le paramètre fonctionnel f ni à spécifier la loi de l'erreur ε . Nous proposons dans un premier temps une forme récursive pour l'estimateur de la matrice d'intérêt. Dans le cas particulier

où l'on considère $H = 2$ tranches, il est possible d'obtenir une expression analytique de l'estimateur de la direction de θ . Nous donnons alors une forme récursive pour cet estimateur. Puis nous montrons des propriétés asymptotiques pour cet estimateur (convergence presque sûre avec la vitesse correspondante, normalité asymptotique). Nous illustrons aussi sur des simulations le bon comportement numérique de l'approche récursive proposée. Un avantage majeur de l'utilisation de la forme récursive est que les temps de calculs des estimateurs sont beaucoup plus courts que ceux obtenus avec la forme non récursive, en particulier lorsque la dimension de x est grande.

2.2 Estimation récursive du paramètre θ

2.2.1 Méthode SIR

Nous avons vu au chapitre 1 que la matrice $\Gamma = \mathbb{V}(\mathbb{E}[x | \mathbb{T}(y)])$ est dégénérée dans toute direction Σ -orthogonale à θ . Ainsi le vecteur propre θ associé à la valeur propre non nulle de $\Sigma^{-1}\Gamma$ est colinéaire à θ . Ce vecteur propre $\tilde{\theta}$ est donc une direction EDR.

Etude du cas où $H = 2$ permettant d'obtenir une expression analytique de $\tilde{\theta}$ et de son estimateur $\tilde{\theta}_n$.

Nous nous intéresserons au cas particulier où l'on ne considère que deux tranches notées s_1 et s_2 . Lorsque $H = 2$, la matrice Γ s'écrit :

$$\Gamma = p_1 z_1 z_1' + p_2 z_2 z_2', \quad (2.2.1)$$

avec $z_h = m_h - \mu$ (pour $h = 1, 2$) où m_h et p_h sont définis au chapitre 1 et $\mu = \mathbb{E}[x]$. Nous pouvons alors montrer que la valeur propre non nulle λ de $\Sigma^{-1}\Gamma$ et le vecteur propre $\tilde{\theta}$ correspondant (colinéaire à θ) s'écrivent sous la forme :

$$\lambda = \frac{p_1}{p_2} z_1' \Sigma^{-1} z_1 \quad \text{et} \quad \tilde{\theta} = \Sigma^{-1}(z_1 - z_2). \quad (2.2.2)$$

Les calculs de $\tilde{\theta}$ et de λ sont détaillés en Annexe A de l'article. Nous pouvons ainsi déduire des relations (2.2.2) des estimateurs λ_n et $\tilde{\theta}_n$ de λ et $\tilde{\theta}$:

$$\lambda_n = \frac{p_{1,n}}{p_{2,n}} z_{1,n}' \Sigma_n^{-1} z_{1,n} \quad \text{et} \quad \tilde{\theta}_n = \Sigma_n^{-1}(z_{1,n} - z_{2,n}), \quad (2.2.3)$$

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

2.2.2 Estimateur récursif de la matrice d'intérêt $\Sigma^{-1}\Gamma$

Lorsque l'on s'intéresse à la forme non récursive des estimateurs, nous disposons d'un échantillon d'observations $\{(x_i, y_i), i = 1, \dots, n\}$ de variables aléatoires (x, y) indépendantes et identiquement distribuées issues du modèle (2.1.1). 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) . Nous supposons ici que la nouvelle observation (x_n, y_n) est telle que $y_n \in s_{h^*}$ avec $h^* = 1$ ou 2 .

Nous avons obtenu la forme récursive de l'estimateur $\Sigma_n^{-1}\Gamma_n$ de la matrice d'intérêt.

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

en posant

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

où Φ_n , Σ_n^{-1} , Γ_n , ρ_n et $p_{h,n}$ sont donnés dans l'article.

2.2.3 Estimateur récursif de la direction de θ

Nous nous plaçons maintenant dans le cas où $H = 2$. Dans ce cas, nous avons donné en (2.2.3) l'expression analytique des formes non récursives des estimateurs λ_n et $\tilde{\theta}_n$. Nous allons maintenant en déduire leurs formes récursives. Nous supposons encore ici que la nouvelle observation (x_n, y_n) est telle que $y_n \in s_{h^*}$ avec $h^* = 1$ ou 2 . Nous obtenons la forme récursive de l'estimateur $\tilde{\theta}_n$ de $\tilde{\theta}$:

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

où $\Phi_{h^*,n}$ et $n_{h,n}$ sont donnés dans l'article.

Nous donnons aussi la forme récursive de l'estimateur λ_n de λ dans l'article.

2.3 Résultats asymptotiques

Nous nous focalisons dans cette section sur des résultats de convergence de l'estimateur $\tilde{\theta}_n$ de la direction EDR dans le cas où $H = 2$. Les hypothèses suivantes sont nécessaires pour établir la convergence presque sûre et la normalité asymptotique de l'estimateur proposé.

- (A1) Les observations (x_i, y_i) , $i = 1, \dots, n$, sont échantillonnées de manière indépendante à partir du modèle (2.1.1).
- (A2) Le support de y est partitionné en deux tranches fixes s_1 et s_2 telles que $p_h \neq 0$ pour $h = 1, 2$.

En utilisant la loi du logarithme itéré de Hartman-Wintner et appliquant l'équation de Riccati pour la matrice inverse, nous montrons le théorème suivant.

Théorème 2.1. *Sous les hypothèses (H), (A1) et (A2), nous avons*

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

En utilisant le Théorème Central Limite, la méthode Delta ainsi que les résultats asymptotiques de SIR décrits dans Saracco (1997), nous montrons le théorème suivant.

Théorème 2.2. *Sous les hypothèses (H), (A1) et (A2), nous avons :*

$$\sqrt{n}(\tilde{\theta}_n - \tilde{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma^{-1} \Delta_3 \Sigma^{-1}), \quad (2.3.1)$$

où l'expression de la matrice Δ_3 est donnée dans l'article.

Il serait possible d'obtenir de manière similaire des résultats de convergence de λ_n vers λ . Les démonstrations de ces deux théorèmes figurent respectivement en Annexes B et C de l'article.

2.4 Quelques résultats de simulations

Nous avons fait une étude sur des données simulées à l'aide de deux modèles de régression : un modèle homoscédastique et un modèle hétéroscédastique ont été considérés, ces deux modèles sont classiquement utilisés dans les simulations décrites dans la littérature sur la méthode SIR. Cette étude a montré le bon comportement de l'estimateur proposé. Les résultats nous montrent que plus le nombre d'observations est important, plus la qualité d'estimation est bonne. De même, plus la structure de la matrice de variances-covariances Σ est complexe, plus il est nécessaire de disposer d'un échantillon de taille raisonnable pour estimer convenablement la direction de θ . Les résultats illustrent bien la normalité asymptotique de l'estimateur récursif proposé.

Les simulations ont été faites à l'aide du logiciel R.

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

Dans la TABLE 1 de l'article, nous avons mesuré les temps de calculs des estimateurs $\tilde{\theta}_n$ de la direction de θ pour n allant de N_0 à N pour les deux approches SIR récursive (nommée SIR récursif) et SIR non récursive (nommée SIR classique) (basées sur $H = 2$ tranches). Nous constatons que l'approche SIR récursive est bien plus rapide en moyen pour calculer les $(N - N_0)$ estimateurs $\tilde{\theta}_n$ que l'approche SIR non récursive. Nous avons donné en (2.2.3) une expression analytique de la forme non récursive de l'estimateur $\tilde{\theta}_n$ (où nous considérons que $H = 2$ tranches). Nous avons mesuré les temps de calculs des estimateurs $\tilde{\theta}_n$ pour n allant de N_0 à N pour cet estimateur (nommé SIR 2 tranches). Pour comparer des temps de calculs entre les approches SIR classique, SIR récursive et SIR non récursive (basées sur $H = 2$ tranches), nous fournissons ici un nouveau tableau qui n'est pas présent dans l'article, voir TAB. 2.1.

		$p = 5$	$p = 20$	$p = 40$
$N = 300$	SIR classique	0.99 (0.008)	1.97 (0.014)	4.10 (0.014)
	SIR 2 tranches	0.42 (0.004)	1.05 (0.010)	1.93 (0.009)
	SIR récursif	0.11(0.001)	0.13 (0.001)	0.17 (0.004)
$N = 600$	SIR classique	2.11 (0.012)	4.57 (0.013)	9.81 (0.020)
	SIR 2 tranches	0.95 (0.009)	2.64 (0.012)	5.10 (0.009)
	SIR récursif	0.22 (0.001)	0.26 (0.001)	0.35 (0.006)
$N = 900$	SIR classique	3.35 (0.019)	7.66 (0.019)	16.76 (0.051)
	SIR 2 tranches	1.63 (0.032)	4.71 (0.005)	9.39 (0.043)
	SIR récursif	0.33 (0.002)	0.39 (0.007)	0.54 (0.009)

TAB. 2.1: 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 ou non : moyennes et écarts-types entre parenthèses calculés sur $\mathcal{B} = 500$ répliques d'échantillons avec différentes valeurs de p

Nous voyons clairement que l'approche SIR 2 tranches non récursive est environ 2 fois plus rapide que l'approche SIR classique. L'approche SIR récursive est plus de 4 fois plus rapide que SIR 2 tranches, lorsque $p = 5$. 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 non récursive (SIR classique et SIR 2 tranches).

Nous obtenons alors un avantage majeur de cette approche récursive : sa rapidité par rapport aux méthodes SIR non récursive et SIR classique. Cet avantage est d'autant plus important que la dimension p de x est grande. Remarquons 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 non récursive et SIR classique.

Remarque : nous avons aussi obtenu la forme récursive de la matrice d'intérêt $\Sigma_n^{-1}\Gamma_n$ dans la section 2.2.2 de ce chapitre. Nous ne l'avons pas utilisée ici pour comparer des temps de calculs. Cependant, au chapitre suivant nous présentons des résultats de simulations une comparaison des temps de calculs entre les approches SIR classique (nommée SIR), SIR récursive (nommée recursive SIR) et deux nouvelles approches que l'on a développées dans l'article "Recursive and non recursive version for SIR and SIRoneslice (a new one slice-based SIR approach)" et qui fait l'objet du chapitre 3 de cette première partie. Notons enfin que dans le chapitre suivant, on considère un nombre H quelconque de tranches.

Estimation récursive en régression inverse par tranches (sliced inverse regression)

Title: Recursive estimation for sliced inverse regression

Thi Mong Ngoc Nguyen^{1,2} et Jérôme Saracco^{1,2,3}

Résumé :

L'intérêt des méthodes récursives est de prendre en compte l'arrivée temporelle des informations et d'affiner ainsi au fil du temps les estimations. L'idée est d'utiliser les estimations calculées sur la base de données initiales et de les remettre à jour en tenant uniquement compte des nouvelles données arrivant dans la base. Le gain en terme de temps de calcul peut être très intéressant et les applications d'une telle approche sont nombreuses. Dans cet article, nous nous intéressons à la méthode SIR (sliced inverse regression, que l'on peut traduire par régression inverse par tranches) qui permet d'estimer le paramètre θ dans un modèle semi-paramétrique de régression du type $y = f(x'\theta, \varepsilon)$ sans avoir à estimer le paramètre fonctionnel f ni à spécifier la loi de l'erreur ε . Dans le cas particulier où l'on considère $H = 2$ tranches, il est possible d'obtenir une expression analytique de l'estimateur de la direction de θ . Nous proposons dans cet article une forme récursive pour cet estimateur. Nous donnons des propriétés asymptotiques de cet estimateur (convergence presque sûre et normalité asymptotique). Nous illustrons aussi sur des simulations le bon comportement numérique de l'approche récursive proposée. Un avantage majeur de l'utilisation de la forme récursive est que les temps de calculs des estimateurs sont beaucoup plus courts que ceux obtenus avec la forme non récursive, en particulier lorsque la dimension de x est grande.

Abstract: In this paper, we study a recursive estimation procedure for sliced inverse regression. When the number H of slices is equal to two, we obtain an analytic expression of the estimator of the direction of the parameter θ in the semiparametric regression model $y = f(x'\theta, \varepsilon)$, which does not require the estimation of the link function f . We propose a recursive estimation procedure for this estimator. We establish some asymptotic properties of the estimator. A simulation study illustrates the good numerical behavior of the estimator. The recursive approach has the advantage to be computationally faster than the non recursive one.

Mots-clés : estimation récursive, modèle semi-paramétrique, Sliced Inverse Regression (SIR).

Keywords: recursive estimation, semiparametric regression model, Sliced Inverse Regression (SIR).

Classification AMS 2000 : primaire 62H12, 62J99 ; secondaire 62L12

¹ Université Bordeaux I, Institut de Mathématiques de Bordeaux, UMR CNRS 5251
351 cours de la libération, 33405 Talence Cedex, France.

E-mail : Thi.Mong.Ngoc.Nguyen@math.u-bordeaux1.fr ; Jerome.Saracco@math.u-bordeaux1.fr

² Equipe CQFD, INRIA Bordeaux Sud-Ouest, France.

³ Université Montesquieu-Bordeaux 4, GREThA, UMR CNRS 5113

1. Introduction

Les modèles de régression sont très utiles pour modéliser la liaison 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 omniprésentes : les modèles paramétriques et les modèles non paramétriques. Ces deux types de modèles sont caractérisés par des avantages et des défauts spécifiques. Les modèles paramétriques permettent une interprétation claire de l'impact de la variable explicative sur la variable à expliquer. Cependant, le choix d'un bon modèle paramétrique au vu des données n'est pas toujours évident. Ainsi le modèle paramétrique choisi peut ne pas être en adéquation avec les données et peut donc parfois être très "éloigné" de la réalité des données. En conséquence les conclusions en découlant peuvent alors être erronées. Les modèles non paramétriques ont été proposés afin de surmonter ce problème de sélection du bon modèle paramétrique. Ils offrent davantage de flexibilité car aucune hypothèse paramétrique n'est imposée dans le modèle, seules des hypothèses de régularité de la fonction de lien entre x et y sont imposées. Cependant il faut estimer la fonction de lien le plus souvent au moyen de procédures de calculs intensifs (en particulier en ce qui concerne la recherche des paramètres de lissage), ce qui est beaucoup plus lourd en temps de calcul. De plus, l'interprétation de la fonction de lien n'est pas toujours évidente.

Les modèles semi-paramétriques ont 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. Ces modèles dépendent généralement d'un paramètre de dimension fini, noté θ , ainsi que d'une fonction de lien f à estimer. Dans cet article, nous nous intéressons au modèle semi-paramétrique de régression proposé par Duan et Li [3] lorsque la variable à expliquer y est à valeurs dans \mathbb{R} et la covariable x appartient à \mathbb{R}^p :

$$y = f(\theta'x, \varepsilon), \quad (1.1)$$

où :

- (i) le paramètre θ est un vecteur inconnu de \mathbb{R}^p ;
- (ii) le bruit ε est un terme d'erreur aléatoire indépendant de x , aucune hypothèse n'est faite sur la distribution de ε ;
- (iii) la fonction de lien f est un paramètre fonctionnel à valeur dans \mathbb{R} , inconnu et arbitraire.

Dans le cadre de ce modèle, le paramètre θ n'est pas totalement identifiable, seule la direction de θ est identifiable. On parlera de direction EDR (comme "effective dimension reduction"). Notons que Li [7] a généralisé le modèle (1.1) en considérant une fonction f avec un nombre fini $d < p$ d'arguments $\theta_1'x, \dots, \theta_d'x$. Dans ce modèle multi-indices, seul le sous-espace linéaire de \mathbb{R}^p engendré par les vecteurs $\theta_1, \dots, \theta_d$ est identifiable, on parle alors d'espace EDR.

Duan et Li [3] et Li [7] ont proposé une méthode permettant d'estimer la direction EDR (ou l'espace EDR) ne nécessitant pas l'estimation de la fonction de lien f du modèle semi-paramétrique (1.1) : leur méthode s'appelle SIR, acronyme de "Sliced Inverse Regression" que l'on peut traduire par régression inverse par tranches. Les techniques d'estimation développées jusqu'à présent pour la méthode SIR ne sont pas récursives. D'une manière générale, l'avantage des méthodes récursives est de prendre en compte l'arrivée temporelle des informations et d'affiner ainsi au fil du temps les algorithmes d'estimation mis en œuvre. Un intérêt majeur de ces méthodes

est qu'il n'est pas nécessaire de relancer le calcul de l'estimateur sur la totalité des données à chaque fois que la base de données est complétée par de nouvelles observations. L'idée est ici d'utiliser les estimations calculées sur la base de données initiales et de les remettre à jour en tenant uniquement compte des nouvelles données arrivant dans la base. Le gain en terme de temps de calcul peut être très intéressant et les applications d'une telle approche sont nombreuses, le lecteur pourra se reporter par exemple aux ouvrages de Dufflo [4], [5] pour une présentation de méthodes récursives aléatoires.

Dans cet article, nous proposons une procédure d'estimation récursive de la direction de θ qui ne nécessite pas l'estimation du paramètre fonctionnel f du modèle semi-paramétrique (1.1) en adaptant la méthode SIR au cadre récursif. Dans la section 2, nous rappelons les fondements de la méthode SIR, puis nous donnons une expression analytique de l'estimateur de la direction de θ dans le cas particulier où l'on considère $H = 2$ tranches (pour construire l'estimateur). Nous proposons alors une forme récursive pour cet estimateur. Nous donnons ensuite à la section 3 quelques propriétés asymptotiques de l'estimateur proposé. Des simulations numériques illustrant le bon comportement de la méthode sont décrites à la section 4. Elles montrent que les temps de calculs obtenus avec la forme récursive de l'estimateur sont bien plus courts que ceux obtenus avec la forme non récursive. Enfin, pour conclure, nous donnons à la section 5 quelques remarques finales et extensions possibles de ce travail. Les démonstrations et développements théoriques sont donnés en Annexes.

2. Estimation récursive du paramètre θ

Nous présentons tout d'abord brièvement la méthode SIR et nous rappelons la définition de l'estimateur de la direction EDR (pour un nombre $H > 1$ quelconque de tranches) comme solution d'un problème aux valeurs propres d'une certaine matrice d'intérêt. Nous précisons ensuite une forme analytique de cet estimateur lorsque $H = 2$. Nous proposons alors dans un premier temps une forme récursive pour l'estimateur de la matrice d'intérêt. Puis nous donnons une forme récursive de l'estimateur de la direction EDR lorsque $H = 2$.

2.1. Méthode SIR

La régression inverse par tranches ou méthode SIR est une méthode de régression semi-paramétrique reposant sur un argument géométrique. Elle a été introduite par Li [7] et Duan et Li [3]. Cette méthode repose sur une propriété de la fonction de régression inverse (c'est-à-dire qu'au lieu de régresser y sur x , ce sont des propriétés de la régression de x sur y qui vont être utilisées). L'avantage de cette inversion de rôle est que la dimension du problème a été réduite : nous avons en effet maintenant p problèmes de dimension 1, la régression inverse permettant de régresser chaque coordonnée de x sur y . Le coût à payer est de rajouter une hypothèse probabiliste sur la distribution de la variable explicative x :

(H) *La variable explicative x possède une distribution de probabilité non dégénérée telle que, pour tout $b \in \mathbb{R}^p$, l'espérance conditionnelle $\mathbb{E}[b'x \mid \theta'x]$ est linéaire en $\theta'x$.*

Cette hypothèse, encore appelée condition de linéarité, est vérifiée lorsque la variable explicative x suit une distribution elliptique, en particulier lorsque la distribution de x est multivariée normale. Avant

de poursuivre la présentation de la méthode SIR, il apparaît intéressant de discuter rapidement cette hypothèse **(H)** sur laquelle est fondée la théorie de SIR et qui n'est pas une hypothèse usuelle pour les modèle de régression. Remarquons tout d'abord qu'elle est très difficile à vérifier a priori vu qu'elle dépend de la direction inconnue θ . Par contre, il est possible de tester l'ellipticité de x . Notons que diverses techniques ont été proposées pour "forcer" les données à être elliptiques, voir par exemple la méthode proposée par Cook et Nachtsheim [2]. Lorsque la dimension de x est grande, il devient très difficile de forcer les données à être elliptiques avec ce type de méthodes. Cependant, Hall et Li [6] ont montré par un argument Bayésien que la condition fondamentale **(H)** est presque sûrement vérifiée lorsque $p \rightarrow \infty$. Ils ont alors précisé comment ce résultat permet d'élargir le champ d'applicabilité de la régression inverse par tranches. On peut aussi mentionner que Li [7] présente des résultats obtenus par simulations (sur des échantillons pour lesquels la variable explicative x n'est pas elliptique) montrant une robustesse de la méthode SIR à des écarts à l'hypothèse **(H)**. Enfin, pour conclure cette discussion, on renvoie le lecteur à l'article de Chen et Li [1] où cette hypothèse est plus longuement discutée.

Posons $\mu = \mathbb{E}[x]$ et $\Sigma = \mathbb{V}(x)$. Nous donnons maintenant le théorème de caractérisation de la direction de θ établi par Li [7].

Theorem 2.1. *Dans le cadre du modèle (1.1) et sous l'hypothèse **(H)**, la courbe de régression inverse centrée, $y \mapsto \mathbb{E}[x | \mathbb{T}(y)] - \mu$, appartient au sous-espace linéaire de \mathbb{R}^p engendré par θ , où \mathbb{T} est une transformation monotone de y (qui correspondra au "tranchage" précisé ultérieurement).*

Une conséquence directe de ce théorème est que la matrice de variances-covariances de cette courbe, $\Gamma = \mathbb{V}(\mathbb{E}[x | \mathbb{T}(y)])$, est dégénérée dans toute direction Σ -orthogonale à θ . Ainsi le vecteur propre $\tilde{\theta}$ associé à la valeur propre non nulle de $\Sigma^{-1}\Gamma$ est colinéaire à θ . Ce vecteur propre $\tilde{\theta}$ est donc une direction EDR.

Remarque 2.1. *Dans le cadre d'une généralisation du modèle (1.1) à un modèle à plusieurs indices, $\theta'_1 x, \dots, \theta'_d x$ avec $d < p$ (voir Li [7]), les vecteurs propres $\tilde{\theta}_1, \dots, \tilde{\theta}_d$, associés aux d plus grandes valeurs propres non nulles de la matrice $\Sigma^{-1}\Gamma$ appartiennent au sous-espace linéaire engendré par les vecteurs $\theta_1, \dots, \theta_d$, c'est à dire appartient l'espace EDR.*

En pratique, nous disposons d'un échantillon d'observations $\mathcal{S} = \{(x_i, y_i), i = 1, \dots, N\}$. Nous allons maintenant rappeler la méthode d'estimation (non récursive) telle qu'elle a été introduite par Duan et Li [3].

On note \bar{x}_N et Σ_N la moyenne et la variance empirique de l'échantillon $\{x_i, i = 1, \dots, N\}$, données respectivement par

$$\bar{x}_N = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{et} \quad \Sigma_N = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}_N)(x_i - \bar{x}_N)'.$$

Il reste alors à estimer la matrice Γ et donc à choisir une transformation \mathbb{T} telle que cette matrice soit facilement estimable. Un choix particulier de \mathbb{T} a été proposé par Duan et Li [3] : le "tranchage" qui est une discrétisation de y fondée sur un découpage du support de y en H tranches distinctes s_1, \dots, s_H . Dans ce cadre, en notant $p_h = \mathbb{P}(y \in s_h)$ et $m_h = \mathbb{E}[x | y \in s_h]$ pour $h = 1, \dots, H$, la matrice Γ s'écrit très simplement sous la forme :

$$\Gamma = \sum_{h=1}^H p_h (m_h - \mu)(m_h - \mu)'.$$

L'estimation des différentes grandeurs intervenant dans la matrice Γ ne pose aucune difficulté : il suffit de substituer les moments empiriques aux moments théoriques. Nous obtenons alors l'estimateur Γ_N :

$$\Gamma_N = \sum_{h=1}^H p_{h,N} (m_{h,N} - \bar{x}_N) (m_{h,N} - \bar{x}_N)',$$

où

$$p_{h,N} = \frac{1}{N} \sum_{i=1}^N \mathbb{I}[y_i \in s_h] = \frac{N_h}{N} \quad \text{et} \quad m_{h,N} = \frac{1}{N_h} \sum_{i \in s_h} x_i$$

sont respectivement la proportion empirique des y_i tombant dans la tranche s_h , et la moyenne empirique des x_i tels que y_i appartienne à la tranche s_h . Ensuite, il suffit de déterminer le vecteur propre $\tilde{\theta}_N$ associé à la plus grande valeur propre de $\Sigma_N^{-1} \Gamma_N$. On dit que $\tilde{\theta}_N$ est un estimateur de la direction EDR.

Des résultats asymptotiques sont disponibles dans la littérature pour l'estimateur SIR non récursif de la direction de θ . Sans être exhaustif, on peut mentionner que la vitesse de convergence de la direction EDR estimée vers la vraie direction est $n^{-1/2}$, voir Li [7] et que la normalité asymptotique des éléments propres d'intérêt (projecteur propre sur l'espace EDR estimé, valeurs propres et vecteurs propres principaux, i.e. estimateurs des directions EDR) a été obtenue par Saracco [10].

Etude du cas où $H = 2$ permettant d'obtenir une expression analytique de l'estimateur $\tilde{\theta}_N$. Dans cet article, nous nous intéresserons au cas particulier où l'on ne considère que deux tranches 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 . En effet, lorsque $H = 2$, la matrice Γ s'écrit :

$$\Gamma = p_1 z_1 z_1' + p_2 z_2 z_2', \quad (2.1)$$

avec $z_h = m_h - \mu$ pour $h = 1, 2$. On peut alors montrer que la valeur propre non nulle λ de $\Sigma^{-1} \Gamma$ et le vecteur propre $\tilde{\theta}$ correspondant (colinéaire à θ) s'écrivent sous la forme :

$$\lambda = \frac{p_1}{p_2} z_1' \Sigma^{-1} z_1 \quad \text{et} \quad \tilde{\theta} = \Sigma^{-1} (z_1 - z_2). \quad (2.2)$$

Les calculs de $\tilde{\theta}$ et de λ sont détaillés en Annexe 1. Nous pouvons ainsi déduire des relations (2.2) des estimateurs λ_N et $\tilde{\theta}_N$ de λ et $\tilde{\theta}$:

$$\lambda_N = \frac{p_{1,N}}{p_{2,N}} z_{1,N}' \Sigma_N^{-1} z_{1,N} \quad \text{et} \quad \tilde{\theta}_N = \Sigma_N^{-1} (z_{1,N} - z_{2,N}), \quad (2.3)$$

où $z_{h,N} = m_{h,N} - \bar{x}_N$ pour $h = 1, 2$.

Dans la sous-section suivante, nous allons tout d'abord proposer une écriture récursive pour l'estimateur de la matrice d'intérêt $\Sigma^{-1} \Gamma$ pour un nombre H quelconque de tranches. Ensuite, nous déterminerons la forme récursive de l'estimateur de la direction θ lorsque $H = 2$.

2.2. Estimateur récursif de la matrice d'intérêt $\Sigma^{-1}\Gamma$

Lorsque l'on s'intéresse à la forme non récursive des estimateurs, nous disposons d'un échantillon d'observations $\{(x_i, y_i), i = 1, \dots, N\}$ de variables aléatoires (x, y) indépendantes et identiquement distribuées issues du modèle (1.1). 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) .

Dans un premier temps, nous donnerons les formes récursives de \bar{x}_N , Σ_N et Σ_N^{-1} , puis celle de Γ_N . Ensuite, nous décrirons la forme récursive de la matrice d'intérêt $\Sigma_N^{-1}\Gamma_N$.

2.2.1. Forme récursive de \bar{x}_N , Σ_N et Σ_N^{-1}

La forme récursive de la moyenne empirique \bar{x}_N des N observations x_1, \dots, x_N est la suivante :

$$\bar{x}_N = \frac{1}{N} \sum_{i=1}^N x_i = \frac{N-1}{N} \bar{x}_{N-1} + \frac{1}{N} x_N = \bar{x}_{N-1} + \frac{1}{N} \Phi_N \quad (2.4)$$

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 :

$$\begin{aligned} \Sigma_N &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}_N)(x_i - \bar{x}_N)' \\ &= \frac{N-1}{N} \Sigma_{N-1} + \frac{N-1}{N^2} (x_N - \bar{x}_{N-1})(x_N - \bar{x}_{N-1})' \\ &= \frac{N-1}{N} \Sigma_{N-1} + \frac{N-1}{N^2} \Phi_N \Phi_N' \end{aligned} \quad (2.5)$$

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

$$\Sigma_N^{-1} = \frac{N}{N-1} \Sigma_{N-1}^{-1} - \frac{N}{(N-1)(N + \rho_N)} \Sigma_{N-1}^{-1} \Phi_N \Phi_N' \Sigma_{N-1}^{-1} \quad (2.6)$$

avec $\rho_N = \Phi_N' \Sigma_{N-1}^{-1} \Phi_N$.

2.2.2. Forme récursive de Γ_N

On se place ici dans le cas général où le nombre H de tranches est quelconque ($H \geq 2$). Afin de donner la forme récursive de l'estimateur Γ_N de la matrice Γ , nous supposons que la nouvelle observation (x_N, y_N) est telle que y_N appartienne à la tranche s_{h^*} .

Notons tout d'abord que la forme récursive de l'estimateur $p_{h,N}$ de p_h est

$$p_{h,N} = \begin{cases} \frac{N-1}{N} p_{h^*,N-1} + \frac{1}{N} & \text{si } h = h^*, \\ \frac{N-1}{N} p_{h,N-1} & \text{sinon.} \end{cases} \quad (2.7)$$

La forme récursive de $m_{h,N}$ est donnée, pour $\Phi_{h^*,N} = x_N - m_{h^*,N-1}$, par

$$m_{h,N} = \begin{cases} m_{h^*,N-1} + \frac{1}{N_{h^*,N-1}+1} \Phi_{h^*,N} & \text{si } h = h^*, \\ m_{h,N-1} & \text{sinon.} \end{cases} \quad (2.8)$$

A partir des relations (2.4) et (2.8), on obtient la forme récursive de $z_{h,N} = m_{h,N} - \bar{x}_N$

$$z_{h,N} = \begin{cases} z_{h^*,N-1} - \frac{1}{N} \Phi_N + \frac{1}{N_{h^*,N-1}+1} \Phi_{h^*,N} & \text{si } h = h^*, \\ z_{h,N-1} - \frac{1}{N} \Phi_N & \text{sinon.} \end{cases} \quad (2.9)$$

En écrivant Γ_N sous la forme

$$\Gamma_N = \sum_{h \neq h^*} p_{h,N} z_{h,N} z'_{h,N} + p_{h^*,N} z_{h^*,N} z'_{h^*,N},$$

nous pouvons déduire des relations (2.7) et (2.9) la forme récursive suivante pour Γ_N :

$$\begin{aligned} \Gamma_N &= \frac{N-1}{N} \Gamma_{N-1} - \frac{N-1}{N^2} \sum_{h=1}^H p_{h,N-1} (z_{h,N-1} \Phi'_N + \Phi_N z'_{h,N-1}) \\ &\quad + \frac{N-1}{N^3} \Phi_N \Phi'_N + \frac{N-1}{N} p_{h^*,N-1} A_{h^*,N} + \frac{1}{N} B_{h^*,N} B'_{h^*,N} \end{aligned} \quad (2.10)$$

avec

$$\begin{aligned} A_{h^*,N} &= \frac{1}{N_{h^*,N-1}+1} (z_{h^*,N-1} \Phi'_{h^*,N} + \Phi_{h^*,N} z'_{h^*,N-1}) \\ &\quad - \frac{1}{N(N_{h^*,N-1}+1)} (\Phi_N \Phi'_{h^*,N} + \Phi_{h^*,N} \Phi'_N) + \frac{1}{(N_{h^*,N-1}+1)^2} \Phi_{h^*,N} \Phi'_{h^*,N} \end{aligned}$$

$$\text{et } B_{h^*,N} = z_{h^*,N-1} - \frac{1}{N} \Phi_N + \frac{1}{N_{h^*,N-1}+1} \Phi_{h^*,N}.$$

2.2.3. Forme récursive de la matrice d'intérêt $\Sigma_N^{-1} \Gamma_N$

A partir des relations (2.6) et (2.10), on peut obtenir une forme récursive pour $\Sigma_N^{-1} \Gamma_N$:

$$\Sigma_N^{-1} \Gamma_N = \Sigma_{N-1}^{-1} \Gamma_{N-1} + \Sigma_{N-1}^{-1} C_{h^*,N} - \frac{1}{N + \rho_N} \Sigma_{N-1}^{-1} \Phi_N \Phi'_N \Sigma_{N-1}^{-1} [\Gamma_{N-1} + C_{h^*,N}] \quad (2.11)$$

en posant

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

2.3. Estimateur récursif de la direction de θ

Nous nous plaçons maintenant dans le cas où $H = 2$. Dans ce cas, nous avons donné en (2.3) l'expression analytique des formes non récursives des estimateurs λ_N et $\tilde{\theta}_N$. Nous allons maintenant en déduire leurs formes récursives. Nous supposons encore ici que la nouvelle observation (x_N, y_N) est telle que $y_N \in s_{h^*}$ avec $h^* = 1$ ou 2 .

2.3.1. Forme récursive de $\tilde{\theta}_N$

D'après (2.6) et (2.9), nous obtenons la forme récursive de l'estimateur $\tilde{\theta}_N$ de $\tilde{\theta}$:

$$\begin{aligned} \tilde{\theta}_N &= \frac{N}{N-1} \tilde{\theta}_{N-1} - \frac{N}{(N-1)(N+\rho_N)} \Sigma_{N-1}^{-1} \Phi_N \Phi_N' \tilde{\theta}_{N-1} \\ &\quad - \frac{(-1)^{h^* N}}{(N_{h^*, N-1} + 1)(N-1)} \left(\Sigma_{N-1}^{-1} - \frac{1}{N+\rho_N} \Sigma_{N-1}^{-1} \Phi_N \Phi_N' \Sigma_{N-1}^{-1} \right) \Phi_{h^*, N}. \end{aligned} \quad (2.12)$$

2.3.2. 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 ici l'expression de $F(\cdot)$ qui, malgré quelques simplifications, reste lourde à écrire et ne présente pas d'intérêt pour la suite de cet article.

3. Résultats asymptotiques

Nous nous focalisons dans cette section sur des résultats de convergence de l'estimateur $\tilde{\theta}_N$ de la direction EDR dans le cas où $H = 2$. Avant de donner ces résultats, nous posons les deux hypothèses suivantes :

- (A1) Les observations $(x_i, y_i), i = 1, \dots, N$, sont échantillonnées de manière indépendante à partir du modèle (1.1).
- (A2) 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 .

Theorem 3.1. *Sous les hypothèses (H), (A₁) et (A₂), nous avons*

$$\|\tilde{\theta}_N - \tilde{\theta}\| = \mathcal{O}\left(\sqrt{\frac{\log(\log N)}{N}}\right) \quad p.s.$$

Theorem 3.2. *Sous les hypothèses (H), (A₁) et (A₂), nous avons :*

$$\sqrt{N}(\tilde{\theta}_N - \tilde{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma^{-1} \Delta_3 \Sigma^{-1}), \quad (3.1)$$

où l'expression de la matrice Δ_3 est donnée en (C.4).

Les démonstrations de ces deux théorèmes figurent respectivement en Annexes B et C.

4. Quelques résultats de simulations

Dans cette partie, nous allons étudier, sur des simulations, le comportement numérique de l'estimateur récursif $\tilde{\theta}_N$ que nous avons proposé lorsque l'on prend $H = 2$ tranches. Cet estimateur $\tilde{\theta}_N$ est un estimateur de la vraie direction de θ du modèle. Nous prendrons alors comme mesure de la qualité de l'estimation la valeur du cosinus carré 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.

Nous présentons tout d'abord les modèles simulés et nous précisons les procédures d'estimation mises en œuvre. Nous donnons dans un second temps des résultats de comparaisons des temps de calculs entre les approches SIR récursive et non récursive, comparaison mettant en exergue la rapidité de l'approche récursive. Nous décrivons ensuite des résultats de simulations sur des échantillons particuliers et sur de nombreuses répliques d'échantillons, résultats montrant la convergence de l'estimateur vers la vraie direction θ pour différents jeux de paramètres (modèle, dimension p , matrice Σ , taille N d'échantillon). Nous terminons en illustrant le normalité asymptotique de l'estimateur $\tilde{\theta}_N$ proposé.

Modèles simulés et procédure d'estimation. Dans ces simulations, nous avons considéré les deux modèles de régression suivants :

$$\begin{aligned} (M1) : & y = (\theta'x)^3 + \varepsilon, \\ (M2) : & y = (\theta'x) \exp(-\theta'x/2) + (\theta'x)\varepsilon, \end{aligned}$$

où x suit la loi multinormale $\mathcal{N}_p(0, \Sigma)$, $\theta = (1, -1, 0, \dots, 0)' \in \mathbb{R}^p$ et le terme d'erreur ε , indépendant de x , suit la loi normale $\mathcal{N}(0, 1)$. Pour la matrice Σ , nous avons considéré le cas où $\Sigma = I_p$ et le cas où Σ est une matrice symétrique définie positive "quelconque" générée aléatoirement¹. Nous avons fait varier la dimension p de x : $p = 5, 10, 20$ ou 40 . Le modèle (M1) est un modèle

¹ Pour générer une matrice Σ "quelconque", nous avons tout d'abord généré p^2 réalisations d'une loi uniforme sur $[-1, 1]$ que nous avons rangées dans une matrice B . Puis nous avons pris $\Sigma = BB' + 0.1I_p$, le second terme permettant d'éviter des problèmes numériques dans l'inversion de Σ .

homoscédastique ne présentant pas de difficultés particulières pour l'estimation de la direction de θ . Le modèle (M2) est hétéroscédastique et l'estimation de la direction de θ est moins aisée avec des données issues d'un tel modèle. Ces deux modèles sont classiquement utilisés dans les simulations décrites dans la littérature sur la méthode SIR. Nous présentons à la Figure 1 deux nuages de $N = 200$ points $\{(x_i'\theta, y_i), i = 1, \dots, N\}$ simulés à partir des modèles (M1) et (M2) pour $p = 10$ et $\Sigma = I_p$. Pour l'ensemble des simulations, nous avons choisi un ratio bruit sur signal (défini comme le quotient entre la variance des ε_i et la variance des y_i) d'environ 6% (resp. 25%) pour (M1) (resp. pour (M2)), quelle que soit la matrice de variances-covariances Σ (matrice identité ou quelconque) et quelle que soit la dimension p .

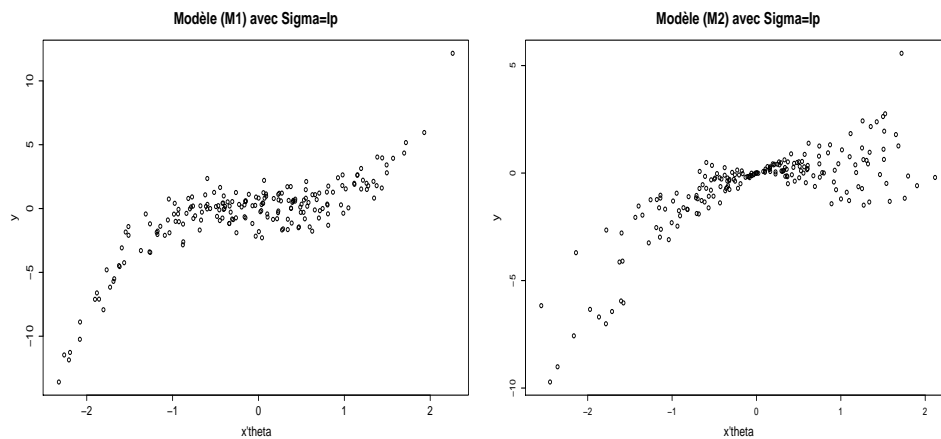


FIGURE 1: Nuage des points $\{(x_i'\theta, y_i), i = 1, \dots, 200\}$ simulés à partir des modèles (M1) et (M2) pour $p = 10$ et $\Sigma = I_p$

Les simulations ont été faites à l'aide du logiciel R, les codes correspondants sont disponibles auprès des auteurs.

Nous avons simulé, à partir des modèles (M1) et (M2), $\mathcal{B} = 500$ échantillons de taille 1000 avec successivement $p = 5, 10, 20$ et 40 , et pour $\Sigma = I_p$ puis pour des matrices Σ "quelconques". Pour chaque jeu de paramètres (modèle, dimension p , matrice Σ) et 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 $p + 1$. Les N_0 premiers couples (x_i, y_i) de l'échantillon permettent de calculer les grandeurs nécessaires à l'initialisation de l'algorithme de calculs. Le choix de $N_0 = p + 1$ s'explique par le fait que l'on a besoin de calculer la matrice $\Sigma_{N_0-1}^{-1}$ qui est de dimension (p, p) . 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. Notons que le choix du seuil n'a pas d'influence sur la qualité de l'estimation pourvu que l'on ait les deux fréquences $p_{h,N}$ non nulles, ceci était déjà vrai pour la version non récursive de l'estimateur SIR.

Comparaison des temps de calculs entre les approches SIR récursive et non récursive. Nous

avons mesuré pour les approches SIR récursive et SIR non récursive (basées sur $H = 2$ tranches) les temps de calculs en secondes des estimateurs $\tilde{\theta}_n$ de la direction de θ pour n allant de N_0 à N . Dans la Table 1, nous donnons les valeurs des moyennes et des écarts-types calculés sur $\mathcal{B} = 500$ répliques d'échantillons issus du modèle (M2) avec différentes valeurs de p et deux types de matrice Σ (identité ou "quelconque"). Les constats que l'on peut faire suite à la lecture de cette table sont les suivants :

- L'approche SIR récursive est bien plus rapide en moyenne pour calculer les $(N - N_0)$ estimateurs $\tilde{\theta}_n$ que l'approche SIR non récursive. Cet avantage est d'autant plus important que la dimension p de x est grande. Ceci provient très certainement du coût (en terme de temps de calculs) de l'inversion de l'estimateur de la matrice Σ de dimension $p \times p$ suivie du calcul des éléments propres d'une matrice de dimension $p \times p$. Lorsque $p = 5$, l'approche récursive est environ 10 fois plus rapide, lorsque $p = 20$, elle est plus de 15 fois plus rapide, et lorsque $p = 40$, elle est plus de 20 fois plus rapide.
- 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 non récursive.
- La matrice de variances-covariances Σ (matrice identité ou quelconque) n'influence pas les temps des calculs des estimateurs.

Des calculs ont été faits de manière identique avec le modèle (M1) et les temps de calculs qui en découlent sont similaires.

			$p = 5$	$p = 20$	$p = 40$
I_p	$N = 300$	SIR classique	1.00 (0.018)	1.99 (0.054)	4.03 (0.012)
		SIR récursif	0.11 (0.001)	0.13 (0.002)	0.17 (0.005)
	$N = 600$	SIR classique	2.01 (0.014)	4.55 (0.093)	10.09 (0.097)
		SIR récursif	0.22 (0.002)	0.26 (0.002)	0.35 (0.013)
	$N = 900$	SIR classique	3.35 (0.022)	7.55 (0.014)	16.57 (0.025)
		SIR récursif	0.33 (0.004)	0.39 (0.004)	0.53 (0.007)
Σ	$N = 300$	SIR classique	0.99 (0.008)	1.97 (0.014)	4.10 (0.014)
		SIR récursif	0.11(0.001)	0.13 (0.001)	0.17 (0.004)
	$N = 600$	SIR classique	2.11 (0.012)	4.57 (0.013)	9.81 (0.020)
		SIR récursif	0.22 (0.001)	0.26 (0.001)	0.35 (0.006)
	$N = 900$	SIR classique	3.35 (0.019)	7.66 (0.019)	16.76 (0.051)
		SIR récursif	0.33 (0.002)	0.39 (0.007)	0.54 (0.009)

TABLE 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 issus du modèle (M2) avec différentes valeurs de p et deux types matrice Σ (identité ou "quelconque")

Etude de quelques échantillons particuliers avec l'approche SIR récursive. A la Figure 2 (resp. Figure 3), nous considérons tout d'abord des échantillons particuliers simulés à partir du modèle (M1) (resp. du modèle (M2)) avec $p = 10$ et $\Sigma = I_p$ ou $\Sigma = \Sigma_1$. Sur ces deux figures, nous avons représenté l'évolution de la mesure de qualité de l'estimateur $\tilde{\theta}_N$ en fonction de N . Nous voyons clairement que plus le nombre d'observations est important, plus la qualité d'estimation

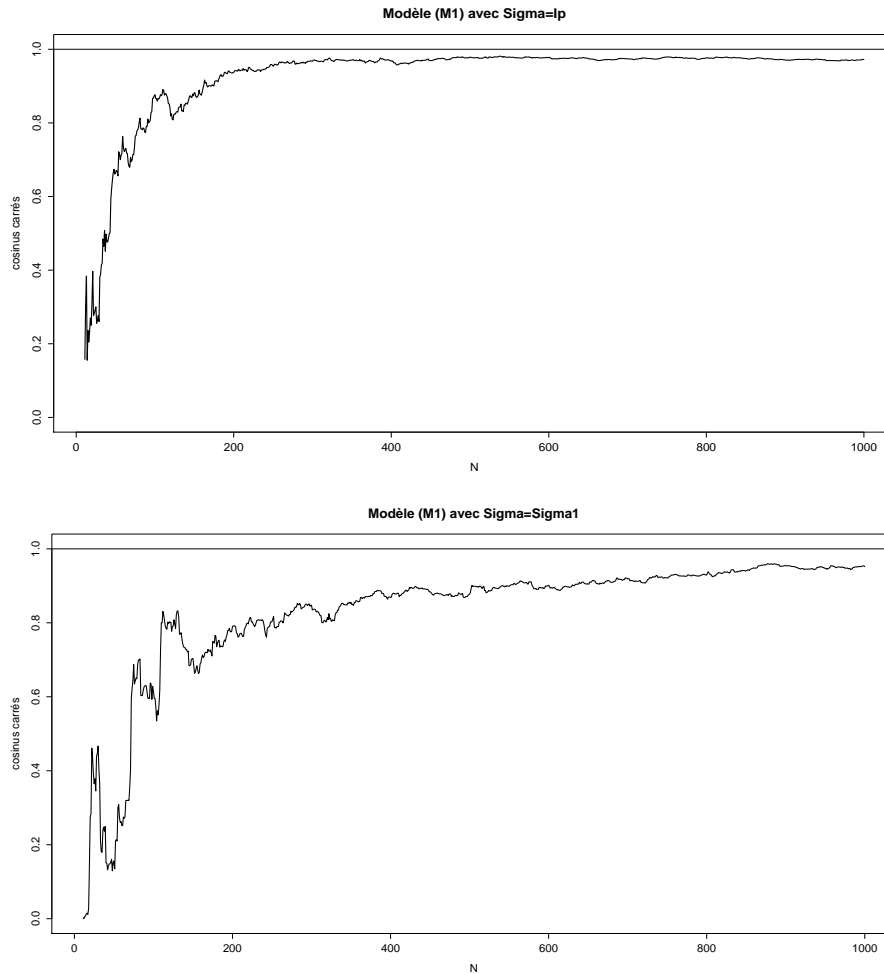


FIGURE 2: Evolution, en fonction de N , de la qualité de l'estimateur $\tilde{\theta}_N$ pour un échantillon issu du modèle (M1) avec $p = 10$ et différentes matrices $\Sigma (= I_p$ ou $\Sigma_1)$

est bonne. Pour le modèle (M1) avec $\Sigma = I_p$, les valeurs du $\cos^2(\tilde{\theta}_N, \theta)$ se rapprochent de 1 très rapidement (voir à la Figure 2 le graphique du haut). Lorsque l'on change de matrice de variances-covariances en prenant $\Sigma = \Sigma_1$, la convergence de $\cos^2(\tilde{\theta}_N, \theta)$ vers 1 est plus lente, mais la qualité d'estimation est tout à fait correcte (supérieure à 0.85) lorsque la taille de l'échantillon N est supérieure 300 (voir à la Figure 2 le graphique du bas). Une explication est que, pour obtenir une bonne estimation de la matrice Σ_1 dont la structure est plus complexe que celle de la matrice I_p , une taille d'échantillon plus importante est nécessaire. A la Figure 3, on observe le même type de phénomènes que ceux décrits pour le modèle (M1). Il est à noter que le modèle (M2) est plus complexe (modèle hétéroscédastique) que le modèle (M1). Il apparaît ainsi assez naturel que la convergence de $\cos^2(\tilde{\theta}_N, \theta)$ vers 1 soit plus lente avec ces deux échantillons issus du modèle (M2). Il est cependant utile de rappeler que ces graphiques ont été obtenus à partir d'échantillons

simulés particuliers, et que, pour un même modèle et un même jeu de paramètres, les évolutions des cosinus carrés peuvent être sensiblement différentes pour de petites tailles N d'échantillon. Pour ces raisons-là, nous présentons ci-après une étude sur $\mathcal{B} = 500$ réplifications d'échantillons.

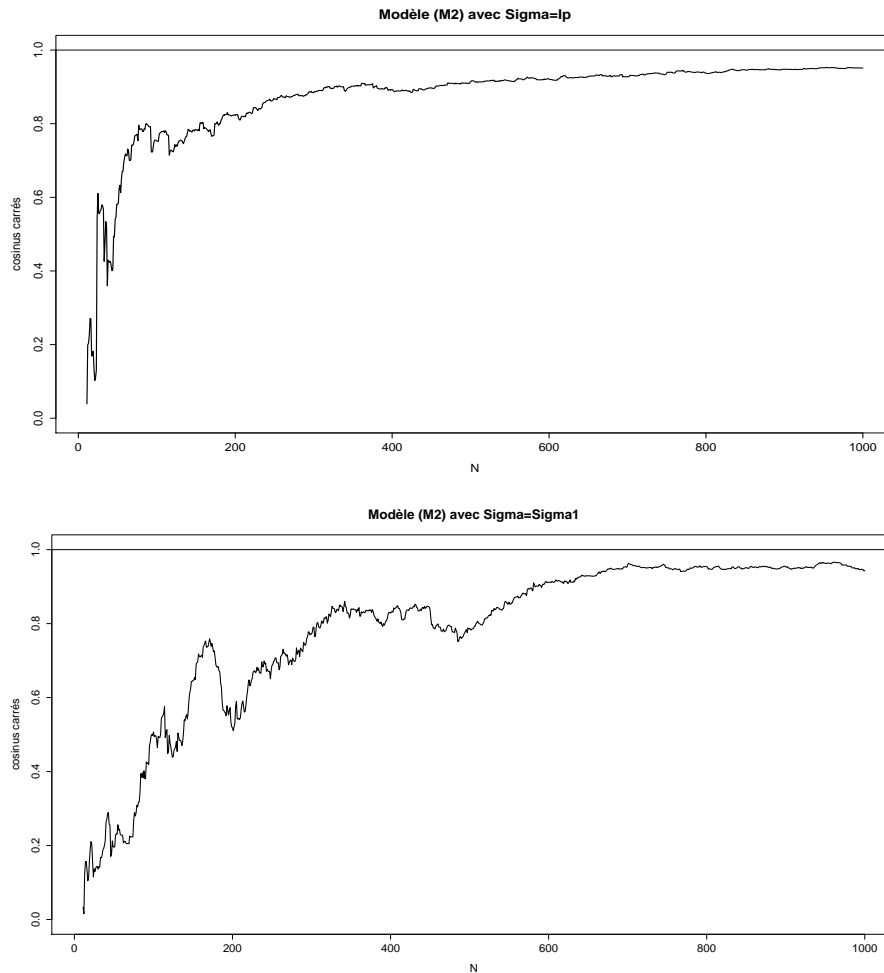


FIGURE 3: Evolution, en fonction de N , de la qualité de l'estimateur $\tilde{\theta}_N$ pour un échantillon issu du modèle (M2) avec $p = 10$ et différentes matrices $\Sigma (= I_p$ ou $\Sigma_1)$

Etude sur $\mathcal{B} = 500$ réplifications d'échantillons avec l'approche SIR récursive. Nous avons représenté à la Figure 4 (resp. Figure 5) les boxplots, pour différentes valeurs de $N (=50, 100, 150, 200, 300, 500, 700, 1000)$, des mesures de qualité de $\tilde{\theta}_N$ calculées sur les $\mathcal{B} = 500$ échantillons issus du modèle (M1) (resp. du modèle (M2)) avec $p = 10$ et différentes matrices $\Sigma = I_p$ et $\Sigma = \Sigma_1$. Nous observons le même type de phénomène que ceux décrits dans le cas des échantillons particuliers aux Figures 2 et 3. Plus précisément, plus la taille N de l'échantillon est grande, meilleurs sont les estimations (avec des valeurs de $\cos^2(\tilde{\theta}_N, \theta)$ de plus en plus proche de 1 et une dispersion de plus en plus faible). De même, plus la structure de la matrice de variances-

covariances Σ est complexe, plus il est nécessaire de disposer d'un échantillon de taille raisonnable pour estimer convenablement la direction de θ .

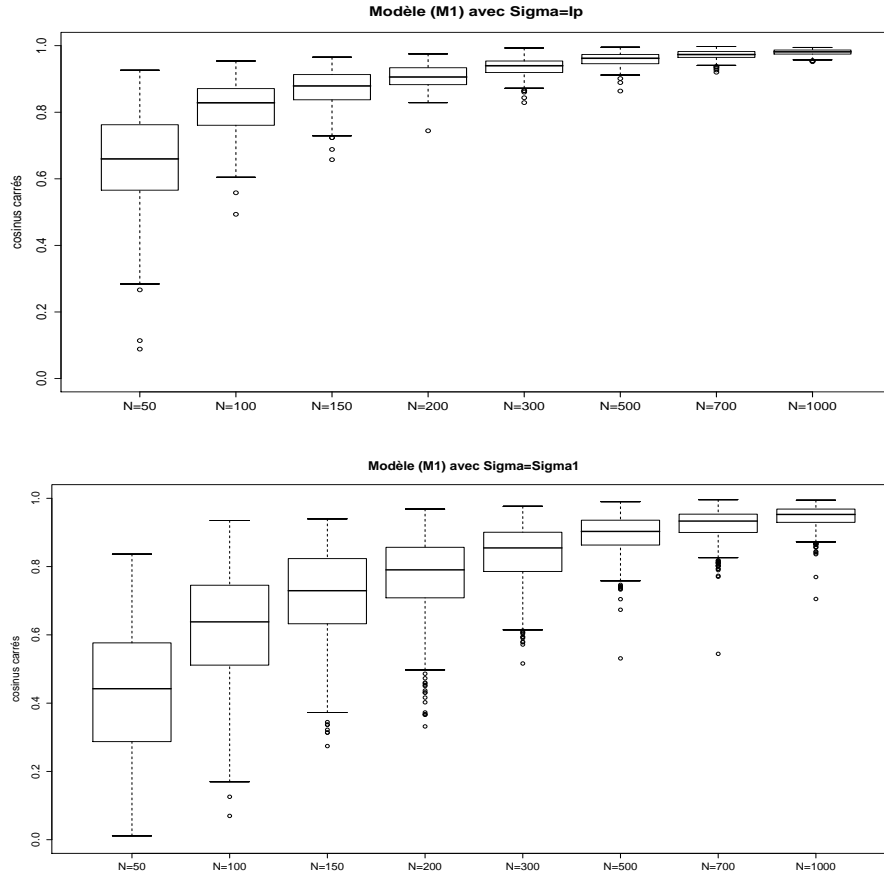


FIGURE 4: Boxplots, pour différentes valeurs de N , des mesures de qualité de $\tilde{\theta}_N$ calculées sur $\mathcal{B} = 500$ échantillons issus du modèle (M1) avec $p = 10$ et différentes matrices Σ ($= I_p$ ou Σ_1)

Etude de l'effet de la dimension p de la covariable x sur la qualité de l'estimation avec l'approche SIR récursive. La dimension p de la covariable a aussi naturellement une certaine influence sur la qualité d'estimation de la direction de θ . Pour illustrer ce point, nous avons représenté à la Figure 6 (resp. Figure 7), 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 (M1) (resp. modèle (M2)) avec différentes matrices Σ (matrice identité ou "quelconque"). On observe clairement que plus la dimension de la covariable est importante, plus les valeurs des cosinus carrés $\cos^2(\tilde{\theta}_N, \theta)$ ont tendance à être plus petites. Ce phénomène est d'autant plus important que la taille N des échantillons est faible et que la structure de la matrice de variances-covariances Σ est complexe.

Illustration de la normalité asymptotique de l'estimateur. A la Figure 8, nous nous intéressons la normalité asymptotique de l'estimateur $\tilde{\theta}_N$. Pour cela, nous avons simulé 3000 réali-

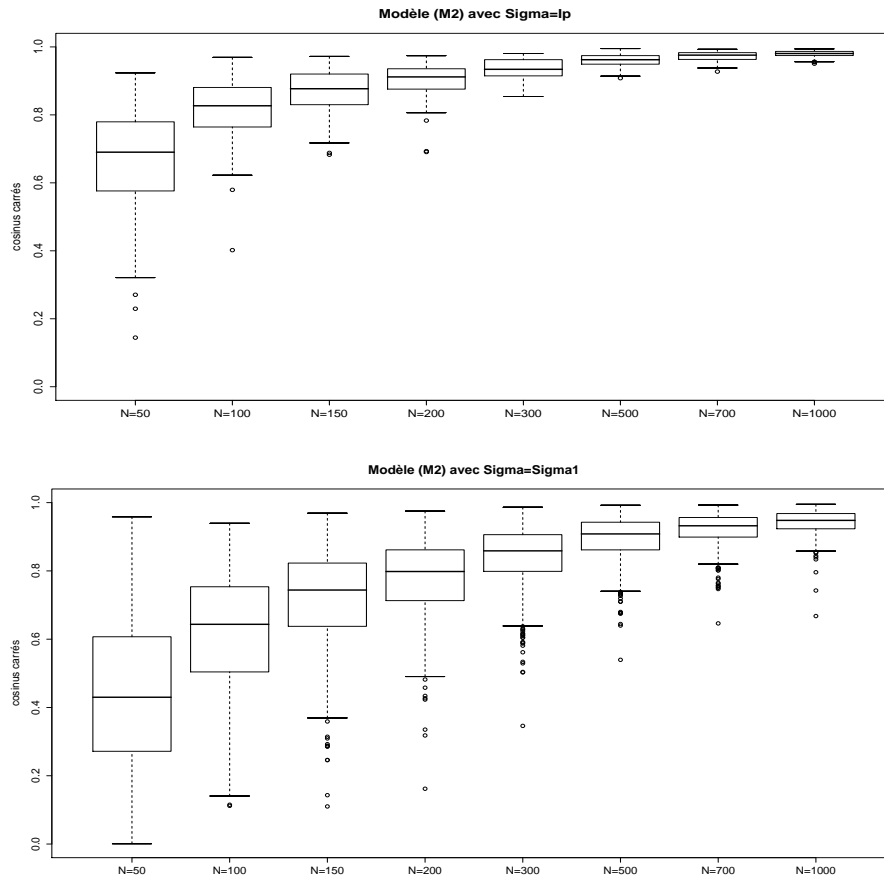


FIGURE 5: Boxplots, pour différentes valeurs de N , des mesures de qualité de $\tilde{\theta}_N$ calculées sur $\mathcal{B} = 500$ échantillons issus du modèle $(M2)$ avec $p = 10$ et différentes matrices $\Sigma (= I_p$ ou $\Sigma_1)$

sations de $\tilde{\theta}_N$ à partir d'échantillons de taille $N = 1000$ issus du modèle $(M2)$ avec $p = 10$ et une matrice de variances-covariances Σ quelconque. Nous avons tracé sur un même graphique la densité de la loi normale centrée réduite (en pointillés) ainsi que l'histogramme des $\left\{ \sqrt{N} \left(\frac{\tilde{\theta}_{N,j}^{(l)} - \theta_j}{\gamma_j} \right), l = 1, \dots, 3000 \right\}$ où $\tilde{\theta}_{N,j}^{(l)}$ correspond à la j ème composante de l'estimateur $\tilde{\theta}_N$ calculé sur le l ème échantillon, θ_j étant la j ème composante de θ , et γ_j le j ème élément diagonal à la puissance 1/2 de la matrice de variances-covariances donnée en (3.1). Nous avons choisi arbitrairement quatre composantes θ_j de θ ($j = 1, 2, 3$ et 10). Nous voyons clairement sur les quatre graphiques de la Figure 8 que la densité de la loi $\mathcal{N}(0, 1)$ se superpose bien aux différents histogrammes, ce qui illustre bien la normalité asymptotique de l'estimateur proposé. Par souci de concision, nous n'avons pas fourni ici les résultats des simulations obtenus à partir de modèle $(M1)$ ou d'autres matrices de variances-covariances Σ (en particulier $\Sigma = I_p$) ou encore d'autres dimensions p . Pour ces différents cas, nous avons encore observé graphiquement une bonne superposition des histogrammes avec la densité de la loi normale centrée réduite. Notons que

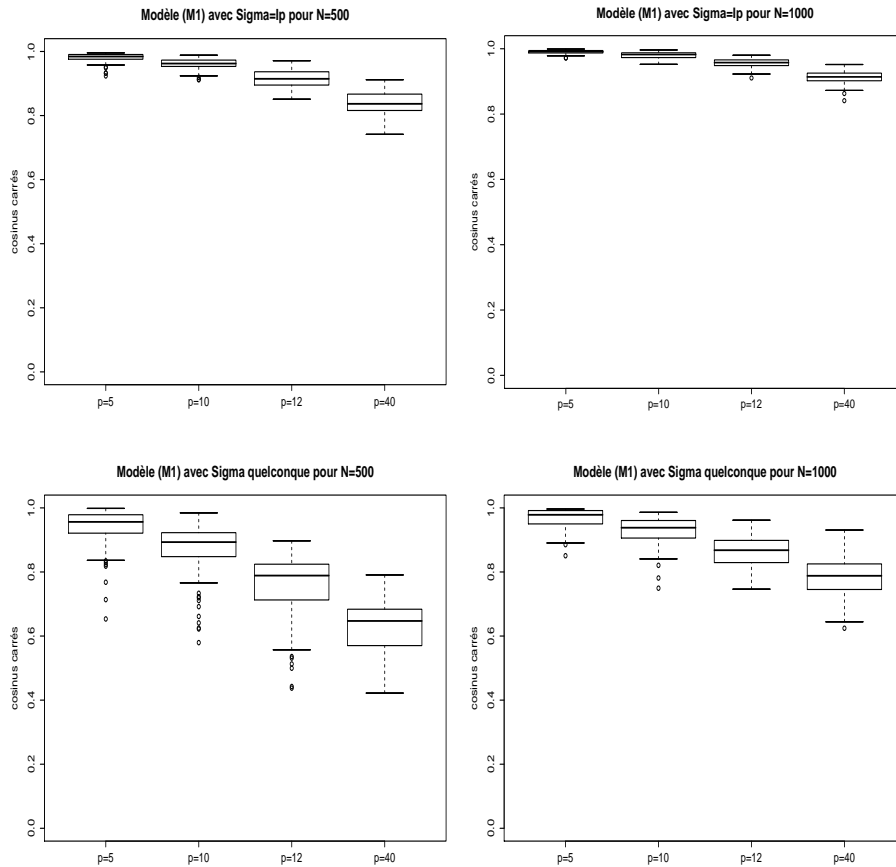


FIGURE 6: Boxplots, en fonction de la dimension p , des mesures de qualité de $\tilde{\theta}_N$ calculées sur $\mathcal{B} = 500$ échantillons issus du modèle (M1) avec différentes matrices $\Sigma (= I_p$ ou “quelconque”), pour deux tailles d’échantillons $N=500$ ou 1000

lorsque la dimension p de x est grande, il faut prendre une taille N plus importante (de l’ordre de 2000 pour $p = 40$) pour obtenir une bonne adéquation.

5. Conclusion

Dans cet article, nous nous sommes intéressés à une approche récursive de la méthode SIR dont l’objectif majeur est l’estimation de la direction du vecteur de paramètre θ du modèle (1.1). Dans un premier temps, nous avons donné la forme récursive de l’estimateur de la matrice d’intérêt, $\Sigma_N^{-1} \Gamma_N$, sur laquelle repose la méthode SIR pour un nombre H quelconque de tranches. Dans le cas particulier où l’on considère $H = 2$ tranches, nous avons précisé une expression analytique de l’estimateur $\tilde{\theta}_N$, et nous avons ensuite proposé une forme récursive pour cet estimateur de la direction de θ . Nous avons donné quelques propriétés asymptotiques de cet estimateur dont la convergence presque sûre (avec sa vitesse) et la normalité asymptotique. Nous avons enfin illustré

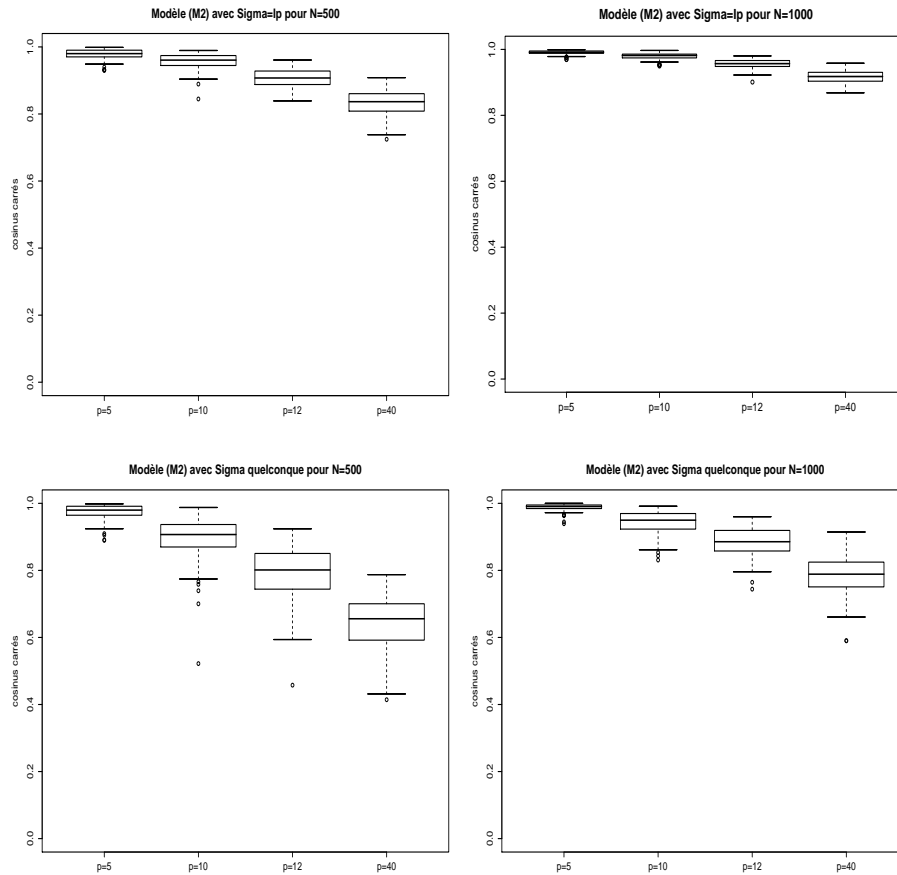


FIGURE 7: Boxplots, en fonction de la dimension p , des mesures de qualité de $\tilde{\theta}_N$ calculés sur $\mathcal{B} = 500$ échantillons issus du modèle (M2) avec différentes matrices Σ ($= I_p$ ou “quelconque”), pour deux tailles d’échantillons $N=500$ ou 1000

sur des simulations (en considérant différents modèles, différentes dimensions p et différentes matrices de variances-covariances Σ) le bon comportement numérique de l’approche récursive proposée. Un avantage majeur de cette approche récursive est sa rapidité par rapport à la méthode SIR non récursive. Les programmes utilisés ont été écrits en R et sont disponibles auprès des auteurs.

La suite naturelle de ce travail consiste en l’estimation simultanée de la direction de θ et de la fonction de lien f dans le cadre d’un modèle du type $y = f(\theta'x) + \varepsilon$. Pour cela, nous allons considérer l’estimateur de Nadaraya-Watson récursif (voir par exemple [5]) pour estimer la fonction de régression f , et le combiner à l’estimateur récursif $\tilde{\theta}_N$ de θ étudié dans cet article :

$$\hat{f}_N(z) = \frac{1}{\sum_{i=1}^N \frac{1}{h_i} \mathbf{K}\left(\frac{z - \hat{z}_i}{h_i}\right)} \sum_{i=1}^N \frac{1}{h_i} \mathbf{K}\left(\frac{z - \hat{z}_i}{h_i}\right) y_i,$$

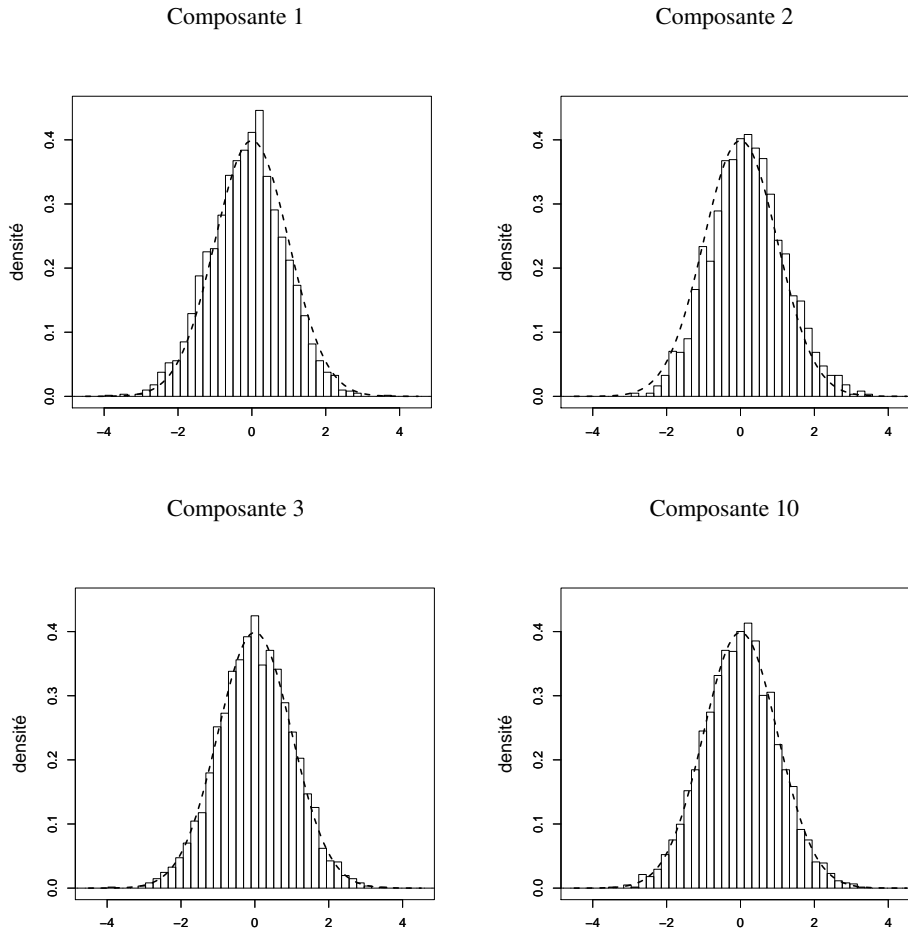


FIGURE 8: Illustration de la normalité asymptotique pour 4 composantes de $\tilde{\theta}_N$ (le graphe de la densité de la loi normale centrée réduite, en pointillé, est superposé à l'histogramme) pour le modèle (M2) avec $p = 10$, $\Sigma = \Sigma_1$ et $N = 1000$

où \mathbf{K} est un noyau, la suite (h_N) correspond au choix de la largeur de fenêtre, et où z et \hat{z}_i sont respectivement les indices $z = \theta'x$ et $\hat{z}_i = \tilde{\theta}'_i x_i$. Ce travail est en cours d'investigation. Une autre extension possible et intéressante de ce travail est de sortir du cadre i.i.d. en considérant un cadre dépendant, comme par exemple le cadre autorégressif. Cette extension n'est pas directe et va nécessiter une nouvelle formulation de l'hypothèse de linéarité (**H**). Notons qu'à notre connaissance, ceci n'a pas encore été étudié dans la littérature pour la version non récursive de la méthode SIR.

Remerciements. Qu'il nous soit permis de remercier l'Éditeur en Chef du Journal de la SFdS, ainsi que les deux relecteurs anonymes : leurs commentaires, leurs critiques et leurs suggestions constructives nous ont permis d'améliorer substantiellement la qualité de cet article.

Références

- [1] Chen, C .H. et Li, K. C. (1998). Can SIR be as popular as multiple linear regression ? *Statistica Sinica*, **8**, 289-316.
- [2] Cook, R. D. et Nachtsheim, C. J. (1994). Re-weighting to achieve elliptically contoured covariates in regression. *Journal of the American Statistical Association*, **89**, 592-599.
- [3] Duan, N. and Li, K. C. (1991). Slicing regression : a link-free regression method. *The Annals of Statistics*, **19**, 505-530.
- [4] Duflo, M. (1990). *Méthodes récursives aléatoires*. Techniques Stochastiques, Masson, Paris.
- [5] Duflo, M. (1997). *Random Iterative Models*, Springer-Verlag, Berlin.
- [6] Hall, P. et Li, K. C. (1993). On almost linearity of low dimensional projections from high dimensional data. *The Annals of Statistics*, **21**, 867-889.
- [7] Li, K. C. (1991). Sliced inverse regression for dimension reduction, with dicussion. *Journal of the American Statistical Association*, **86**, 316-342.
- [8] Magnus, J. R. and Neudecker, H. (1979). The commutation matrix : some properties and applications. *Annals of Statistics*, **7**(2), 381-394.
- [9] Magnus, J. R. and Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley & Sons, Ltd., Chichester.
- [10] Saracco, J. (1997). An asymptotic theory for Sliced Inverse Regression. *Communications in Statistics - Theory and methods*, **26**, 2141-2171.
- [11] Stout, W.F. (1974). *Almost sure convergence* A Series of Monographs and Textbooks in Probability and Mathematical Statistics.

Annexe A: expression analytique du vecteur propre $\tilde{\theta}$ et de la valeur propre λ de $\Sigma^{-1}\Gamma$ pour $H = 2$

Démonstration. Lorsque $H = 2$, nous avons

$$p_1 z_1 + p_2 z_2 = p_1 m_1 + p_2 m_2 - (p_1 + p_2)\mu = p_1 m_1 + p_2 m_2 - \mu = \mu - \mu = 0. \quad (\text{A.1})$$

Vérifions que $\tilde{\theta} = \Sigma^{-1}(z_1 - z_2)$ est bien vecteur propre de $\Sigma^{-1}\Gamma$ et déterminons la valeur propre associée. En utilisant (2.1), on a

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

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

De plus, en utilisant (A.1), nous avons : $p_1 q_1 = -p_2 q_{12}$ et $p_2 q_2 = -p_1 q_{12}$. Nous en déduisons que

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

Or en utilisant (A.1) 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 : $\Sigma^{-1}\Gamma\tilde{\theta} = \lambda \tilde{\theta}$. Le vecteur $\tilde{\theta}$ est donc bien vecteur propre associé à la valeur propre λ de la matrice $\Sigma^{-1}\Gamma$. \square

Annexe B: démonstration du Théorème 3.1

Nous rappelons ici tout d'abord le théorème de Harman-Winter (appelé aussi loi du logarithme itéré de Hartman-Wintner, voir Stout [11] page 136 pour une démonstration de ce théorème).

Soit $\{Z_i, i \geq 1\}$ une suite de variables aléatoires indépendantes et identiquement distribuées.

Notons $T_n = \sum_{i=1}^n Z_i$ pour $n \geq 1$, si $\mathbb{E}Z_1^2 < \infty$ et $\mathbb{E}Z_1 = 0$, alors

$$\limsup T_n / (2n\mathbb{E}Z_1^2 \log(\log n))^{1/2} = 1 \quad p.s.$$

Démonstration. Lorsque $H = 2$, nous avons $\tilde{\theta}_N = \Sigma_N^{-1}(z_{1,N} - z_{2,N})$ et $\tilde{\theta} = \Sigma^{-1}(z_1 - z_2)$. On en déduit que

$$\tilde{\theta}_N - \tilde{\theta} = (\Sigma_N^{-1} - \Sigma^{-1})z_{1,N} - (\Sigma_N^{-1} - \Sigma^{-1})z_{2,N} + \Sigma^{-1}(z_{1,N} - z_1) - \Sigma^{-1}(z_{2,N} - z_2).$$

Nous obtenons alors la majoration suivante :

$$\begin{aligned} \|\tilde{\theta}_N - \tilde{\theta}\|^2 &\leq 4 \|z_{1,N}\|^2 \|\Sigma_N^{-1} - \Sigma^{-1}\|^2 + 4 \|z_{2,N}\|^2 \|\Sigma_N^{-1} - \Sigma^{-1}\|^2 \\ &\quad + 4 \|\Sigma^{-1}\|^2 \|z_{1,N} - z_1\|^2 + 4 \|\Sigma^{-1}\|^2 \|z_{2,N} - z_2\|^2. \end{aligned} \quad (\text{B.1})$$

Dans la suite de la démonstration, nous obtenons tout d'abord la vitesse de convergence de $(z_{h,N})$ en précisant la vitesse de convergence de (\bar{x}_N) et de $(m_{h,N})$ où $h = 1, 2$. Puis en appliquant l'équation de Riccati pour la matrice inverse, nous obtenons la vitesse de convergence de (Σ_N^{-1}) . Enfin, nous en déduisons la vitesse de convergence de $(\hat{\theta}_N)$.

Pour une matrice M , nous noterons par $\lambda_{\max}(M)$ sa plus grande valeur propre et par $\lambda_{\min}(M)$ sa plus petite valeur propre.

Étape 1 : étude de la vitesse de convergence de $(z_{h,N})$

Pour $h = 1, 2$, nous avons :

$$z_{h,N} - z_h = (m_{h,N} - m_h) - (\bar{x}_N - \mu).$$

Utilisons la majoration suivante :

$$\|z_{h,N} - z_h\|^2 \leq 2 \|m_{h,N} - m_h\|^2 + 2 \|(\bar{x}_N - \mu)\|^2 \quad (\text{B.2})$$

Précisons tout d'abord la vitesse de convergence de (\bar{x}_N) et celle de $(m_{h,N})$.

Étude de la vitesse de convergence de (\bar{x}_N)

Nous pouvons déduire de la loi du logarithme itéré de Hartman-Wintner (le théorème 3.2.9 page 136 de Stout [11]) que

$$\limsup \frac{\|S_N - N\mu\|^2}{2N \log(\log N)} \leq \lambda_{\max}(\Sigma) \quad \text{p.s.}$$

où $S_N = \sum_{i=1}^N x_i$. Il en découle que

$$\limsup \left(\frac{N}{2 \log(\log N)} \right) \|\bar{x}_N - \mu\|^2 \leq \lambda_{\max}(\Sigma) \quad \text{p.s.}$$

Ce qui entraîne

$$\|\bar{x}_N - \mu\|^2 = \mathcal{O} \left(\frac{\log(\log N)}{N} \right) \quad \text{p.s.} \quad (\text{B.3})$$

Étude de la vitesse de convergence de $m_{h,N}$

Nous avons

$$m_{h,N} - m_h = m_{h,N} - \frac{N}{N_h} \mathbb{E}[X \mathbb{I}_{[Y \in s_h]}] + \frac{N}{N_h} \mathbb{E}[X \mathbb{I}_{[Y \in s_h]}] - m_h$$

où on rappelle que $m_{h,N} = \frac{1}{N_h} \sum_{i=1}^N x_i \mathbb{I}_{[y_i \in s_h]}$ et $N_h = \sum_{i=1}^N \mathbb{I}_{[y_i \in s_h]}$.

En majorant, nous obtenons

$$\|m_{h,N} - m_h\|^2 \leq 2 \left\| m_{h,N} - \frac{N}{N_h} \mathbb{E}[X \mathbb{I}_{[Y \in s_h]}] \right\|^2 + 2 \left\| \frac{N}{N_h} \mathbb{E}[X \mathbb{I}_{[Y \in s_h]}] - m_h \right\|^2. \quad (\text{B.4})$$

Intéressons-nous tout d'abord au premier terme de la majoration. En remarquant que

$$N_h m_{h,N} = \sum_{i=1}^N x_i \mathbb{I}_{[y_i \in s_h]} = \sum_{i=1}^N \varepsilon_i(h) + N \mathbb{E}[X \mathbb{I}_{[Y \in s_h]})]$$

où $\varepsilon_i(h) = x_i \mathbb{I}_{[y_i \in s_h]} - \mathbb{E}[x_i \mathbb{I}_{[y_i \in s_h]}] = x_i \mathbb{I}_{[y_i \in s_h]} - \mathbb{E}[X \mathbb{I}_{[Y \in s_h]}]$, nous avons

$$M_N(h) = N_h m_{h,N} - N \mathbb{E}[X \mathbb{I}_{[Y \in s_h]}] = \sum_{i=1}^N \varepsilon_i(h).$$

La suite $(\varepsilon_N(h))$ est constituée de vecteurs aléatoires indépendantes et de même loi avec $\mathbb{E}[\varepsilon_N(h)] = 0$ et

$$\mathbb{E}[\varepsilon_N(h) \varepsilon_N'(h)] \leq \Sigma + \mu \mu'.$$

Nous déduisons à nouveau de la loi du logarithme itéré de Hartman-Wintner que

$$\|M_N(h)\|^2 = \mathcal{O}\left(N \log(\log N)\right) \text{ p.s.}$$

Par suite

$$\|N_h m_{h,N} - N \mathbb{E}[X \mathbb{I}_{[Y \in s_h]}\|^2 = \mathcal{O}\left(N \log(\log N)\right) \text{ p.s.}$$

donc

$$\|m_{h,N} - \frac{N}{N_h} \mathbb{E}[X \mathbb{I}_{[Y \in s_h]}\|^2 = \mathcal{O}\left(\frac{N}{N_h^2} \log(\log N)\right) \text{ p.s.}$$

De plus, nous avons également

$$\frac{N_h}{N} \longrightarrow p_h \text{ p.s.} \quad \text{avec} \quad p_h = \mathbb{P}(Y \in s_h). \quad (\text{B.a})$$

Sous l'hypothèse (A2) : $p_1 > 0$ et $p_2 > 0$. Donc $N_h \sim p_h N$ et nous obtenons finalement

$$\|m_{h,N} - \frac{N}{N_h} \mathbb{E}[X \mathbb{I}_{[Y \in s_h]}\|^2 = \mathcal{O}\left(\frac{\log(\log N)}{N}\right) \text{ p.s.} \quad (\text{B.5})$$

Étudions maintenant le second terme de la majoration. Comme $m_h p_h = \mathbb{E}[X \mathbb{I}_{[Y \in s_h]}]$, nous avons

$$\frac{N}{N_h} \mathbb{E}[X \mathbb{I}_{[Y \in s_h]}] - m_h = \frac{N}{N_h} m_h p_h - m_h = \frac{m_h}{N_h} (N p_h - N_h). \quad (\text{B.b})$$

En outre, comme $N_h = \sum_{i=1}^N Z_i(h) + N p_h$, où $Z_i(h) = \mathbb{I}_{[y_i \in s_h]} - \mathbb{E}[\mathbb{I}_{[y_i \in s_h]}]$, nous avons

$$N_h - N p_h = M_N = \sum_{i=1}^N Z_i(h).$$

D'après la loi du logarithme itéré de Hartman-Wintner, nous avons

$$\|N_h - N p_h\|^2 = \mathcal{O}\left(N \log(\log N)\right) \text{ p.s.}$$

et en utilisant le fait que $N_h \sim p_h N$ (d'après (B.a)), nous déduisons de (B.b) que

$$\left\| \frac{N}{N_h} \mathbb{E}[X \mathbb{I}_{[Y \in s_h]}] - m_h \right\|^2 = \mathcal{O}\left(\frac{\log(\log N)}{N}\right) \text{ p.s.} \quad (\text{B.6})$$

Finalement, en utilisant les vitesses obtenues en (B.5) et (B.6), nous déduisons de (B.4) que

$$\|m_{h,N} - m_h\|^2 = \mathcal{O}\left(\frac{\log(\log N)}{N}\right) \quad \text{p.s.} \quad (\text{B.7})$$

Obtention de la vitesse de $(z_{h,N})$

À partir des vitesses de convergence données en (B.3) et (B.7), nous trouvons par (B.2) que

$$\|z_{h,N} - z_h\|^2 = \mathcal{O}\left(\frac{\log(\log N)}{N}\right) \quad \text{p.s.} \quad (\text{B.8})$$

Étape 2 : étude de la vitesse de convergence de (Σ_N^{-1})

D'après l'équation de Riccati pour l'inversion de matrice donnée à la page 96 de Duflo [5], nous pouvons écrire Σ_N^{-1} sous la forme

$$\Sigma_N^{-1} = \Sigma^{-1} - \Sigma^{-1}(\Sigma_N - \Sigma)\Sigma^{-1} + R_N$$

où $R_N = \Sigma^{-1}(\Sigma - \Sigma_N)\Sigma_N^{-1}(\Sigma - \Sigma_N)\Sigma^{-1}$.

Il est alors immédiat que

$$\|\Sigma_N^{-1} - \Sigma^{-1}\|^2 \leq 2\|R_N\|^2 + 2\|\Sigma^{-1}(\Sigma_N - \Sigma)\Sigma^{-1}\|^2. \quad (\text{B.9})$$

Étude de la vitesse de convergence de (Σ_N)

Comme

$$\begin{aligned} \Sigma_N &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}_N)(x_i - \bar{x}_N)' \\ &= \frac{1}{N} \sum_{i=1}^N (x_i - \mu)(x_i - \mu)' - (\bar{x}_N - \mu)(\bar{x}_N - \mu)' \end{aligned}$$

nous avons

$$\Sigma_N - \Sigma = \frac{1}{N} \sum_{i=1}^N e_i - (\bar{x}_N - \mu)(\bar{x}_N - \mu)',$$

où $e_i = (x_i - \mu)(x_i - \mu)' - \Sigma$.

Par conséquent, nous obtenons

$$\|\Sigma_N - \Sigma\|^2 \leq \frac{2}{N^2} \left\| \sum_{i=1}^N e_i \right\|^2 + 2\|\bar{x}_N - \mu\|^4. \quad (\text{B.10})$$

Posons $M_N(u) = u' \left(\sum_{i=1}^N e_i \right) u = \sum_{i=1}^N u' e_i u = \sum_{i=1}^N e_i(u)$, où $e_i(u) = u' e_i u = u'(x_i - \mu)(x_i - \mu)' u - u' \Sigma u$.

Nous avons

$$\mathbb{E}[e_i(u)] = \mathbb{E}[u'(x_i - \mu)(x_i - \mu)' u - u' \Sigma u] = u' \Sigma u - u' \Sigma u = 0.$$

Nous avons également

$$\begin{aligned} \mathbb{E}[(e_i(u))^2] &= \mathbb{E}[(u'(x_i - \mu)(x_i - \mu)' u - u' \Sigma u)^2] \\ &= \mathbb{E}[u'(x_i - \mu)(x_i - \mu)' u u' (x_i - \mu)(x_i - \mu)' u] - (u' \Sigma u)^2 \\ &\leq \mathbb{E}[(u'(x_i - \mu)(x_i - \mu)' u)^2] \\ &\leq \mathbb{E}[\|x_i - \mu\|^4] \times \|u\|^4. \end{aligned}$$

Nous pouvons déduire une nouvelle fois par la loi du logarithme itéré de Hartman-Wintner que

$$\limsup \frac{\|\sum_{i=1}^N e_i(u)\|^2}{2N \log(\log N)} \leq \tau^4 \times \|u\|^4 \quad \text{p.s.}$$

où $\tau^4 = \mathbb{E}[\|x_i - \mu\|^4]$.

Ceci étant vrai pour tout u , il en découle que

$$\left\| \sum_{i=1}^N e_i \right\|^2 = \mathcal{O}\left(N \log(\log N)\right) \quad \text{p.s.} \quad (\text{B.11})$$

À partir des vitesses de convergence obtenues en (B.3) et (B.11), nous déduisons de (B.10) que

$$\begin{aligned} \|\Sigma_N - \Sigma\|^2 &= \mathcal{O}\left(\frac{1}{N^2} N \log(\log N)\right) + \mathcal{O}\left(\left(\frac{\log(\log N)}{N}\right)^2\right) \quad \text{p.s.} \\ &= \mathcal{O}\left(\frac{\log(\log N)}{N}\right) \quad \text{p.s.} \end{aligned} \quad (\text{B.12})$$

Étude de la vitesse de convergence de (R_N)

Nous déduisons de l'expression de R_N que

$$\begin{aligned} \|R_N\| &\leq \lambda_{\max}(\Sigma_N^{-1}) \Sigma^{-1} (\Sigma - \Sigma_N)^2 \Sigma^{-1} \\ &\leq \frac{1}{\lambda_{\min}(\Sigma_N)} \Sigma^{-1} (\Sigma - \Sigma_N)^2 \Sigma^{-1} \\ &\leq \frac{1}{\lambda_{\min}(\Sigma_N)} \times \frac{\|\Sigma_N - \Sigma\|^2}{(\lambda_{\min}(\Sigma))^2}. \end{aligned}$$

À partir de (B.12), nous obtenons

$$\|R_N\| = \mathcal{O}\left(\frac{\log(\log N)}{N}\right) \quad \text{p.s.} \quad (\text{B.13})$$

Obtention de la vitesse de (Σ_N^{-1})

À partir des vitesses de convergence obtenues en (B.12) et (B.13), il est immédiat par (B.9) que

$$\|\Sigma_N^{-1} - \Sigma^{-1}\|^2 = \mathcal{O}\left(\frac{\log(\log N)}{N}\right) \quad \text{p.s.} \quad (\text{B.14})$$

Étape 3 : obtention de la vitesse de convergence de $\tilde{\theta}_N$

Finalement, à partir des vitesses de convergence obtenues en (B.8) et (B.14), nous déduisons de (B.1) que

$$\|\tilde{\theta}_N - \tilde{\theta}\|^2 = \mathcal{O}\left(\frac{\log(\log N)}{N}\right) \quad \text{p.s.}$$

d'où

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

□

Ce qui achève la preuve du Théorème 3.1.

Annexe C: démonstration du Théorème 3.2

Nous donnons ici tout d'abord quelques rappels sur la transformation de vectorisation, "vec", ainsi que sur le produit de Kronecker.

Soit $M = [m_1, \dots, m_p]$ une matrice de dimensions $q \times p$ où les m_k , $k = 1, \dots, p$, 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 :

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

Soient M et N deux matrices de dimensions respectives $q \times p$ et $r \times s$, alors le produit de Kronecker de M et N est la matrice partitionnée de dimensions $qr \times ps$ définie par : $M \otimes N = [m_{j,k}N]$ avec $j = 1, \dots, q$ et $k = 1, \dots, p$.

Une propriété importante reliant le produit de Kronecker et la transformation "vec" est :

$$\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$$

où les dimensions des matrices A , B et C sont telles que le produit matriciel est bien défini. D'autres propriétés sur la transformation "vec" et le produit de Kronecker peuvent être trouvées dans Magnus et Neudecker [8], [9].

Démonstration. Pour débiter la démonstration, nous rappelons que $\tilde{\theta}_N = \Sigma_N^{-1}(z_{1,N} - z_{2,N})$ et $\tilde{\theta} = \Sigma^{-1}(z_1 - z_2)$. Nous cherchons la loi asymptotique de $\sqrt{N}(\tilde{\theta}_N - \tilde{\theta}) = \Sigma_N^{-1}Z$, où $Z = \sqrt{N}(z_{1,N} - z_{2,N} - \Sigma_N \tilde{\theta})$.

En utilisant la définition de $z_{1,N}$, $z_{2,N}$, z_1 et z_2 , nous avons :

$$\begin{aligned} Z &= \sqrt{N}\{m_{1,N} - m_{2,N} - \Sigma_N \Sigma^{-1}(m_1 - m_2)\} \\ &= \sqrt{N}\{m_{1,N} - m_{2,N} - (\Sigma_N - \Sigma + \Sigma)\Sigma^{-1}(m_1 - m_2)\} \\ &= \sqrt{N}\{(m_{1,N} - m_1) - (m_{2,N} - m_2) - (\Sigma_N - \Sigma)\Sigma^{-1}(m_1 - m_2)\}. \end{aligned}$$

En utilisant la propriété reliant le produit de Kronecker et la transformation "vec", nous avons :

$$\begin{aligned} (\Sigma_N - \Sigma)\Sigma^{-1}(m_1 - m_2) &= \text{vec}(I_p(\Sigma_N - \Sigma)\Sigma^{-1}(m_1 - m_2)) \\ &= ((m_1 - m_2)' \Sigma^{-1} \otimes I_p)\text{vec}(\Sigma_N - \Sigma). \end{aligned}$$

Finalement, nous obtenons :

$$Z = \sqrt{N}\{(m_{1,N} - m_1) - (m_{2,N} - m_2) - D^* \text{vec}(\Sigma_N - \Sigma)\}, \quad (\text{C.1})$$

où $D^* = (m_1 - m_2)' \Sigma^{-1} \otimes I_p$.

Dans la suite de la démonstration, nous obtenons tout d'abord la loi de Z en deux étapes en utilisant le théorème central limite et la Delta méthode. Puis, dans une troisième étape, nous en déduisons la loi de $\sqrt{N}(\tilde{\theta}_N - \tilde{\theta})$.

Étape 1 : application du théorème central limite

Notons U_i le vecteur aléatoire de dimension $(2 + 2p + p + p^2)$ défini par :

$$U_i = (\mathbb{I}_{1(i)}, \mathbb{I}_{2(i)}, x_i' \mathbb{I}_{1(i)}, x_i' \mathbb{I}_{2(i)}, x_i', \text{vec}(x_i x_i'))',$$

où $\mathbb{I}_{h(i)} := \mathbb{I}[y_i \in s_h]$, avec $h = 1, 2$.

Sous les hypothèses du théorème, les vecteurs U_i , $i = 1, \dots, N$ sont indépendants et identiquement distribués d'espérance μ_U et de matrice de covariances Σ_U de la forme :

$$\mu_U = \mathbb{E}[U_i] = (p_1, p_2, \tilde{m}'_1, \tilde{m}'_2, \mu', \text{vec}(\Sigma + \mu \mu'))',$$

et

$$\Sigma_U = \mathbb{V}(U_i) = \begin{pmatrix} B_{11} & B_{12} & B_{13} & B_{14} \\ B'_{12} & B_{22} & B_{23} & B_{24} \\ B'_{13} & B'_{23} & B_{33} & B_{34} \\ B'_{14} & B'_{24} & B'_{34} & B_{44} \end{pmatrix},$$

où

$$\tilde{m}_1 := \mathbb{E}[x \mathbb{I}[y \in s_1]] \text{ et } \tilde{m}_2 := \mathbb{E}[x \mathbb{I}[y \in s_2]],$$

$$B_{11} = \begin{bmatrix} p_1(1-p_1) & -p_2 p_1 \\ -p_1 p_2 & p_2(1-p_2) \end{bmatrix},$$

$$B_{12} = \begin{bmatrix} (1-p_1)\tilde{m}'_1 & -p_1\tilde{m}'_2 \\ -p_2\tilde{m}'_1 & (1-p_2)\tilde{m}'_2 \end{bmatrix},$$

$$B_{13} = \begin{bmatrix} \tilde{m}'_1 - p_1 \mu' \\ \tilde{m}'_2 - p_2 \mu' \end{bmatrix},$$

$$B_{14} = \begin{bmatrix} \mathbb{E}[(x' \otimes x') \mathbb{I}[y \in s_1]] - p_1 \text{vec}(\Sigma + \mu \mu')' \\ \mathbb{E}[(x' \otimes x') \mathbb{I}[y \in s_2]] - p_2 \text{vec}(\Sigma + \mu \mu')' \end{bmatrix},$$

$$B_{22} = \begin{bmatrix} \mathbb{E}[x x' \mathbb{I}[y \in s_1]] - \tilde{m}_1 \tilde{m}'_1 & -\tilde{m}_1 \tilde{m}'_2 \\ -\tilde{m}_2 \tilde{m}'_1 & \mathbb{E}[x x' \mathbb{I}[y \in s_2]] - \tilde{m}_2 \tilde{m}'_2 \end{bmatrix},$$

$$B_{23} = \begin{bmatrix} \mathbb{E}[x x' \mathbb{I}[y \in s_1]] - \tilde{m}_1 \mu' \\ \mathbb{E}[x x' \mathbb{I}[y \in s_2]] - \tilde{m}_2 \mu' \end{bmatrix},$$

$$B_{24} = \begin{bmatrix} \mathbb{E}[x(x' \otimes x') \mathbb{I}[y \in s_1]] - \tilde{m}_1 \text{vec}(\Sigma + \mu \mu')' \\ \mathbb{E}[x(x' \otimes x') \mathbb{I}[y \in s_2]] - \tilde{m}_2 \text{vec}(\Sigma + \mu \mu')' \end{bmatrix},$$

$$B_{33} = \Sigma,$$

$$B_{34} = \mathbb{E}[x(x' \otimes x')] - \mu \text{vec}(\Sigma + \mu \mu')'.$$

$$B_{44} = \mathbb{E}[(x x') \otimes (x x')] - \text{vec}(\Sigma + \mu \mu') \text{vec}(\Sigma + \mu \mu')',$$

Définissons le vecteur $\bar{U} = \frac{1}{N} \sum_{i=1}^N U_i = (p_{1,N}, p_{2,N}, \tilde{m}'_{1,N}, \tilde{m}'_{2,N}, \bar{x}', \frac{1}{N} \sum_{i=1}^N \text{vec}(x_i x_i'))'$ où $\tilde{m}_{h,N} = \frac{1}{N} \sum_{i=1}^N x_i \mathbb{I}_{h(i)}$ pour $h = 1, 2$. Par le théorème central limite multidimensionnel, nous obtenons la distribution asymptotique suivante

$$\sqrt{N}(\bar{U} - \mu_U) \xrightarrow{\mathcal{L}} \mathcal{N}_{2+3p+p^2}(0, \Sigma_U).$$

Étape 2 : application de la Delta méthode

A partir de \bar{U} , nous formons le vecteur colonne

$$\bar{U}_1 = (p_{1,N}, p_{2,N}, m'_{1,N}, m'_{2,N}, \bar{x}', \frac{1}{N} \sum_{i=1}^N \text{vec}(x_i x_i'))',$$

où nous rappelons que $m_{h,N} = \frac{1}{N p_{h,N}} \sum_{i=1}^N x_i \mathbb{I}_{h(i)} = \frac{1}{p_{h,N}} \tilde{m}_{h,N}$ pour $h = 1, 2$. Par la Delta méthode, nous allons déduire la distribution asymptotique de \bar{U}_1 à partir de celle de \bar{U} .

Soit la fonction h_1 définie par

$$\begin{aligned} h_1 : \mathbb{R}^{2+3p+p^2} &\longrightarrow \mathbb{R}^{2+3p+p^2} \\ u = (a_1, a_2, b'_1, b'_2, c', d')' &\longmapsto (a_1, a_2, b'_1/a_1, b'_2/a_2, c', d')' \end{aligned}$$

En remarquant que $\bar{U}_1 = h_1(\bar{U})$ et que $\mu_{U_1} := h_1(\mu_U) = (p_1, p_2, m'_1, m'_2, \mu', \text{vec}(\Sigma + \mu \mu'))'$, nous obtenons par la Delta méthode

$$\sqrt{N}(h_1(\bar{U}) - h_1(\mu_U)) \xrightarrow{\mathcal{L}} \mathcal{N}_{2+3p+p^2}(0, J_1 \Sigma_U J_1') \quad (\text{C.2})$$

avec $J_1 = \partial h_1 / \partial u |_{\mathbb{E}}$, où la notation $g |_{\mathbb{E}}$ désigne la valeur de la fonction g en l'espérance de son argument.

Notons I_s la matrice identité de dimensions $s \times s$ et $0_{s_1, s_2}$ la matrice nulle de dimension $s_1 \times s_2$. Après quelques calculs matriciels, nous obtenons

$$J_1' = \partial h_1' / \partial u |_{\mathbb{E}} = \begin{pmatrix} I_2 & A_1 & 0_{2,p} & 0_{2,p^2} \\ 0_{2p,2} & A_2 & 0_{2p,p} & 0_{2p,p^2} \\ 0_{p,2} & 0_{p,2p} & I_p & 0_{p,p^2} \\ 0_{p^2,2} & 0_{p^2,2p} & 0_{p^2,p} & I_{p^2} \end{pmatrix},$$

$$\text{avec } A_1 = \begin{bmatrix} -\tilde{m}'_1/p_1^2 & 0_{1,p} \\ 0_{1,p} & -\tilde{m}'_2/p_2^2 \end{bmatrix} \text{ et } A_2 = \begin{bmatrix} \frac{1}{p_1} I_p & 0_{p,p} \\ 0_{p,p} & \frac{1}{p_2} I_p \end{bmatrix}.$$

Nous en déduisons l'expression de la matrice de covariance de la distribution asymptotique de \bar{U}_1 :

$$\Delta_1 = J_1 \Sigma_U J_1' = \begin{pmatrix} B_{11} & 0_{2,2p} & B_{13} & B_{14} \\ 0_{2p,2} & B_{22}^* & B_{23}^* & B_{24}^* \\ B'_{13} & B'_{23}^* & B_{33} & B_{34} \\ B'_{14} & B'_{24}^* & B'_{34} & B_{44} \end{pmatrix}.$$

où $B_{22}^* = A_1' B_{11} A_1 + A_1' B_{12} A_2 + A_2 B'_{12} A_1 + A_2 B_{22} A_2$; $B_{23}^* = A_1' B_{13} + A_2 B_{23}$ et $B_{24}^* = A_1' B_{14} + A_2 B_{24}$.

Considérons maintenant la fonction h_2 :

$$\begin{aligned} h_2 : \mathbb{R}^{2+3p+p^2} &\longrightarrow \mathbb{R}^{p+p+p^2} \\ u = (a_1, a_2, b'_1, b'_2, c', \text{vec}(d))' &\longmapsto (b'_1, b'_2, (\text{vec}(d) - \text{vec}(cc')))' \end{aligned}$$

En remarquant que $h_2(\bar{U}_1) = \begin{pmatrix} m_{1,N} \\ m_{2,N} \\ \text{vec}(\Sigma_N) \end{pmatrix}$ et $h_2(\mu_{U_1}) = \begin{pmatrix} m_1 \\ m_2 \\ \text{vec}(\Sigma) \end{pmatrix}$, nous déduisons par la Delta méthode que

$$\sqrt{N} \left(\begin{pmatrix} m_{1,N} \\ m_{2,N} \\ \text{vec}(\Sigma_N) \end{pmatrix} - \begin{pmatrix} m_1 \\ m_2 \\ \text{vec}(\Sigma) \end{pmatrix} \right) \xrightarrow{\mathcal{L}} \mathcal{N}_{2p+p^2}(0, \Delta_2),$$

où

$$\Delta_2 = J_2 \Delta_1 J_2' \quad (\text{C.3})$$

avec $J_2' = \partial h_2' / \partial u |_{\mathbb{E}}$. Après quelques calculs de dérivées vectorielles, nous obtenons

$$J_2' = \begin{pmatrix} 0_{2,p} & 0_{2,p} & 0_{2,p^2} \\ I_p & 0_{p,p} & 0_{p,p^2} \\ 0_{p,p} & I_p & 0_{p,p^2} \\ 0_{p,p} & 0_{p,p} & -(I_p \otimes \mu' + \mu' \otimes I_p) \\ 0_{p^2,p} & 0_{p^2,p} & I_{p^2} \end{pmatrix}.$$

Puisque $Z = \sqrt{N} \{ (m_{1,N} - m_1) - (m_{2,N} - m_2) - D^* \text{vec}(\Sigma_N - \Sigma) \}$, introduisons la matrice de dimension $p \times (2p + p^2)$ suivante : $E = \begin{pmatrix} I_p & -I_p & -D^* \end{pmatrix}$. Nous avons alors

$$Z = E \sqrt{N} \left(\begin{pmatrix} m_{1,N} \\ m_{2,N} \\ \text{vec}(\Sigma_N) \end{pmatrix} - \begin{pmatrix} m_1 \\ m_2 \\ \text{vec}(\Sigma) \end{pmatrix} \right)$$

La loi asymptotique de Z est donc

$$Z \xrightarrow{\mathcal{L}} \mathcal{N}_{2p+p^2}(0, \Delta_3),$$

où

$$\Delta_3 = E \Delta_2 E'. \quad (\text{C.4})$$

Étape 3 : loi asymptotique de $\sqrt{N}(\tilde{\theta}_N - \tilde{\theta})$

Nous avons $\sqrt{N}(\tilde{\theta}_N - \tilde{\theta}) = \Sigma_N^{-1} Z$. Vu que Σ_N^{-1} converge en probabilité vers Σ^{-1} et que Z est asymptotiquement normal de moyenne nulle et de matrice de covariances Δ_3 , nous avons finalement

$$\sqrt{N}(\tilde{\theta}_N - \tilde{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma^{-1} \Delta_3 \Sigma^{-1}).$$

□

Chapitre 3

Approche récursive ou non de SIR basée sur une seule tranche

Dans ce chapitre, nous allons développer une approche (récursive et non récursive) de la méthode SIR basée sur l'utilisation de l'information contenue dans une seule tranche (optimale en un certain sens), dans le cas d'un nombre H quelconque de tranches. Nous proposons aussi une façon choisir le nombre H de tranches. Pour cette approche, des résultats théoriques ont été obtenus. Le bon comportement des estimateurs (récursifs et non récursifs) a été mis en évidence sur des simulations numériques. Comme précédemment, les temps de calculs de la version récursive de l'estimateur sont beaucoup plus courts que ceux de la version non récursive ou ceux de la méthode SIR classique.

L'approche proposée dans ce chapitre a donné lieu à l'article intitulé "Recursive and non recursive versions for SIR and SIRoneslice (a new one slice- based SIR approach)" écrit en collaboration avec Bernard Bercu et Jérôme Saracco. Cet article est soumis et disponible à la fin de ce chapitre.

3.1 Principe

Nous considérons à nouveau le modèle semi-paramétrique à un seul indice décrit au chapitre 1 lorsque la covariable x appartient à \mathbb{R}^p et la variable à expliquer y est à valeur dans \mathbb{R} . Nous rappelons que ce modèle s'écrit sous la forme

$$y = f(\beta'x) + \varepsilon, \quad \text{où } \varepsilon \perp x, \quad (3.1.1)$$

la distribution de ε est inconnue et arbitraire, f est une fonction inconnue à valeurs dans \mathbb{R} . Une réduction de dimension est incluse dans ce modèle par le biais d'un indice $x'\beta$. Nous avons deux objectifs principaux dans ce chapitre. Tout d'abord, nous proposons un estimateur de la direction de β basé sur l'information contenue dans une seule tranche (optimale en un certain sens qu'il faudra choisir parmi les H tranches). Nous appelons cette nouvelle méthode "SIRoneslice". Puis, nous fournissons la version récursive de SIRoneslice. Nous donnons ensuite un résultat asymptotique de l'approche SIRoneslice. Une étude sur simulations montre la bonne performance numérique de la méthode SIRoneslice et présente clairement l'avantage principal de l'utilisation de la version récursive de SIR et SIRoneslice du point de vue des temps de calculs.

3.2 Approche SIR basée sur une seule tranche : “SIRoneslice”

Dans le cadre du modèle (3.1.1) et sous la condition de linéarité (LC), pour chaque tranche h , nous avons

$$\mathbb{E}[x \mid y \in s_h] = \mu + k_h \Sigma \beta,$$

où

$$k_h = \frac{\mathbb{E}[(x - \mu)' \beta \mid y \in s_h]}{\beta' \Sigma \beta}.$$

Posons $z_h := \mathbb{E}[x \mid y \in s_h] - \mu = m_h - \mu$ pour $h = 1, \dots, H$. Nous en déduisons que

$$z_h = k_h \Sigma \beta \quad \text{et} \quad k_h = \frac{z_h' \beta}{\beta' \Sigma \beta}. \quad (3.2.1)$$

Ainsi z_h appartient au sous-espace de dimension 1 engendré par $\Sigma \beta$ si $k_h \neq 0$. À partir de (3.2.1), pour toute tranche h telle que le scalaire k_h est non nul, nous pouvons estimer la direction EDR d'un échantillon d'observations $\{(x_i, y_i), i = 1, \dots, n\}$, en utilisant l'estimateur suivant :

$$\hat{b}_{h,n} := \hat{\Sigma}_n^{-1} \hat{z}_{h,n},$$

avec $\hat{z}_{h,n} = \hat{m}_{h,n} - \bar{x}_n$.

Notons que, contrairement à l'approche SIR classique, nous n'utilisons pas l'information des H tranches dans l'approche SIRoneslice. Nous ne nous concentrons que sur une tranche h pour laquelle la valeur de k_h est non nulle. Dans certaines situations, par exemple, lorsque le modèle de régression est partiellement symétrique, l'estimateur $\hat{b}_{h,n}$ obtenu à partir d'une certaine tranche h est plus performant que l'estimateur SIR classique. Définissons maintenant comment trouver une tranche h° optimale en un certain sens.

Dans la suite, nous supposons que β est telle que $\|\beta\|_{\Sigma} = 1$. Ainsi, le terme k_h est donné par $k_h = (m_h - \mu)' \beta = z_h' \beta$ mais dépend du paramètre inconnu β . Toutefois, en utilisant $\|\beta\|_{\Sigma} = 1$, nous déduisons de (3.2.1) que

$$(k_h)^2 = \|z_h\|_{\Sigma^{-1}}^2.$$

À partir de ce résultat, nous pouvons proposer une tranche optimale pour l'approche SIRoneslice définie par :

$$h^\circ = \arg \max_h (k_h)^2.$$

La version sur population de l'estimateur (Σ -normalisé) correspondant est la suivante :

$$\beta_{h^\circ} := \frac{\Sigma^{-1} z_{h^\circ}}{\|z_{h^\circ}\|_{\Sigma^{-1}}}.$$

Nous obtenons ensuite la version sur échantillon de l'estimateur correspondant (qui est $\hat{\Sigma}_n$ -normalisé) :

$$\hat{\beta}_{\hat{h}_n^\circ, n} := \frac{\hat{\Sigma}_n^{-1} \hat{z}_{\hat{h}_n^\circ, n}}{\|\hat{z}_{\hat{h}_n^\circ, n}\|_{\hat{\Sigma}_n^{-1}}},$$

où $\hat{h}_n^o = \arg \max_h (\hat{k}_{h,n})^2$ avec $(\hat{k}_{h,n})^2 = \|\hat{z}_{h,n}\|_{\hat{\Sigma}_n^{-1}}^2$.

Nous pouvons aussi considérer un estimateur alternatif de la direction β (qui n'est pas $\hat{\Sigma}_n$ -normalisé) :

$$\hat{b}_{\hat{h}_n^o,n} := \hat{\Sigma}_n^{-1} \hat{z}_{\hat{h}_n^o,n}.$$

Par la Loi des Grandes Nombres, nous pouvons démontrer que $\hat{\beta}_{\hat{h}_n^o,n}$ converge presque sûrement vers $\pm\beta$.

3.3 Version récursive de SIRoneslice

Nous proposons une version récursive de l'estimateur $\hat{b}_{\hat{h}_n^o,n}$:

$$\begin{aligned} \hat{b}_{\hat{h}_n^o,n} &= \frac{n}{n-1} \hat{\Sigma}_{n-1}^{-1} \hat{z}_{\hat{h}_n^o,n-1} - \frac{1}{n-1} \hat{\Sigma}_{n-1}^{-1} \Phi_n - \frac{1}{(n-1)(n+\rho_n)} \hat{\Sigma}_{n-1}^{-1} \Phi_n \Phi_n' \hat{\Sigma}_{n-1}^{-1} (n \hat{z}_{\hat{h}_n^o,n-1} - \Phi_n) \\ &+ \frac{1}{n-1} \Phi_{\hat{h}_n^o,n}' \left(\hat{\Sigma}_{n-1}^{-1} - \frac{1}{(n+\rho_n)} \hat{\Sigma}_{n-1}^{-1} \Phi_n \Phi_n' \hat{\Sigma}_{n-1}^{-1} \right) \mathbb{I}_{[h^*=\hat{h}_n^o]}, \end{aligned}$$

où h^* désigne la tranche contenant la nouvelle observation (x_n, y_n) .

Nous avons obtenu un résultat asymptotique pour l'estimateur proposé. Les hypothèses suivantes sont nécessaires pour établir la convergence presque sûre de notre estimateur.

- (A1) Les observations $(x_i, y_i), i = 1, \dots, n$, sont échantillonnées de manière indépendante à partir du modèle (3.1.1).
- (A2) Le support de y est partitionné en H tranches fixes s_1, \dots, s_H telles que $p_h \neq 0$ pour $h = 1, \dots, H$.

En utilisant la loi du logarithme itéré de Hartman-Wintner et appliquant l'équation de Riccati pour la matrice inverse, nous montrons le théorème suivant.

Théorème 3.1. *Sous la condition de linéarité (LC) et les hypothèses (A₁) et (A₂), nous avons*

$$\|\hat{b}_{\hat{h}_n^o,n} - b_{h^o}\| = \mathcal{O} \left(\sqrt{\frac{\log(\log n)}{n}} \right) \quad p.s.$$

La démonstration de ce théorème figure en ‘‘Appendix’’ de l'article.

3.4 Etude sur simulations

Nous illustrons sur des simulations les bonnes performances numériques de la méthode SIRoneslice. Nous précisons également que le nombre H de tranches a une influence sur la qualité d'estimation. Nous proposons alors un critère ‘‘bootstrap naïf’’ pour le choix du nombre H de tranches. Avec un nombre H de tranches choisi par le critère ‘‘bootstrap naïf’’ proposé, nous avons fait une étude pour comparer les deux méthodes SIRoneslice et SIR. Les résultats nous montrent que la qualité de l'estimateur de la méthode SIRoneslice est meilleure que celle de la méthode SIR sauf dans le cas où le modèle ne présente aucune dépendance symétrique (les deux méthodes sont alors

très proches). Nous présentons ensuite l'évolution, en fonction de nombre n d'observation, de la qualité de l'estimateur SIRoneslice en version récursive. Les résultats nous montrent le bon comportement de l'estimateur récursif proposé dans le cas de différents types de modèles (avec ou sans dépendance symétrique). Nous montrons clairement un avantage majeur de l'utilisation de la version récursive des méthodes SIR et SIRoneslice du point de vue des temps de calculs. Les différents résultats de simulations sont présentés par le biais de graphiques et de tableaux dans l'article et y sont abondamment commentés.

Les simulations ont été faites à l'aide du logiciel R.

Recursive and non recursive versions for SIR and SIRoneslice (a new one slice-based SIR approach)

Bernard BERCU^{1,2}, Thi Mong Ngoc NGUYEN^{1,3}, Jérôme SARACCO^{1,3,4}

¹ Institut de Mathématiques de Bordeaux, UMR CNRS 5251, Université de Bordeaux,
351 cours de la libération, 33405 Talence Cedex, France.

² INRIA Bordeaux Sud-Ouest, ALEA team, France.

³ INRIA Bordeaux Sud-Ouest, CQFD team, France.

⁴ Institut Polytechnique de Bordeaux, 1 avenue du Dr Albert Schweitzer, 33402 Talence Cedex, France.

e-mail : {Bernard.Bercu,Thi.Mong.Ngoc.Nguyen,Jerome.Saracco}@math.u-bordeaux1.fr

ABSTRACT

We consider a semiparametric single index regression model involving a p -dimensional quantitative covariable x and a real dependent variable y . A dimension reduction is included in this model via an index $x'\beta$. Sliced inverse regression (SIR) is a well-known method to estimate the direction of the euclidean parameter β which is based on a “slicing step” of y in the population and sample versions. The goal of this paper is twice. We first propose an estimator of the direction of β based on the use of only one “optimal” slice chosen among the H slices. We call this new method SIRoneslice. Then we provide the recursive versions of the SIR and SIRoneslice estimators. We give an asymptotic result for SIRoneslice approach. Simulation study shows good numerical performances of SIRoneslice method and clearly exhibits the main advantage of using recursive versions of the SIR and SIRoneslice methods from a computational times point of view. Some extensions are discussed in concluding remarks. The proposed methods and criterion have been implemented in R and the corresponding codes are available from the authors.

Keywords: *recursive estimation, semiparametric regression model, sliced inverse regression (SIR).*

1 Introduction

In statistical applications, high-dimensional data became common. In a regression framework, let $y \in \mathbb{R}$ be the response variable and $x \in \mathbb{R}^p$ be the regressor. In order to study the relationship

between y and x , it is usual to impose assumptions for a specific structure on the mean regression function of y on x such as linearity or additivity. To cope with dimensionality, in a dimension reduction setting, many authors suppose that x can be replaced by a linear combination of its components, $\beta'x$, without losing information on the conditional distribution of y given x . This assumption can be expressed as

$$y \perp x \mid \beta'x \tag{1}$$

where the notation $u_1 \perp u_2 \mid u_3$ means that the random variable u_1 is independent of the random variable u_2 given any values for the random variable u_3 . An example where (1) holds is the following single index model with an additive error:

$$y = f(\beta'x) + \varepsilon, \tag{2}$$

where $\varepsilon \perp x$, the distribution of ε is arbitrary and unknown, and f is an unknown real-valued function. In this model, sufficient dimension reduction of the regression is achieved and leads to summary plot of y versus $\beta'x$ which provides graphical modelling information. Then, it is possible to nonparametrically estimate the link function f on the reduced data, more efficiently in the sense that the curse of dimensionality has been overcome. Finally, when (1) holds, it straightforwardly also holds for any vector colinear to β . Let us call $E := \text{span}(\beta)$ the dimension reduction subspace, also named effective dimension reduction (EDR) space following Duan and Li (1991) or Li (1991) in their initial presentation of sliced inverse regression (SIR).

Methods based on the use of inverse regression are available in literature for estimating the EDR space. In order for inverse regression to be useful in estimating the EDR space, some of them, like SIR and related methods (for instance principal Hessian direction introduced by Li (1992) or sliced average variance estimation discussed by Cook (2000)) place certain conditions on the marginal distribution of the covariable x . In this paper, we will focus on SIR. Let us then now recall the theory of SIR and its necessary assumption on x . The SIR approach relies on the following assumption often called linearity condition (LC):

$$\text{For all } b \in \mathbb{R}^p, \mathbb{E}[b'x \mid \beta'x] \text{ is linear in } x'\beta. \tag{3}$$

The LC is required to hold only for the true direction EDR. Since β is unknown, in practice it is not possible to verify a priori this assumption. Hence, we can assume that LC holds for all possible vector β , which is equivalent to elliptical symmetry of the distribution of x . Recall that the well-known multivariate normal distribution is an example of such a distribution. Finally

Hall and Li (1993) mentioned that the LC is not a severe restriction because this LC holds to a good approximation in many problems as the dimension p of the predictors increases, see also Chen and Li (1998) or Cook and Ni (2005) for interesting discussions on the LC.

Let us consider a montone transformation T . Under model (1) and LC, Duan an Li (1991) showed that the centered inverse regression curve satisfies:

$$\mathbb{E}[x | T(y)] - \mu \in \text{Span}(\Sigma\beta), \quad (4)$$

where $\mu := \mathbb{E}[x]$ and $\Sigma := \mathbb{V}(x)$. This result implies that the space spanned by the centered inverse curve, $\{\mathbb{E}[x | T(y)] - \mathbb{E}[x] : y \in \mathcal{Y}\}$ where \mathcal{Y} is the support of response variable y , is a subspace of the EDR space but it does not guarantee equality. For instance, pathological model has been identified in the literature and are called symme-tric dependent model: that is model for which the centered inverse regression curve is degenerated.

From (4), the centered inverse regression curve can be used to recover the EDR space (if the model is not pathological). Indeed, a direct consequence of this result is that the covariance matrix of this curve,

$$\Gamma := \mathbb{V}(\mathbb{E}[x | T(y)]),$$

is degenerate in any direction Σ -orthogonal to β . Therefore, the eigenvector b associated with the non null eigenvalue of $\Sigma^{-1}\Gamma$ is an EDR direction, that is $\text{Span}(b) = E$.

In the slicing step of SIR, the range of y is partitioned into H slices $\{s_1, \dots, s_H\}$. With such slicing, the covariance matrix Γ can be straightforwardly written this way: $\Gamma := \sum_{h=1}^H p_h(m_h - \mu)(m_h - \mu)'$ where $p_h = P(y \in s_h)$ and $m_h = \mathbb{E}[x | y \in s_h]$. Let us consider a random sample $\{(x_i, y_i), i = 1, \dots, n\}$ from model (1). By substituting empirical versions of μ , Σ , p_h and m_h for their theoret-ical counterparts, we obtain an estimated basis of E spanned by the eigenvector \hat{b}_{SIR} associated with the largest eigenvalue of the estimate $\hat{\Sigma}_n^{-1}\hat{\Gamma}_n$ of $\Sigma^{-1}\Gamma$ where

$$\hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)(x_i - \bar{x}_n)' \quad \text{and} \quad \hat{\Gamma}_n = \sum_{h=1}^H \hat{p}_{h,n}(\hat{m}_{h,n} - \bar{x}_n)(\hat{m}_{h,n} - \bar{x}_n)',$$

with $\bar{x}_n = \sum_{i=1}^n x_i/n$, $\hat{p}_{h,n} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}[y_i \in s_h] = \frac{\hat{n}_{h,n}}{n}$ and $\hat{m}_{h,n} = \frac{1}{\hat{n}_{h,n}} \sum_{i \in s_h} x_i$, the notation $\mathbb{I}[\cdot]$ designating the indicator function. This approach is the one proposed by Duan and Li (1991) and Li (1991) when they initially introduced SIR approach. The SIR method has been extensively studied by several authors, see for instance Hsing and Carroll (1992), Zhu and Ng (1995), Saracco (1997), Chen and Li (1998), Cook and Ni (2005) among others.

The goal of this paper is twice. We first propose in Section 2 an estimator of the direction of β based on the use of only one “optimal” slice chosen among the H slices. We will call this new method “SIRone-slice”. Then we provide in Section 3 the recursive versions of the SIR and SIRoneslice estimators. The recursive approach (which has never been introduced in SIR and related methods) relies on the use of the estimator defined on the $n - 1$ first observations (x_i, y_i) and the new observation (x_n, y_n) . We give in this section an asymptotic result for SIRoneslice estimator. The proof is deferred in the Appendix. In Section 4, a simulation study shows good numerical performances of SIRoneslice method and clearly exhibits the main advantage of using recursive versions of the SIR and SIRoneslice methods from a computational times point of view. All the methods have been implemented in R. Finally concluding remarks are discussed in Section 5.

2 A “one slice-based SIR” approach

Under model (1) and the LC, $\mathbb{E}[x | T(y)]$ can be written as, for each slice h ,

$$\mathbb{E}[x | y \in s_h] = \mu + k_h \Sigma \beta, \quad (5)$$

where

$$k_h = \frac{\mathbb{E}[(x - \mu)' \beta | y \in s_h]}{\beta' \Sigma \beta}. \quad (6)$$

Let $z_h := \mathbb{E}[x | y \in s_h] - \mu = m_h - \mu$ for $h = 1, \dots, H$. We have respectively from (5) and (6)

$$z_h = k_h \Sigma \beta \quad \text{and} \quad k_h = \frac{z_h' \beta}{\beta' \Sigma \beta}. \quad (7)$$

Then z_h falls along a line spanned by $\Sigma \beta$ since k_h is nonzero. From (7), if the scalar k_h is non null, we can estimate an EDR direction from a random sample $\{(x_i, y_i), i = 1, \dots, n\}$, using the following estimator:

$$\hat{b}_{h,n} := \hat{\Sigma}_n^{-1} \hat{z}_{h,n},$$

with $\hat{z}_{h,n} = \hat{m}_{h,n} - \bar{x}_n$. If k_h is non null, $\hat{b}_{h,n}$ is a \sqrt{n} -consistent estimator of the EDR direction since it can easily be shown that $\hat{b}_{h,n}$ converges to $k_h \beta$ at root n rate.

Note that, contrary to the usual SIR approach, we do not use in this SIRoneslice approach the information from all the H slices. We only focus on one slice \tilde{h} for which $k_{\tilde{h}}$ is non null. In some situations, for instance when the regression model is partially symmetric, the estimator $\hat{b}_{\tilde{h},n}$ obtained for the “best” slice \tilde{h} performs better than \hat{b}_{SIR} . This point is illustrated in the simulation study described in Section 4. One choice of this best slice will be discussed hereafter.

In the following, let us assume that β is such that $\|\beta\|_{\Sigma} = 1$. Then the term k_h is given by $k_h = (m_h - \mu)' \beta = z_h' \beta$ but depends on the unknown index parameter β . However, from (7), we have $\beta = \frac{1}{k_h} \Sigma^{-1} z_h$, and using $\|\beta\|_{\Sigma} = 1$, we get $\frac{1}{(k_h)^2} z_h' \Sigma^{-1} \Sigma \Sigma^{-1} z_h = 1$ and then

$$(k_h)^2 = \|z_h\|_{\Sigma^{-1}}^2,$$

which can be easily estimated from the sample data. From this result, we can propose an “optimal” slice for the SIRoneslice approach defined by:

$$h^o = \arg \max_h (k_h)^2, \quad (8)$$

and the corresponding population version of the estimator is then

$$\beta_{h^o} := \frac{\Sigma^{-1} z_{h^o}}{\|z_{h^o}\|_{\Sigma^{-1}}}.$$

From a computational point of view, it is easy to estimate each k_h from a sample by substituting z_h and Σ by their sample versions $\hat{z}_{h,n}$ and $\hat{\Sigma}_n$, then we get: $(\hat{k}_{h,n})^2 = \|\hat{z}_{h,n}\|_{\hat{\Sigma}_n^{-1}}^2$. From these values, we can define

$$\hat{h}_n^o = \arg \max_h (\hat{k}_{h,n})^2. \quad (9)$$

With this choice of slice, the corresponding $\hat{\Sigma}_n$ -normalized estimator is

$$\hat{\beta}_{\hat{h}_n^o, n} := \frac{\hat{\Sigma}_n^{-1} \hat{z}_{\hat{h}_n^o, n}}{\|\hat{z}_{\hat{h}_n^o, n}\|_{\hat{\Sigma}_n^{-1}}}. \quad (10)$$

By the law of large numbers, it can be shown that $\hat{\beta}_{\hat{h}_n^o, n}$ converges almost surely to $\pm\beta$.

Straightforwardly, an alternative estimator of the direction of β (which is not $\hat{\Sigma}_n$ -normalized) is given by

$$\hat{b}_{\hat{h}_n^o, n} := \hat{\Sigma}_n^{-1} \hat{z}_{\hat{h}_n^o, n}. \quad (11)$$

3 Recursive versions of SIR and “one slice-based SIR” approaches

We first present in Section 3.1 the recursive approach and we give recursive expressions of \bar{x}_n , $\hat{\Sigma}_n$, $\hat{\Sigma}_n^{-1}$, $\hat{p}_{h,n}$ and $\hat{z}_{h,n}$. Then, we define in Section 3.2 the recursive SIR estimator of the EDR direction as the major eigenvector of the recursive expression of the matrix of interest $\hat{\Sigma}_n^{-1} \hat{\Gamma}_n$. Then, we propose in Section 3.3 the recursive version of the SIRoneslice estimator. The almost surely convergence at $\sqrt{\log(\log n)/n}$ rate of the SIRoneslice estimator is obtained in Section 3.4.

3.1 Description of the recursive approach

Let us consider that the sample $\{(x_i, y_i), i = 1, \dots, n\}$ is splitted into two subsets : the subsample of the first $(n - 1)$ observations $\{(x_i, y_i), i = 1, \dots, n - 1\}$ and the new observation (x_n, y_n) .

Let us first give the recursive form of the estimators \bar{x}_n , $\hat{\Sigma}_n$ and $\hat{\Sigma}_n^{-1}$ of μ , Σ and Σ^{-1} .

The recursive expression of the empirical mean \bar{x}_n is the following:

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i = \frac{n-1}{n} \bar{x}_{n-1} + \frac{1}{n} x_n = \bar{x}_{n-1} + \frac{1}{n} \Phi_n \quad (12)$$

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

In the first term we can observe the presence of \bar{x}_{n-1} , the empirical mean of the first $(n - 1)$ observations x_i , and in the second one, the presence of the n th observation x_n . Similarly, the recursive form of the empirical covariance matrix $\hat{\Sigma}_n$ is given by:

$$\begin{aligned} \hat{\Sigma}_n &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)(x_i - \bar{x}_n)' \\ &= \frac{n-1}{n} \hat{\Sigma}_{n-1} + \frac{n-1}{n^2} (x_n - \bar{x}_{n-1})(x_n - \bar{x}_{n-1})' \\ &= \frac{n-1}{n} \hat{\Sigma}_{n-1} + \frac{n-1}{n^2} \Phi_n \Phi_n'. \end{aligned} \quad (13)$$

It can be shown, via Riccati equation, that the inverse of $\hat{\Sigma}_n$ has the following recursive form:

$$\hat{\Sigma}_n^{-1} = \frac{n}{n-1} \hat{\Sigma}_{n-1}^{-1} - \frac{n}{(n-1)(n+\rho_n)} \hat{\Sigma}_{n-1}^{-1} \Phi_n \Phi_n' \hat{\Sigma}_{n-1}^{-1} \quad (14)$$

where $\rho_n = \Phi_n' \hat{\Sigma}_{n-1}^{-1} \Phi_n$.

Let us now give the recursive form of the estimators $\hat{p}_{h,n}$ and $\hat{m}_{h,n}$ of p_h and m_h . Let us denote by h^* the slice containing y_n , the n th observation of y . We can now give the recursive expression of the estimator $\hat{p}_{h,n}$ of p_h :

$$\hat{p}_{h,n} = \begin{cases} \frac{n-1}{n} \hat{p}_{h^*,n-1} + \frac{1}{n} & \text{if } h = h^*, \\ \frac{n-1}{n} \hat{p}_{h,n-1} & \text{otherwise.} \end{cases} \quad (15)$$

The recursive form of $\hat{m}_{h,n}$ is given by:

$$\hat{m}_{h,n} = \begin{cases} \hat{m}_{h^*,n-1} + \frac{1}{\hat{n}_{h^*,n-1}+1} \Phi_{h^*,n} & \text{if } h = h^*, \\ \hat{m}_{h,n-1} & \text{otherwise,} \end{cases} \quad (16)$$

where $\Phi_{h^*,n} = x_n - \hat{m}_{h^*,n-1}$.

From (12) and (16), we obtain the recursive form of $\hat{z}_{h,n}$:

$$\hat{z}_{h,n} = \begin{cases} \hat{z}_{h^*,n-1} - \frac{1}{n}\Phi_n + \frac{1}{\hat{n}_{h^*,N-1}+1}\Phi_{h^*,n} & \text{if } h = h^*, \\ \hat{z}_{h,n-1} - \frac{1}{n}\Phi_n & \text{otherwise.} \end{cases} \quad (17)$$

3.2 Recursive SIR

After to have the recursive forms of \bar{x}_n , $\hat{p}_{h,n}$, $\hat{\Sigma}_n^{-1}$ and $\hat{z}_{h,n}$, we describe the recursive form of the matrix interest $\hat{\Sigma}_n^{-1}\hat{\Gamma}_n$ so that determine the eigenvector associated with the largest eigenvalue of this matrix.

In writing the form $\hat{\Gamma}_n$

$$\hat{\Gamma}_n = \sum_{h \neq h^*} \hat{p}_{h,n} \hat{z}_{h,n} \hat{z}'_{h,n} + \hat{p}_{h^*,n} \hat{z}_{h^*,n} \hat{z}'_{h^*,n},$$

we can deduce from the relations (15) and (17) the following recursive form for $\hat{\Gamma}_n$:

$$\begin{aligned} \hat{\Gamma}_n &= \frac{n-1}{n} \hat{\Gamma}_{n-1} - \frac{n-1}{n^2} \sum_{h=1}^H \hat{p}_{h,n-1} (\hat{z}_{h,n-1} \Phi'_n + \Phi_n \hat{z}'_{h,n-1}) \\ &\quad + \frac{n-1}{n^3} \Phi_n \Phi'_n + \frac{n-1}{n} \hat{p}_{h^*,n-1} A_{h^*,n} + \frac{1}{n} B_{h^*,n} B'_{h^*,n} \end{aligned} \quad (18)$$

with

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

and $B_{h^*,n} = \hat{z}_{h^*,n-1} - \frac{1}{n}\Phi_n + \frac{1}{\hat{n}_{h^*,n-1}+1}\Phi_{h^*,n}$.

From (14) and (18), we obtain the recursive form of the matrix interest $\hat{\Sigma}_n^{-1}\hat{\Gamma}_n$:

$$\hat{\Sigma}_n^{-1}\hat{\Gamma}_n = \hat{\Sigma}_{n-1}^{-1}\hat{\Gamma}_{n-1} + \hat{\Sigma}_{n-1}^{-1}C_{h^*,n} - \frac{1}{n+\rho_n} \hat{\Sigma}_{n-1}^{-1} \Phi_n \Phi'_n \hat{\Sigma}_{n-1}^{-1} [\hat{\Gamma}_{n-1} + C_{h^*,n}] \quad (19)$$

where

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

Finally, we obtain an estimated basis of E spanned by the eigenvector $\hat{b}_{\text{recursive SIR}}$ associated with the largest eigenvalue of the estimate $\hat{\Sigma}_n^{-1}\hat{\Gamma}_n$ of $\Sigma^{-1}\Gamma$

3.3 A recursive version of the SIRoneslice approach

We gave in (10) and (11) the non-recursive expression of the estimators $\hat{\beta}_{\hat{h}_n^o, n}$ and $\hat{b}_{\hat{h}_n^o, n}$. From (14) and (17), it is possible to obtain their corresponding recursive expression. We only detail in this article the recursive expression of $\hat{b}_{\hat{h}_n^o, n} = \hat{\Sigma}_n^{-1}\hat{z}_{\hat{h}_n^o, n}$:

$$\begin{aligned}\hat{b}_{\hat{h}_n^o, n} &= \frac{n}{n-1}\hat{\Sigma}_{n-1}^{-1}\hat{z}_{\hat{h}_n^o, n-1} - \frac{1}{n-1}\hat{\Sigma}_{n-1}^{-1}\Phi_n - \frac{1}{(n-1)(n+\rho_n)}\hat{\Sigma}_{n-1}^{-1}\Phi_n\Phi_n'\hat{\Sigma}_{n-1}^{-1}(n\hat{z}_{\hat{h}_n^o, n-1} - \Phi_n) \\ &+ \frac{1}{n-1}\Phi_{\hat{h}_n^o, n}'\left(\hat{\Sigma}_{n-1}^{-1} - \frac{1}{(n+\rho_n)}\hat{\Sigma}_{n-1}^{-1}\Phi_n\Phi_n'\hat{\Sigma}_{n-1}^{-1}\right)\mathbb{I}_{[h^*=\hat{h}_n^o]},\end{aligned}$$

where h^* still designates the slice containing the n th observation.

In order to keep this document clear and to a reasonable size, we do not provide the recursive expression of $\hat{\beta}_{\hat{h}_n^o, n}$ since this expression is a cumbersome and very complicated expression.

3.4 An asymptotic result

The asymptotic result needs the following two additional assumptions:

- (A1) The observations $(x_i, y_i), i = 1, \dots, n$, are sampled independently from the model (2).
- (A2) The support of y is partitioned into H fixed slices s_1, \dots, s_H such that $p_h \neq 0, \forall h = 1, \dots, H$.

Theorem 3.1 *Under the condition (LC) and assumptions (A₁) and (A₂), we have*

$$\|\hat{b}_{\hat{h}_n^o, n} - b_{h^o}\| = \mathcal{O}\left(\sqrt{\frac{\log(\log n)}{n}}\right) \quad a.s.$$

The proof of this Theorem is given in the Appendix.

4 Simulation study

4.1 Presentation of the simulation study

In this simulation study, we will consider the four methods described in the previous sections: SIR, SIRoneslice, and their recursive versions (recursive SIR and recursive SIRoneslice). First, we compare in Section 4.2 the computational times of the four methods and we will observe that the recursive versions are the fastest ones. Let us recall here that SIR and recursive SIR (resp. SIRoneslice and recursive SIRoneslice) applied on the same data $\{(x_i, y_i), i = 1, \dots, n\}$ provide the same the estimations, only the manner to calculate the estimate differs (recursive

version or non recursive one). For this reason, we only compare the quality of the estimations obtained from SIR and SIRoneslice in Section 4.3. We also provide in this subsection a naive bootstrap criterion in order to select the number H of slices for SIRoneslice. Finally in Section 4.4, we exhibit the evolution of the quality of the recursive SIRoneslice estimator according to the sample size n .

In the simulation study, we consider the following regression model in order to generate simulated datasets:

$$y = (x'\beta)^2 \exp(x'\beta/A) + \varepsilon, \quad (20)$$

where x follows a p -dimensional centered normal distribution with covariance matrix Σ defined below, ε follows the normal distribution $\mathcal{N}(0, \sigma)$ with $\sigma = 1.5$, and $\beta = (1, -1, 2, -2, 0, \dots, 0)'$. The covariance matrix Σ is arbitrary chosen as follows: $\Sigma = \Lambda'\Lambda + 0.5I_p$ where the p^2 components of the $p \times p$ matrix Λ have been generated from the uniform distribution on $[-2, 2]$, the second term $0.5I_p$ allows to avoid numerical problem of inversion of Σ . The parameter A has an influence on the form on the dependence between the index $x'\beta$ and y .

- When the value of A is small (for instance $A = 1$), the model presents “no symmetric dependence” since the exponential part in (20) is predominant. Since each slice is informative on the direction of β , SIR which is based on the H slices will perform well to recover the EDR direction as well as the proposed SIRoneslice which is based on only one (“optimal”) slice.
- When the value of A is medium (for instance $A = 2.5$), the model presents a “moderate symmetric dependence”, the influence of the exponential part tends to disappears in favor of the squared polynomial part. In this case, SIR may not perform well contrary to the SIRoneslice which can recover information on the EDR direction from the “best” selected slice which does not potentially suffer of symmetric dependence.
- When the value of A is high (for instance $A = 5$), the model presents a “strong symmetric dependence” since the squared polynomial part in (20) is predominant. Each slice suffers of symmetry dependence and then the two approaches, SIR and SIRoneslice, will not properly recover the EDR direction, however SIRoneslice may possibly provide suitable estimation obtained from an “optimal” selected slice which escape from symmetric dependence if the number of slices is large enough.

These comments will be illustrated in Section 4.3. From this model, we will generate $n \times p$ datasets for various values of A ($=1, 2.5$ and 5), p ($=5, 10, 15$ and 20) and n ($=300, 600, 900$ and 1200). For each simulated sample, we estimate the direction of β with SIR, recursive SIR, SIRoneslice and/or recursive SIRoneslice. Let us denote by \hat{b} the corresponding estimated EDR direction. Since only the direction of β is identifiable, we use the following efficiency measure in order to evaluate the numerical quality of the estimator:

$$\cos^2(\hat{b}, \beta) = \frac{(\hat{b}'\Sigma\beta)^2}{(\hat{b}'\Sigma\hat{b})(\beta'\Sigma\beta)}.$$

The closer this squared cosine of the angle between \hat{b} and β is to one, the better is the estimation.

All the methods (SIR, SIRoneslice, recursive SIR and recursive SIRoneslice) have been implemented in R. The simulation study has been made with this software. The corresponding codes are available from the authors.

4.2 Comparison of computational times between non recursive and recursive approaches of SIR and SIRoneslice

In this part of the simulation study we only focus on the computational times of the four estimators SIR, recursive SIR, SIRoneslice and recursive SIRoneslice. For each method, we measure in seconds the computational time needed to calculate the corresponding estimators \hat{b}_n of the EDR direction for n going from $N_0 = 30$ in N . For various values of p and N , we generate $\mathcal{B} = 100$ replicated samples from model (20) with $A = 2.5$. Then for each method and each simulated sample, we estimated the EDR direction for various values of H .

In Table 1, we set the number H of slices to 10 and we give the means of computational times and the corresponding standard deviations evaluated on the $\mathcal{B} = 100$ replicated samples for different values of p ($=5, 10, 15, 20$) and N ($=300, 600, 900$ and 1200). In Table 2, we set the size N to 900 and we give these means and the corresponding standard deviations evaluated on the $\mathcal{B} = 100$ replicated samples for different values of p ($=5, 10, 15, 20$) and H ($=5, 10, 15$ and 20). From the reading of these tables, one can exhibit the following comments.

- From Table 1, one can observe that SIR and Sironeslice provide nearly the same computational times in mean with a slight advantage for SIRoneslice which does not require an eigenvalue decomposition. Not surprisingly, the larger are the dimension p or the size N , the larger is the mean of computational times. The recursive versions of SIR and Sironeslice provide very smaller mean of computational times in comparison with their non recursive

versions. Moreover the recursive SIRoneslice method is clearly the fastest one: recursive SIRoneslice seems to be twice faster than recursive SIR. One explanation of this gain in term of computational time is certainly due to the recursive calculation of the $p \times p$ matrix inverse $\widehat{\Sigma}_n$. Note that when the dimension p increases, the computational time in mean of recursive methods only slightly increase contrary to the non recursive ones: for instance, the computational time is multiplied by around 2 from $p = 5$ to $p = 20$ for the non recursive approaches, whereas it is multiplied by around 1.4 (resp. 1.1) for recursive SIR (resp. recursive SIRoneslice). For all the methods, the computational times increase at the same rate according to size N .

- From Table 2, the main information is that the computational times in mean are penalized by the number H of slices. The recursive approaches are always widely the fastest ones. One can however mention that for SIR and recursive SIR, the computational times are multiplied by around 2 from $H = 5$ to $H = 20$ for all values of p while these computational times are multiplied by around 3 for SIRoneslice and recursive SIRoneslice. This phenomenon is explained by the cost of the research of the “optimal” slice which spends more times for large value of H . Nevertheless the advantage of the recursive approaches remain very important in terms of computational times: for instance when $p = 20$ and $H = 20$, the recursive SIRoneslice is still around 12 times (resp. around 1.8 times) faster than SIR (resp. recursive SIR).

4.3 SIRoneslice versus SIR

In this part of this simulation, the goal is to study the numerical behavior of SIRoneslice and to compare it with SIR. To do this, we generate $\mathcal{B} = 500$ datasets of size $n = 300$ from model (20) with $p = 10$ and for different values of A ($=1, 2.5, 5$). For each simulated sample, we estimate the EDR direction with the non recursive versions of SIR and SIRoneslice. We first illustrate the efficiency of the choice of the “optimal” slice \hat{h}_n^o by our proposed criterion on an example when $H = 5$ (arbitrary fixed) and $A = 1$. Then we consider several values of H in order to evaluate the possible influence of this tuning parameter on the quality of the estimators for various values of A . From the obtained results, the parameter H needs to be properly chosen and therefore we provide a naive bootstrap choice for the number of slices for SIRoneslice. Finally we provide simulation results of comparison of SIRoneslice versus SIR when the tuning parameter H is chosen by the bootstrap criterion.

		$p = 5$	$p = 10$	$p = 15$	$p = 20$
$N = 300$	SIR	1.790 (0.055)	2.318 (0.012)	2.881 (0.016)	3.486 (0.011)
	recursive SIR	0.502 (0.008)	0.547 (0.009)	0.608 (0.010)	0.708 (0.011)
	SIRoneslice	1.671 (0.007)	2.113 (0.008)	2.608 (0.010)	3.126 (0.016)
	recursive SIRoneslice	0.255 (0.010)	0.260 (0.006)	0.262 (0.002)	0.282 (0.005)
$N = 600$	SIR	3.908 (0.043)	5.099 (0.017)	6.366 (0.009)	7.740 (0.014)
	recursive SIR	1.044 (0.017)	1.147 (0.019)	1.258 (0.014)	1.478 (0.018)
	SIRoneslice	3.580 (0.016)	4.683 (0.014)	5.795 (0.011)	6.976 (0.016)
	recursive SIRoneslice	0.526 (0.006)	0.536 (0.004)	0.546 (0.006)	0.580 (0.002)
$N = 900$	SIR	6.120 (0.009)	8.128 (0.022)	10.235 (0.030)	12.523 (0.030)
	recursive SIR	1.594 (0.015)	1.735 (0.017)	1.921 (0.027)	2.245 (0.021)
	SIRoneslice	5.665 (0.028)	7.487 (0.016)	9.358 (0.015)	11.327 (0.013)
	recursive SIRoneslice	0.796 (0.007)	0.816 (0.006)	0.826 (0.003)	0.881 (0.004)
$N = 1200$	SIR	8.522 (0.014)	11.380 (0.014)	14.390 (0.058)	17.901 (0.045)
	recursive SIR	2.137 (0.018)	2.325 (0.029)	2.564 (0.019)	2.952 (0.024)
	SIRoneslice	7.909 (0.009)	10.537 (0.016)	13.200 (0.013)	16.253 (0.046)
	recursive SIRoneslice	1.071 (0.005)	1.098 (0.010)	1.113 (0.005)	1.187 (0.008)

Table 1: Computational times (in seconds) for calculating estimators \hat{b}_n of the direction of β (for n going from $N_0 = 30$ to N) with SIR, recursive SIR, SIRoneslice and recursive SIRoneslice methods (with $H = 10$): mean and standard deviation in parentheses, calculated on $\mathcal{B} = 100$ replicated samples from the model (20) with $A = 2.5$ and different values of p and N

		$p = 5$	$p = 10$	$p = 15$	$p = 20$
$H = 5$	SIR	4.888 (0.053)	5.824 (0.152)	7.365 (0.060)	9.063 (0.014)
	recursive SIR	1.421 (0.013)	1.554 (0.017)	1.750 (0.018)	2.077 (0.020)
	SIRoneslice	3.534 (0.010)	4.686 (0.007)	5.976 (0.010)	7.318 (0.009)
	recursive SIRoneslice	0.481 (0.001)	0.496 (0.005)	0.504 (0.002)	0.540 (0.005)
$H = 10$	SIR	6.120 (0.009)	8.128 (0.022)	10.235 (0.030)	12.523 (0.030)
	recursive SIR	1.594 (0.015)	1.735 (0.017)	1.921 (0.027)	2.245 (0.021)
	SIRoneslice	5.665 (0.028)	7.487 (0.016)	9.358 (0.015)	11.327 (0.013)
	recursive SIRoneslice	0.796 (0.007)	0.816 (0.006)	0.826 (0.003)	0.881 (0.004)
$H = 15$	SIR	7.489 (0.007)	10.081 (0.022)	12.832 (0.054)	15.697 (0.096)
	recursive SIR	1.766 (0.016)	1.918 (0.019)	2.319 (0.023)	2.470 (0.006)
	SIRoneslice	7.594 (0.050)	9.992 (0.047)	12.457 (0.043)	15.031 (0.091)
	recursive SIRoneslice	1.092 (0.009)	1.117 (0.006)	1.135 (0.008)	1.207 (0.005)
$H = 20$	SIR	9.046 (0.014)	12.251 (0.028)	15.646 (0.032)	19.231 (0.061)
	recursive SIR	1.922 (0.014)	2.082 (0.021)	2.325 (0.021)	2.728 (0.011)
	SIRoneslice	9.628 (0.013)	12.660 (0.042)	15.760 (0.013)	19.006 (0.083)
	recursive SIRoneslice	1.394 (0.005)	1.432 (0.006)	1.459 (0.009)	1.544 (0.007)

Table 2: Computational times (in seconds) for calculating estimators \hat{b}_n of the direction of β (for n going from $N_0 = 30$ to N) with SIR, recursive SIR, SIRoneslice and recursive SIRoneslice methods for different values of H : mean and standard deviation in parentheses, calculated on $\mathcal{B} = 100$ replicated samples from the model (20) with $A = 2.5$, $N = 900$ and different values of p

Illustration of the efficiency of the choice of \hat{h}_n^o for SIRoneslice. We only present here an example results obtained from $\mathcal{B} = 500$ replicated samples from the model (20) with $A = 1$ and $n = 300$. For SIRoneslice, we arbitrary set $H = 5$ slices. We plot on Figure 1(a) the boxplots of the $\mathcal{B} = 500$ values of $\hat{k}_{h,n}^2$ for $h = 1, \dots, 5$. Clearly for all the \mathcal{B} simulated samples, the value of $\hat{h}_n^o = \arg \max_h (\hat{k}_{h,n}^2)$ is 5. In Figure 1(b), we represent the boxplots of the square cosines $\cos^2(\hat{b}_{h,n}, \beta)$ for $h = 1, \dots, 5$. When $h = \hat{h}_n^o = 5$, the corresponding boxplot is very close to one contrary to those associated with the other values of h . Hence, one can observe that the proposed choice for the “optimal” slice seems to be efficient on this example. We do not exhibit other examples (with different values for A and H) since they provide same conclusions.

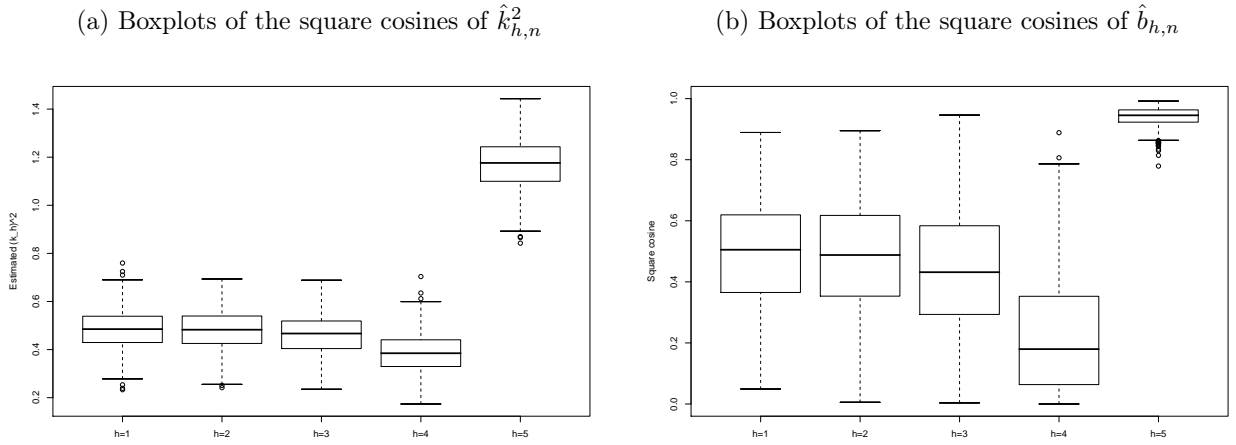


Figure 1: Illustration of the efficiency of the choice of \hat{h}_n^o for SIRoneslice using model (20) with $A = 1$, $n = 300$, $p = 10$ and for a fixed number of slices $H = 5$

Influence of the number H of slices on SIR and SIRoneslice estimates. We consider here $\mathcal{B} = 500$ replicated samples of size $n = 300$ from the model (20) with $p = 10$ and $A = 1, 2.5$ or 5 . For each simulated sample, we estimate the EDR direction with SIRoneslice and SIR for various values of H ($=3, 4, \dots, 10, 15, 20, 25, 30$). For each value of A , we calculate the mean of the quality measures of each estimator over the $\mathcal{B} = 500$ samples, see Tables 3-5. Moreover we indicate in these tables the percent of times when SIRoneslice dominates SIR.

From the reading of Tables 3-5, one can observe that:

- When the model presents “no symmetric dependence” ($A = 1$), SIRoneslice and SIR methods does not seem to be very sensible to the number H of slices. This is particularly true

for SIR and this has been already mentioned by several authors. Even if the square cosine (in mean) obtained with SIRoneslice are greater than 0.89 for most values of H , a good choice for H (here between 6 and 10) will provide in mean square cosine greater than 0.94.

H	SIRoneslice	SIR	Percent of “SIRoneslice>SIR”
3	0.890	0.888	57.8
4	0.924	0.921	55.0
5	0.935	0.934	52.2
6	0.947	0.946	51.2
7	0.946	0.947	47.2
8	0.946	0.950	43.4
9	0.948	0.952	37.6
10	0.944	0.953	32.8
15	0.925	0.953	17.4
20	0.903	0.952	11.0
25	0.886	0.951	8.0
30	0.866	0.949	4.8

Table 3: Means of quality measures obtained with SIRoneslice and SIR over $\mathcal{B} = 500$ samples generated from model (20) with $A = 1$, $n = 300$, $p = 10$ for different values of H

With this kind of non symmetric dependent model, all slices bring information on the EDR direction. When the number H of slices becomes large, the number of observations in each slice decreases and straightforwardly the quality measures of the SIRoneslice estimates (based only on one slice) will be penalized whereas the SIR estimates still provide high quality measures. Finally note that for values of H lower than 7 (that is for moderate values of H according to sample size n), half the time SIRoneslice performs better than SIR.

- When the model present a moderate symmetric dependence ($A = 2.5$) or a strong symmetric dependence ($A = 5$), the two methods are sensitive to the choice of the parameter H . In Table 5, for instance the mean of square cosine varies from 0.286 (when $H = 3$) to 0.714 (when $H = 30$) for SIRoneslice. Therefore a good choice of H appears to be important from a computational point of view. We propose in the next paragraph a naive bootstrap choice for H which provides suitable results.

The symmetric dependence of the model leads to non informative slices to retrieve the EDR direction. When the number of slices is large, we have more chance to obtain an informative slice which is not affected by the symmetric dependence. Note that when $A = 2.5$ or 5, most of the time SIRoneslice performs better than SIR (with percents around 75% for $A = 5$ and

$H \geq 8$) and provides quality measures close to 0.71 (in mean, with $H = 30$) versus 0.57 for SIR.

H	SIRoneslice	SIR	Percent of “SIRoneslice>SIR”
3	0.610	0.605	56.0
4	0.704	0.682	64.2
5	0.781	0.755	75.6
6	0.804	0.774	74.8
7	0.822	0.789	73.6
8	0.846	0.814	72.2
9	0.865	0.841	69.8
10	0.877	0.857	70.6
15	0.891	0.873	61.2
20	0.885	0.870	59.0
25	0.876	0.870	48.6
30	0.862	0.862	47.2

Table 4: Means of quality measures obtained with SIRoneslice and SIR over $\mathcal{B} = 500$ samples generated from model (20) with $A = 2.5$, $n = 300$, $p = 10$ for different values of H

H	SIRoneslice	SIR	Percent of “SIRoneslice>SIR”
3	0.286	0.287	47.0
4	0.382	0.357	56.0
5	0.418	0.374	65.4
6	0.487	0.441	66.0
7	0.533	0.477	71.0
8	0.593	0.504	78.0
9	0.641	0.540	76.0
10	0.566	0.487	72.4
15	0.661	0.536	78.8
20	0.687	0.523	77.4
25	0.701	0.556	79.8
30	0.714	0.574	76.6

Table 5: Means of quality measures obtained with SIRoneslice and SIR over $\mathcal{B} = 500$ samples generated from model (20) with $A = 5$, $n = 300$, $p = 10$ for different values of H

A naive bootstrap choice for the number H of slices for SIRoneslice. We propose hereafter a criterion in order to choose an “optimal” number H of slices for SIRoneslice. This criterion could also be used for SIR. The idea is to provide a value of H such that the quality measure of the estimator ($\hat{b}_{\hat{h}_n, n}$ for SIRoneslice which relies on H) remains stable over bootstrap replications of the available sample $s^{(r)} = \{(x_i, y_i), i = 1, \dots, n\}$.

Let B^* be the number of bootstrap replications. For $r = 1, \dots, B^*$, let us consider $\{(x_i^{(r)}, y_i^{(r)}), i = 1, \dots, n\}$ a non parametric bootstrap replication. According to Efron (1982) naive bootstrap es-

timate of the mean of the expectation of the quality measure $\cos^2(\hat{b}_{\hat{h}_n^{(r)},n}, \beta)$ is defined by:

$$\hat{g}_n(H) = \frac{1}{B^*} \sum_{r=1}^{B^*} \hat{q}_n^{(r)}$$

with $\hat{q}_n^{(r)} = \frac{\left(\hat{b}_{\hat{h}_n^{(r)},n}^{(r)'} \hat{\Sigma}_n \hat{b}_{\hat{h}_n^{(r)},n} \right)^2}{\left(\hat{b}_{\hat{h}_n^{(r)},n}^{(r)'} \hat{\Sigma}_n \hat{b}_{\hat{h}_n^{(r)},n}^{(r)} \right) \left(\hat{b}_{\hat{h}_n^{(r)},n}' \hat{\Sigma}_n \hat{b}_{\hat{h}_n^{(r)},n} \right)}$ where $\hat{b}_{\hat{h}_n^{(r)},n}^{(r)}$ is the SIRone-slice estimator based on the bootstrap replication sample $s^{(r)}$. In practice this criterion will be computed for values of H from H_{\min} to H_{\max} (with H_{\min} and H_{\max} chosen by the user). An optimal number of slices can then be defined as: $\hat{H}_n = \arg \max_H \hat{g}_n(H)$.

To illustrate the numerical behavior of this proposed criterion, we first consider a simulated sample of size $n = 300$ from model (20) with $A = 2.5$ and $p = 10$. We set $B^* = 200$ and choose $H_{\min} = 3$ and $H_{\max} = 25$. In Figure 2, for each value of H in $[H_{\min}; H_{\max}]$ the boxplot of the values $\hat{q}_n^{(r)}$, $r = 1, \dots, B^*$ are plotted. One can see that the dispersion of the $\hat{q}_n^{(r)}$'s are large for too small or too large values of H . The solid line links the points $(H, \hat{g}_n(H))$ in order to easily see the values of the criterion. One can observe that the criterion chooses in this example $\hat{H}_n = 7$. Note that the corresponding boxplot of the $\hat{q}_n^{(r)}$'s shows a small dispersion of these values. Let us finally mention that our R code provides the numerical values of $\hat{g}_n(H)$ for all H and the optimal value \hat{H}_n if the user is not interested in graphical representation of the criterion.

In order to show the efficiency of the criterion, we generate 50 samples of size $n = 300$ from model (20) with $p = 10$ and $A = 2.5$. We keep the same parameter values for the criterion: $B^* = 200$, $H_{\min} = 3$ and $H_{\max} = 25$. For each simulated sample, we estimate the EDR direction with SIRoneslice using the optimal value \hat{H}_n obtained with the bootstrap criterion. We also estimate this direction for all values of H in $[H_{\min}; H_{\max}]$, and we select the value of H which provides the maximum value of the quality measure. In Figure 3, we plot the boxplot of the square cosines obtained with SIRoneslice based on \hat{H}_n slices (named ‘‘Bootstrap criterion’’ in the graphic), the boxplot of the ‘‘best’’ square cosines as defined above (named ‘‘with best value for H ’’ in the graphic) and the boxplot of the square cosines obtained with SIRoneslice based on all possible values of H in $[H_{\min}; H_{\max}]$ (named ‘‘for all H in the graphic’). Let us mention that the selection of the ‘‘best’’ number H of slices can only be done in a simulation study when the true direction β is known whereas our bootstrap criterion can always be used in practice. One can observe that the quality measures obtained with the proposed bootstrap criterion are very close to the ‘‘best’’ ones. Note that if the choice of H is arbitrary made by the user, the quality of the corresponding estimated direction could behave worse, see the large dispersion of the

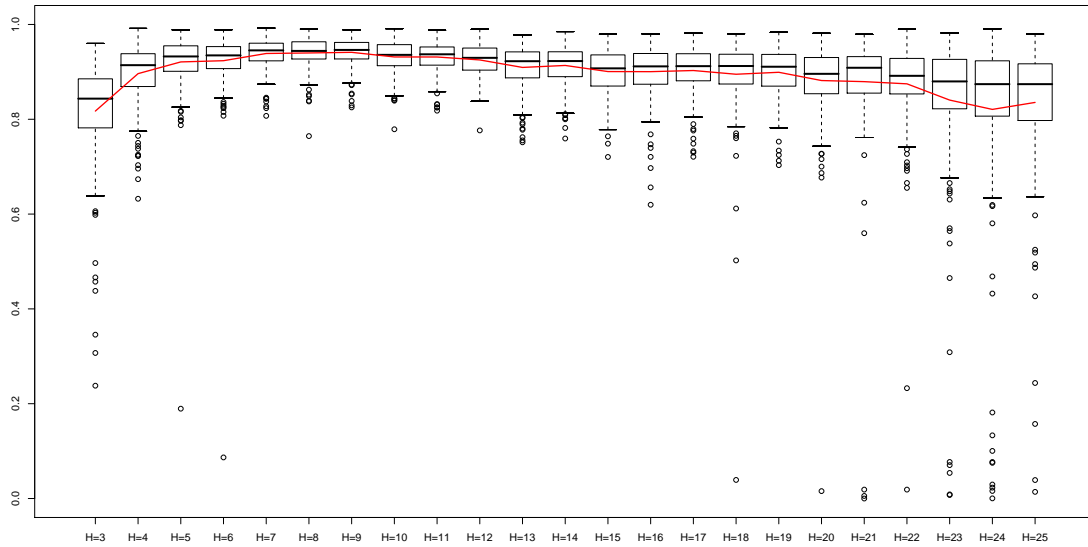


Figure 2: An example of the bootstrap criterion on a simulated sample from model (20) with $n = 300$, $p = 10$ and $A = 2.5$

corresponding boxplot. From these simulation results, the proposed naive bootstrap criterion to select H seems to be useful in practice and we use it in the sequel.

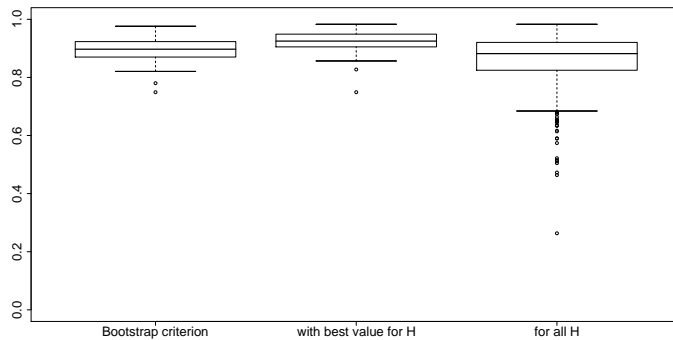


Figure 3: Illustration of the efficiency of the proposed bootstrap criterion via a simulation study using model (20) with $n = 300$, $p = 10$ and $A = 2.5$

SIRoneslice versus SIR. We consider $\mathcal{B} = 500$ replicated samples of size $n = 300$ from the model (20) with $p = 10$ and $A = 1, 2.5$ or 5 . For each simulated sample, we estimate the EDR direction with SIR and SIRoneslice. Note that for the two methods, the number H of slices

has been chosen by the proposed bootstrap criterion introduced in the previous paragraph (with $H_{\min} = 3$ and $H_{\max} = 25$).

For each value of A , we plot in Figure 4(a), (c) and (e) the boxplots of the $\mathcal{B} = 500$ values of the corresponding square cosines of SIR and SIRoneslice estimates. In addition, we represent, in Figure 4(b), (d) and (f), the scatter plots of the square cosines of the SIR estimates versus those of SIRoneslice estimates. From this figure, one can point out the following comments.

- When there is no symmetric dependence ($A = 1$), the two methods provide very good estimations with square cosines greater than 0.8. Moreover the quality measures for SIRoneslice and SIR are very close.
- When there is moderate symmetric dependence ($A = 2.5$), SIRoneslice appraises to be better than SIR. We clearly see on the scatter plot that the great majority of the points are over the first bisecting line.
- When there is strong symmetric dependence ($A = 5$) the quality measures of the SIRoneslice estimates are largely better than those obtained with SIR. The corresponding scatter plot confirms a uniformly advantage in favor of SIRoneslice.

4.4 Recursive SIRoneslice simulation results

We consider here a sample of size $n = 1200$ from model (20) with $p = 10$ and $A = 1, 2.5$ or 5 . We arbitrary set $H = 15$ slices. For each simulated sample, we estimate for $n = N_0 = 60$ until $N = 1200$, the EDR direction with the recursive SIRoneslice method and we calculate the corresponding quality measures $\cos^2(\hat{b}_{\hat{h}_n^o, n}, \beta)$.

In Figure 5, we represent the evolution of the quality measures $\cos^2(\hat{b}_{\hat{h}_n^o, n}, \beta)$ according to n . Not surprisingly, we clearly observe that the more important the sample size n , the greater the quality measure. As previously mentioned, the parameter A has influence on the dependence between the index $x'\beta$ and y in (20) and then effects the quality estimator $\hat{b}_{\hat{h}_n^o, n}$. For instance, in Figure 5, the estimations are clearly better for small sample size when the value of A is small.

In Figure 6, we provide, for some various values of N ($= 100, 200, 300, 600, 900$ and 1200), the boxplots of $\cos^2(\hat{b}_{\hat{h}_n^o, n}, \beta)$ calculated on $\mathcal{B} = 500$ samples simulated from model (20) with $p = 10$ and $A = 2.5$. We observe the same kind of phenomenon as those described in the corresponding case on Figure 5. Since the sample size n is sizeable, the quality of estimations are very good.

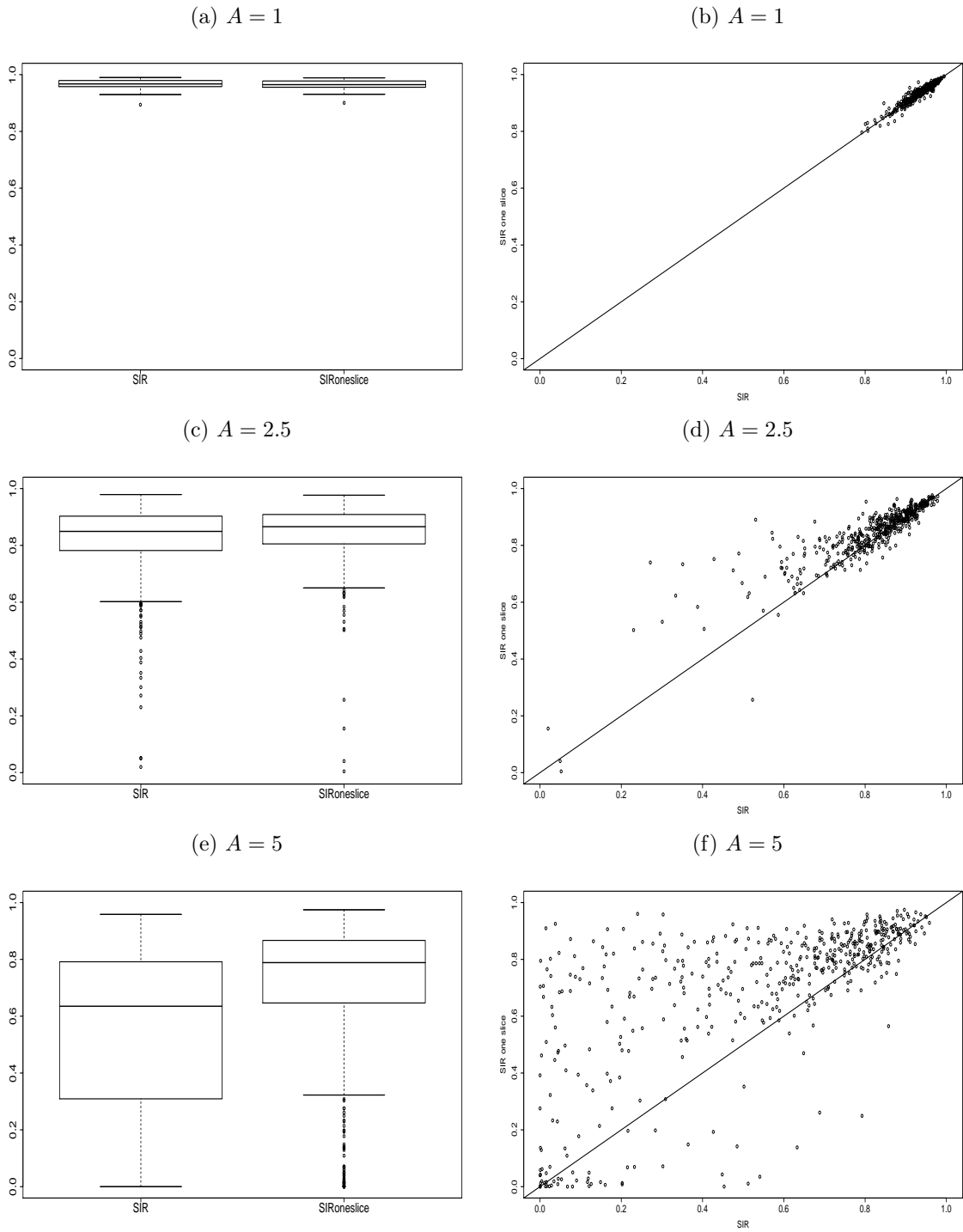


Figure 4: Simulation results obtained with SIR and SIRoneslice for $\mathcal{B} = 500$ samples generated from model (20) with $n = 300$, $p = 10$ and different values of A : boxplots and scatter plots of square cosines

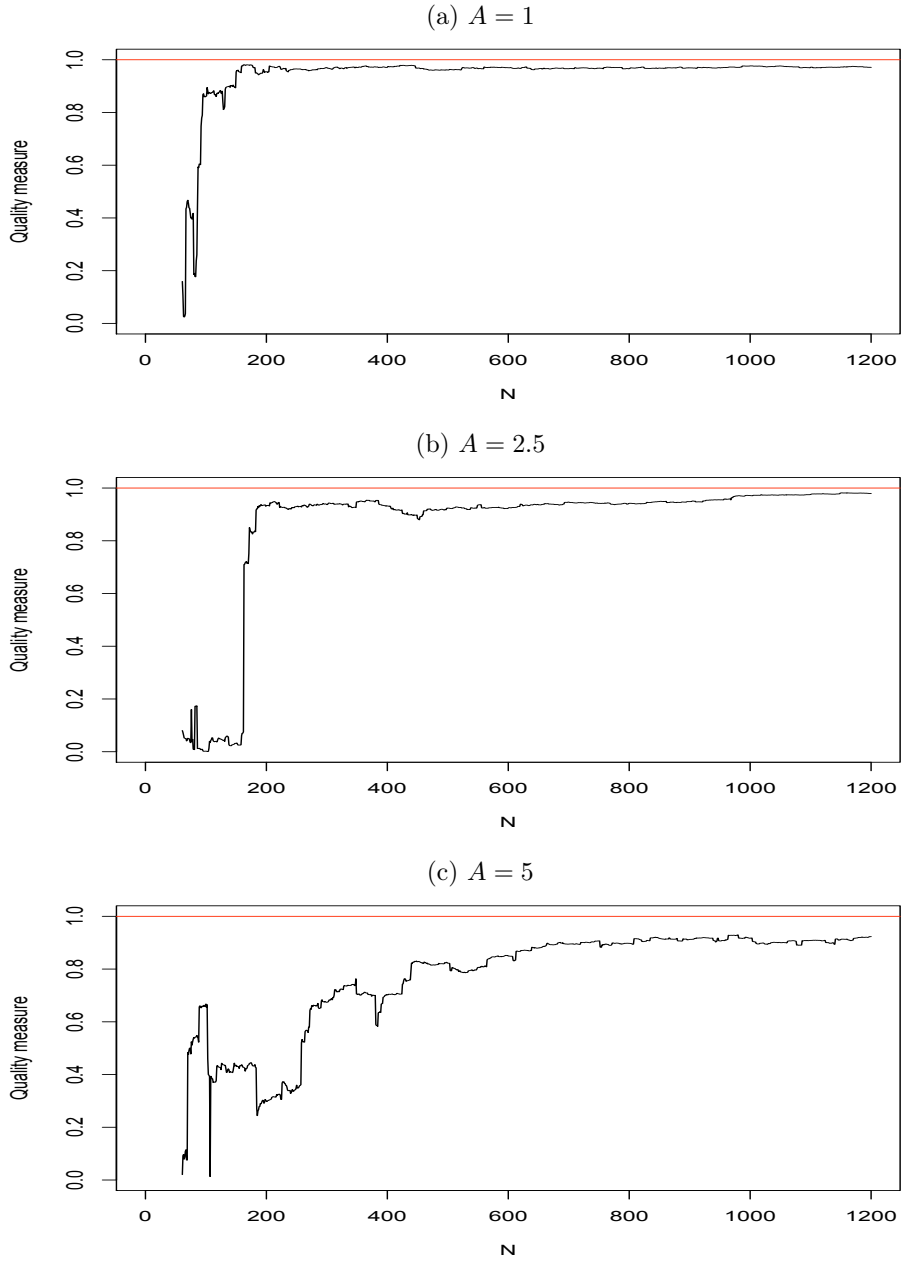


Figure 5: Evolution, according to n , of the quality measures $\cos^2(\hat{b}_{\hat{h}_n^o, n}, \beta)$ obtained with recursive SIRoneslice for three simulated samples from model (20) with $p = 10$ and various values of A .

For example, the boxplots of the $\cos^2(\hat{b}_{\hat{h}_n^o, n}, \beta)$'s are very concentrated around the value 0.95 when $n \geq 600$.

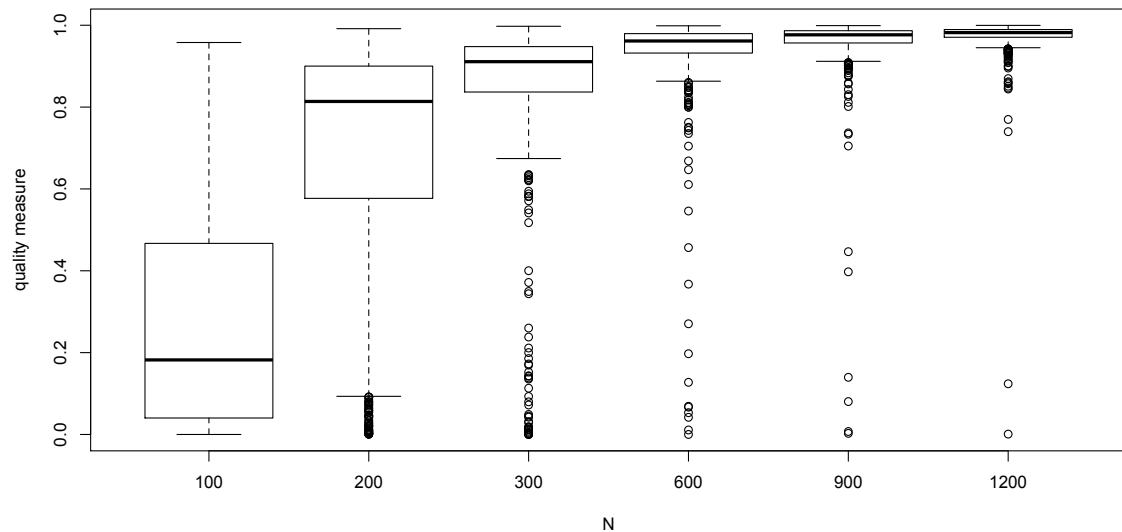


Figure 6: Simulation results obtained with recursive SIRoneslice for $\mathcal{B} = 500$ samples generated from model (20) with $p = 10$ and $A = 2.5$: boxplots of square cosines for various values of n

5 Concluding remarks

In this paper, we first proposed an estimator of the direction of β based on the use of only one “optimal” slice chosen among the H slices of the slicing step. We called this method SIRoneslice. Then, we proposed recursive versions of the SIR and SIRoneslice estimators. We showed the almost sure convergence with its rate of the SIRoneslice estimator. Finally, we illustrated on simulations the good numerical performances of SIRoneslice method. We also proposed a naive bootstrap criterion in order to choose the number H of slices. We clearly exhibited the main advantage of using recursive versions of the SIR and SIRoneslice methods from a computational times point of view. The SIRoneslice, recursive SIR and recursive SIRoneslice methods have been implemented in R and the corresponding codes are available from the authors.

Interesting extensions of this work concern the development of recursive versions for other dimension reduction approaches such as SAVE (see for instance Cook, 2000), pHd (see for example Li, 1992) or SIR_α (see for instance Li, 1991, or Gannoun and Saracco, 2003). Moreover in order to save computational time when the dependent variable y is multivariate, recursive estimators could be also introduced in multivariate SIR methods (see for example Li et al., 2003, Saracco, 2005 or Barreda et al., 2007).

Appendix: proof of Theorem 3.1

For $h = 1, \dots, H$, we have $\hat{b}_{h,n} - b_h = (\hat{\Sigma}_n^{-1} - \Sigma^{-1})\hat{z}_{h,n} + \Sigma^{-1}(\hat{z}_{h,n} - z_h)$ where $b_h = \Sigma^{-1}z_h$ and $\hat{b}_{h,n} = \hat{\Sigma}_n^{-1}\hat{z}_{h,n}$. Then we can deduce the following inequality:

$$\|\hat{b}_{h,n} - b_h\|^2 \leq 2 \|\hat{z}_{h,n}\|^2 \|\hat{\Sigma}_n^{-1} - \Sigma^{-1}\|^2 + 2 \|\Sigma^{-1}\|^2 \|\hat{z}_{h,n} - z_h\|^2. \quad (21)$$

In the following demonstration, we first specify the rate of convergence of $\hat{z}_{h,n}$ based on the rates of convergence of \bar{x}_n and $\hat{m}_{h,n}$. Then, applying equation of Riccati for the inverse matrix, we obtain the rate of convergence of $\hat{\Sigma}_n^{-1}$. Finally, we deduce the rate of convergence of $\hat{b}_{h,n}$.

For a matrix M , we denote by $\lambda_{max}(M)$ its largest eigenvalue and by $\lambda_{min}(M)$ its smallest eigenvalue.

Step 1: Study the rate of convergence of $(\hat{z}_{h,n})$

For $h = 1, \dots, H$, we have $\hat{z}_{h,n} - z_h = (\hat{m}_{h,n} - m_h) - (\bar{x}_n - \mu)$. Let us consider the following inequality:

$$\|\hat{z}_{h,n} - z_h\|^2 \leq 2 \|\hat{m}_{h,n} - m_h\|^2 + 2 \|\bar{x}_n - \mu\|^2. \quad (22)$$

Study the rate of convergence of \bar{x}_n

From the Hartman-Wintner law of the iterated logarithm (see Theorem 3.2.9, page 136 of Stout,1974) we have

$$\limsup \frac{\|S_n - n\mu\|^2}{2n \log(\log n)} \leq \lambda_{max}(\Sigma) \quad \text{a.s.}$$

where $S_n = \sum_{i=1}^n x_i$. It follows that

$$\limsup \left(\frac{n}{2 \log(\log n)} \right) \|\bar{x}_n - \mu\|^2 \leq \lambda_{max}(\Sigma) \quad \text{a.s.}$$

and then we get

$$\|\bar{x}_n - \mu\|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \quad (23)$$

Study the rate of convergence of $\hat{m}_{h,n}$

We have $\hat{m}_{h,n} - m_h = \hat{m}_{h,n} - \frac{n}{\hat{n}_{h,n}} \mathbb{E}[x \mathbb{I}_{[y \in s_h]}] + \frac{n}{\hat{n}_{h,n}} \mathbb{E}[x \mathbb{I}_{[y \in s_h]}] - m_h$ with $\hat{m}_{h,n} = \frac{1}{\hat{n}_{h,n}} \sum_{i=1}^n x_i \mathbb{I}_{[y_i \in s_h]}$

and $\hat{n}_{h,n} = \sum_{i=1}^n \mathbb{I}_{[y_i \in s_h]}$. Similarly to the study of \bar{x}_n , we get

$$\|\hat{m}_{h,n} - m_h\|^2 \leq 2 \left\| \hat{m}_{h,n} - \frac{n}{\hat{n}_{h,n}} \mathbb{E}[x \mathbb{I}_{[y \in s_h]}] \right\|^2 + 2 \left\| \frac{n}{\hat{n}_{h,n}} \mathbb{E}[x \mathbb{I}_{[y \in s_h]}] - m_h \right\|^2. \quad (24)$$

Let us focus on the first term. We have

$$\hat{n}_{h,n}\hat{m}_{h,n} = \sum_{i=1}^n x_i \mathbb{I}_{[y_i \in s_h]} = \sum_{i=1}^n \varepsilon_i(h) + n\mathbb{E}[x\mathbb{I}_{[y \in s_h]}]$$

where $\varepsilon_i(h) = x_i \mathbb{I}_{[y_i \in s_h]} - \mathbb{E}[x_i \mathbb{I}_{[y_i \in s_h]}] = x_i \mathbb{I}_{[y_i \in s_h]} - \mathbb{E}[x\mathbb{I}_{[y \in s_h]}]$. Let $M_n(h) := \sum_{i=1}^n \varepsilon_i(h)$. The sequence $(\varepsilon_n(h))$ is a sequence of independent random variables with $\mathbb{E}[\varepsilon_i(h)] = 0$ and $\mathbb{E}[\varepsilon_i(h)\varepsilon_i'(h)] \leq \Sigma + \mu\mu'$. We deduce again from the Hartman-Wintner law of the iterated logarithm that

$$\|M_n(h)\|^2 = \mathcal{O}\left(n \log(\log n)\right) \quad \text{a.s.}$$

It follows that

$$\left\| \hat{n}_{h,n}\hat{m}_{h,n} - n\mathbb{E}[x\mathbb{I}_{[y \in s_h]}] \right\|^2 = \mathcal{O}\left(n \log(\log n)\right) \quad \text{a.s.}$$

Since we have $\frac{\hat{n}_{h,n}}{n} \xrightarrow{\text{a.s.}} p_h$ where $p_h = \mathbb{P}(y \in s_h) > 0$ (see assumption (A2)), we finally obtain

$$\left\| \hat{m}_{h,n} - \frac{n}{\hat{n}_{h,n}} \mathbb{E}[x\mathbb{I}_{[y \in s_h]}] \right\|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \quad (25)$$

Let us now study the second term. Since $m_h p_h = \mathbb{E}[x\mathbb{I}_{[y \in s_h]}]$, we have

$$\frac{n}{\hat{n}_{h,n}} \mathbb{E}[x\mathbb{I}_{[y \in s_h]}] - m_h = \frac{n}{\hat{n}_{h,n}} m_h p_h - m_h = \frac{m_h}{\hat{n}_{h,n}} (np_h - \hat{n}_{h,n}). \quad (26)$$

Moreover, as $\hat{n}_{h,n} = \sum_{i=1}^n Z_i(h) + np_h$ with $Z_i(h) = \mathbb{I}_{[y_i \in s_h]} - \mathbb{E}[\mathbb{I}_{[y_i \in s_h]}]$, we have

$$\hat{n}_{h,n} - np_h = M_n = \sum_{i=1}^n Z_i(h).$$

By the Hartman-Wintner law of the iterated logarithm, we get

$$\left\| \hat{n}_{h,n} - np_h \right\|^2 = \mathcal{O}\left(n \log(\log n)\right) \quad \text{a.s.}$$

Then using the fact that $\hat{n}_{h,n} \sim p_h n$, we deduce from (26) that

$$\left\| \frac{n}{\hat{n}_{h,n}} \mathbb{E}[x\mathbb{I}_{[y \in s_h]}] - m_h \right\|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \quad (27)$$

Finally, using the rates obtained in (25) and (27), we deduce from (24) that

$$\| \hat{m}_{h,n} - m_h \|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \quad (28)$$

Study of the rate of convergence of $\hat{z}_{h,n}$

Using (23) and (28), we obtain from (22) that

$$\| \hat{z}_{h,n} - z_h \|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \quad (29)$$

Step 2: Study the rate of convergence of $(\widehat{\Sigma}_n^{-1})$

According to the Riccati equation for the matrix inverse (see for instance Duflo (1997), page 96), we can write $\widehat{\Sigma}_n^{-1}$ as follows $\widehat{\Sigma}_n^{-1} = \Sigma^{-1} - \Sigma^{-1}(\widehat{\Sigma}_n - \Sigma)\Sigma^{-1} + R_n$ where $R_n = \Sigma^{-1}(\Sigma - \widehat{\Sigma}_n)\widehat{\Sigma}_n^{-1}(\Sigma - \widehat{\Sigma}_n)\Sigma^{-1}$. Straightforwardly, we get

$$\|\widehat{\Sigma}_n^{-1} - \Sigma^{-1}\|^2 \leq 2\|R_n\|^2 + 2\|\Sigma^{-1}(\widehat{\Sigma}_n - \Sigma)\Sigma^{-1}\|^2. \quad (30)$$

Study of the rate of convergence of $\widehat{\Sigma}_n$

Since $\widehat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)(x_i - \bar{x}_n)'$ we can write $\widehat{\Sigma}_n - \Sigma = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)' - (\bar{x}_n - \mu)(\bar{x}_n - \mu)'$, where $e_i = (x_i - \mu)(x_i - \mu)' - \Sigma$. Therefore, we obtain

$$\|\widehat{\Sigma}_n - \Sigma\|^2 \leq \frac{2}{n^2} \left\| \sum_{i=1}^n e_i \right\|^2 + 2\|\bar{x}_n - \mu\|^4. \quad (31)$$

Let $M_N(u) = u' \left(\sum_{i=1}^n e_i \right) u = \sum_{i=1}^n u' e_i u = \sum_{i=1}^n e_i(u)$, where $e_i(u) = u' e_i u = u'(x_i - \mu)(x_i - \mu)'u - u'\Sigma u$.

We have $\mathbb{E}[e_i(u)] = \mathbb{E}[u'(x_i - \mu)(x_i - \mu)'u - u'\Sigma u] = u'\Sigma u - u'\Sigma u = 0$ and

$$\begin{aligned} \mathbb{E}[(e_i(u))^2] &= \mathbb{E}[(u'(x_i - \mu)(x_i - \mu)'u - u'\Sigma u)^2] \\ &= \mathbb{E}[u'(x_i - \mu)(x_i - \mu)'u u'(x_i - \mu)(x_i - \mu)'u] - (u'\Sigma u)^2 \\ &\leq \mathbb{E}[u'(x_i - \mu)(x_i - \mu)'u^2] \\ &\leq \mathbb{E}[\|x_i - \mu\|^4] \times \|u\|^4. \end{aligned}$$

We can deduce again by the Hartman-Wintner law of the iterated logarithm that

$$\limsup \frac{\left\| \sum_{i=1}^n e_i(u) \right\|^2}{2n \log(\log n)} \leq \tau^4 \times \|u\|^4 \quad \text{a.s.}$$

where $\tau^4 = \mathbb{E}[\|x_i - \mu\|^4]$. It follows that

$$\left\| \sum_{i=1}^n e_i \right\|^2 = \mathcal{O}\left(n \log(\log n)\right) \quad \text{a.s.} \quad (32)$$

From the rates obtained in (23) and (32), we deduce from (31) that

$$\begin{aligned} \|\widehat{\Sigma}_n - \Sigma\|^2 &= \mathcal{O}\left(\frac{1}{n^2} n \log(\log n)\right) + \mathcal{O}\left(\left(\frac{\log(\log n)}{n}\right)^2\right) \quad \text{a.s.} \\ &= \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \end{aligned} \quad (33)$$

Study of the rate of convergence of R_n

We deduce from the expression of R_n that

$$\begin{aligned} \| R_n \| &\leq \lambda_{max}(\widehat{\Sigma}_n^{-1})\Sigma^{-1}(\Sigma - \widehat{\Sigma}_n)^2\Sigma^{-1} \\ &\leq \frac{1}{\lambda_{min}(\widehat{\Sigma}_n)}\Sigma^{-1}(\Sigma - \widehat{\Sigma}_n)^2\Sigma^{-1} \\ &\leq \frac{1}{\lambda_{min}(\widehat{\Sigma}_n)} \times \frac{\| \widehat{\Sigma}_n - \Sigma \|^2}{(\lambda_{min}(\Sigma))^2}. \end{aligned}$$

From (33), we obtain

$$\| R_n \| = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \quad (34)$$

Study of the rate of convergence of $\widehat{\Sigma}_n^{-1}$

Using (33) and (34), we straightforwardly deduce from (30) that

$$\| \widehat{\Sigma}_n^{-1} - \Sigma^{-1} \|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \quad (35)$$

Setp 3: Study of the rate of convergence of $\hat{b}_{\hat{h}_n^o, n}$

Finally, from the rates of convergence obtained in (29) et (35), we deduce from (21) that, $\forall h = 1, \dots, H$, we have

$$\| \hat{b}_{h, n} - b_h \|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \quad (36)$$

Obviously, this result still remains true for $h = h^o$. Let us assume that $\exists ! h^o$ such that $k_{h^o} > k_h$. In addition, since we have $\forall h, \hat{k}_{h_n, n} \rightarrow k_h$, then $\exists n_o$ such that $\forall n > n_o, (\hat{k}_{h^o, n})^2 > (\hat{k}_{h, n})^2$. According to the definitions of h^o and \hat{h}_n^o given in (8) and (9), we have $\hat{h}_n^o \rightarrow h^o$. From this result and (36), we obtain

$$\| \hat{b}_{\hat{h}_n^o, n} - b_{h^o} \|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.}$$

This completes the proof of the Theorem.

References

- Barreda, L., Gannoun, A. and Saracco, J. (2007). Some extensions of multivariate sliced inverse regression. *J. Stat. Comput. Simul.*, **77**(1-2), 1-17.
- Chen, C.-H., Li, K.-C. (1998). Can SIR be as popular as multiple linear regression? *Statist. Sinica*, **8**(2), 289-316.

- Cook, R. D. (2000). SAVE: A method for dimension reduction and graphics in regression. *Comm. Statist. Theory Methods*, **29**, 2109-2121.
- Cook, R. D. and Ni, L. (2005). Sufficient dimension reduction via inverse regression: a minimum discrepancy approach. *Journal of American Statistical Association*, **100**, 410-418.
- Duan, N. and Li, K. C. (1991). Slicing regression: a link-free regression method. *The Annals of Statistics*, **19**, 505-530.
- Duflo, M. (1997). *Random Iterative Models*, Springer-Verlag, Berlin.
- Efron, B. (1982) *The jackknife, the bootstrap and other resampling plans*. CBMS-NSF Regional Conference Series in Applied Mathematics, 38. Society for Industrial and Applied Mathematics (SIAM), Philadelphia.
- Gannoun, A., Saracco, J. (2003). An asymptotic theory for SIR_α method. *Statistica Sinica*, **13**, 297-310.
- Hall, P. and Li, K.-C. (1993). On almost linearity of low-dimensional projections from high-dimensional data. *Ann. Statist.*, **21**(2), 867-889.
- Hsing, T. and Carroll, R. J. (1992). An asymptotic theory for Sliced Inverse regression. *The Annals of Statistics*, **20**, 1040-1061.
- Li, K. C. (1991). Sliced inverse regression for dimension reduction, with dicussion. *Journal of the American Statistical Association*, **86**, 316-342.
- Li, K.-C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma. *J. Amer. Statist. Assoc.*, **87**, 1025-1039.
- Li, K. C., Aragon, Y., Shedden, K. and Thomas Agnan, C. (2003). Dimension reduction for multivariate response data. *Journal of the American Statistical Association*, **98**, 99-109.
- Saracco, J. (1997). An asymptotic theory for Sliced Inverse Regression. *Communications in Statistics - Theory and methods*, **26**, 2141-2717.
- Saracco, J. (2005). Asymptotics for pooled marginal slicing estimator based on SIR a approach. *Journal of multivariate Analysis*, **96**, 117-135.
- Stout, W. F. (1974). *Almost sure convergence*. Probability and Mathematical Statistics, Vol. 24. Academic Press , New York-London, 1974.
- Zhu, L. X. and Ng, K. W. (1995). Asymptotics of sliced inverse regression. *The Annals of Statistics*, **24**, 1053-1068.

Bibliographie

- [1] Aragon, Y. and Saracco, J. (1997). Sliced Inverse Regression (SIR) : an appraisal of small sample alternatives to slicing. *Computation and Statistics*, **12**, 109-130.
- [2] Barreda, L., Gannoun, A., Saracco, J. (2007). Some extensions of multivariate SIR. *Journal of Statistical Computation and Simulation*, **77** (1-2), 1-17.
- [3] Bernard-Michel, C., Douté, S., Fauvel, M., Gardes, L. et 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.
- [4] Bernard-Michel, C., Gardes, L. et Girard, S. (2009). Gaussian Regularized Sliced Inverse Regression. *Statistics and Computing*, **19**, 85-98.
- [5] Bernard-Michel, C., Gardes, L. et Girard, S. (2009). A Note on Sliced Inversion Regression with Regularizations. *Biometrics*, **64**, 982-984.
- [6] Brillinger, D. R. (1983). A generalized linear model with “gaussian” regressor variables. *Festschrift for Erich L. Lehmann in Honor of His SixtyFifth birthday*. 97-114.
- [7] Cook et Nachtsheim (1994). Re-weighting to achieve elliptically contoured covariates in regression. *J. Amer. Statist. Assoc.* **89**, 592-599.
- [8] Cook, R. D. and Weisberg, S. (1991). Discussion of “Sliced inverse regression”. *Journal of American Statistical Association*, **86**, 328-332.
- [9] Duan, N. and Li, K. C. (1991). Slicing regression : a link-free regression method. *The Annals of Statistics*, **19**, 505-530.
- [10] Duflo, M. (1990). *Méthodes récursives aléatoires*. Techniques Stochastiques, Masson, Paris.
- [11] Duflo, M. (1997). *Random Iterative Models*, Springer-Verlag, Berlin.
- [12] Ferré, L. (1997). Choix de dimension en régression inverse par tranches (SIR). *C. R. Acad. Sci. Paris*, t. **322**, série I.
- [13] Ferré, L. (1998). Determining the dimension in Sliced Inverse Regression and related methods. *Journal of American Statistical Association*, **93**, 132-140.
- [14] 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.

- [15] Gannoun, A., Girard, S., Guinot, C., Saracco, J. (2004). Sliced inverse regression in reference curves estimation. *Computational Statistics and Data Analysis*, **46** (1), 103-122.
- [16] Gannoun, A., Saracco, J. (2003a). An asymptotic theory for SIR_α method. *Statistica Sinica*, **13**, 297-310.
- [17] 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.
- [18] Hall, P. et Li, K. C. (1993). On almost linearity of low dimensional projections from high dimensional data. *The Annals of Statistics*, **21**, 867-889.
- [19] Hsing, T. and Carroll, R. J. (1992). An asymptotic theory for Sliced Inverse regression. *The Annals of Statistics*, **20**, 1040-1061.
- [20] Kuentz, V. et Saracco, J. (2009). Cluster-based sliced inverse regression. *Journal of Korean Statistical Society*.
- [21] Li, K.C., Aragon, Y., Shedden, K. et Thomas-Agnon, C. (2003). Dimension Reduction for Multivariate Response Data. *Journal of the American Statistical Association*, **98** (461), 99 -109.
- [22] Li, K. C. (1991). Sliced inverse regression for dimension reduction, with dicussion. *Journal of the American Statistical Association*, **86**, 316-342.
- [23] Liquet, B. et Saracco, J. (2007). Pooled marginal slicing approach via SIR_α with discrete covariables. *Computational Statistics*, **4**, 599-617.
- [24] 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.
- [25] Magnus, J. R. et Neudecker, H. (1979). The commutation matrix : some properties and applications. *Annals of Statistics*, **7**(2), 381-394.
- [26] Magnus, J. R. et Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley & Sons, Ltd., Chichester.
- [27] Révész, P. (1977). How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. *Math. Op. Statist. Ser. Statistics*, **8**, 119-126.
- [28] Saracco, J. (1997). An asymptotic theory for Sliced Inverse Regression. *Communications in Statistics - Theory and methods*, **26**, 2141-2717.
- [29] Saracco, J. (2001). Pooled Slicing methodes versus Slicing methods. *Communications in Statistics - Simulation and Computation*, **30**, 489-511.
- [30] Saracco, J. (2005). Asymptotics for pooled marginal slicing estimator based on SIR_α approach. *Journal of multivariate Analysis*, **96**, 117-135.
- [31] Schott, J. R. (1994). Determining the dimensionality in Sliced Inverse Regression. *Journal of American Statistical Association*, **89**, 141-148.

- [32] Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.
- [33] Yang, S. S. (1977). General distribution theory of the concomitants of order statistics. *The Annals of Statistics*, **5**, 996-1002.
- [34] Zhu, L. X. and Ng, K. W. (1995). Asymptotics of sliced inverse regression. *The Annals of Statistics*, **24**, 1053-1068.
- [35] Zhu, L. et Yu, Z. (2007) On spline approximation of sliced inverse regression. *Science in China Series A : Mathematics*, **Vol. 50, No. 9**, 1289-1302.

Deuxième partie
Estimation fonctionnelle

Introduction

Dans cette partie, nous présentons des estimateurs non paramétriques d'une densité de probabilité puis d'une fonction de régression.

Dans le premier chapitre, nous donnons tout d'abord la construction de l'estimateur à noyau classique (l'estimateur de Parzen-Rosenblatt) d'une densité de probabilité. Ensuite, nous présentons l'estimateur récursif d'une densité de probabilité introduit par Wolverton-Wagner, puis Deheuvels. Nous précisons également quelques résultats de convergence de ces estimateurs.

Nous poursuivons dans le second chapitre en présentant l'estimateur classique à noyau de la fonction de régression introduit par Nadaraya-Watson. Nous nous intéressons ensuite à l'estimateur de Nadaraya-Watson récursif ainsi qu'à l'estimateur de Révész. Enfin, nous rappelons les principales propriétés asymptotiques de ces estimateurs.

Le troisième chapitre, intitulé "Application en valvométrie" présente les travaux que j'ai réalisés dans le cadre d'une collaboration avec Gilles Durrieu et Mohamedou Sow du laboratoire EPOC (Environnements et Paléoenvironnements OCéaniques). Nous nous plaçons dans le cadre d'une application réelle dont l'objectif est d'étudier, à partir de données de valvométrie mesurées sur des bivalves, les perturbations des rythmes qui peuvent induire une perturbation du milieu comme une pollution. Sur ces données, nous avons comparé le comportement numérique des trois estimateurs présentés au second chapitre. Les résultats nous montrent qu'au sens de la variance, les estimateurs récursifs sont statistiquement meilleurs que la version non récursive. Ce résultat suggère d'utiliser l'estimateur récursif (voir Durrieu et al. 2009).

Chapitre 4

Estimation non paramétrique d'une densité

Dans ce chapitre, nous allons présenter brièvement quelques estimateurs non paramétriques d'une densité. Ce chapitre n'est pas un panorama exhaustif de tous les travaux faits et publiés dans cette thématique. Nous nous focalisons essentiellement sur les estimateurs suivants : histogramme, histogramme mobile, estimateur de Parzen-Rosenblatt, estimateur récursif à noyau : estimateur de Wolverton-Wagner et estimateur de Deheuvels. Nous rappelons également quelques unes de leurs propriétés asymptotiques.

4.1 Histogramme

L'estimateur le plus rudimentaire pour estimer une densité est l'histogramme des fréquences. Supposons que l'on ait X_1, \dots, X_n , n observations indépendantes d'une même loi de probabilité de densité g , où g est à support compact $[a, b]$. Estimer cette densité g par la méthode de l'histogramme revient à approcher g par une fonction en escaliers. Pour cela, on découpe $[a, b]$ en k classes $[\alpha_j, \alpha_{j+1}[$ où $j = 1, \dots, k$, avec $\alpha_1 = a$ et $\alpha_{k+1} = b$. L'estimateur histogramme s'écrit alors : $\forall x \in [a, b]$, $\exists j \in \{1, \dots, k\}$ tel que, $\forall x \in [\alpha_j, \alpha_{j+1}[$,

$$\check{g}_n(x) = \frac{\check{g}_j}{\alpha_{j+1} - \alpha_j},$$

où \check{g}_j est la fréquence empirique du nombre d'observations appartenant à la classe correspondante, c'est-à-dire

$$\check{g}_j = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[\alpha_j, \alpha_{j+1}[}(X_i).$$

Ce que l'on peut encore écrire plus concisément sous la forme suivante : $\forall x \in [a, b]$,

$$\check{g}_n(x) = \sum_{j=1}^k \frac{\check{g}_j}{\alpha_{j+1} - \alpha_j} \mathbf{1}_{[\alpha_j, \alpha_{j+1}[}(x),$$

ce qui est équivalent à

$$\check{g}_n(x) = \frac{1}{n} \sum_{j=1}^k \frac{\sum_{i=1}^n \mathbf{1}_{[\alpha_j, \alpha_{j+1}[}(X_i)}{(\alpha_{j+1} - \alpha_j)} \mathbf{1}_{[\alpha_j, \alpha_{j+1}[}(x).$$

Pour simplifier les notations, on supposera maintenant les classes de même largeur, c'est-à-dire que $\forall j = 1 \dots, k; \alpha_{j+1} - \alpha_j = b_n$. Il est aisé de remarquer que \check{g}_n est une densité de probabilité. Si l'on pense à la convergence de cet estimateur, il est clair que \check{g}_n sera d'autant plus proche de la vraie densité g que le nombre d'observations sera important et que les largeurs de classe seront plus étroites, d'où la nécessité d'imposer que $b_n \rightarrow 0$ quand $n \rightarrow \infty$. En revanche, il ne faut pas que b_n tende trop vite vers 0, sinon on pourrait avoir des classes ne contenant aucune observation, et donc une fonction en escalier \check{g}_n avec des marches d'ordonnée nulle, très éloignée de la réalité. Il faut donc que, certes b_n tende vers 0 avec n , et que, malgré cela, il "tombe" de plus en plus d'observations dans chaque classe, ce que l'on peut résumer par la condition :

$$nb_n \rightarrow \infty \quad \text{quand } n \rightarrow \infty.$$

Cet estimateur histogramme \check{g}_n a un défaut majeur. Considérons la classe $C_j = [\alpha_j, \alpha_{j+1}[$, et imaginons que le point $x \in C_j$ où l'on veut estimer $g(x)$ par $\check{g}_n(x)$ se situe près de l'extrémité α_j . Alors, toutes les observations de la classe C_j interviennent dans le calcul de \check{g}_n , mais on se rend compte qu'une observation située près de α_{j+1} sera prise en compte, alors qu'elle est assez éloignée de x , et qu'une observation située tout près de x dans la classe C_{j-1} n'entre pas en ligne de compte dans le calcul de \check{g}_n .

Pour remédier à cet inconvénient, on peut alors utiliser l'histogramme mobile, qui est un translaté de l'histogramme de manière à ce que l'observation x où l'on estime, se retrouve au centre d'une classe, plus précisément au centre de la classe $[x - h_n, x + h_n[$ où h_n désigne la demi-largeur d'une classe. L'estimateur histogramme mobile s'écrit alors :

$$\tilde{g}_n(x) = \frac{1}{2nh_n} \sum_{i=1}^n \mathbf{1}_{[x-h_n, x+h_n[}(X_i).$$

Remarquons que $x - h_n \leq X_i < x + h_n$ si et seulement si $-1 \leq h_n^{-1}(X_i - x) < 1$. On peut donc écrire que

$$\tilde{g}_n(x) = \frac{1}{2nh_n} \sum_{i=1}^n \mathbf{1}_{[-1, 1[}\left(\frac{X_i - x}{h_n}\right).$$

En posant $\mathbf{K}(x) = \frac{1}{2} \mathbf{1}_{[-1, 1[}(x)$, l'estimateur s'écrit alors :

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

Pour plus de détails concernant cette section, le lecteur pourra par exemple se référer aux articles suivants : Bosq et Lecoutre (1987, 1992), Praksa Rao (1983) et Carbon et Francq (1995).

4.2 Estimation par la méthode du noyau

L'estimateur précédemment construit peut encore être amélioré. En effet, maintenant que la classe est centrée en x , on peut tout de même remarquer que pour l'estimateur histogramme mobile, toutes les observations de cette classe ont le même rôle quant au calcul de $\tilde{g}_n(x)$. Il serait plus judicieux de penser que plus une observation est proche de x , plus elle doit intervenir dans le calcul de $\tilde{g}_n(x)$. L'idée alors la plus naturelle est de pondérer les observations en mettant d'autant plus de poids qu'on se trouve proche de x , et d'autant moins qu'on s'en trouve éloigné.

On a déjà vu un exemple de fonction de poids, notée \mathbf{K} au paragraphe précédent. C'était une densité de probabilité correspondant à la loi uniforme sur l'intervalle $[-1, 1]$. Cette fonction de poids est trop "brutale" et elle ne répond pas à nos préoccupations. On choisira donc des fonctions de poids dans une classe plus large de densités, comprenant notamment des densités à support non borné, et ayant un seul mode à l'origine comme par exemple la loi normale centrée réduite.

4.2.1 Estimateur de Parzen-Rosenblatt

Soit (X_n) une suite de variables aléatoires indépendantes et de même loi, de densité de probabilité g . On veut estimer g à partir d'un échantillon (X_1, X_2, \dots, X_n) .

Soit (h_n) une suite strictement positive, appelée **fenêtre** telle que

$$h_n \xrightarrow[n \rightarrow \infty]{} 0 \quad \text{et} \quad nh_n \xrightarrow[n \rightarrow \infty]{} +\infty.$$

Soit \mathbf{K} une fonction positive, bornée, appelée **noyau** telle que

$$\int_{\mathbb{R}} \mathbf{K}(x) dx = 1, \quad \int_{\mathbb{R}} x \mathbf{K}(x) dx = 0 \quad \text{et} \quad \int_{\mathbb{R}} \mathbf{K}^2(x) dx = \tau^2.$$

Nous pouvons choisir par exemple un des trois noyaux suivants

Noyau Uniforme :

$$K_{Unif}(x) = \begin{cases} \frac{1}{2} & \text{si } |x| \leq 1, \\ 0 & \text{sinon.} \end{cases}$$

Noyau d'Epanechnikov :

$$K_{Epa}(x) = \begin{cases} \frac{3}{4}(1-x^2) & \text{si } |x| \leq 1, \\ 0 & \text{sinon.} \end{cases}$$

Noyau Gaussien :

$$K_{Gau}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

On peut estimer g par l'estimateur de **Parzen-Rosenblatt** (voir Rosenblatt (1956) et Parzen (1962)).

$$\hat{g}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n \mathbf{K}\left(\frac{X_i - x}{h_n}\right). \quad (4.2.1)$$

Cet estimateur a été largement étudié par de nombreux auteurs, citons par exemple Wolverton et Wagner (1969), Roussas (2000, 2001), Bosq et al. (1999) et Lu (2001).

Une première justification concernant la forme de l'estimateur de **Parzen-Rosenblatt**, a été donnée précédemment. En voici une seconde (voir Deheuvels et Hominal, 1979), basée sur la fonction de répartition empirique associée à (X_n) , définie par

$$G_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}}.$$

La dérivée par différentiation de G_n est donnée par

$$\begin{aligned} \hat{g}_n(x) &= \frac{1}{h_n} \left(G_n\left(x + \frac{h_n}{2}\right) - G_n\left(x - \frac{h_n}{2}\right) \right), \\ &= \frac{1}{nh_n} \sum_{i=1}^n \mathbf{1}_{\left\{x - \frac{h_n}{2} \leq X_i \leq x + \frac{h_n}{2}\right\}}, \\ &= \frac{1}{nh_n} \sum_{i=1}^n \mathbf{1}_{\left\{\frac{1}{2} \leq \frac{X_i - x}{h_n} \leq \frac{1}{2}\right\}}, \\ &= \frac{1}{nh_n} \sum_{i=1}^n \mathbf{K}\left(\frac{X_i - x}{h_n}\right) \end{aligned}$$

où \mathbf{K} est le noyau uniforme.

Quant aux propriétés de convergence, on peut montrer que $\hat{g}_n(x)$ est un estimateur asymptotiquement sans biais de $g(x)$. On peut aussi montrer que, pour que $\mathbb{V}(\hat{g}_n(x)) \rightarrow 0$, il faut que $h_n \rightarrow 0$ et $nh_n \rightarrow +\infty$. L'estimateur à noyau avec un noyau autre que le noyau uniforme, présente l'avantage d'être un estimateur lissé par rapport à l'histogramme qui était constant par morceaux, et de convergence plus rapide. En effet, pour le meilleur choix de h_n , l'erreur quadratique moyenne $\mathbb{E}[\hat{g}_n(x) - g(x)]^2$ est de l'ordre de $n^{-4/5}$, alors que pour l'histogramme, $\mathbb{E}[\hat{g}_n(x) - g(x)]^2$ est de l'ordre de $n^{-2/3}$, (Bosq et Lecoutre, 1992).

Sous des conditions données sur h_n et \mathbf{K} , Rosenblatt (1971) a montré que $\hat{g}_n(x)$ converge en moyenne quadratique vers $g(x)$ et que

$$\mathbb{V}(\hat{g}_n(x)) \sim \frac{\tau^2}{nh_n} g(x) \quad \text{pour} \quad n \rightarrow \infty.$$

Dans le cas d'observations multidimensionnelles à valeur dans \mathbb{R}^p , $p \in \mathbb{N}^*$ (pour plus de détails voir Devroye et Györfi (1985)), \mathbf{K} est une densité sur \mathbb{R}^p et l'estimateur (4.2.1) devient,

$$\forall x \in \mathbb{R}^p, \quad \hat{g}_n(x) = \frac{1}{nh_n^p} \sum_{i=1}^n \mathbf{K}\left(\frac{X_i - x}{h_n}\right).$$

4.2.2 Estimateur récursif à noyau

On peut préférer utiliser un estimateur récursif à noyau de la densité de probabilité g . Un estimateur récursif à noyau est donné par la relation récursive suivante :

$$\hat{g}_n(x) = (1 - \gamma_n)\hat{g}_{n-1}(x) + \frac{\gamma_n}{h_n} \mathbf{K} \left(\frac{x - X_n}{h_n} \right)$$

où (γ_n) et (h_n) sont deux suites de réels positifs qui tendent vers 0. Voici deux exemples d'estimateurs récursifs.

Estimateur de Wolverton-Wagner. Il correspond au choix $\gamma_n = 1/n$ et s'écrit :

$$\hat{g}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i} \mathbf{K} \left(\frac{x - X_i}{h_i} \right). \quad (4.2.2)$$

Cet estimateur récursif a été introduit par Wolverton et Wagner (1969) et étudié par Yamato (1971).

Estimateur de Deheuvels. Dans le cas où on pose $\gamma_n = h_n/S_n$ avec $S_n = \sum_{i=1}^n h_i$, Deheuvels (1973) a proposé un autre estimateur récursif de la densité de la forme :

$$\check{g}_n(x) = \frac{1}{S_n} \sum_{i=1}^n \mathbf{K} \left(\frac{x - X_i}{h_i} \right).$$

Quelques résultats de convergence pour l'estimateur de Wolverton-Wagner

Yamato (1971) a démontré en particulier (voir le théorème 2, page 4) que pour tout couple $(\mathbf{K}(y), (h_n))$ appartenant à C_2 sous ensemble de C_1 pour lequel les suites (h_n) sont décroissantes, $\hat{g}_n(x)$ converge en moyenne quadratique vers $g(x)$. Si de plus $(\mathbf{K}(y), (h_n))$ appartenant à C'_2 sous ensemble de C_2 pour lequel les suites (h_n) vérifient :

$$\frac{h_n}{n} \sum_{i=1}^n \frac{1}{h_i} \xrightarrow{n \rightarrow \infty} \alpha$$

avec $\alpha \in]0, 1[$, alors il a montré (voir le théorème 3, page 6) que

$$\lim_{n \rightarrow \infty} nh_n \nabla(\hat{g}_n(x)) = \alpha \tau^2 g(x).$$

Le théorème suivant nous donne la convergence presque sûre (voir le théorème 1-1°, page 1219 de Deheuvels (1974)) et la normalité asymptotique de $\hat{g}_n(x)$.

Théorème 4.1. *Si g est dérivable à dérivée bornée et si la fenêtre est telle que $h_n = n^{-\alpha}$ avec $0 < \alpha < 1$, on a, pour tout $x \in \mathbb{R}$,*

$$\lim_{n \rightarrow \infty} \hat{g}_n(x) = g(x) \quad p.s.$$

De plus, si $1/5 < \alpha < 1$, on a

$$\sqrt{nh_n} (\hat{g}_n(x) - g(x)) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\tau^2 g(x)}{1 + \alpha} \right).$$

Le théorème de limite centrale a été obtenu par Duflo (1997) (voir page 233) avec $1/3 < \alpha < 1$, et par Bercu et Chafaï (2007) (voir page 130) avec $1/5 < \alpha < 1$.

Quelques résultats de convergence pour l'estimateur de Deheuvels

Parmi de nombreux autres résultats, Deheuvels a démontré que $\check{g}_n(x)$ converge en moyenne quadratique vers $g(x)$ et qu'en particulier (voir le théorème 1 page 1119 de Deheuvels (1973)) :

$$\mathbb{V}(\check{g}_n(x)) \sim \frac{1}{S_n} \tau^2 g(x).$$

Le théorème suivant nous donne la convergence presque sûre (voir le théorème 1-3°, page 1219 de Deheuvels (1974)) et la normalité asymptotique de $\hat{g}_n(x)$.

Théorème 4.2. *Si g est dérivable à dérivée bornée et si la fenêtre est telle que $h_n = n^{-\alpha}$ avec $0 < \alpha < 1$, on a, pour tout $x \in \mathbb{R}$,*

$$\lim_{n \rightarrow \infty} \check{g}_n(x) = g(x) \quad p.s.$$

De plus, si $1/5 < \alpha < 1$, on a

$$\sqrt{nh_n} (\check{g}_n(x) - g(x)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \tau^2 g(x)).$$

En utilisant les martingales comme dans le théorème précédent, on peut obtenir de même un théorème de limite centrale pour $\check{g}_n(x)$.

Choix du noyau et de la fenêtre : Ces choix ne peuvent s'effectuer que par l'utilisation de certains critères. Sans entrer dans tous les détails, il s'avère que le choix du noyau n'a pas d'influence majeure s'il est choisi dans une classe raisonnable. En revanche, le choix de la fenêtre h_n est crucial. En général, h_n est obtenue par des techniques de validation croisée. Nous allons plus détailler ces choix dans le chapitre suivant.

Chapitre 5

Estimation non paramétrique d'une fonction de régression

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 f non nécessairement linéaire de X , c'est-à-dire de trouver une expression de la forme :

$$Y = f(X) + \varepsilon,$$

où le terme ε correspond à une erreur ou un bruit de modélisation aléatoire. On peut alors chercher à déterminer f comme solution du problème de minimisation :

$$\min \mathbb{E}[Y - f(X)]^2.$$

Si le bruit ε est intégrable et centré et s'il est indépendant de la variables aléatoire X , on a

$$f(X) = \mathbb{E}[Y | X].$$

Dans ce chapitre, nous allons présenter brièvement quelques estimateurs non paramétriques d'une fonction de régression. Ce chapitre n'est pas un panorama exhaustif de tous les travaux faits et publiés dans cette thématique. Nous nous focalisons essentiellement sur les estimateurs suivants : régressogramme, régressogramme mobile, régression par la méthode du noyau, estimateur de Nadaraya-Watson, estimateur de Nadaraya-Watson récursif et estimateur de Révész . Nous rappelons également quelques unes de leurs propriétés asymptotiques.

5.1 Modèle de régression

Un des modèles les plus fréquemment utilisés en statistique paramétrique ou non-paramétrique est le modèle de régression. 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) . Pour tout $1 \leq i \leq n$, on a donc

$$Y_i = f(X_i) + \varepsilon_i \tag{5.1.1}$$

où la suite (ε_i) est constituée de variables aléatoires indépendantes et de même loi normale centrée et de variance σ^2 avec $\sigma^2 > 0$. On suppose également que les suites (X_i) et (ε_i) sont indépendantes.

Nous considérons le problème délicat posé par l'estimation de la fonction de régression, sans hypothèse particulière sur la loi du couple (X, Y) autre que celle de l'existence de la fonction de lien f , supposée suffisamment régulière, ainsi que de moments supérieurs d'ordre convenable pour X et Y .

Nous allons présenter dans la suite des estimateurs de la fonction de régression. Pour cela, on va maintenant écrire la fonction de régression sous la forme d'un ratio. Soit (X, Y) un couple de variables aléatoires réelles admettant une densité jointe sur \mathbb{R}^2 notée $g_{X,Y}$. On note g la densité marginale de X . La variable aléatoire Y est supposée intégrable, i.e. $\mathbb{E}[|Y|] < \infty$. Nous pouvons alors définir proprement la fonction de régression ou l'espérance conditionnelle de Y sachant $X = x$, par

$$f(x) = \mathbb{E}[Y | X = x] = \frac{\int_{\mathbb{R}} yg_{X,Y}(x, y)dy}{\int_{\mathbb{R}} g_{X,Y}(x, y)dy} = \frac{r(x)}{g(x)}.$$

Cette expression n'est bien sûr valable que si g ne s'annule pas en x . Le problème de l'estimation de f est de type non-paramétrique, i.e. la fonction de régression appartient à un ensemble non-paramétrique ou infini-dimensionnel. Cette fonction $f(x)$ réalise, pour tout x fixé, la meilleure approximation de Y sachant $X = x$, au sens des moindres carrés, en supposant Y de carré intégrable.

5.2 Régressogramme

D'un point de vue pratique, à partir de n observations $(X_1, Y_1), \dots, (X_n, Y_n)$, on va estimer $f(x) = \mathbb{E}[Y | X = x]$, en fabriquant des classes C_j où se situent ou non les X_i . Pour la classe C_j où se trouve le point x , on effectue une moyenne des Y_i correspondants aux X_i de cette classe C_j . Pour tout $x \in C_j$, on estime $f(x)$ par :

$$\check{f}_n(x) = \frac{\sum_{i=1}^n \mathbf{1}_{C_j}(X_i)Y_i}{\sum_{i=1}^n \mathbf{1}_{C_j}(X_i)},$$

qui est proposé par Tukey (1961). Il est clair, comme pour l'histogramme, que le régressogramme $\check{f}_n(x)$ est constant sur chaque classe C_j . Pour l'améliorer, on va suivre alors la même démarche que pour l'histogramme.

Comme pour l'histogramme mobile, la première amélioration consiste à centrer la classe sur le point x où l'on estime la fonction de régression. Ce régressogramme mobile s'écrit

alors

$$\tilde{f}_n(x) = \frac{\sum_{i=1}^n \mathbf{1}_{[x-h_n, x+h_n]}(X_i) Y_i}{\sum_{i=1}^n \mathbf{1}_{[x-h_n, x+h_n]}(X_i)}.$$

Il en découle que

$$\tilde{f}_n(x) = \frac{\sum_{i=1}^n \mathbf{1}_{[-1,1]}(X_i - x/h_n) Y_i}{\sum_{i=1}^n \mathbf{1}_{[-1,1]}(X_i - x/h_n)}.$$

Cet estimateur a été étudié par de nombreux auteurs, voir par exemple Bosq (1970), Geffroy (1975), Sabry (1978) et Collomb (1978a, 1978b).

5.3 Régression par la méthode du noyau

Par analogie au cas de l'histogramme, une nouvelle étape consiste à remplacer le noyau uniforme sur l'intervalle $[-1; 1]$ par un noyau quelconque du type décrit dans le chapitre précédent. La régression par la méthode du noyau s'écrit alors

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

Dans le cas d'observations multidimensionnelles à valeur dans \mathbb{R}^p (pour plus détails voir Härdle (1989)), cet estimateur est inchangé.

Nous présentons maintenant le célèbre estimateur à noyau de la fonction de régression introduit par Nadaraya (1964) et Watson (1964). Ensuite, nous nous intéresserons à l'estimateur de Nadaraya-Watson récursif (Duflo, 1997) puis à l'estimateur récursif de Révész (Révész, 1977 et Mokkaï et al., 2009). Nous rappellerons aussi les principales propriétés asymptotiques de ces estimateurs.

5.4 L'estimateur de Nadaraya-Watson

Le premier estimateur rencontré dans la littérature est l'estimateur à noyau de Nadaraya-Watson (1964), noté estimateur **NW**. Il est construit à partir d'une fonction noyau \mathbf{K} et d'une fenêtre (h_n) , de manière analogue à l'estimateur à noyau de la densité de probabilité g introduit par Parzen (1962) et Rosenblatt (1956).

La fenêtre (h_n) désigne une suite de nombres réels strictement positifs vérifiant $h_n \rightarrow 0$ et $n h_n \rightarrow \infty$ lorsque $n \rightarrow \infty$. Le noyau est une fonction mesurable, positive et

bornée telle que

$$\int_{\mathbb{R}} K(x) dx = 1, \quad \int_{\mathbb{R}} x\mathbf{K}(x) dx = 0, \quad \int_{\mathbb{R}} |x| \mathbf{K}(x) dx < +\infty, \quad \int_{\mathbb{R}} \mathbf{K}^2(x) dx = \tau^2.$$

L'estimateur **NW** se présente sous la forme d'une moyenne locale pondérée des valeurs Y_i . Il est défini par

$$\hat{f}_n(x) = \begin{cases} \frac{\sum_{i=1}^n \mathbf{K}\left(\frac{x - X_i}{h_n}\right) Y_i}{\sum_{i=1}^n \mathbf{K}\left(\frac{x - X_i}{h_n}\right)} & \text{si } \sum_{i=1}^n \mathbf{K}\left(\frac{x - X_i}{h_n}\right) \neq 0, \\ \frac{1}{n} \sum_{i=1}^n Y_i & \text{sinon.} \end{cases} \quad (5.4.1)$$

Le noyau \mathbf{K} détermine la forme du voisinage autour du point x , tandis que la fenêtre (h_n) contrôle la taille de ce voisinage, c'est-à-dire implicitement le nombre d'observations prises en compte pour effectuer la moyenne locale. Intuitivement, il est naturel que la fenêtre (h_n) soit prépondérante pour la consistance de l'estimateur **NW**.

5.4.1 Consistance de l'estimateur Nadaraya-Watson

L'estimateur à noyau de la régression est donc dépendant du choix de deux paramètres, la fenêtre (h_n) et le noyau \mathbf{K} . Nous verrons dans les sections suivantes que le paramètre crucial est la fenêtre pour obtenir de bonnes propriétés asymptotiques. Toutefois, le noyau ne doit pas être négligé, il permet de réduire le biais de notre estimateur en s'appuyant sur les propriétés de régularité de la courbe de régression. Dans cette section, nous déterminerons les conditions sur la fenêtre et le noyau nécessaires à la consistance de l'estimateur **NW**.

Nous obtenons la consistance des estimateurs du type **NW**, via la décomposition biais-variance suivante,

$$\mathbb{E} \left[(\hat{f}_n(x) - f(x))^2 \right] = \mathbb{V} \left(\hat{f}_n(x) \right) + \left(\mathbb{E}[\hat{f}_n(x)] - f(x) \right)^2. \quad (5.4.2)$$

Une simple étude des critères de convergence vers zéro du biais et de la variance ci-dessus nous précisera les conditions nécessaires à la consistance de l'estimateur **NW**.

5.4.1.1 Calcul de la variance

Nous débutons l'étude de l'estimateur **NW** par le calcul de sa variance et son expression asymptotique. Soit

$$\sigma^2(x) = \mathbb{V}(Y | X = x) = \frac{1}{g(x)} \int y^2 g_{X,Y}(x, y) dy - (f(x))^2.$$

Proposition 5.1. (Collomb 1976, 1977a ou Blondin, 2004) On suppose que $\mathbb{E}[Y^2] < \infty$. À chaque point de continuité x des fonctions $f(x)$, $g(x)$ et $\sigma^2(x)$, tel que $g(x) > 0$, on a

$$\mathbb{V}(\hat{f}_n(x)) = \frac{1}{nh_n} \times \left\{ \frac{\sigma^2(x)}{g(x)} \int_{\mathbb{R}} \mathbf{K}^2(u) du \right\} (1 + o(1)),$$

où le terme $o(1)$ tend vers 0 lorsque $n \rightarrow \infty$.

Si la fenêtre (h_n) satisfait les conditions $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ lorsque $n \rightarrow \infty$, la variance de l'estimateur **NW** tend vers zéro. Dans l'expression asymptotique des termes de variance de l'estimateur à noyau, nous retrouvons invariablement la quantité

$$\tau^2 = \int_{\mathbb{R}} \mathbf{K}^2(u) du = \|\mathbf{K}\|_2^2.$$

Pour s'assurer de la finitude de cette intégrale, nous pouvons choisir la fonction noyau \mathbf{K} à variation bornée sur \mathbb{R} et à support compact. En vue de l'optimalité asymptotique, la variance minimale sera obtenue en minimisant la quantité ci-dessus suivant \mathbf{K} dans une certaine classe de noyaux fixée. Pour une expression explicite des noyaux de variance minimale, nous citons l'article de Gasser, Müller et Mammitzsch (1985).

Remarque 5.1. Afin de s'assurer de l'existence de la variance conditionnelle $\sigma^2(x)$, l'hypothèse $\mathbb{E}[Y^2] < \infty$ est indispensable.

Pour plus détails, nous pouvons citer par exemple les articles de Collomb (1977a, 1981) et les thèse de Collomb (1976), Blondin (2004).

5.4.1.2 Calcul du biais

Le traitement du biais est purement analytique et repose essentiellement sur un développement de Taylor. Il nous faut supposer certaines conditions de régularité sur les fonctions f et g qui détermineront l'ordre du biais asymptotique en fonction du paramètre de lissage (h_n) . L'estimateur **NW** se présente sous la forme d'un quotient aléatoire, c'est pourquoi on utilise généralement comme terme de centrage l'approximation suivante

$$\tilde{\mathbb{E}}[\hat{f}_n(x)] = \frac{\mathbb{E}[\hat{r}_n(x)]}{\mathbb{E}[\hat{g}_n(x)]}.$$

Cette formule est plus facile à manipuler et permet notamment la linéarisation de la déviation $\tilde{d}_n(x) = \hat{f}_n(x) - \tilde{\mathbb{E}}[\hat{f}_n(x)]$.

Afin de justifier le choix du terme de centrage, Nadaraya (1989) a montré que lorsque Y est borné et $nh_n \rightarrow \infty$,

$$\mathbb{E}[\hat{f}_n(x)] = \tilde{\mathbb{E}}[\hat{f}_n(x)] + \mathcal{O}((nh_n)^{-1}), \quad (5.4.3)$$

lorsque $\mathbb{E}[Y^2] < \infty$ et $nh_n^2 \rightarrow \infty$,

$$\mathbb{E}[\hat{f}_n(x)] = \tilde{\mathbb{E}}[\hat{f}_n(x)] + \mathcal{O}((n^{1/2}h_n)^{-1}). \quad (5.4.4)$$

Supposons que la variable Y est bornée, de telle sorte que (5.4.3) soit vérifiée, le biais asymptotique de l'estimateur **NW** est donné par une proposition suivante :

Proposition 5.2. (Collomb 1976, 1977a ou Blondin, 2004) Supposons que f et g sont de class $C^2(\mathbb{R})$ et que le noyau \mathbf{K} est d'ordre 2, i.e. tel que

$$\int_{\mathbb{R}} K(z) dz = 1, \quad \int_{\mathbb{R}} z\mathbf{K}(z) dz = 0, \quad \int_{\mathbb{R}} z^2 \mathbf{K}(z) dz < \infty.$$

Nous avons alors, lorsque $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$,

$$\mathbb{E}[\hat{f}_n(x)] - f(x) = \frac{h_n^2}{2} \left(\left\{ f''(x) + 2f'(x) \frac{g'(x)}{g(x)} \right\} \int_{\mathbb{R}} z^2 \mathbf{K}(z) dz \right) (1 + o(1)).$$

Notons que le terme $o(1)$ ci-dessus se décompose comme suit $\left\{ \mathcal{O}(h_n) + \mathcal{O}((nh_n)^{-1}) \right\}$, d'après (5.4.3).

Lorsque la fonction de régression admet des conditions de régularité supplémentaires, il est possible de réduire le biais asymptotique de l'estimateur **NW** en utilisant un noyau d'ordre supérieur. Soit q un entier naturel fixé, le biais asymptotique de l'estimateur **NW** est donné par

Proposition 5.3. (Collomb 1976, 1977a ou Blondin, 2004) Supposons que f et g sont de class $C^q(\mathbb{R})$ et que le noyau \mathbf{K} est d'ordre q , i.e. tel que

$$\int_{\mathbb{R}} K(z) dz = 1, \quad \int_{\mathbb{R}} z^j \mathbf{K}(z) dz = 0, \quad j = 1, \dots, q-1, \quad \text{et} \quad \int_{\mathbb{R}} z^q \mathbf{K}(z) dz < \infty.$$

Nous avons alors, lorsque $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$,

$$\mathbb{E}[\hat{f}_n(x)] - f(x) = \frac{h_n^q}{q!} \left(\left\{ f^{(q)}(x) + 2f^{(q-1)}(x) \frac{g'(x)}{g(x)} \right\} \int_{\mathbb{R}} z^q \mathbf{K}(z) dz \right) (1 + o(1)).$$

Pour plus détails, nous pouvons citer par exemple les articles de Collomb (1977a, 1981), Müller et Song (1993), Linton et Nielsen (1994), Müller (1997), Choi et al. (2000) et Hall et Miller (2003); les thèse de Collomb (1976) et Blondin (2004).

5.4.2 Normalité asymptotique

La première démonstration de la normalité asymptotique de l'estimateur **NW** est due à Schuster (1972). On se réfère également aux théorèmes 1.3 et 1.4 p. 117-120 de Nadaraya (1989) et au théorème 4.2.1 p. 99 de Härdle (1990), qui reposent sur d'autres méthodes de démonstration. Le noyau \mathbf{K} est supposé borné, à support compact et de classe $C^2(\mathbb{R})$. La fenêtre est choisie égale à $cn^{-1/5}$.

Théorème 5.1. Härdle (1990) On suppose que Y possède un moment d'ordre > 2 . Les fonctions g et f sont supposées deux fois continûment dérivables sur \mathbb{R} . À chaque point de continuité x de $\sigma^2(x)$, tel que $g(x) > 0$, on a

$$\sqrt{nh_n} \left(\hat{f}_n(x) - f(x) \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(B(x), v^2(x) \right),$$

$$\text{où } v^2(x) = \frac{\tau^2 \sigma^2(x)}{g(x)} \quad \text{et} \quad B(x) = \left\{ f''(x) + 2f'(x) \frac{g'(x)}{g(x)} \right\} \times \int_{\mathbb{R}} u^2 \mathbf{K}(u) du.$$

5.5 Nadaraya-Watson récursif et Révész récursif

5.5.1 Les estimateurs récursifs

L'estimateur de Nadaraya-Watson récursif (Duflo, 1997), noté **NWR**, est défini par

$$\tilde{f}_n(x) = \begin{cases} \frac{\sum_{i=1}^n \frac{1}{h_i} \mathbf{K}\left(\frac{x - X_i}{h_i}\right) Y_i}{\sum_{i=1}^n \frac{1}{h_i} \mathbf{K}\left(\frac{x - X_i}{h_i}\right)} & \text{si } \sum_{i=1}^n \frac{1}{h_i} \mathbf{K}\left(\frac{x - X_i}{h_i}\right) \neq 0, \\ \frac{1}{n} \sum_{i=1}^n Y_i & \text{sinon.} \end{cases} \quad (5.5.1)$$

On propose également d'utiliser l'estimateur récursif de Révész (introduit par Révész (1993, 1977) et étudié par Mokkadem et al., 2009), noté **R**, qui est défini par

$$\check{f}_n(x) = \check{f}_{n-1}(x) + \frac{1}{nh_n} \mathbf{K}\left(\frac{x - X_n}{h_n}\right) (Y_n - \check{f}_{n-1}(x)). \quad (5.5.2)$$

5.5.2 Propriétés asymptotiques

Nous rappelons ici les principales propriétés asymptotiques des estimateurs **NW**, **NWR** et **R**. Nous choisissons d'utiliser la fenêtre $h_n = n^{-\alpha}$ avec $0 < \alpha < 1$. Nous ajoutons deux conditions de régularité : la fonction de lien f et la densité g sont bornées et deux fois continûment dérivables sur \mathbb{R} et $\mathbb{E}(Y^2) < \infty$.

Théorème 5.2 (NW). *Pour tout $\alpha \in [1/5, 1[$ et pour tout $x \in \mathbb{R}$ tel que $g(x) > 0$, on a*

1.

$$\hat{f}_n(x) \xrightarrow{p.s.} f(x).$$

2.

$$\sqrt{nh_n} \left(\hat{f}_n(x) - f(x) \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\sigma^2 \tau^2}{g(x)} \right).$$

Duflo (1997) a étudié les propriétés asymptotiques de l'estimateur récursif de Nadaraya-Watson lorsque la fenêtre h_n est choisie égale à $n^{-\alpha}$. Elle a montré dans le cadre du modèle $Y = f(X) + \varepsilon$ que

Théorème 5.3 (NWR). *Pour tout $\alpha \in]1/3, 1[$ et pour tout $x \in \mathbb{R}$ tel que $g(x) > 0$, on a*

1.

$$\tilde{f}_n(x) \xrightarrow{p.s.} f(x).$$

2.

$$\sqrt{nh_n} \left(\tilde{f}_n(x) - f(x) \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\sigma^2 \tau^2}{g(x)(1 + \alpha)} \right).$$

Révész (1977) a étudié les propriétés asymptotiques de l'estimateur récursif de Révész lorsque la fenêtre h_n est choisie égale à $n^{-\alpha}$. Il a montré que

Théorème 5.4 (R). *Pour tout $\alpha \in]1/2, 1[$ et pour tout $x \in \mathbb{R}$ tel que $2g(x) > 1 - \alpha$, on a*

1.

$$\check{f}_n(x) \xrightarrow{p.s.} f(x).$$

2.

$$\sqrt{nh_n} (\check{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\sigma^2 \tau^2 g(x)}{2g(x) - (1 - \alpha)} \right).$$

Il est possible d'estimer la variance de la loi normale limite. Pour cela, la densité marginale g est estimée par $\hat{g}_n(x)$ donnée en (4.2.1) pour **NW** et par $\hat{g}_n(x)$ donnée en (4.2.2) pour **NWR** et **R**. La variance σ^2 est estimée respectivement pour **NW**, **NWR** et **R** par

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{f}_n(x))^2,$$

$$\tilde{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \tilde{f}_n(x))^2,$$

$$\check{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \check{f}_n(x))^2.$$

Chapitre 6

Application en valvométrie

6.1 Introduction

Les activités humaines sont responsables d'importants rejets d'agents polluants dans le milieu naturel. Ces polluants entraînent la dégradation de nombreux biotopes, perturbant les écosystèmes et posant également des problèmes en termes de santé publique. Des réglementations et des contrôles sur la qualité des eaux ont été mis en place. Parmi ces contrôles, les bioindicateurs sont de plus en plus utilisés et sont très efficaces par leurs capacités à révéler la présence de traces (concentrations très faibles) de contaminant. Nous utilisons ici comme moyen de surveillance du milieu la valvométrie. La valvométrie (mesure de l'activité des valves de mollusques) est une technique qui permet d'enregistrer les réactions de bivalves, face aux changements de la qualité de l'eau dans laquelle ils vivent. Le premier enregistrement publié de l'activité valvaire de mollusques bivalves date à notre connaissance de 1909 (Marceauc, 1909). Depuis, des dizaines de systèmes ont été publiés. Les deux plus connus, parce qu'ils ont trouvé une application commerciale, sont le Mosselmonitor (Kramer et Foekema, 2001) et le Dreissena monitor (Borcherding, 2006). Ifremer a aussi développé un système avec la société Micrel NKE.

Nous voulons caractériser et traiter statistiquement le signal obtenu et le traiter sur de longues séries. Un des buts est de faire du suivi à très long terme afin de comprendre le fonctionnement et la trajectoire temporelle des systèmes environnementaux et d'appréhender leur résilience, leurs réponses à des événements perturbateurs et leur dynamique d'ajustement.

Les mollusques bivalves ventilent tout au long de la journée pour se nourrir et respirer. Ils sont équipés de récepteurs qui leur permettent en permanence d'estimer la qualité de l'eau dans laquelle ils baignent, de façon à pouvoir réagir immédiatement face à une eau qu'ils jugent nocive pour leur intégrité. Le cas des mollusques bivalves est particulièrement intéressant en tant qu'espèce bioindicatrice car ce sont des animaux sédentaires qui peuvent être témoins de changement locaux de la qualité de l'eau. Le but est de caractériser son activité valvaire (qui résume à elle seule l'ensemble du comportement locomoteur de ce type d'animal) en fonction des paramètres du milieu afin de discriminer ensuite un comportement atypique lié à une contamination aquatique. Au final, nous recherchons à caractériser une signature comportementale spécifique d'un milieu, de la nature d'un contaminant ou d'un facteur du milieu. Ces

facteurs peuvent être soit des algues toxiques (Bassin d’Arcachon), soit des déchets de l’activité anthropique comme les métaux lourds (lagon sud de Nouvelle Calédonie, usine Goro-Nickel), les pesticides ou les HAPs, soit liés au réchauffement global dans un fjord (Arctique, Ny-Alesund).

Nous avons utilisé un système de bivalves basé à Arcachon, nous avons travaillé en particulier sur les efflorescences d’algues toxiques produisant des toxines diarrhéiques (Diarrheic Shellfish Poisoning toxins) ou paralysantes (Paralytic Shellfish Poisoning toxins). Elles sont en très forte augmentation le long des côtes françaises et elles causent de lourdes pertes économiques dans le secteur conchylicole et ostréicole dues aux fermetures administratives de commercialisation des coquillages.

Notre travail consiste dans un premier temps à comprendre d’abord comment les facteurs du milieu règlent la rythmicité biologique pour discriminer ensuite un comportement qui n’est pas régi par ces facteurs, mais résulte d’une perturbation du milieu ambiant.

6.2 Modèle et Estimateurs

Il est possible de caractériser chaque animal par une équation sur une journée. Nous proposons un modèle de régression non paramétrique et comparons trois estimateurs non paramétriques, récursif ou non, de la fonction de lien. Ce travail est fait en routine sur des données acquises au niveau de la jetée d’EYRAC (Arcachon) et du récif IORO (Nouvelle Calédonie). L’objectif final est d’utiliser la valvométrie comme un système de biosurveillance de la qualité du milieu pour suivre l’impact potentiel d’une nouvelle mine de Nickel et de cobalt. Les enregistrements et les résultats du traitement statistique sont accessibles sur le site web “L’oeil du mollusque” (http://www.domino.u-bordeaux.fr/molluscan_eye).

Précisément, nous disposons d’un échantillon composé de n couples indépendants de variables aléatoires $(T_1, Y_1), \dots, (T_n, Y_n)$ et nous considérons le modèle de régression non paramétrique donné, pour $i = 1, \dots, n$, par

$$Y_i = f(T_i) + \varepsilon_i. \tag{6.2.1}$$

Dans ce modèle intervient une fonction de lien f inconnue à estimer qui exprime la valeur moyenne de l’écartement valvaire de nos bivalves en fonction du temps T et un terme aléatoire d’erreur ε de loi inconnue et indépendant de T . Nous proposons trois estimateurs non paramétriques de la fonction f . Le premier estimateur est l’estimateur de Nadaraya-Watson **NW**. On propose également d’utiliser l’estimateur de Nadaraya-Watson récursif **NWR** et l’estimateur récursif de Révész **R**.

Nous évaluons ici les trois estimateurs non paramétriques sur les données recueillies depuis septembre 2007 au niveau du site IORO en Nouvelle Calédonie. Les mesures sont collectées par deux cartes électroniques qui gèrent à la fois l’acquisition (toutes les 1,6 secondes) sur un groupe de 16 bénitiers et le transfert des données. Ainsi ce dispositif génère tous les jours, pour chaque bénitier, 54000 couples de valeurs (T_i, Y_i) qui sont le temps en heure et l’écartement valvaire en mm. Les résultats étant similaires en considérant les 16 huîtres de la jetée d’EYRAC à Arcachon, nous présentons seulement les résultats de modèle du site IORO.

6.3 Enregistrement et Acquisition des données

Nous collons sur les valves des bivalves des électrodes (électro-aimant) légères (200 mg hors enrobage) et non invasives (l'animal est libre de se positionner et se déplacer) qui permettent de mesurer en permanence leur état d'ouverture. Ces électrodes sont gérées par une carte analogique immergée à côté des animaux en boîtier étanche. Le tout est relié à une seconde carte électronique en surface pilotée sous Linux. Les données brutes sont ensuite transférées tous les jours par le réseau de la téléphonie mobile (GPRS) et internet (FTP) sur les stations de travail du laboratoire à Arcachon. La fréquence d'échantillonnage est fixée à une mesure toutes les 0.1 s mais cette fréquence peut être changée. Le tout consomme 2 watts et est alimentable par panneaux solaires. Pour avoir un nombre d'animaux représentatif, nous travaillons sur des groupes de 16 animaux. Nous disposons donc d'un total de 864 000 couples de points par jour, tous les jours et plus de 300 millions de mesures effectuées par an.

Une fois les données brutes arrivées au laboratoire, elles sont modélisées et traitées statistiquement. Deux heures après le transfert depuis le terrain, les résultats sont ensuite mis en accès public et ou professionnel sur le site web (*l'Oeil du Mollusque* http://www.domino.u-bordeaux.fr/molluscan_eye) (voir FIG. 6.1).

Le premier site équipé est situé devant la Station de Biologie Marine d'Arcachon. Il fonctionne depuis mars 2006. Ce sont des huîtres que nous utilisons entre autres pour tenter de détecter l'arrivée de contaminant dans le bassin d'Arcachon. Pour le 2ème site, nous avons installé 16 bénitiers dans le lagon sud de Nouvelle Calédonie. L'idée est de participer à la création d'un Observatoire de l'Environnement, en relation avec la province sud de Nouvelle Calédonie pour suivre l'impact potentiel d'une nouvelle mine de nickel et de cobalt.

Nous donnons ci-dessous deux exemples des enregistrements qui représentent le comportement journalier des bivalves. Les deux figures du haut correspondent à des enregistrements de bénitiers (récif en Nouvelle Calédonie) et les deux autres sont des enregistrements d'huîtres (jetée EYRAC à Arcachon). Sur chaque figure, nous avons reporté, en abscisse, le temps sur 24 heures et en ordonnée l'écartement valvaire en mm (voir FIG. 6.2).

6.4 Résultats

Le choix de la fenêtre est crucial pour nos trois estimateurs. En pratique, ce paramètre est choisi comme un compromis entre la variance et le biais de l'estimation. Une importante littérature est consacrée à ce sujet, et en particulier aux méthodes de sélection automatique par minimisation d'un critère. Nous utilisons comme critère la méthode de la validation croisée (Härdle et Marron, 1985 et Härdle, 1990) qui consiste à minimiser par rapport à h la fonction

$$CV(h) = \sum_{i=1}^n (Y_i - \hat{m}_{(-i)}(T_i; h))^2$$

où $\hat{m}_{(-i)}(T_i; h)$ désigne un estimateur (**NW**, **NWR** ou **R**) de la fonction de lien au point T_i calculé sur l'échantillon privé du couple (T_i, Y_i) .

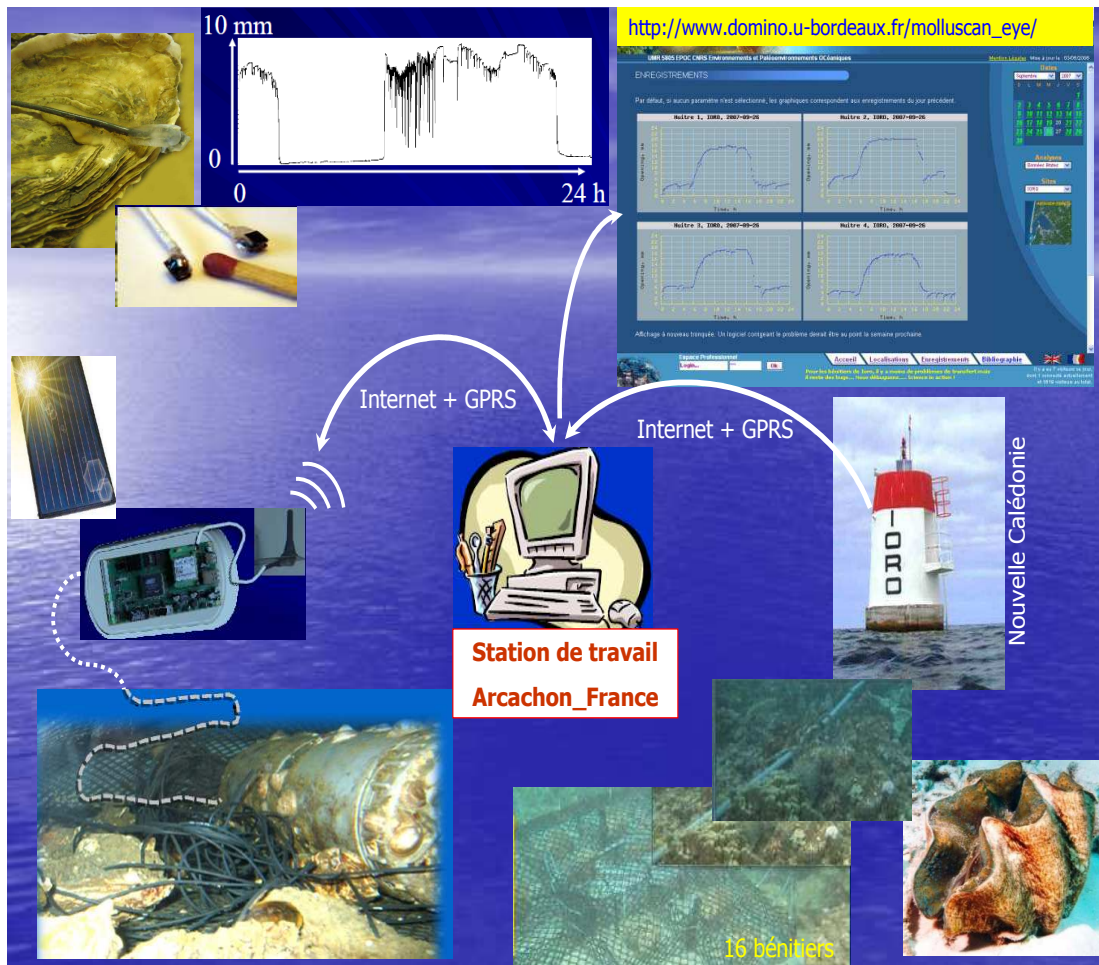


FIG. 6.1: Enregistrement et Acquisition des données

Enfin, nous déterminons les largeurs des fenêtres en utilisant la méthode de validation croisée (voir FIG. 6.3). Ensuite, nous avons choisi le noyau Gaussien pour les estimateurs **NW** et **NWR**. Pour l'estimateur **R**, l'article de Révész (1977) recommande le choix du noyau uniforme.

La figure FIG. 6.3 représente le critère $CV(h) = CV(\alpha)$ par la méthode de validation croisée. Nous voyons clairement que les fonctions CV obtenues numériquement sont toutes convexes pour nos 3 estimateurs. Les fenêtres estimées sont $h = n^{-\alpha_{cv}}$ où les α_{cv} pour chaque estimateur sont respectivement égales à $n^{-0,47}$ pour **NW**, $n^{-0,50}$ pour **NWR** et $n^{-0,99}$ pour **R**.

Avec ces choix de noyaux et de fenêtres, nous obtenons un très bon ajustement des modèles de régression aux données (voir FIG. 6.4).

La figure FIG. 6.4 représente le comportement journalier pour les trois estimateurs. Nous obtenons un très bon ajustement des estimateurs aux données brutes, avec les

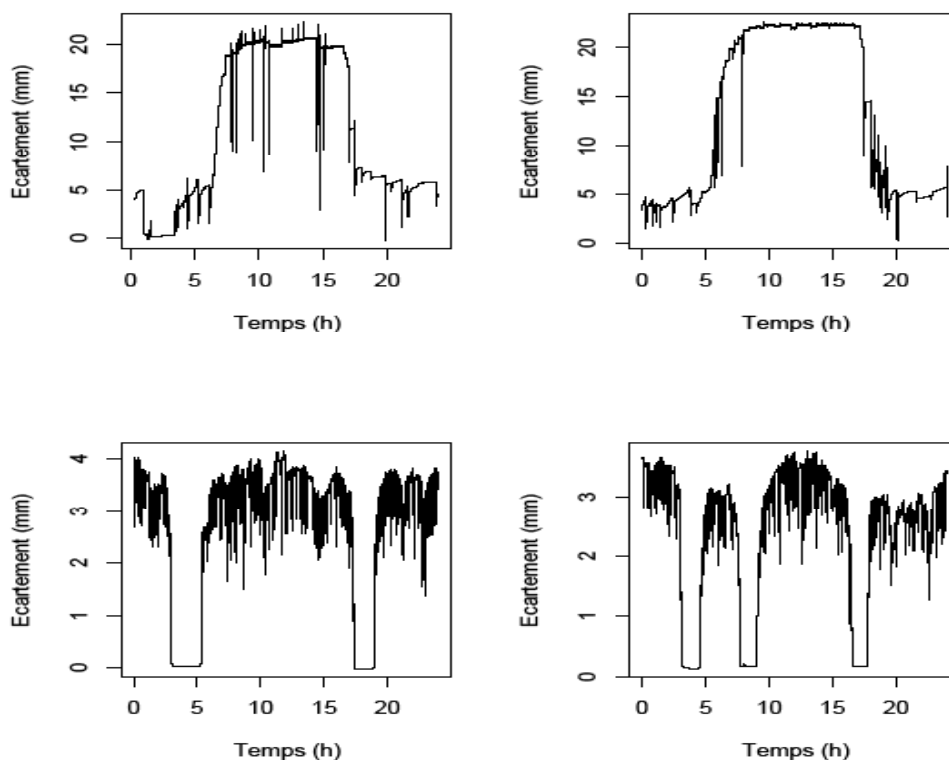


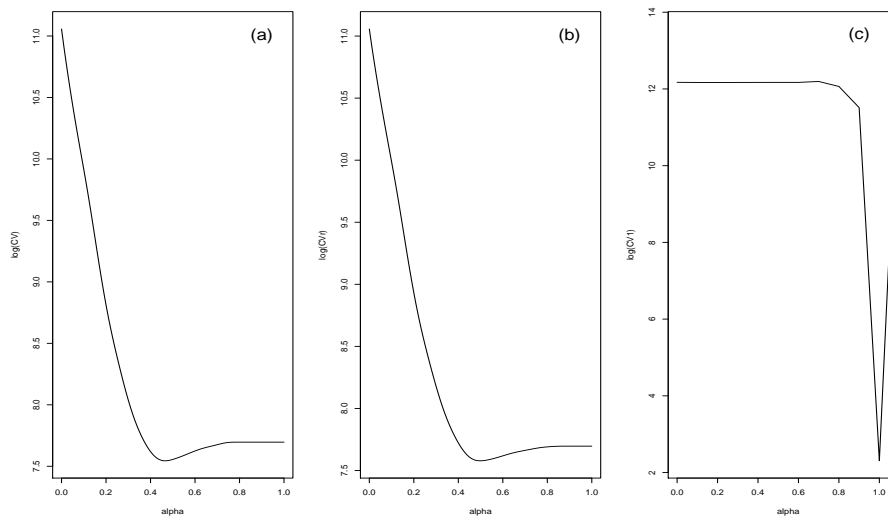
FIG. 6.2: Globalité des enregistrements de bénitiers (deux figures du haut) et d'huîtres (deux figures en bas).

choix de noyaux et de fenêtre proposés plus haut. Les deux courbes des données bruts et des données calculées sont superposées.

Afin de comparer facilement les trois estimateurs, nous présentons la figure FIG. 6.5 qui représente les résidus des estimateurs. Nous voyons clairement que l'estimateur de Révész est nettement meilleur que les deux autres.

Les propriétés asymptotiques des estimateurs sont illustrées numériquement sur un bénitier en se fixant arbitrairement deux temps entre 0 et 24 h. La figure FIG.6.6 illustre les résultats théoriques concernant la normalité asymptotique. En effet, cette figure nous montre un très bon ajustement des estimateurs convenablement normalisés de **NW**, **NWR** et **R** avec une loi normale centrée pour le temps $t_1 = 5$ h (FIG. 6.6 a1-b1-c1) et le temps $t_2 = 23$ h (FIG. 6.6 a2-b2-c2) en considérant $N = 475$ jours.

En comparant nos trois estimateurs au sens du critère de la variance, nous montrons que significativement $\mathbb{V}(\mathbf{R}) < \mathbb{V}(\mathbf{NWR}) < \mathbb{V}(\mathbf{NW})$ ($p < 0.05$). Ainsi, ce résultat nous suggère d'utiliser plutôt un estimateur récursif de type **NWR** ou **R** au sens de la variance asymptotique minimale. Cependant, le temps calcul de l'estimateur **R** est plus important.



(a) : NW, $\alpha_{cv} = 0.47$, (b) : NWR, $\alpha_{cv} = 0.5$, (c) : R, $\alpha_{cv} = 0.99$

FIG. 6.3: Estimation des tailles de fenêtres par Validation Croisée.

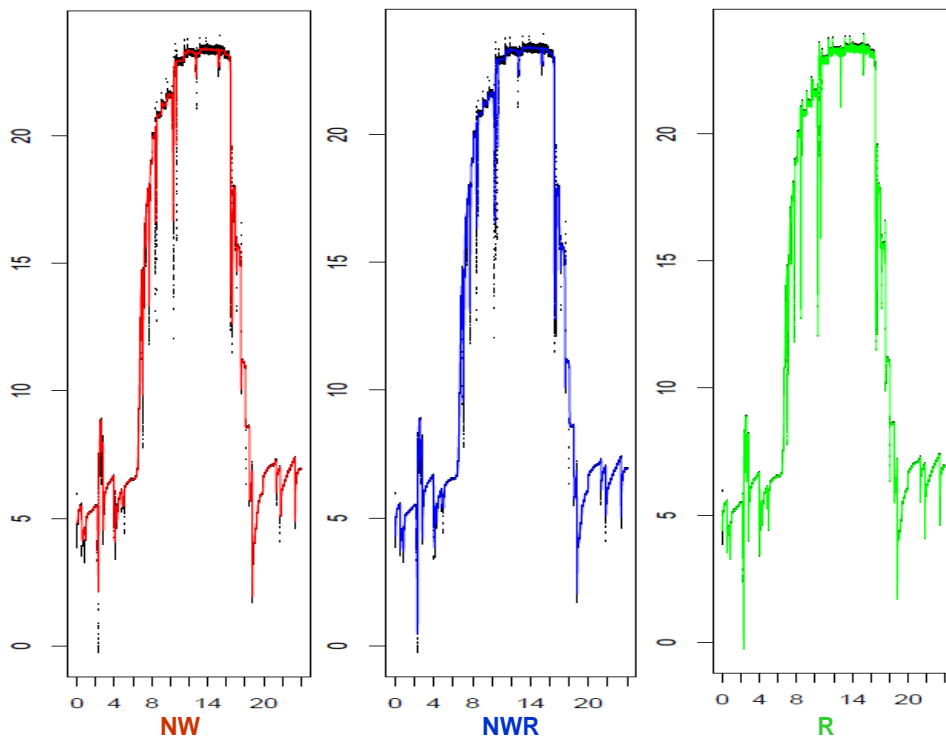


FIG. 6.4: Comportement journalier présenté pour les 3 estimateurs pour le bémittier numéro 7 (récif en Nouvelle Calédonie).

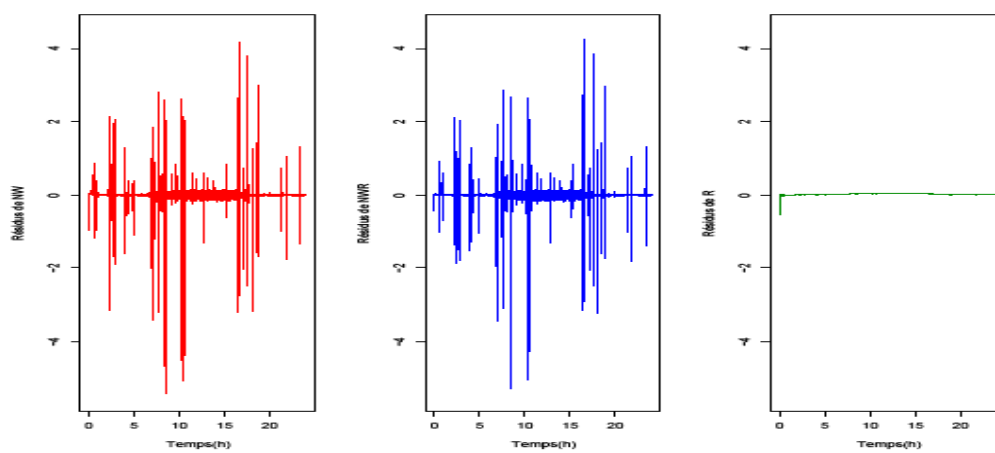


FIG. 6.5: Résidus des estimateurs

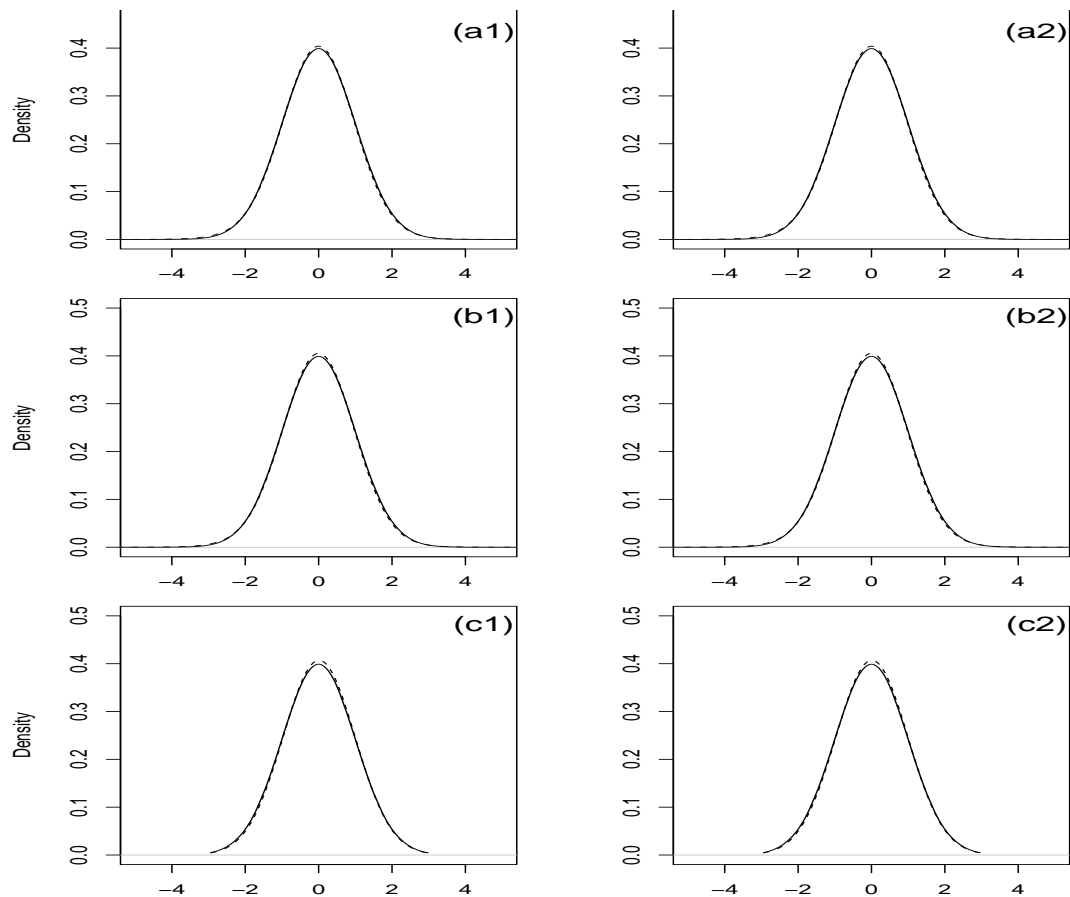


FIG. 6.6: Superposition de la densité de la loi limite des estimateurs de **NW**, **NWR** et **R** (courbes en ligne continue) avec une loi normale centrée (courbes en pointillés) pour le temps $t_1 = 5$ h (Figure 6.6 a1-b1-c1) et le temps $t_2 = 23$ h (Figure 6.6 a2-b2-c2) en considérant $N = 475$ jours.

Bibliographie

- [1] Banon, G. (1976) Sur un estimateur non paramétrique de la densité de probabilité . *Revue de statistique appliquée*, tome 24, **no. 4**, p. 61- 73
- [2] Bercu, B., Chafaï, D. (2007) Modélisation stochastique et simulation. *Dunod*.
- [3] Bosq, D. et Lecoutre, J.P. (1987). Théorie de l'estimation fonctionnelle, *ed. Economica*.
- [4] Carbon, M. et Francq, C. (1995). Estimation non paramétrique de la densité et de la régression - prévision non paramétrique, *La Revue de Modulad, ISSN 1145-895X* , **No.15**, 1-25.
- [5] Choi, E., Hall, P. and Rousson, V. (2000). Data sharpening methods for bias reduction in nonparametric regression. *Ann. Statist.*, **28.5**, 1339-1355.
- [6] Collomb, G. (1977a). Quelques propriétés de la méthode du noyau pour l'estimation non-paramétrique de la régression en un point fixé. *Comptes Rendus Acad. Scu Paris*, **285**, 289-292.
- [7] Collomb, G. (1981). Estimation non-paramétrique de la régression : Revue bibliographique. *International Statistics Review*, **Vol 49, No. 1**, 75-93.
- [8] Deheuvels, P. (1973) Sur l'Estimation séquentielle de la densité. *C. R. de l'Académie des Sciences de Paris, Serie A*, **276** :1119-1121.
- [9] Deheuvels, P. (1974) Sur une famille d'estimateur de la densité d'une variable aléatoire. *C. R. de l'Académie des Sciences de Paris Serie A*, **276** :1013-1015.
- [10] Deheuvels, P. (1977). Estimation non paramétrique de la densité par histogrammes généralisés, *Revue de Statistique Appliquées*, XXV, 5-42.
- [11] Devroye, L., Györfi, L. (1985). Nonparametric Density Estimation : the L1 View, Wiley, New York, Russian translation : Mir, Moscow.
- [12] Duffo, M. (1997) Random Iterative Models. *Collection mathématiques et applications*, Springer, 385 pages.
- [13] Durrieu, G., Nguyen, T.M.N. et Sow, M. (2009). Comparaison d'estimateurs de régression non paramétrique : application en valvométrie. *41èmes Journées de Statistique*, SFdS, HAL- INRIA.
- [14] Eggermont, P.P.B. and LaRiccia, V.N. (2001). *Maximum Penalized Likelihood Estimation*. Springer, New-York.
- [15] Gasser, T. et Müller, H.G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc.*, **B, 47**, 238-252.
- [16] Granovsky, B.L., Müller, H.G. and Pfeifer, C. (1995). Some remark on optimal kernel functions. *Statist. Decisions* , **13**, 101-116.

- [17] Hall, P. (1984). Asymptotic properties of integrated square error and cross validation for kernel estimation of a regression function. *Z. Wahrsch. Verw. Gebiete*, **67**, 175-196.
- [18] Hall, P. and Müller, H.G. (2003). Order-preserving nonparametric regression with applications to conditional distribution and quantile function estimation. *Journal of American Statistical Association*, **98**, 598-608.
- [19] Härdle, W. and Tsybakov, A.B. (1988). Robust nonparametric regression with simultaneous scale curve estimation. *Ann. Statist.* , **16.1**, 120-135.
- [20] Härdle, W. and Marron, J.S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.* , **13.4**, 1465-1481.
- [21] Härdle, W. and Kelly, G. (1987). Nonparametric kernel regression estimation optimal choice of bandwidth. *Statist.* , **18.1**, 21-35.
- [22] Härdle, W., Hall, P. and Marron, J.S. (1988). How far are automatically chosen regression smoothing parameters from their optimum. *J. Amer. Statist. Assoc.*, **83**, 86-95.
- [23] Härdle, W., Hall, P. and Marron, J.S. (1992). Regression smoothing parameters that are not far from their optimum. *J. Amer. Statist. Assoc.*, **87**, 227-233.
- [24] Härdle, W. (1990) Applied nonparametric regression. *Econometric Society Monographs*, 333 pages.
- [25] Härdle, W. and Marron, J.S (1985) Optimal bandwidth selection in nonparametric regression function estimation. *The Annals of Statistics*, **13**, 4, 1465-1481.
- [26] Jones, M.C., Marron, J.S. and Sheather (1996). A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.*, **19**, 401-407.
- [27] Linton, O. and Nielsen, J.P. (1994). A multiplicative bias reduction method for nonparametric regression. *Stat. Probab. Letters*, **19**, 181-187.
- [28] Mammen, E. and Marrin, J.S. (1997). Mass centred kernel smoothers. *Biometrika*, **84**, 765-777.
- [29] Mokkadem, A., Pelletier, M. and Slaoui, Y (2009) Revisiting Révész stochastic approximation method for the estimation of a regression function. *Alea* **6**, 63-114.
- [30] Müller, H.-G. and Song, K.-S. (1993). Identity reproducing multivariate nonparametric regression. *J. of Mult. Analysis*, **46**, 273-253.
- [31] Müller, H.G. (1997). Density adjusted kernel smoothers for random design nonparametric regression. *Stat. Probab. Letters*, **19**, 181-187.
- [32] Nadaraya, E.A. (1964) On estimating regression. *Theory of Probability and its Applications* **10**, 186-190.
- [33] Nadaraya, E. A. (1989). *Nonparametric Estimation of Probability Densities and Regression Curves*. Kluwer, Dordrecht.
- [34] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, **33**, 1065-1076.
- [35] Révész, P. (1977). How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. *Math. Operationsforsch. Statist., Ser. Statistics*, **8**, 119-126.

-
- [36] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, **27**, 832-837.
- [37] Rozenblatt, M. (1971) Curve Estimates. *Ann. Math. Stat.*, **42**, pp. 1815-1842.
- [38] Schuster, E.F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of points. *Annals of Mathematical Statistics*, **43.1**, 84-88.
- [39] Silverman, B.W (1986). Density estimation for statistics and data analysis, Chapman & Hall.
- [40] Stone, C.J. (1977). Consistent nonparametric regression. *Ann. Statist.*, **5.4**, 595-645.
- [41] Tran, D., Ciret, P., Ciutat, A., Durrieu, G. et Massabuau, J.C. (2003). Estimation of potential and limits of bivalve closure response to detect contaminants : application to cadmium. *Environmental Toxicology and Chemistry*, **22(4)**, 914-920.
- [42] Tukey, J.W. (1961). Curves as parameters, and touch estimation. *Proc. 4th Berkeley Symp.*, 681-694.
- [43] Watson, G.S. (1964) Smooth regression analysis. *Sankhya*. **26**, 359-372.
- [44] Wolverton, C. and Wagner, T.J. (1969). Recursive Estimates of Probability Densities. *IEEE Transactions on Systemes Sciences and Cybernetics*. **Vol 5**, 307-308.
- [45] Yamato, H. (1971). Sequential estimation of a continuous propability density function and mode. *Bull. Math. Statist. Jap.*, **Vol. 14**, 1-12.

Troisième partie

Conjonction des deux algorithmes d'estimation

Introduction

Dans cette partie, nous nous intéressons au modèle de régression donné $Y_{n+1} = f(\theta'X_n) + \varepsilon_{n+1}$. Notre objectif est d'étudier le comportement l'estimateur de $f(x'\theta)$. Nous proposons tout d'abord l'estimateur récursif combinant l'estimateur de Nadaraya-Watson récursif de f (qui a été présenté dans la seconde partie) à l'estimateur récursif $\hat{\theta}_n$ de θ (qui a été étudié dans la partie 1). Ensuite, nous étudions quelques propriétés asymptotiques de notre estimateur proposé en donnant deux théorèmes (convergence presque sûre avec la vitesse correspondante et normalité asymptotique). Puis, nous détaillons la démonstration de ces deux théorèmes. Enfin, nous illustrons sur des simulations le bon comportement numérique de l'estimateur récursif proposé. Les différents résultats de simulations sont présentés par le biais de graphiques avec les commentaires associés.

Chapitre 7

Résultats asymptotiques

Nous nous focalisons dans ce chapitre sur les propriétés asymptotiques de notre estimateur $\hat{f}_n(y)$. Dans un premier temps, nous considérons un modèle de régression et proposons un estimateur récursif en combinant l'estimateur à noyau de Nadaraya-Watson récursif à l'estimateur récursif SIR. Ensuite, nous posons les cinq hypothèses nécessaires pour établir la convergence presque sûre ponctuelle avec la vitesse correspondante et la normalité asymptotique de l'estimateur proposé. Puis nous donnons les deux théorèmes, Loi des Grands Nombres et Théorème de la Limite Centrale associés à $\hat{f}_n(y)$. Enfin, nous donnons alors la démonstration du théorème de la Loi des Grands Nombres en précisant sa vitesse de convergence. Finalement, nous achevons ce chapitre par la preuve du Théorème de Limite Centrale.

7.1 Modèle et Estimateurs

Nous disposons d'un échantillon (X_n) constitué de vecteurs aléatoires de \mathbb{R}^p indépendants et de même loi de densité de probabilité f_X . Nous considérons le modèle de régression donné par

$$Y_{n+1} = f(\Phi_n) + \varepsilon_{n+1} \quad (7.1.1)$$

avec

- (i) $\Phi_n = \theta' X_n$ où le paramètre θ est un vecteur inconnu de \mathbb{R}^p dont aucune coordonnée n'est nulle ;
- (ii) La suite (ε_n) est une suite de termes d'erreur aléatoire, indépendante de (X_n) . Nous supposons seulement que $\mathbb{E}[\varepsilon_{n+1} \mid \mathcal{F}_n] = 0$ et $\mathbb{E}[\varepsilon_{n+1}^2 \mid \mathcal{F}_n] = \sigma^2$.
- (iii) La fonction de lien inconnue f est un paramètre fonctionnel à valeurs dans \mathbb{R} .

Dans ce modèle, nous voulons estimer le vecteur θ et le paramètre fonctionnel f . Notons $\hat{\theta}_n$ l'estimateur de θ par la méthode SIR récursif qui a été étudiée dans la première partie. Nous pouvons alors prédire Φ_n par $\hat{\Phi}_n = \hat{\theta}'_{n-1} X_n$. En combinant l'estimateur à noyau de Nadaraya-Watson récursif de f (qui a été présenté dans la seconde partie) à l'estimateur récursif de la direction EDR de θ , nous avons

$$\forall y \in \mathbb{R}, \quad \widehat{f}_n(y) = \frac{1}{\sum_{i=2}^n \frac{1}{h_{i-1}} \mathbf{K} \left(\frac{y - \widehat{\Phi}_{i-1}}{h_{i-1}} \right)} \sum_{i=2}^n \frac{1}{h_{i-1}} \mathbf{K} \left(\frac{y - \widehat{\Phi}_{i-1}}{h_{i-1}} \right) Y_i. \quad (7.1.2)$$

Dans la suite de cette partie, nous allons étudier des propriétés asymptotiques de cet estimateur $\widehat{f}_n(y)$. Nous illustrerons les qualités numériques de cet estimateur par des simulations au chapitre 8.

Notons que nous pouvons aussi écrire $\widehat{f}_n(y)$ sous la forme récursive suivante :

$$\widehat{f}_n(y) = \widehat{f}_{n-1}(y) + \frac{1}{\sum_{i=2}^n \frac{1}{h_{i-1}} \mathbf{K} \left(\frac{y - \widehat{\Phi}_{i-1}}{h_{i-1}} \right)} \frac{1}{h_{n-1}} \mathbf{K} \left(\frac{y - \widehat{\Phi}_{n-1}}{h_{n-1}} \right) (Y_n - \widehat{f}_{n-1}(y)).$$

Afin de simplifier les écritures dans la suite de cette troisième partie, nous écrivons l'estimateur de $f(y)$ donné en (7.1.1) sous la forme

$$\widehat{f}_n(y) = \frac{1}{N_{n-1}(y)} \sum_{i=2}^n \frac{1}{h_{i-1}} \mathbf{K} \left(\frac{y - \widehat{\theta}'_{i-2} X_{i-1}}{h_{i-1}} \right) Y_i, \quad (7.1.3)$$

où $N_{n-1}(y) = \sum_{i=2}^n W_{i-1}(y)$, avec $W_{i-1}(y) = \frac{1}{h_{i-1}} \mathbf{K} \left(\frac{y - \widehat{\theta}'_{i-2} X_{i-1}}{h_{i-1}} \right)$.

Nous avons la décomposition

$$\begin{aligned} \widehat{f}_n(y) - f(y) &= \frac{1}{N_{n-1}(y)} \sum_{i=2}^n W_{i-1}(y) Y_i - f(y) \\ &= \frac{1}{N_{n-1}(y)} \sum_{i=2}^n W_{i-1}(y) \{f(\Phi_{i-1}) + \varepsilon_i - f(y)\} \\ &= \frac{1}{N_{n-1}(y)} \{P_n(y) + Q_{n-1}(y)\}, \end{aligned} \quad (7.1.4)$$

où $P_n(y) = \sum_{i=2}^n W_{i-1}(y) \varepsilon_i$ et $Q_{n-1}(y) = \sum_{i=2}^n W_{i-1}(y) \{f(\Phi_{i-1}) - f(y)\}$.

Les variables aléatoires $P_n(y)$ et $Q_{n-1}(y)$ vont jouer un rôle important dans l'étude des propriétés asymptotiques de notre estimateur.

7.2 Hypothèses

Avant de donner les résultats asymptotiques pour $\hat{f}_n(y)$, nous supposons que les cinq hypothèses suivantes sont satisfaites :

- (H1) La fenêtre (h_n) désigne une suite décroissante de nombres réels strictement positifs vérifiant $h_n \rightarrow 0$ et $n h_n \rightarrow \infty$ lorsque $n \rightarrow \infty$.
 (H2) \mathbf{K} est un noyau à support compact. Le noyau est une fonction mesurable, positive et bornée telle que :

$$\int_{\mathbb{R}} \mathbf{K}(x) dx = 1, \int_{\mathbb{R}} \mathbf{K}^2(x) dx = \nu^2, \int_{\mathbb{R}} \mathbf{K}^4(x) dx = \gamma^4, \int_{\mathbb{R}} x \mathbf{K}(x) dx = 0, \int_{\mathbb{R}} x \mathbf{K}^2(x) dx = 0,$$

$$\int_{\mathbb{R}} x \mathbf{K}^4(x) dx = 0, \int_{\mathbb{R}} x^2 \mathbf{K}(x) dx = \tau^2, \int_{\mathbb{R}} x^2 \mathbf{K}^2(x) dx = \mu^2, \int_{\mathbb{R}} x^2 \mathbf{K}^4(x) dx = \eta^2.$$

- (H3) f est une fonction Lipschitzienne

- (H4) $f_X \in \mathcal{C}^2(\mathbb{R}^p)$ satisfaisant, pour tout $1 \leq i \leq p$, $\sup_{x \in \mathbb{R}^p} \left| \frac{\delta^2}{\delta x_i^2} f_X(x) \right| < +\infty$

- (H5) La suite (X_n) est bornée presque sûrement.

7.3 Théorèmes

En utilisant la Loi des Grands Nombres et le Théorème de la Limite Centrale pour les martingales, nous montrons les deux théorèmes suivants.

Théorème 7.1. Loi des Grands Nombres

Nous supposons que (ε_n) possède un moment conditionnel d'ordre $a > 2$ fini. Alors, sous les cinq hypothèses (H1), ..., (H5), nous avons pour tout $y \in \mathbb{R}$

$$\hat{f}_n(y) \xrightarrow{p.s.} f(y).$$

Plus précisément, si $0 < \alpha < 1/3$,

$$\hat{f}_n(y) - f(y) = \mathcal{O}(n^{-\alpha}) \quad p.s.$$

tandis que, si $1/3 \leq \alpha < 1$

$$\hat{f}_n(y) - f(y) = \mathcal{O}\left(\sqrt{n^{\alpha-1} \log(n)}\right) \quad p.s.$$

Théorème 7.2. Théorème de la Limite Centrale

Nous supposons que (ε_n) possède un moment conditionnel d'ordre $a > 2$ fini. Sous les cinq hypothèses (H1), ..., (H5) et pour tout $\alpha \in]1/3, 1[$, nous avons

$$\sqrt{nh_n} \left(\hat{f}_n(y) - f(y) \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\sigma^2 \nu^2}{(1 + \alpha) h(\theta, y)} \right).$$

7.4 Démonstration des théorèmes

Cette section se divise en deux sous-sections : la démonstration du théorème 7.1 et la démonstration du théorème 7.2. Nous commençons par les études de la vitesse de convergence de $(N_n(y))$, de $(P_n(y))$ et de $(Q_n(y))$ de (7.1.4). Puis, nous donnons la Loi des Grands Nombres avec sa vitesse de convergence. Enfin, le Théorème de la Limite Centrale est donné dans la seconde sous-section.

7.4.1 Démonstration du théorème 7.1 (Loi des Grands Nombres)

7.4.1.1 Etude de la vitesse de convergence de $(N_n(y))$

Nous avons

$$N_n(y) = \sum_{i=1}^n W_i(y) = \sum_{i=1}^n \frac{1}{h_i} \mathbf{K} \left(\frac{y - \hat{\theta}_{i-1} X_i}{h_i} \right). \quad (7.4.1)$$

De plus, si $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n, Y_1, \dots, Y_n)$, nous avons

$$\begin{aligned} \sum_{i=1}^n W_i(y) &= \sum_{i=1}^n \left(W_i(y) - \mathbb{E}[W_i(y) | \mathcal{F}_{i-1}] + \mathbb{E}[W_i(y) | \mathcal{F}_{i-1}] \right) \\ &= M_n^{(N)}(y) + R_n^{(N)}(y), \end{aligned} \quad (7.4.2)$$

où $M_n^{(N)}(y) = \sum_{i=1}^n \left(W_i(y) - \mathbb{E}[W_i(y) | \mathcal{F}_{i-1}] \right)$ et $R_n^{(N)}(y) = \sum_{i=1}^n \mathbb{E}[W_i(y) | \mathcal{F}_{i-1}]$.

Etude de $M_n^{(N)}(y)$

Nous avons $\Delta M_n^{(N)}(y) = M_n^{(N)}(y) - M_{n-1}^{(N)}(y) = W_n(y) - \mathbb{E}[W_n(y) | \mathcal{F}_{n-1}]$. Nous en déduisons que $(M_n^{(N)}(y))$ est une martingale de carré intégrable car le noyau \mathbf{K} est borné. Nous avons donc

$$\begin{aligned} \langle M^{(N)}(y) \rangle_n &= \sum_{i=1}^n \mathbb{E}[(\Delta M_i^{(N)}(y))^2 | \mathcal{F}_{i-1}] \\ &= \sum_{i=1}^n \mathbb{E}[W_i^2(y) | \mathcal{F}_{i-1}] - \sum_{i=1}^n \mathbb{E}^2[W_i(y) | \mathcal{F}_{i-1}]. \end{aligned} \quad (7.4.3)$$

Etude de $\mathbb{E}^2[W_n(y) | \mathcal{F}_{n-1}]$

Nous avons

$$\mathbb{E}[W_n(y) | \mathcal{F}_{n-1}] = \frac{1}{h_n} \int_{\mathbb{R}^p} \mathbf{K} \left(\frac{y - \hat{\theta}_{n-1} x_n}{h_n} \right) f_{X_n}(x_n) dx_n.$$

Nous supposons que $\hat{\theta}_{n-1} \in \mathbb{R}^p$ est différent du vecteur nul. Ceci est bien sûr vrai pour n assez grand car $\hat{\theta}_n$ converge presque sûrement vers θ qui est un vecteur de \mathbb{R}^p dont toutes les coordonnées sont non nulles. Sans perte de généralité, il existe donc au moins une coordonnée du vecteur $\hat{\theta}_{n-1}$ qui est différente de zéro et nous choisissons la première

coordonnée. Nous supposons donc que $\widehat{\theta}_{n-1,1} \neq 0$. Nous effectuons le changement de variable dans \mathbb{R}^p

$$\begin{cases} z = \frac{y - \widehat{\theta}'_{n-1} x_n}{h_n} \\ z_2 = x_{n,2} \\ \cdot \\ \cdot \\ z_p = x_{n,p} \end{cases}$$

Ce système est équivalent à

$$\begin{cases} x_{n,1} = \frac{1}{\widehat{\theta}_{n-1,1}} \left(y - zh_n - \sum_{j=2}^p \widehat{\theta}_{n-1,j} z_j \right) \\ x_{n,2} = z_2 \\ \cdot \\ \cdot \\ x_{n,p} = z_p \end{cases}$$

Le Jacobien de cette transformation vaut

$$J = \frac{-h_n}{\widehat{\theta}_{n-1,1}}.$$

Nous avons donc

$$\begin{aligned} \mathbb{E}[W_n(y) \mid \mathcal{F}_{n-1}] &= \frac{1}{|\widehat{\theta}_{n-1,1}|} \int_{\mathbb{R}^p} \mathbf{K}(z) f_{X_n} \left(\frac{1}{\widehat{\theta}_{n-1,1}} \left(y - zh_n - \sum_{j=2}^p \widehat{\theta}_{n-1,j} z_j \right), z_2, \dots, z_p \right) dz dz_2 \dots dz_p \\ &= \int_{\mathbb{R}} \mathbf{K}(z) h(\widehat{\theta}_{n-1}, y - zh_n) dz \end{aligned}$$

$$\text{où } h(\widehat{\theta}_{n-1}, y) = \frac{1}{|\widehat{\theta}_{n-1,1}|} \int_{\mathbb{R}^{p-1}} f_{X_n} \left(\frac{1}{\widehat{\theta}_{n-1,1}} \left(y - \sum_{j=2}^p \widehat{\theta}_{n-1,j} z_j \right), z_2, \dots, z_p \right) dz_2 \dots dz_p.$$

Il est important de noter que $h(\theta, y)$ correspond exactement à la densité de probabilité de la variable aléatoire $\Phi_n = \theta' X_n$. Nous en déduisons que

$$\begin{aligned} \mathbb{E}[W_n(y) \mid \mathcal{F}_{n-1}] &= \int_{\mathbb{R}} \mathbf{K}(z) \left\{ h(\widehat{\theta}_{n-1}, y) - zh_n h'(\widehat{\theta}_{n-1}, y) + \frac{z^2}{2} h_n^2 h''(\widehat{\theta}_{n-1}, t_{z,n}) \right\} dz \\ &= h(\widehat{\theta}_{n-1}, y) + \frac{h_n^2}{2} \int_{\mathbb{R}} z^2 \mathbf{K}(z) h''(\widehat{\theta}_{n-1}, t_{z,n}) dz \end{aligned}$$

Nous avons supposé que pour tout $1 \leq i \leq p$,

$$\sup_{x \in \mathbb{R}^p} \left| \frac{\partial^2}{\partial x_i^2} f_X(x) \right| < +\infty$$

Ceci entraîne que, pour n assez grand

$$\sup_{y \in \mathbb{R}} \left| h''(\hat{\theta}_{n-1}, y) \right| < +\infty$$

Par suite, il existe $\rho > 0$ telle que

$$\left| \mathbb{E}[W_n(y) \mid \mathcal{F}_{n-1}] - h(\hat{\theta}_{n-1}, y) \right| \leq \frac{\rho \tau^2}{2} h_n^2 \quad \text{p.s.} \quad (7.4.4)$$

Ceci entraîne que

$$\frac{1}{n} \sum_{i=1}^n \left| \mathbb{E}[W_i(y) \mid \mathcal{F}_{i-1}] - h(\hat{\theta}_{i-1}, y) \right| = \mathcal{O}\left(\frac{1}{n} \sum_{i=1}^n h_i^2\right) \quad \text{p.s.}$$

Comme $\sum_{i=1}^n h_i^2 = o(n)$, nous obtenons alors

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[W_i(y) \mid \mathcal{F}_{i-1}] - h(\hat{\theta}_{i-1}, y) \xrightarrow{\text{p.s.}} 0. \quad (7.4.5)$$

Le fait que $\hat{\theta}_n \xrightarrow{\text{p.s.}} \theta$ et la continuité de h entraînent que

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[W_i(y) \mid \mathcal{F}_{i-1}] \xrightarrow{\text{p.s.}} h(\theta, y). \quad (7.4.6)$$

Nous avons également

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}^2[W_i(y) \mid \mathcal{F}_{i-1}] \xrightarrow{\text{p.s.}} h^2(\theta, y).$$

Il est alors immédiat que

$$\sum_{i=1}^n \mathbb{E}^2[W_i(y) \mid \mathcal{F}_{i-1}] = \mathcal{O}(n) \quad \text{p.s.} \quad (7.4.7)$$

Etude de $\mathbb{E}[W_n^2(y) \mid \mathcal{F}_{n-1}]$

Nous avons

$$\mathbb{E}[W_n^2(y) \mid \mathcal{F}_{n-1}] = \frac{1}{h_n^2} \int_{\mathbb{R}^p} \mathbf{K}^2\left(\frac{y - \hat{\theta}'_{n-1} x_n}{h_n}\right) f_{X_n}(x_n) dx_n.$$

De la même manière, nous effectuons le changement de variable dans \mathbb{R}^p . Nous obtenons alors

$$\begin{aligned} \mathbb{E}[W_n^2(y) \mid \mathcal{F}_{n-1}] &= \frac{1}{h_n} \int_{\mathbb{R}} \mathbf{K}^2(z) h(\hat{\theta}_{n-1}, y - zh_n) dz \\ &= \frac{1}{h_n} \int_{\mathbb{R}} \mathbf{K}^2(z) \left\{ h(\hat{\theta}_{n-1}, y) - zh_n h'((\hat{\theta}_{n-1}, y) + \frac{z^2}{2} h_n^2 h''(\hat{\theta}_{n-1}, t_{z,n})) \right\} dz \end{aligned}$$

Nous avons donc

$$\mathbb{E}[W_n^2(y) \mid \mathcal{F}_{n-1}] = \frac{1}{h_n} h(\hat{\theta}_{n-1}, y) \int_{\mathbb{R}} \mathbf{K}^2(z) dz + \frac{h_n}{2} \int_{\mathbb{R}} z^2 \mathbf{K}^2(z) h''(\hat{\theta}_{n-1}, t_{z,n}) dz. \quad (7.4.8)$$

Nous en déduisons qu'il existe $\rho > 0$ telle que :

$$\left| \mathbb{E}[W_n^2(y) \mid \mathcal{F}_{n-1}] - \frac{\nu^2}{h_n} h(\hat{\theta}_{n-1}, y) \right| \leq \frac{\rho \mu^2}{2} h_n \quad \text{p.s.} \quad (7.4.9)$$

Par conséquent, nous avons

$$\frac{1}{n^{\alpha+1}} \sum_{i=1}^n \left| \mathbb{E}[W_i^2(y) \mid \mathcal{F}_{i-1}] - \frac{\nu^2}{h_i} h(\hat{\theta}_{i-1}, y) \right| = \mathcal{O}\left(\frac{1}{n^{\alpha+1}} \sum_{i=1}^n h_i\right) \quad \text{p.s.}$$

Cependant,

$$\frac{1}{n^{\alpha+1}} \sum_{i=1}^n \frac{1}{h_i} \longrightarrow \frac{1}{\alpha+1}.$$

La continuité de h entraîne alors que

$$\frac{1}{n^{\alpha+1}} \sum_{i=1}^n \mathbb{E}[W_i^2(y) \mid \mathcal{F}_{i-1}] \xrightarrow{\text{p.s.}} \frac{\nu^2}{\alpha+1} h(\theta, y). \quad (7.4.10)$$

Obtention de la convergence de $M_n^{(N)}(y)$

Finalement, en utilisant les résultats obtenus en (7.4.7) et (7.4.10), nous déduisons de (7.4.3) que

$$\frac{1}{n^{\alpha+1}} \langle M^{(N)}(y) \rangle_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \frac{\nu^2}{\alpha+1} h(\theta, y). \quad (7.4.11)$$

De plus, la martingale $(M_n^{(N)}(y))$ possède un moment conditionnel d'ordre $a > 2$ fini presque sûrement car le noyau \mathbf{K} est borné. Il découle de la loi des grands nombres pour les martingales (Dacunha-Castelle et Dufflo, 1993) que

$$\frac{M_n^{(N)}(y)}{\langle M^{(N)}(y) \rangle_n} \xrightarrow{\text{p.s.}} 0.$$

Plus précisément,

$$\frac{\left(M_n^{(N)}(y)\right)^2}{\langle M^{(N)}(y) \rangle_n} = \mathcal{O}\left(\log(\langle M^{(N)}(y) \rangle_n)\right) \quad \text{p.s.}$$

De plus, nous déduisons de (7.4.11) que

$$\frac{\log \langle M^{(N)}(y) \rangle_n}{\log(n)} \xrightarrow{\text{p.s.}} (\alpha+1).$$

Par conséquent, nous obtenons

$$\left(M_n^{(N)}(y)\right)^2 = \mathcal{O}\left(n^{\alpha+1} \log(n)\right) \quad \text{p.s.}$$

d'où

$$M_n^{(N)}(y) = \mathcal{O}\left(\sqrt{n^{\alpha+1} \log(n)}\right) \quad \text{p.s.} \quad (7.4.12)$$

Finalement, nous obtenons

$$\frac{1}{n} M_n^{(N)}(y) = \mathcal{O}\left(\sqrt{n^{\alpha-1} \log(n)}\right) \quad \text{p.s.}$$

d'où

$$\frac{M_n^{(N)}(y)}{n} \xrightarrow{\text{p.s.}} 0. \quad (7.4.13)$$

Etude de $R_n^{(N)}(y)$

Nous déduisons de (7.4.6) que

$$\frac{R_n^{(N)}(y)}{n} \xrightarrow{\text{p.s.}} h(\theta, y). \quad (7.4.14)$$

Obtention de la convergence de $\frac{1}{n} \sum_{i=1}^n W_i(y)$

Finalement, en utilisant les résultats obtenus en (7.4.13) et (7.4.14), nous déduisons de (7.4.1) que

$$\frac{1}{n} \sum_{i=1}^n W_i(y) \xrightarrow{\text{p.s.}} h(\theta, y),$$

d'où

$$\widehat{h}_n(\widehat{\theta}_n, y) \xrightarrow{\text{p.s.}} h(\theta, y). \quad (7.4.15)$$

Nous pouvons noter que cette convergence est vraie pour tout $0 < \alpha < 1$.

7.4.1.2 Etude de la vitesse de convergence de $(P_n(y))$

Comme $P_n(y) = \sum_{i=1}^n W_{i-1}(y) \varepsilon_i$, nous obtenons alors $\Delta P_n(y) = W_{n-1}(y) \varepsilon_n$,

$$\mathbb{E}[\Delta P_n(y) \mid \mathcal{F}_{n-1}] = 0,$$

En supposant que $\mathbb{E}[\varepsilon_n^2 \mid \mathcal{F}_{n-1}] = \sigma^2$, nous avons donc

$$\mathbb{E}[\Delta P_n^2(y) \mid \mathcal{F}_{n-1}] = \sigma^2 W_{n-1}^2(y).$$

Nous en déduisons que $(P_n(y))$ est une martingale de carré intégrable car le noyau \mathbf{K} est borné. Nous avons alors

$$\langle P(y) \rangle_n = \sigma^2 \sum_{i=2}^n W_{i-1}^2(y) \quad (7.4.16)$$

De plus, à partir de (7.4.8), nous avons

$$\mathbb{E}[W_n^2(y) \mid \mathcal{F}_{n-1}] = \frac{\nu^2}{h_n} h(\widehat{\theta}_{n-1}, y) + \frac{h_n}{2} \int_{\mathbb{R}} z^2 \mathbf{K}^2(z) h''(\widehat{\theta}_{n-1}, t_{z,n}) dz.$$

Par suite,

$$\langle P(y) \rangle_n = \sigma^2 \sum_{i=2}^n \{W_{i-1}^2(y) - \mathbb{E}[W_{i-1}^2(y) | \mathcal{F}_{i-2}]\} + \sigma^2 \sum_{i=2}^n \mathbb{E}[W_{i-1}^2(y) | \mathcal{F}_{i-2}]$$

Nous en déduisons que

$$\langle P(y) \rangle_n - \sigma^2 \nu^2 \sum_{i=1}^n \frac{1}{h_i} h(\hat{\theta}_{i-1}, y) = \sigma^2 \{M_{n-1}^{(P)}(y) + R_{n-1}^{(P)}(y)\} \quad (7.4.17)$$

où

$$M_n^{(P)}(y) = \sum_{i=1}^n \{W_i^2(y) - \mathbb{E}[W_i^2(y) | \mathcal{F}_{i-1}]\} = \sum_{i=1}^n \{U_i(y) - \mathbb{E}[U_i(y) | \mathcal{F}_{i-1}]\},$$

$$\text{et} \quad R_n^{(P)}(y) = \sum_{i=1}^n \frac{h_i}{2} \int_{\mathbb{R}} z^2 \mathbf{K}^2(z) h''(\hat{\theta}_{i-1}, t_{z,i}) dz,$$

avec $U_n(y) = W_n^2(y)$.

Etude de $M_n^{(P)}(y)$

Il est clair que $(M^{(P)}(y))$ est une martingale de carré intégrable. Nous en déduisons que

$$\langle M^{(P)}(y) \rangle_n = \sum_{i=1}^n \mathbb{E}[U_i^2(y) | \mathcal{F}_{i-1}] - \sum_{i=1}^n \mathbb{E}^2[U_i(y) | \mathcal{F}_{i-1}]. \quad (7.4.18)$$

Etude de $\mathbb{E}^2[U_n(y) | \mathcal{F}_{n-1}]$

À partir de (7.4.10), nous obtenons $\frac{1}{n^{\alpha+1}} \sum_{i=1}^n \mathbb{E}[U_i(y) | \mathcal{F}_{i-1}] \xrightarrow{p.s.} \frac{\nu^2}{\alpha+1} h(\theta, y)$. De plus, il est clair comme précédemment que

$$\frac{1}{n^{2\alpha+1}} \sum_{i=2}^n \mathbb{E}^2[U_i(y) | \mathcal{F}_{i-1}] \xrightarrow{p.s.} \frac{\nu^4}{(2\alpha+1)} h^2(\theta, y),$$

$$\text{car} \quad \frac{1}{n^{2\alpha+1}} \sum_{i=1}^n \frac{1}{h_i^2} \longrightarrow \frac{1}{2\alpha+1}.$$

Il est donc immédiat que

$$\sum_{i=1}^n \mathbb{E}^2[U_i(y) | \mathcal{F}_{i-1}] = \mathcal{O}(n^{1+2\alpha}) \quad \text{p.s.} \quad (7.4.19)$$

Etude de $\mathbb{E}[U_n^2(y) | \mathcal{F}_{n-1}]$

Nous avons

$$\mathbb{E}[U_n^2(y) | \mathcal{F}_{n-1}] = \frac{1}{h_n^4} \int_{\mathbb{R}^p} \mathbf{K}^4\left(\frac{y - \hat{\theta}'_{n-1} x_n}{h_n}\right) f_{X_n}(x_n) dx_n.$$

De la même manière, nous effectuons le changement de variable dans \mathbb{R}^p . Nous obtenons alors

$$\begin{aligned}\mathbb{E}[U_n^2(y) \mid \mathcal{F}_{n-1}] &= \frac{1}{h_n^3} \int_{\mathbb{R}} \mathbf{K}^4(z) h(\widehat{\theta}_{n-1}, y - zh_n) dz \\ &= \frac{1}{h_n^3} \int_{\mathbb{R}} \mathbf{K}^4(z) \left\{ h(\widehat{\theta}_{n-1}, y) - zh_n h'(\widehat{\theta}_{n-1}, y) + \frac{z^2}{2} h_n^2 h''(\widehat{\theta}_{n-1}, t_{z,n}) \right\} dz\end{aligned}$$

Nous avons donc

$$\mathbb{E}[U_n^2(y) \mid \mathcal{F}_{n-1}] = \frac{1}{h_n^3} h(\widehat{\theta}_{n-1}, y) \int_{\mathbb{R}} \mathbf{K}^4(z) dz + \frac{1}{2h_n} \int_{\mathbb{R}} z^2 \mathbf{K}^4(z) h''(\widehat{\theta}_{n-1}, t_{z,n}) dz. \quad (7.4.20)$$

Nous en déduisons qu'il existe $\rho > 0$ telle que :

$$\left| \mathbb{E}[U_n^2(y) \mid \mathcal{F}_{n-1}] - \frac{\gamma^4}{h_n^3} h(\widehat{\theta}_{n-1}, y) \right| \leq \frac{\rho \eta^2}{2} \frac{1}{h_n}$$

Par conséquent, nous avons

$$\frac{1}{n^{3\alpha+1}} \sum_{i=1}^n \left| \mathbb{E}[U_i^2(y) \mid \mathcal{F}_{i-1}] - \frac{\gamma^4}{h_i^3} h(\widehat{\theta}_{i-1}, y) \right| = \mathcal{O}\left(\frac{1}{n^{3\alpha+1}} \sum_{i=1}^n \frac{1}{h_i}\right) \quad \text{p.s.}$$

Cependant,

$$\frac{1}{n^{3\alpha+1}} \sum_{i=1}^n \frac{1}{h_i^3} \longrightarrow \frac{1}{3\alpha+1}.$$

La continuité de h entraîne alors que

$$\frac{1}{n^{3\alpha+1}} \sum_{i=1}^n \mathbb{E}[U_i^2(y) \mid \mathcal{F}_{i-1}] \xrightarrow{\text{p.s.}} \frac{\gamma^4}{3\alpha+1} h(\theta, y). \quad (7.4.21)$$

Il est immédiat que

$$\sum_{i=1}^n \mathbb{E}[U_i^2(y) \mid \mathcal{F}_{i-1}] = \mathcal{O}\left(n^{1+3\alpha}\right) \quad \text{p.s..}$$

Obtention de la convergence de $M_n^{(P)}(y)$

Finalement, en utilisant les résultats obtenus en (7.4.19) et (7.4.21), nous déduisons de (7.4.18) que

$$\frac{1}{n^{3\alpha+1}} \langle M^{(P)}(y) \rangle_n \xrightarrow{\text{p.s.}} \frac{\gamma^4}{3\alpha+1} h(\theta, y). \quad (7.4.22)$$

De plus, la martingale $(M_n^{(P)}(y))$ possède un moment conditionnel d'ordre $a > 2$ fini presque sûre car le noyau \mathbf{K} est borné. Il découle de la loi des grands nombres pour les martingales que

$$\frac{M_n^{(P)}(y)}{\langle M^{(P)}(y) \rangle_n} \xrightarrow{\text{p.s.}} 0.$$

Plus précisément,

$$\frac{\left(M_n^{(P)}(y)\right)^2}{\langle M^{(P)}(y) \rangle_n} = \mathcal{O}\left(\log(\langle M^{(P)}(y) \rangle_n)\right) \quad \text{p.s.}$$

De plus, nous déduisons de (7.4.22) que

$$\frac{\log \langle M^{(P)}(y) \rangle_n}{\log(n)} \xrightarrow{\text{p.s.}} (3\alpha + 1).$$

Par conséquent, nous obtenons

$$\left(M_n^{(P)}(y)\right)^2 = \mathcal{O}\left(n^{3\alpha+1} \log(n)\right) \quad \text{p.s.}$$

d'où

$$M_n^{(P)}(y) = \mathcal{O}\left(\sqrt{n^{3\alpha+1} \log(n)}\right) \quad \text{p.s.} \quad (7.4.23)$$

Etude de $R_n^{(P)}(y)$

Nous avons

$$R_n^{(P)}(y) = \mathcal{O}\left(\sum_{i=1}^n h_i\right) \quad \text{p.s.}$$

Il est donc immédiat que

$$\frac{1}{n} R_n^{(P)}(y) \xrightarrow{\text{p.s.}} 0, \quad (7.4.24)$$

Obtention de la convergence de $\langle P(y) \rangle_n - \sigma^2 \nu^2 \sum_{i=1}^n \frac{1}{h_i} h(\hat{\theta}_{i-1}, y)$

En utilisant les résultats en (7.4.23) et (7.4.24), nous déduisons de (7.4.17) que

$$\langle P(y) \rangle_n - \sigma^2 \nu^2 \sum_{i=1}^n \frac{1}{h_i} h(\hat{\theta}_{i-1}, y) = \mathcal{O}\left(\sqrt{n^{3\alpha+1} \log(n)}\right) + o(n),$$

Ceci entraîne que

$$\frac{\langle P(y) \rangle_n}{n^{\alpha+1}} \xrightarrow{\text{p.s.}} \frac{\sigma^2 \nu^2}{\alpha + 1} h(\theta, y). \quad (7.4.25)$$

De plus, la martingale $(P_n(y))$ possède un moment conditionnel d'ordre $a > 2$ fini presque sûrement car le noyau \mathbf{K} est borné. Il découle de la loi des grands nombres pour les martingales que

$$\frac{P_n(y)}{\langle P(y) \rangle_n} \xrightarrow{\text{p.s.}} 0.$$

Plus précisément,

$$\frac{(P_n(y))^2}{\langle P(y) \rangle_n} = \mathcal{O}\left(\log(\langle P(y) \rangle_n)\right) \quad \text{p.s.}$$

De plus, nous déduisons de (7.4.25) que

$$\frac{\log \langle M^{(P)}(y) \rangle_n}{\log(n)} \xrightarrow{p.s.} (\alpha + 1).$$

Par conséquent, nous obtenons

$$\left(P_n(y) \right)^2 = \mathcal{O}\left(n^{\alpha+1} \log(n) \right) \quad \text{p.s.}$$

d'où

$$P_n(y) = \mathcal{O}\left(\sqrt{n^{\alpha+1} \log(n)} \right) \quad \text{p.s.} \quad (7.4.26)$$

Notons que cette convergence est vraie pour tout $0 < \alpha < 1$.

7.4.1.3 Etude de la vitesse de convergence de $(Q_n(y))$

$$\begin{aligned} Q_n(y) &= \sum_{i=1}^n W_i(y) \left\{ f(\Phi_i) - f(y) \right\} \\ &= \sum_{i=1}^n W_i(y) \left\{ f(\Phi_i) - f(\widehat{\Phi}_i) \right\} + \sum_{i=1}^n W_i(y) \left\{ f(\widehat{\Phi}_i) - f(y) \right\} \\ &= T_n(y) + S_n(y), \end{aligned} \quad (7.4.27)$$

$$\text{où } T_n(y) = \sum_{i=1}^n W_i(y) \left\{ f(\Phi_i) - f(\widehat{\Phi}_i) \right\} \quad \text{et} \quad S_n(y) = \sum_{i=1}^n W_i(y) \left\{ f(\widehat{\Phi}_i) - f(y) \right\}.$$

Etude de $T_n(y)$

Intéressons nous tout d'abord au premier terme de (7.4.27). Comme f est une fonction Lipschitzienne, il existe une constante $c_f > 0$ telle que

$$\begin{aligned} \left| \sum_{i=1}^n W_i(y) \left\{ f(\Phi_i) - f(\widehat{\Phi}_i) \right\} \right| &\leq \sum_{i=1}^n W_i(y) \left| f(\Phi_i) - f(\widehat{\Phi}_i) \right| \\ &\leq c_f \sum_{i=1}^n W_i(y) \left| \Phi_i - \widehat{\Phi}_i \right| \\ &\leq c_f \sum_{i=1}^n W_i(y) \| X_i \| \|\widehat{\theta}_{i-1} - \theta\|. \end{aligned}$$

Nous avons supposé que la suite (X_n) est bornée presque sûrement. Par conséquent, il

existe une constante $c > 0$ telle que

$$\begin{aligned}
 |T_n(y)| &\leq c \sum_{i=1}^n W_i(y) \|\widehat{\theta}_{i-1} - \theta\| \\
 &\leq c \sum_{i=1}^n \left\{ W_i(y) - \mathbb{E}[W_i(y) | \mathcal{F}_{i-1}] + \mathbb{E}[W_i(y) | \mathcal{F}_{i-1}] \right\} \|\widehat{\theta}_{i-1} - \theta\| \\
 &= c \left\{ M_n^T(y) + R_n^T(y) \right\}
 \end{aligned} \tag{7.4.28}$$

où $M_n^T(y) = \sum_{i=1}^n \left\{ W_i(y) - \mathbb{E}[W_i(y) | \mathcal{F}_{i-1}] \right\} \|\widehat{\theta}_{i-1} - \theta\|$

et $R_n^T(y) = \sum_{i=1}^n \mathbb{E}[W_i(y) | \mathcal{F}_{i-1}] \|\widehat{\theta}_{i-1} - \theta\|$.

Etude de $R_n^{(T)}(y)$

Nous avons

$$\begin{aligned}
 R_n^{(T)}(y) &= \sum_{i=1}^n \mathbb{E}[W_i(y) | \mathcal{F}_{i-1}] \|\widehat{\theta}_{i-1} - \theta\| \\
 &= \sum_{i=1}^n \left\{ \mathbb{E}[W_i(y) | \mathcal{F}_{i-1}] - h(\widehat{\theta}_{i-1}, y) + h(\widehat{\theta}_{i-1}, y) \right\} \|\widehat{\theta}_{i-1} - \theta\|
 \end{aligned}$$

En utilisant le résultat obtenu en (7.4.4), nous déduisons que

$$R_n^{(T)}(y) \leq \frac{\rho\tau^2}{2} \sum_{i=1}^n h_i^2 \|\widehat{\theta}_{i-1} - \theta\| + \sum_{i=1}^n h(\widehat{\theta}_{i-1}, y) \|\widehat{\theta}_{i-1} - \theta\| \tag{7.4.29}$$

Nous avons déjà vu dans la première partie de la thèse que

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

Par conséquent, nous avons

$$\begin{aligned}
 \sum_{i=1}^n h_i^2 \|\widehat{\theta}_{i-1} - \theta\| &= \mathcal{O}\left(\sum_{i=3}^n h_i^2 \sqrt{\frac{\log(\log(i))}{i}}\right) \quad \text{p.s.} \\
 &= \mathcal{O}\left(\sqrt{\log(\log(n))} \sum_{i=3}^n \frac{1}{i^{2\alpha+1/2}}\right) \quad \text{p.s.} \\
 &= \mathcal{O}\left(\sqrt{n^{1-4\alpha} \log(\log(n))}\right) \quad \text{p.s.}
 \end{aligned}$$

De plus, nous avons

$$\begin{aligned}
 \sum_{i=1}^n h(\hat{\theta}_{i-1}, y) \|\hat{\theta}_{i-1} - \theta\| &= \mathcal{O}\left(\sum_{i=1}^n \|\hat{\theta}_{i-1} - \theta\|\right) \quad \text{p.s.} \\
 &= \mathcal{O}\left(\sum_{i=3}^n \sqrt{\frac{\log(\log(i))}{i}}\right) \quad \text{p.s.} \\
 &= \mathcal{O}\left(\sqrt{n \log(\log(n))}\right) \quad \text{p.s.}
 \end{aligned}$$

Nous déduisons alors de (7.4.29) que

$$R_n^{(T)}(y) = \mathcal{O}\left(\sqrt{n \log(\log(n))}\right) \quad \text{p.s.} \quad (7.4.30)$$

Etude de $M_n^{(T)}(y)$

Nous avons

$$M_n^{(T)}(y) = \sum_{i=1}^n \left\{ W_i(y) - \mathbb{E}[W_i(y) \mid \mathcal{F}_{i-1}] \right\} \|\hat{\theta}_{i-1} - \theta\|$$

Il est clair que $(M_n^{(T)}(y))$ est une martingale de carré intégrable. Nous en déduisons que

$$\langle M^{(T)}(y) \rangle_n = \sum_{i=1}^n \|\hat{\theta}_{i-1} - \theta\|^2 \left\{ \mathbb{E}[W_i^2(y) \mid \mathcal{F}_{i-1}] - \sum_{i=1}^n \mathbb{E}^2[W_i(y) \mid \mathcal{F}_{i-1}] \right\}. \quad (7.4.31)$$

$$\text{Etude de } \sum_{i=1}^n \|\hat{\theta}_{i-1} - \theta\|^2 \mathbb{E}[W_i^2(y) \mid \mathcal{F}_{i-1}]$$

À partir de (7.4.9), nous en déduisons que

$$\|\hat{\theta}_{n-1} - \theta\|^2 \left| \mathbb{E}[W_n^2(y) \mid \mathcal{F}_{n-1}] - \frac{\nu^2}{h_n} h(\hat{\theta}_{n-1}, y) \right| \leq \frac{\rho\mu^2}{2} \|\hat{\theta}_{n-1} - \theta\|^2 h_n \quad \text{p.s.}$$

Nous avons donc

$$\sum_{i=1}^n \|\hat{\theta}_{i-1} - \theta\|^2 \left| \mathbb{E}[W_i^2(y) \mid \mathcal{F}_{i-1}] - \frac{\nu^2}{h_i} h(\hat{\theta}_{i-1}, y) \right| \leq \frac{\rho\mu^2}{2} \sum_{i=1}^n \|\hat{\theta}_{i-1} - \theta\|^2 h_i \quad \text{p.s.}$$

Cependant,

$$\sum_{i=1}^n \|\hat{\theta}_{i-1} - \theta\|^2 h_i = \mathcal{O}\left(n^{-\alpha} \log(\log(n))\right) \quad \text{p.s.}$$

et

$$\sum_{i=1}^n \|\hat{\theta}_{i-1} - \theta\|^2 \frac{1}{h_i} = \mathcal{O}\left(n^\alpha \log(\log(n))\right) \quad \text{p.s.}$$

Nous obtenons alors que

$$\sum_{i=1}^n \|\widehat{\theta}_{i-1} - \theta\|^2 \mathbb{E}[W_i^2(y) \mid \mathcal{F}_{i-1}] = \mathcal{O}\left(n^\alpha \log(\log(n))\right) \quad \text{p.s.} \quad (7.4.32)$$

$$\text{Etude de } \sum_{i=1}^n \|\widehat{\theta}_{i-1} - \theta\|^2 \mathbb{E}^2[W_i(y) \mid \mathcal{F}_{i-1}]$$

Nous avons

$$\sum_{i=1}^n \|\widehat{\theta}_{i-1} - \theta\|^2 \mathbb{E}^2[W_i(y) \mid \mathcal{F}_{i-1}] = \sum_{i=1}^n \|\widehat{\theta}_{i-1} - \theta\|^2 \left\{ \mathbb{E}^2[W_i(y) \mid \mathcal{F}_{i-1}] - h^2(\widehat{\theta}_{i-1}, y) + h^2(\widehat{\theta}_{i-1}, y) \right\}$$

De plus, à partir de (7.4.4), nous avons

$$\mathbb{E}^2[W_n(y) \mid \mathcal{F}_{n-1}] - h^2(\widehat{\theta}_{n-1}, y) \leq \rho\tau^2 h(\widehat{\theta}_{n-1}, y) h_n^2 + \frac{\rho^2\tau^4}{4} h_n^4$$

Ceci entraîne que

$$\begin{aligned} \sum_{i=1}^n \|\widehat{\theta}_{i-1} - \theta\|^2 \mathbb{E}^2[W_i(y) \mid \mathcal{F}_{i-1}] &\leq \rho\tau^2 \sum_{i=1}^n h(\widehat{\theta}_{i-1}, y) h_i^2 \|\widehat{\theta}_{i-1} - \theta\|^2 \\ &\quad + \frac{\rho^2\tau^4}{4} \sum_{i=1}^n h_i^4 \|\widehat{\theta}_{i-1} - \theta\|^2 + \sum_{i=1}^n h^2(\widehat{\theta}_{i-1}, y) \|\widehat{\theta}_{i-1} - \theta\|^2 \end{aligned}$$

Cependant, nous avons

$$\begin{aligned} \sum_{i=1}^n h(\widehat{\theta}_{i-1}, y) h_i^2 \|\widehat{\theta}_{i-1} - \theta\|^2 &= \mathcal{O}\left(\sum_{i=1}^n h_i^2 \|\widehat{\theta}_{i-1} - \theta\|^2\right) \quad \text{p.s.} \\ &= \mathcal{O}\left(n^{-2\alpha} \log(\log(n))\right) \quad \text{p.s.} \end{aligned}$$

Nous avons également

$$\sum_{i=1}^n h_i^4 \|\widehat{\theta}_{i-1} - \theta\|^2 = \mathcal{O}\left(\sum_{i=1}^n h_i^4 \|\widehat{\theta}_{i-1} - \theta\|^2\right) = \mathcal{O}\left(n^{-4\alpha} \log(\log(n))\right) \quad \text{p.s.}$$

De plus, nous avons

$$\sum_{i=1}^n h^2(\widehat{\theta}_{i-1}, y) \|\widehat{\theta}_{i-1} - \theta\|^2 = \mathcal{O}\left(\sum_{i=1}^n \|\widehat{\theta}_{i-1} - \theta\|^2\right) = \mathcal{O}\left(\log(n) \log(\log(n))\right) \quad \text{p.s.}$$

Par conséquent, nous obtenons

$$\sum_{i=1}^n \|\widehat{\theta}_{i-1} - \theta\|^2 \mathbb{E}^2[W_i(y) \mid \mathcal{F}_{i-1}] = \mathcal{O}\left(\log(n) \log(\log(n))\right) \quad \text{p.s.} \quad (7.4.33)$$

Obtention de la convergence de $(M_n^{(T)}(y))$

En utilisant les résultats obtenus en (7.4.32) et (7.4.33), nous déduisons de (7.4.31) que

$$\langle M^{(T)}(y) \rangle_n = \mathcal{O}\left(n^\alpha \log(\log(n))\right) \text{ p.s.}$$

De plus, la martingale $(M_n^{(T)}(y))$ possède un moment conditionnel d'ordre $a > 2$ fini presque sûrement. Il découle de la loi grands nombres pour les martingales que

$$\frac{M_n^{(T)}(y)}{\langle M^{(T)}(y) \rangle_n} \xrightarrow{\text{p.s.}} 0$$

Plus précisément,

$$(M_n^{(T)}(y))^2 = \mathcal{O}\left(n^\alpha \log(n) \log(\log(n))\right) \text{ p.s.}$$

d'où

$$M_n^{(T)}(y) = \mathcal{O}\left(\sqrt{n^\alpha \log(n) \log(\log(n))}\right) \text{ p.s.} \quad (7.4.34)$$

Obtention de la convergence de $T_n(y)$

Finalement, à partir des résultats donnés en (7.4.30) et (7.4.34), nous déduisons de (7.4.28)

$$|T_n(y)| = \mathcal{O}\left(\sqrt{n \log(\log(n))}\right) \text{ p.s.} \quad (7.4.35)$$

Etude de $S_n(y)$

Étudions maintenant le second terme de (7.4.27). Comme f est une fonction Lipschitzienne et \mathbf{K} est un noyau à support compact, il existe deux constantes $A > 0$ et $c > 0$ telle que

$$\begin{aligned} |S_n(y)| &\leq c \sum_{i=1}^n W_i(y) \left| \hat{\theta}'_{i-1} X_i - y \right| \\ &\leq c \sum_{i=1}^n \frac{1}{h_i} \mathbf{K} \left(\frac{y - \hat{\theta}'_{i-1} X_i}{h_i} \right) \mathbf{1}_{\left[\left| \hat{\theta}'_{i-1} X_i - y \right| \leq Ah_i \right]} \left| \hat{\theta}'_{i-1} X_i - y \right| \\ &\leq cA \sum_{i=1}^n h_i W_i(y) \mathbf{1}_{\left[\left| \hat{\theta}'_{i-1} X_i - y \right| \leq Ah_i \right]} \\ &\leq cA \sum_{i=1}^n h_i W_i(y). \end{aligned}$$

Nous avons donc

$$\left| \sum_{i=1}^n W_i(y) \left(f(\hat{\Phi}_i) - f(y) \right) \right| \leq cA \sum_{i=1}^n h_i \left(W_i(y) - \mathbb{E}[W_i(y) \mid \mathcal{F}_{i-1}] \right) + cA \sum_{i=1}^n h_i \mathbb{E}[W_i(y) \mid \mathcal{F}_{i-1}] \quad (7.4.36)$$

Nous avons vu en (7.4.4) que

$$\left| \mathbb{E}[W_n(y) \mid \mathcal{F}_{n-1}] - h(\widehat{\theta}_{n-1}, y) \right| \leq \frac{\rho\tau^2}{2} h_n^2 \quad \text{p.s.}$$

Ceci entraîne que

$$\left| h_n \mathbb{E}[W_n(y) \mid \mathcal{F}_{n-1}] - h_n h(\widehat{\theta}_{n-1}, y) \right| \leq \frac{\rho\tau^2}{2} h_n^3 \quad \text{p.s.}$$

Par conséquent, nous avons

$$\frac{1}{n^{1-\alpha}} \sum_{i=1}^n \left| h_i \mathbb{E}[W_i(y) \mid \mathcal{F}_{i-1}] - h_i h(\widehat{\theta}_{i-1}, y) \right| = \mathcal{O}\left(\frac{1}{n^{1-\alpha}} \sum_{i=1}^n h_i^3\right) \quad \text{p.s.}$$

Cependant,

$$\frac{1}{n^{1-\alpha}} \sum_{i=1}^n h_i \longrightarrow \frac{1}{1-\alpha} \quad \text{et} \quad \sum_{i=1}^n h_i^3 = o(n^{1-\alpha})$$

La continuité de h entraîne alors que

$$\frac{1}{n^{1-\alpha}} \sum_{i=1}^n h_i \mathbb{E}[W_i(y) \mid \mathcal{F}_{i-1}] \xrightarrow{\text{p.s.}} \frac{h(\theta, y)}{1-\alpha}$$

d'où

$$\sum_{i=1}^n h_i \mathbb{E}[W_i(y) \mid \mathcal{F}_{i-1}] = \mathcal{O}\left(n^{1-\alpha}\right) \quad \text{p.s.} \quad (7.4.37)$$

D'autre part, nous avons également vu en (7.4.2) et (7.4.13) que

$$M_n^{(N)}(y) = \sum_{i=1}^n \left(W_i(y) - \mathbb{E}[W_i(y) \mid \mathcal{F}_{i-1}] \right) = \mathcal{O}\left(\sqrt{n^{\alpha+1} \log(n)}\right) \quad \text{p.s.}$$

Nous déduisons d'un raisonnement analogue à (7.4.13) que

$$\sum_{i=1}^n h_i \left(W_i(y) - \mathbb{E}[W_i(y) \mid \mathcal{F}_{i-1}] \right) = \mathcal{O}\left(\sqrt{n^{1-\alpha} \log(n)}\right) \quad \text{p.s.} \quad (7.4.38)$$

À partir de (7.4.37) et (7.4.38), nous déduisons de (7.4.36) que

$$|S_n(y)| = \mathcal{O}\left(n^{1-\alpha}\right) \quad \text{p.s.} \quad (7.4.39)$$

Obtention de la convergence de $Q_n(y)$

En utilisant les résultats donnés par (7.4.35) et (7.4.39), nous déduisons de (7.4.27) que

$$Q_n(y) = \mathcal{O}\left(\sqrt{n \log(\log(n))}\right) + \mathcal{O}\left(n^{1-\alpha}\right) \quad \text{p.s.} \quad (7.4.40)$$

En particulier, si $0 < \alpha < 1/2$,

$$Q_n(y) = \mathcal{O}\left(n^{1-\alpha}\right) \quad \text{p.s.}$$

tandis que, si $1/2 \leq \alpha < 1$,

$$Q_n(y) = \mathcal{O}\left(\sqrt{n \log(\log(n))}\right) \quad \text{p.s.}$$

7.4.1.4 Obtention de la convergence de $\hat{f}_n(y)$

Finalement, en utilisant les vitesses de convergence obtenues en (7.4.15), (7.4.26) et (7.4.40), nous déduisons de (7.1.4) que

$$\hat{f}_n(y) \xrightarrow{p.s.} f(y).$$

Plus précisément, nous avons la vitesse de convergence

$$\hat{f}_n(y) - f(y) = \mathcal{O}\left(\sqrt{n^{\alpha-1} \log(n)}\right) + \mathcal{O}\left(n^{-\alpha}\right) + \mathcal{O}\left(\sqrt{n^{-1} \log(\log(n))}\right) \quad \text{p.s.}$$

En particulier, si $0 < \alpha < 1/3$,

$$\hat{f}_n(y) - f(y) = \mathcal{O}\left(n^{-\alpha}\right) \quad \text{p.s.}$$

tandis que, si $1/3 \leq \alpha < 1$,

$$\hat{f}_n(y) - f(y) = \mathcal{O}\left(\sqrt{n^{\alpha-1} \log(n)}\right) \quad \text{p.s.}$$

7.4.2 Démonstration du théorème 7.2 (Théorème de Limite Centrale)

Nous savons que $(P_n(y))$ est une martingale de carré intégrable. De plus, nous avons déjà vu en (7.4.25) que

$$\frac{\langle P(y) \rangle_n}{n^{\alpha+1}} \xrightarrow{p.s.} \frac{\sigma^2 \nu^2 h(\theta, y)}{\alpha + 1}.$$

Il est nécessaire vérifier la condition de Lindeberg afin d'utiliser le théorème central limite pour les martingale (Dacunha-Castelle et Duflo, 1993). Plus précisément, nous allons montrer que, pour tout $\varepsilon > 0$,

$$\frac{1}{n^{\alpha+1}} \sum_{i=1}^n \mathbb{E} \left[|\Delta P_i(y)|^2 \mathbf{1}_{|\Delta P_i(y)| \geq \varepsilon \sqrt{n^{\alpha+1}}} \mid \mathcal{F}_{i-1} \right] \xrightarrow{p.s.} 0$$

Nous avons

$$\Delta P_i(y) = W_{i-1}(y) \varepsilon_i = \frac{1}{h_{i-1}} \mathbf{K} \left(\frac{y - \hat{\theta}'_{i-2} X_{i-1}}{h_{i-1}} \right) \varepsilon_i$$

Puisque \mathbf{K} est borné,

$$\begin{aligned} \mathbf{1} \left\{ |\Delta P_i(y)| \geq \varepsilon \sqrt{n^{\alpha+1}} \right\} &\leq \mathbf{1} \left\{ n^\alpha \|\mathbf{K}\|_\infty |\varepsilon_i| \geq \varepsilon \sqrt{n^{\alpha+1}} \right\} \\ &\leq \mathbf{1} \left\{ |\varepsilon_i| \geq \frac{\varepsilon}{\|\mathbf{K}\|_\infty} \sqrt{n^{1-\alpha}} \right\} \end{aligned}$$

Par conséquent, si $\xi = \frac{\varepsilon}{\|\mathbf{K}\|_\infty}$ et si $\sup_n \mathbb{E} \left[|\varepsilon_n|^{2+\gamma} \mid \mathcal{F}_{n-1} \right] < \infty$ avec $\gamma > 0$, on a

$$\begin{aligned} \mathbb{E} \left[|\Delta P_i(y)|^2 \mathbf{1}_{\{|\Delta P_i(y)| \geq \varepsilon \sqrt{n^{\alpha+1}}\}} \mid \mathcal{F}_{i-1} \right] &\leq W_{i-1}^2(y) \mathbb{E} \left[\varepsilon_i^2 \mathbf{1}_{\{|\varepsilon_i| \geq \xi \sqrt{n^{1-\alpha}}\}} \mid \mathcal{F}_{i-1} \right], \\ &\leq \frac{W_{i-1}^2(y)}{\left(\xi \sqrt{n^{1-\alpha}}\right)^\gamma} \mathbb{E} \left[|\varepsilon_i|^{2+\gamma} \mid \mathcal{F}_{i-1} \right], \\ &\leq \sup_n \mathbb{E} \left[|\varepsilon_n|^{2+\gamma} \mid \mathcal{F}_{n-1} \right] \frac{W_{i-1}^2(y)}{\left(\xi \sqrt{n^{1-\alpha}}\right)^\gamma}. \end{aligned}$$

De plus, à partir de (7.4.25), nous déduisons de (7.4.16) que

$$\sum_{i=1}^n W_{i-1}^2(y) = \mathcal{O}\left(n^{1+\alpha}\right) \text{ p.s.}$$

Par conséquent, nous obtenons

$$\frac{1}{n^{\alpha+1}} \sum_{i=1}^n \mathbb{E} \left[|\Delta P_i(y)|^2 \mathbf{1}_{|\Delta P_i(y)| \geq \varepsilon \sqrt{n^{\alpha+1}}} \mid \mathcal{F}_{i-1} \right] = \mathcal{O}\left(n^{(\alpha-1)\gamma/2}\right) \text{ p.s.}$$

Il est alors immédiat que, pour tout $0 < \alpha < 1$,

$$\frac{1}{n^{\alpha+1}} \sum_{i=1}^n \mathbb{E} \left[|\Delta P_i(y)|^2 \mathbf{1}_{|\Delta P_i(y)| \geq \varepsilon \sqrt{n^{\alpha+1}}} \mid \mathcal{F}_{i-1} \right] \xrightarrow{\text{p.s.}} 0$$

La condition de Lindeberg est ainsi vérifiée. Par conséquent, d'après le théorème central limite pour les martingales, nous avons

$$\frac{P_n(y)}{\sqrt{n^{\alpha+1}}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2 \nu^2 h(\theta, y)}{(1+\alpha)}\right) \quad (7.4.41)$$

En utilisant le résultat donné en (7.4.15), nous avons

$$\frac{N_{n-1}(y)}{n} \xrightarrow{\text{p.s.}} h(\theta, y) \quad (7.4.42)$$

A partir des résultats donnés en (7.4.41) et (7.4.42), nous obtenons que

$$\sqrt{nh_n} \frac{P_n(y)}{N_{n-1}(y)} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2 \nu^2}{(1+\alpha)h(\theta, y)}\right) \quad (7.4.43)$$

$$\text{car } \sqrt{nh_n} = \frac{n}{\sqrt{n^{\alpha+1}}}.$$

En utilisant les vitesses de convergence données en (7.4.42) et (7.4.40), nous obtenons donc

$$\sqrt{nh_n} \frac{Q_{n-1}(y)}{N_{n-1}(y)} = \mathcal{O}\left(\sqrt{n^{-\alpha} \log(\log(n))}\right) + \mathcal{O}\left(\sqrt{n^{1-3\alpha}}\right) \text{ p.s.}$$

En particulier, si $1/3 < \alpha < 1/2$,

$$\sqrt{nh_n} \frac{Q_{n-1}(y)}{N_{n-1}(y)} = \mathcal{O}\left(\sqrt{n^{1-3\alpha}}\right) \quad \text{p.s.}$$

tandis que, si $1/2 \leq \alpha < 1$,

$$\sqrt{nh_n} \frac{Q_{n-1}(y)}{N_{n-1}(y)} = \mathcal{O}\left(\sqrt{n^{-\alpha} \log(\log(n))}\right) \quad \text{p.s.}$$

d'où pour tout $1/3 < \alpha < 1$

$$\sqrt{nh_n} \frac{Q_{n-1}(y)}{N_{n-1}(y)} \xrightarrow{\text{p.s.}} 0. \quad (7.4.44)$$

A partir des résultats donnés en (7.4.43) et (7.4.44), nous trouvons finalement par (7.1.4) que, pour tout $1/3 < \alpha < 1$,

$$\sqrt{nh_n} \left(\widehat{f}_n(y) - f(y) \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\sigma^2 \nu^2}{(1 + \alpha)h(\theta, y)} \right).$$

Chapitre 8

Quelques résultats de simulation

Dans ce chapitre, nous allons étudier, sur des simulations, le comportement numérique de l'estimateur récursif combinant l'estimateur de Nadaraya-Watson récursif de f à l'estimateur SIR récursif $\hat{\theta}_n$ que nous avons proposé lorsque l'on prend $H = 2$ tranches. Nous parlerons tout d'abord du modèle simulé et nous préciserons les procédures d'estimation mises en œuvre. Nous donnons dans un second temps les résultats de simulations montrant la convergence des estimateurs (l'estimateur récursif $\hat{\theta}_n$ converge presque sûrement vers la vraie direction θ , puis l'estimateur récursif $\hat{f}_n(x'_0\hat{\theta}_n)$ converge presque sûrement vers $f(x'_0\theta)$) pour différents jeux de paramètres (modèle, dimension p de la covariable de x , valeur σ^2 de la variance du bruit ε , valeur x_0 de la covariable pour laquelle on a fait l'estimation et puissance α de la fenêtre $h_n = n^{-\alpha}$). Nous terminons en illustrant la normalité asymptotique de l'estimateur $\hat{f}_n(x'_0\hat{\theta}_n)$.

8.1 Modèles simulés et procédure d'estimation

Dans ces simulations, nous avons considéré les deux modèles de régression suivants :

$$\begin{aligned}(M1) : & y = (\theta'x) \exp(-\theta'x) + \varepsilon, \\(M2) : & y = (\theta'x)^3 + \varepsilon,\end{aligned}$$

où x suit la loi multinormale $\mathcal{N}_p(0, 1)$, $\theta = (1, -1, 0 \dots, 0)' / \sqrt{2} \in \mathbb{R}^p$ et le terme d'erreur ε , indépendant de x , suit la loi normale $\mathcal{N}(0, \sigma^2)$. Nous avons fait varier la dimension p de x : $p = 2, 5, 10, 20$ ou 40 ainsi que la valeur σ^2 de la variance du bruit ε .

Nous avons utilisé le noyau Gaussien ($\mathbf{K}_{Gauss}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$) et la fenêtre $h_n = n^{-\alpha}$ avec $0 < \alpha < 1$.

Nous présentons à la figure FIG. 8.1 deux nuages de $n = 1000$ points $\{(x'_i\theta, y_i), i = 1, \dots, n\}$ simulés à partir des modèles (M1) et (M2) pour $p = 10$ avec $\sigma = 1$. Pour l'ensemble des simulations, nous avons choisi un ratio bruit sur signal (défini comme le quotient entre la variance des ε_i et la variance des y_i) d'environ 3% (resp. 6%) pour (M1) (resp. (M2)), quelle que soit la dimension p .

Les simulations ont été faites à l'aide du logiciel MATLAB. Nous avons simulé, à partir des modèles (M1) et (M2), $\mathcal{B} = 100$ échantillons de taille 1000 avec successivement

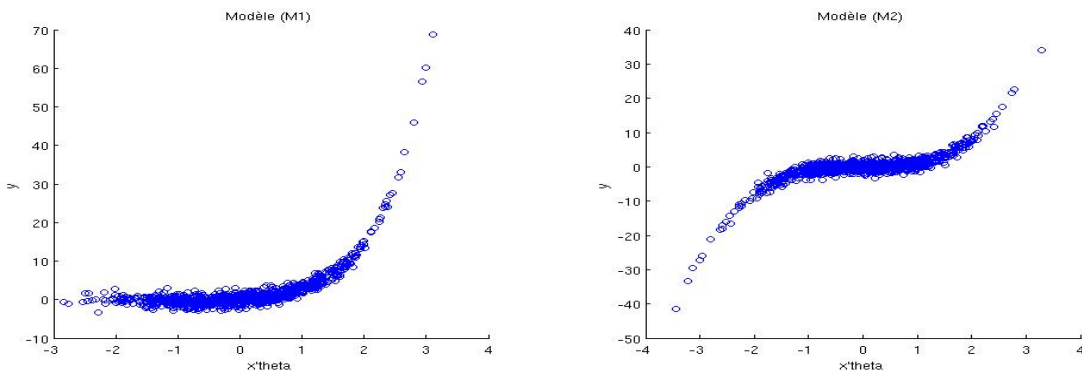


FIG. 8.1: Nuages des points $\{(x'_i\theta, y_i), i = 1, \dots, 1000\}$ simulés à partir des modèles (M1) et (M2) pour $p = 10$ avec $\sigma = 1$

$p = 2, 5, 10, 20$ et 40 , et pour différentes valeurs de $\sigma (= 0.001, 0.01, 0.1, 0.5$ ou $1)$. Pour chaque jeu de paramètres (modèle, dimension p , valeur σ , valeur x_0 et puissance α de la fenêtre) et pour chaque échantillon simulé, nous avons calculé pour $n = N_0$ jusqu'à $n = 1000$, l'estimateur récursif $\hat{\theta}_n$ ainsi que la qualité correspondant, à savoir $\cos^2(\hat{\theta}_n, \theta)$, puis l'estimateur récursif combinant l'estimateur de Nadaraya-Watson récursif de f à l'estimateur récursif $\hat{\theta}_n$.

8.2 Qualité des estimateurs

Dans cette section, nous étudions tout d'abord la qualité des estimateurs sur des échantillons particuliers sous des modèles différents. Puis, nous regarderons l'effet de la dimension p de la covariable x sur la qualité de l'estimation. Nous présentons ensuite une étude sur $\mathcal{B} = 100$ répliquions d'échantillons afin de montrer que la valeur de la covariable x choisie et la variance du bruit σ ont de l'influence sur la qualité de l'estimateur de $f(x'\theta)$. Enfin, nous montrons que le choix de la fenêtre h_n (par le biais du choix de la valeur α) est sans réelle surprise un point crucial pour la qualité de l'estimation.

Etude de la qualité des estimateurs sur deux modèles

À la Figure FIG. 8.2, nous considérons tout d'abord des échantillons particuliers simulés à partir du modèle (M1) (resp. du modèle (M2)) avec la dimension $p = 5$ et $\sigma = 1$. Pour l'estimateur non paramétrique de f , nous avons pris différentes valeurs de $x \in \mathbb{R}^5$ ($x_0 = (0, \dots, 0)$ ou $x_1 = (1, 0, \dots, 0)$), le noyau Gaussien et $\alpha = 0.1$ ou $\alpha = 0.5$. Sur cette figure, nous avons représenté l'évolution de la mesure de qualité de l'estimateur de $\hat{\theta}_n$ et de $\hat{f}_n(x'\hat{\theta}_n)$ en fonction de n . Nous voyons clairement que plus le nombre d'observations est important, plus la qualité d'estimation est bonne pour les deux modèles. Nous observons sur le graphique du haut que les valeurs de $\cos^2(\hat{\theta}_n, \theta)$ se rapprochent de 1 très rapidement. Pour les deux graphiques du bas, nous voyons clairement que la courbe bleu (qui représente $\hat{f}_n(x'\hat{\theta}_n)$) tend bien vers les pointillés rouges (qui représentent valeur "cible" $f(x'\theta)$), ce qui correspond à nos attentes quant à la vérification expérimentale de la Loi des Grands Nombres pour l'estimateur de

$f(x'\theta)$ avec deux valeurs différentes de x .

Notons que le choix du modèle n'a pas d'influence sur la qualité de l'estimateur $f(x'\theta)$. En revanche, il semble que la valeur choisie pour x a un effet sur la qualité de notre estimateur proposé. Nous présentons plus tard une étude sur $\mathcal{B} = 100$ réplifications d'échantillons afin de préciser cet effet.

Etude de l'effet de la dimension p sur la qualité des estimateurs proposés.

Nous avons présenté à la figure FIG. 8.3, l'évolution, en fonction de n , de la qualité de l'estimateur de θ et de $f(x'\theta)$ pour un échantillon issu du modèle (M2) avec $\sigma = 1$ et différentes dimensions p ($= 2, 10, 20$ ou 40). Nous considérons ici deux valeurs de $x \in \mathbb{R}^p$ ($x_0 = (0, \dots, 0)$ ou $x_1 = (1, 0, \dots, 0)$). Pour l'estimateur non paramétrique de f , nous avons pris le noyau Gaussien et une largeur de fenêtre avec $\alpha = 0.35$. Nous observons clairement que plus la dimension p de la covariable est importante, plus la convergence de $\cos^2(\hat{\theta}_n, \theta)$ vers 1 est lente. La qualité d'estimation est tout à fait correcte (supérieure à 0.85) lorsque la taille n de l'échantillon est supérieure à 300 (resp. supérieure à 400) pour $p \leq 40$ (resp. pour $p > 40$). Par contre, la dimension p ne semble pas avoir d'influence sur la qualité de l'estimateur de $f(x'\theta)$. Plus précisément pour chaque valeur x considérée, la qualité de l'estimateur de $f(x'\theta)$ ne change pas lorsque la dimension p varie. Il est utile de rappeler que l'estimateur de $f(x'\theta)$ a été obtenu par la combinaison de l'estimateur de Nadaraya-Watson récursif de f avec celui de l'estimateur récursif $\hat{\theta}_n$. Ce qui explique que la qualité de l'estimateur $\hat{f}_n(x'\hat{\theta}_n)$ dépende de la celle de $\hat{\theta}_n$. Les erreurs d'estimation sur $\hat{\theta}_n$ semblent être "lissées" ("atténuées") lors de l'estimation non paramétrique de f .

Remarquons que ces graphiques dans les deux figures FIG. 8.2 et FIG. 8.3 ont été obtenus à partir d'échantillons simulés particuliers, et que, pour un même jeu de paramètres, les évolutions des qualités de l'estimateur de θ et de $f(x'\theta)$ peuvent être sensiblement différents pour de petites tailles n d'échantillon. Pour ces raisons là, nous présentons ci-après une étude sur $\mathcal{B} = 100$ réplifications d'échantillons.

Dans la suite, nous allons nous plaçons dans le cas où la covariable x est de dimension $p = 2$.

Etude de l'effet de la valeur de x considérée sur la qualité des estimateurs proposés.

Nous avons présenté à la figure FIG. 8.4 les boxplots, pour différentes valeurs de n ($= 10, 50, 100, 300, 500, 700, 1000$), des estimateurs de $f(x'\theta)$ calculés sur les $\mathcal{B} = 100$ échantillons issus du modèle (M2) pour différentes valeurs de $x \in \mathbb{R}^2$ ($= (0, 0), (0, 1), (0.5, -1)$ ou $(1, -1)$) avec $\sigma = 0.1$ ou $= 0.5$. Nous avons pris le noyau Gaussien et une largeur de fenêtre $\alpha = 0.35$ pour l'estimateur non paramétrique de f . Nous observons le même type de phénomène que ceux décrits dans le cas des échantillons particuliers aux figures FIG. 8.2 et FIG. 8.3. Plus précisément, plus la taille n de l'échantillon est grande, meilleures sont les estimations avec des valeurs de $\hat{f}_n(x'\hat{\theta}_n)$ de plus en plus proches des "pointillés rouges" et une dispersion de plus en plus faible. Nous voyons clairement que plus la valeur choisie pour x est éloignée du point $(0, 0)$, plus la convergence de $\hat{f}_n(x'\hat{\theta}_n)$ vers $f(x'\theta)$ est lente quelle que soit la valeur de σ . Une explication est que, en observant le nuage des points du modèle (M2), plus la valeur $x'\theta$ est éloignée de 0, plus le nombre d'observations diminue autour de ce point (le modèle (M2) représente une fonction symétrique, c'est pourquoi, nous avons présenté les boxplots avec les valeurs de x comme $x (= (0, 0), (0, 1), (0.5, -1)$ ou $(1, -1)$). Nous

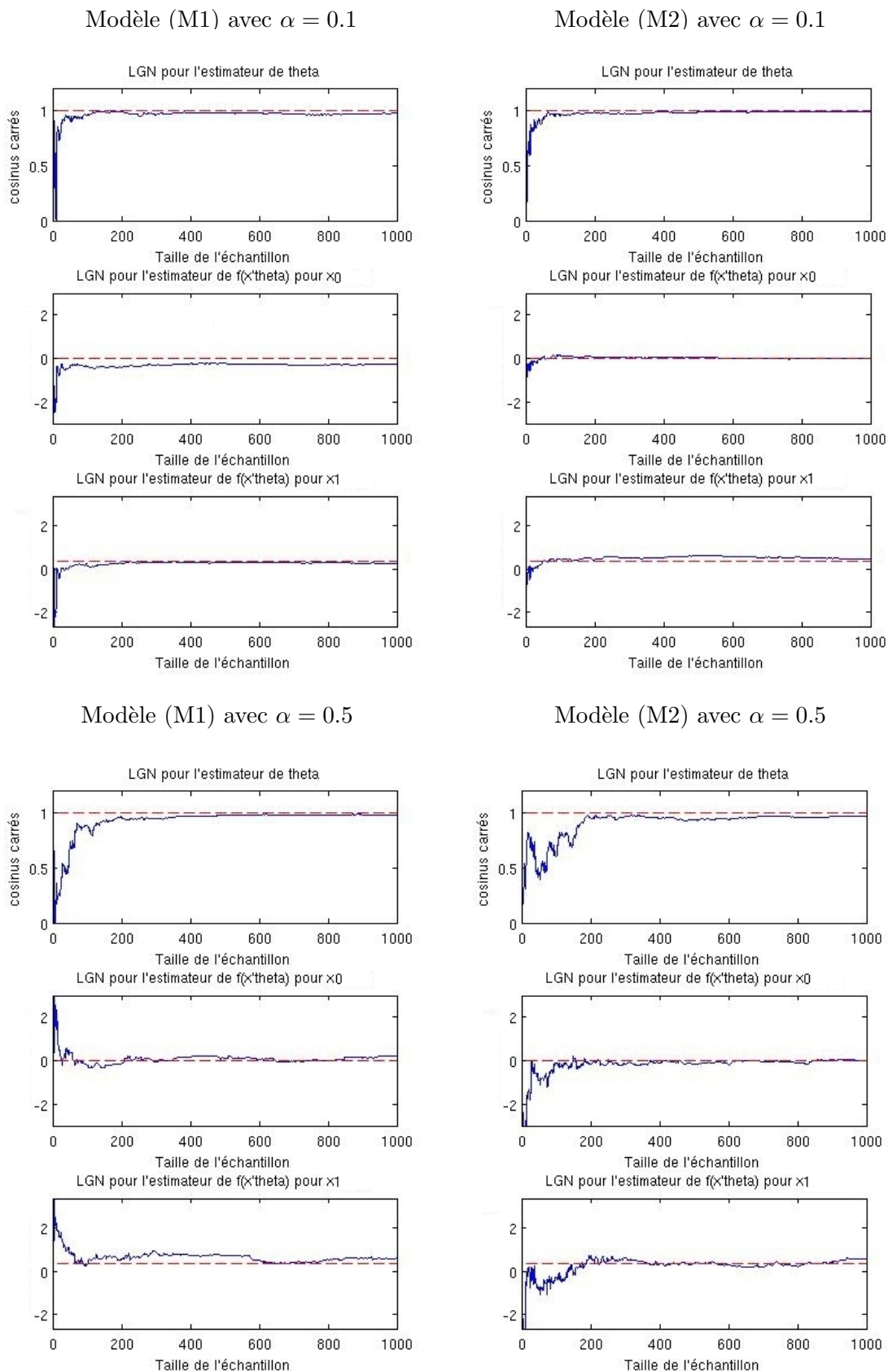


FIG. 8.2: Evolution, en fonction de n , de la qualité de l'estimateur de θ et de $f(x'\theta)$ avec $x \in \mathbb{R}^5$ ($x_0 = (0, \dots, 0)$ ou $x_1 = (1, 0, \dots, 0)$) pour un échantillon issu du modèle (M1) ou (M2) avec $p = 5$, $\sigma = 1$, $\alpha = 0.1$ ou 0.5

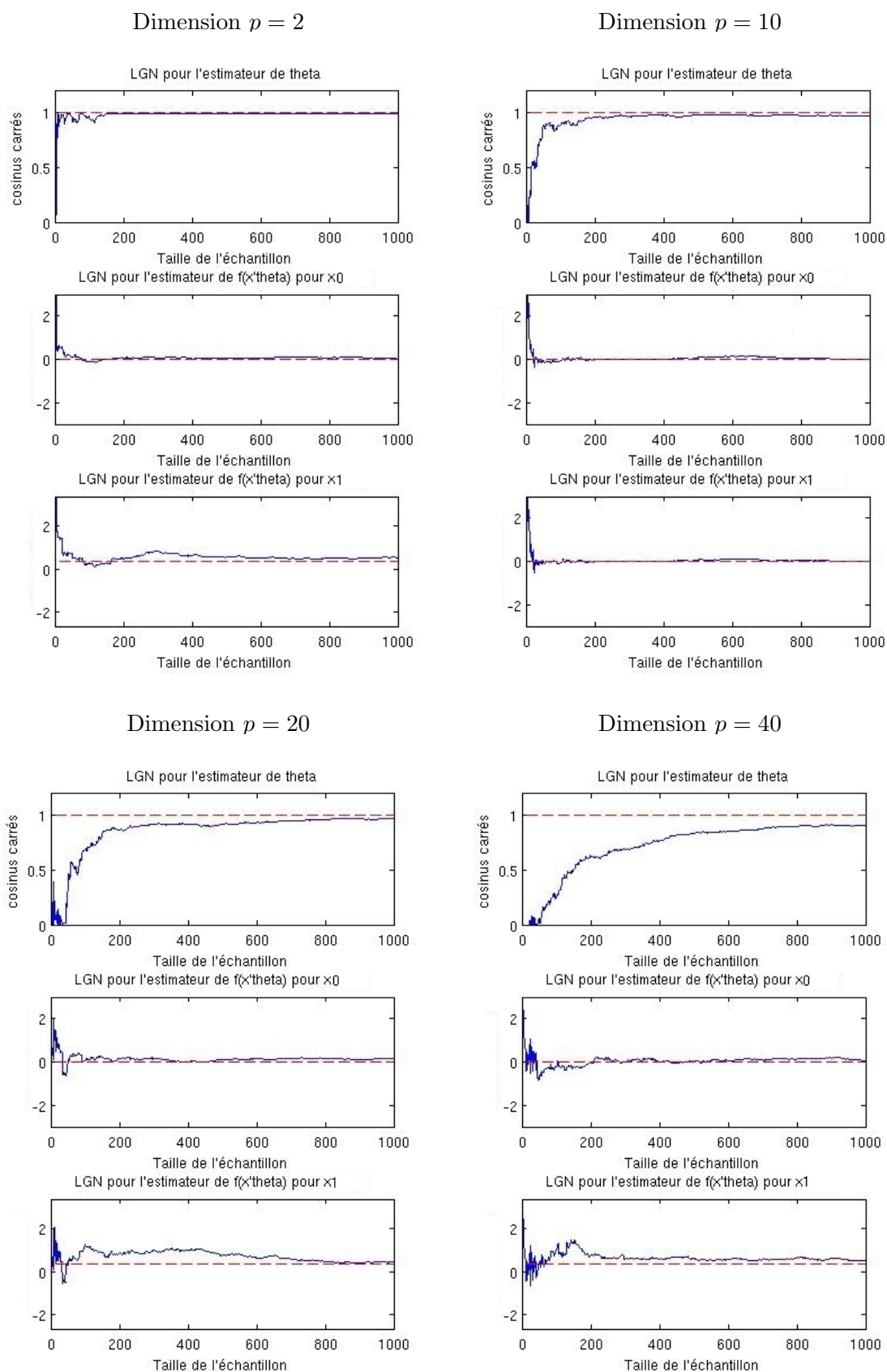


FIG. 8.3: Evolution, en fonction de n , de la qualité de l'estimateur de θ et de $f(x'\theta)$ avec $x \in \mathbb{R}^p$ ($x_0 = (0, \dots, 0)$ ou $x_1 = (1, 0, \dots, 0)$) pour un échantillon issu du modèle (M2) avec différentes dimensions p ($= 2, 10, 20$ ou 40), $\sigma = 1$ et $\alpha = 0.35$

remarquons aussi que la valeur σ a de l'influence sur la qualité de l'estimateur proposé. Pour illustrer ce point, nous pouvons nous repartir à la figure FIG. 8.5.

Etude de l'effet de la variance du bruit σ sur la qualité des estimateurs proposés.

Sans réelle surprise, la variance σ du bruit a aussi une certaine influence sur la qualité d'estimation de $f(x'\theta)$. À la figure FIG. 8.5, nous présentons les boxplots, pour différentes valeurs de n ($= 10, 50, 100, 300, 500, 700, 1000$), des estimateurs de $f(x'\theta)$ calculés sur les $\mathcal{B} = 100$ échantillons issus du modèle (M2) pour deux valeurs de x ($= (0, 0)$ ou $(0.5, -1)$) avec différentes valeurs σ ($= 0.001, 0.1, 0.5$ ou 0.5). Pour l'estimateur non paramétrique de f nous avons pris le noyau Gaussien et une largeur de fenêtre avec $\alpha = 0.35$. Nous voyons clairement que plus σ est petite, meilleures sont les estimations (avec des valeurs estimées de plus en plus proches de $f(x'\theta)$ et une dispersion de plus en plus faible). Il y a donc bien sûr sans surprise, une influence du choix de la variance du bruit sur la qualité des estimateurs proposés.

Etude de l'effet du paramètre de lissage h_n sur la qualité des estimateurs proposés.

Comme dans la plupart des problèmes d'estimation fonctionnelle, la performance d'un estimateur dépend fortement du choix de la largeur de fenêtre. Afin d'illustrer ce point, nous présentons à la figure FIG. 8.6 les boxplots, pour différentes valeurs de n ($= 10, 50, 100, 300, 500, 700, 1000$), des estimateurs de $f(x'\theta)$ calculés sur les $\mathcal{B} = 100$ échantillons issus du modèle (M2) pour deux valeurs de x ($= (0, 0)$ ou $(1, 0)$) et $\sigma = 0.1$. Nous avons pris le noyau Gaussien et différentes valeurs de α ($= 0.2, 0.35, 0.7$ ou 0.9) pour l'estimateur non paramétrique de f . Nous observons le même type de phénomène que ceux décrits dans les autres graphiques. Plus précisément, plus la taille n de l'échantillon est grande, meilleurs sont les estimateurs. Nous voyons clairement que la puissance α de la fenêtre $h_n = n^{-\alpha}$ a l'influence sur la performance de l'estimateur de $f(x'\theta)$. Plus précisément, lorsque $\alpha = 0.35$ les estimateurs sont meilleures que celui avec $\alpha = 0.9$ et celui avec $\alpha = 0.2$. Lorsque la largeur de la fenêtre h_n est trop grande (α est trop petit), nous avons le problème du surlissage.

8.3 Illustration de la normalité asymptotique de l'estimateur

À la figure FIG. 8.7, nous nous intéressons à la normalité asymptotique de l'estimateur $\hat{f}_n(x'\hat{\theta}_n)$. Pour cela, nous avons simulé 1000 réalisations de $\hat{f}_n(x'\hat{\theta}_n)$ à partir d'échantillons de taille $n = 1000$ issus du modèle (M1) avec $p = 2$. Pour l'estimateur, nous avons pris le noyau Gaussien et la fenêtre $h_n = n^{-0.45}$ et $h_n = n^{-0.9}$. Nous avons tracé sur un même graphique l'histogramme de Z_n (en bleu) ainsi que la densité de la loi normale centrée réduite (en rouge), où

$$Z_n = \sqrt{nh_n} \left(\frac{\hat{f}_n(x'\hat{\theta}_n) - f(x'\theta)}{\sqrt{\frac{\sigma^2 \nu^2}{(1+\alpha)h(\theta, z)}}} \right)$$

rappelons que $\nu^2 = \int_{\mathbb{R}} \mathbf{K}^2(x) dx$, $\sigma^2 = \mathbb{E} [\varepsilon_i^2 \mid \mathcal{F}_{i-2}]$ et $h(\theta, x'\theta)$ et la densité de probabilité de Φ .

8.3. Illustration de la normalité asymptotique de l'estimateur

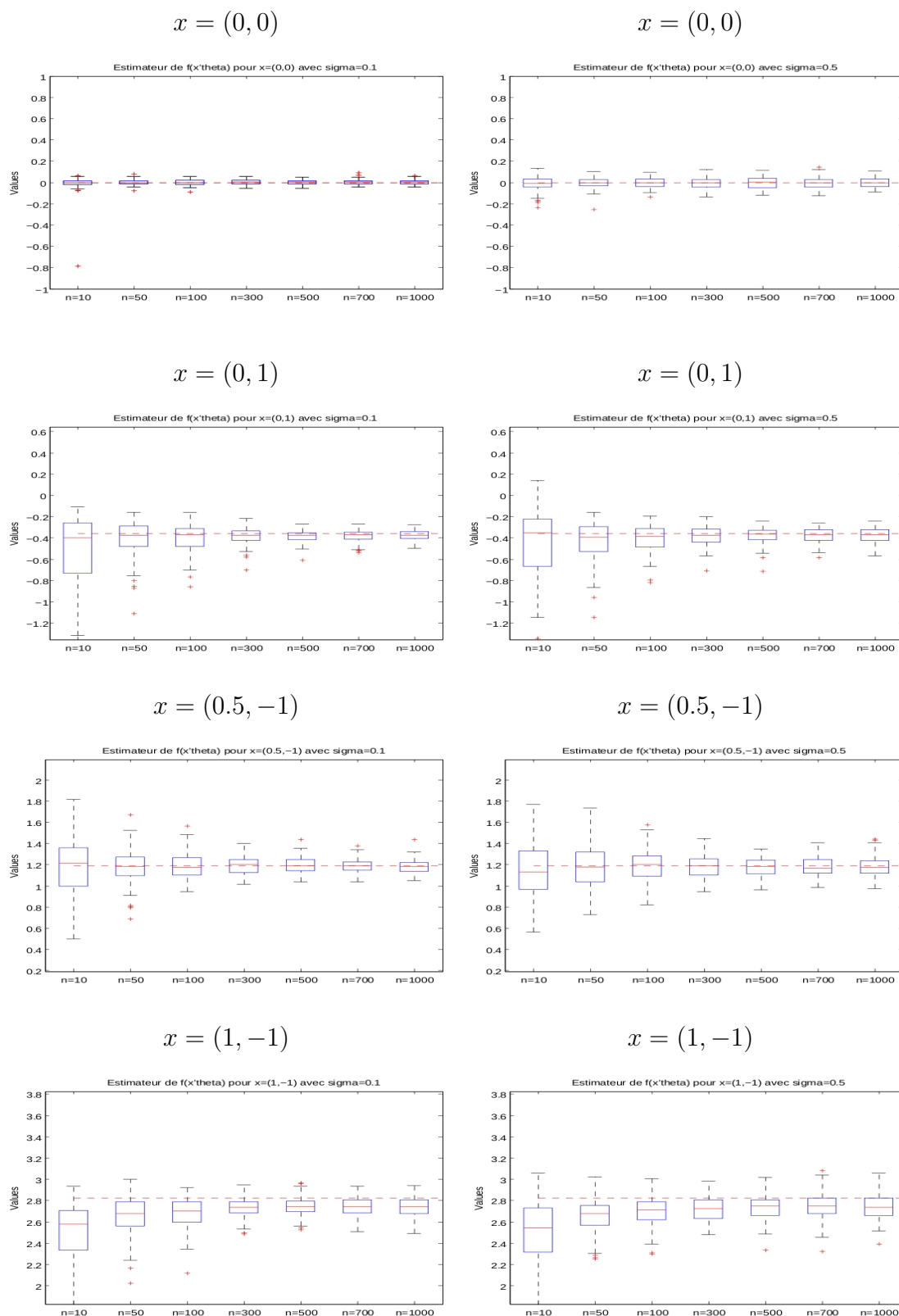


FIG. 8.4: Boxplots, en fonction de n , des estimateurs de $f(x'\theta)$ calculés sur $\mathcal{B} = 100$ échantillons issus du modèle (M2) pour différentes valeurs de $x \in \mathbb{R}^2 (= (0, 0), (0, 1), (0.5, -1)$ ou $(1, -1)$ avec $\sigma = 0.1$ ou $= 0.5$ et $\alpha = 0.35$

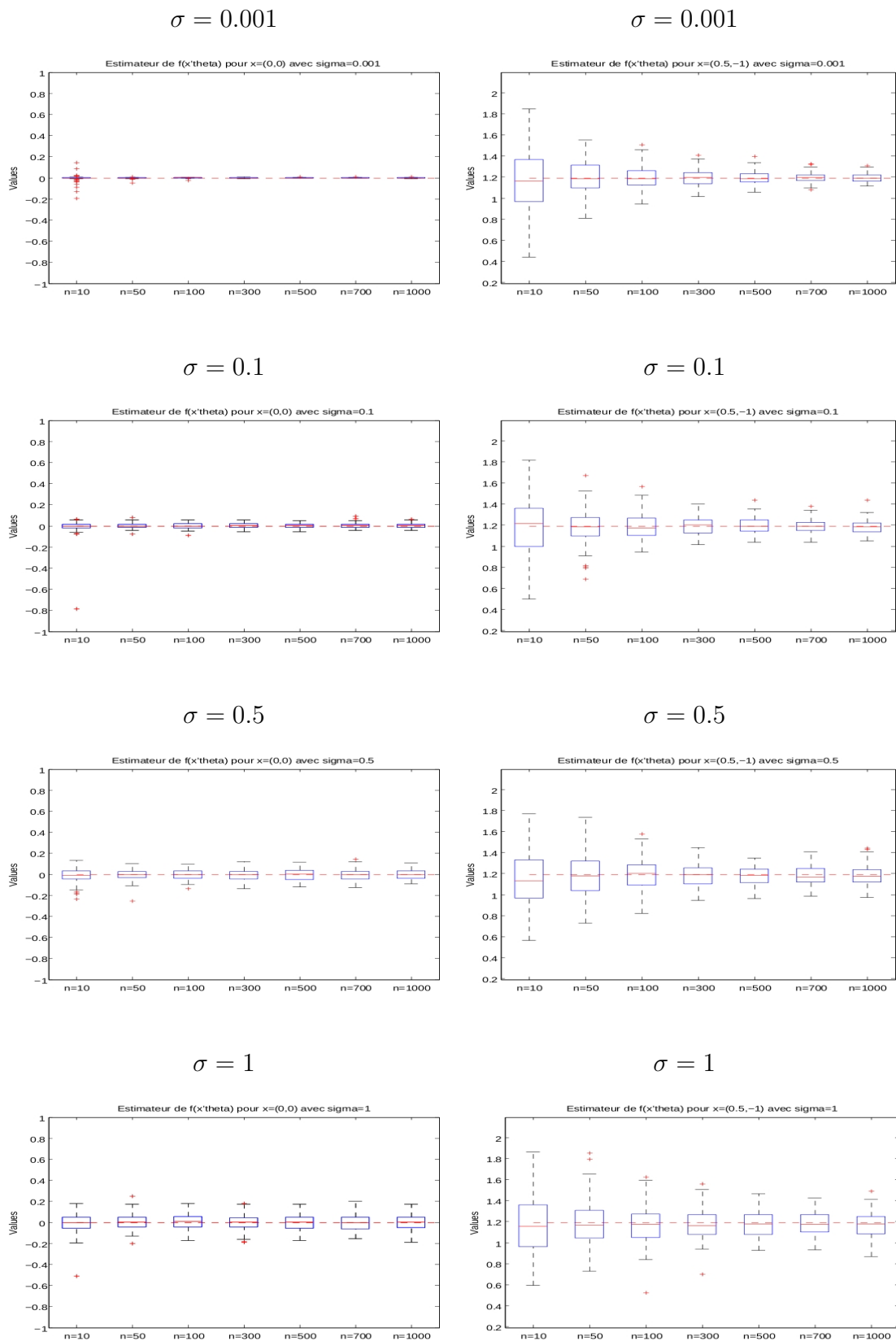


FIG. 8.5: Boxplots, en fonction de n , des estimateurs de $f(x'\theta)$ calculés sur $\mathcal{B} = 100$ échantillons issus du modèle (M2) pour deux valeurs de $x \in \mathbb{R}^2$ ($= (0, 0)$ ou $(0.5, -1)$) avec différentes valeurs de σ ($= 0.001, 0.1, 0.5$ ou 1) et $\alpha = 0.35$.

8.3. Illustration de la normalité asymptotique de l'estimateur

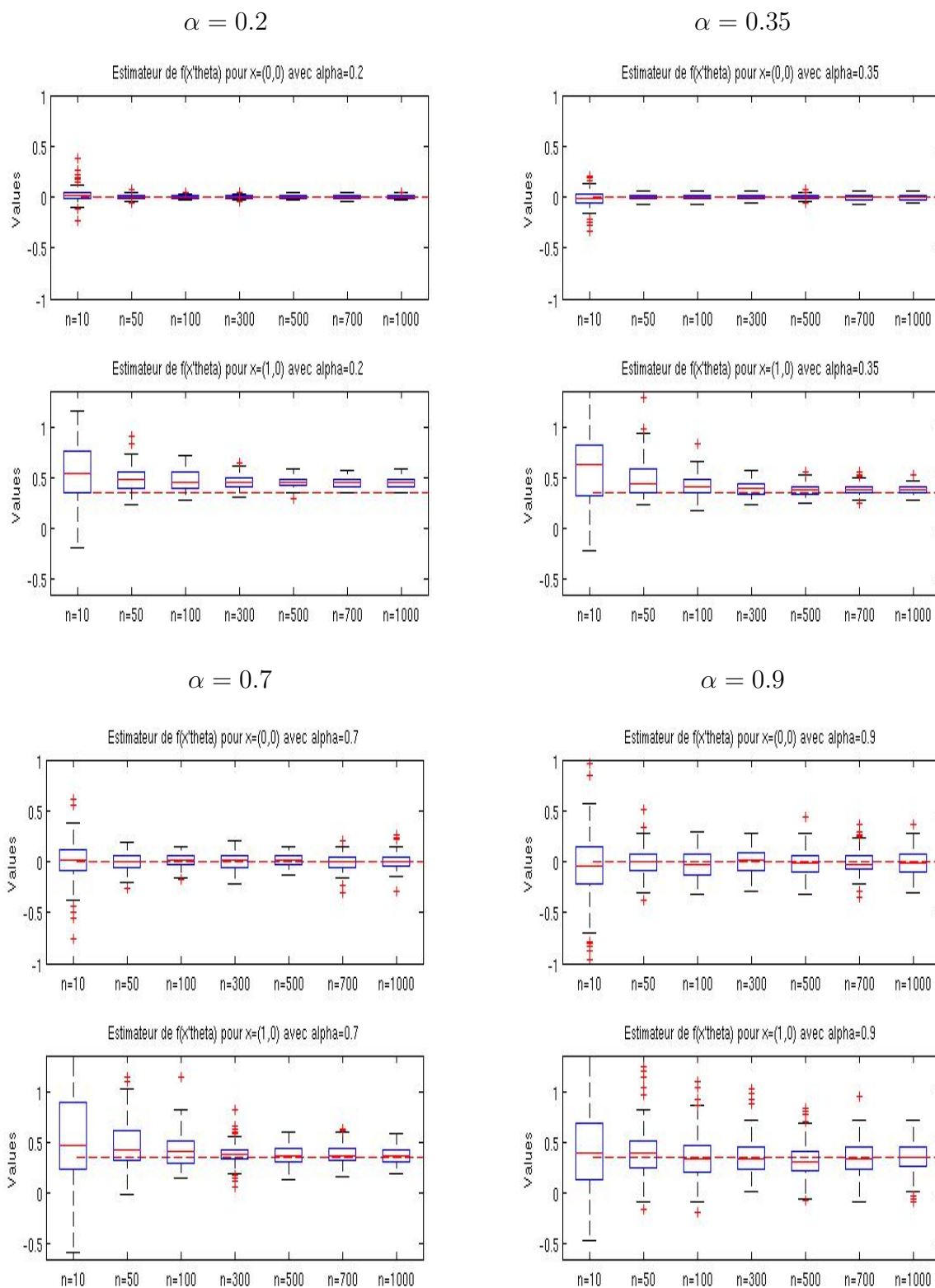


FIG. 8.6: Boxplots, en fonction de n , des estimateurs de $f(x'\theta)$ calculés sur $\mathcal{B} = 100$ échantillons issus du modèle (M2) avec deux valeurs de $x \in \mathbb{R}^2 (= (0, 0)$ ou $(1, 0))$, $\sigma = 0.1$ et différentes valeurs de $\alpha (= 0.2, 0.35, 0.7$ ou $0.9)$

Nous voyons clairement que les deux histogrammes ont bien un comportement gaussien représenté par la courbe rouge : la densité de la loi $\mathcal{N}(0,1)$ se superpose bien aux différents histogrammes. Ceci illustre bien la normalité asymptotique de l'estimateur proposé. Nous avons aussi obtenu d'autres graphiques avec différents noyaux et différentes fenêtres qui montrent aussi bien la normalité asymptotique de $\hat{f}_n(x'\hat{\theta}_n)$.

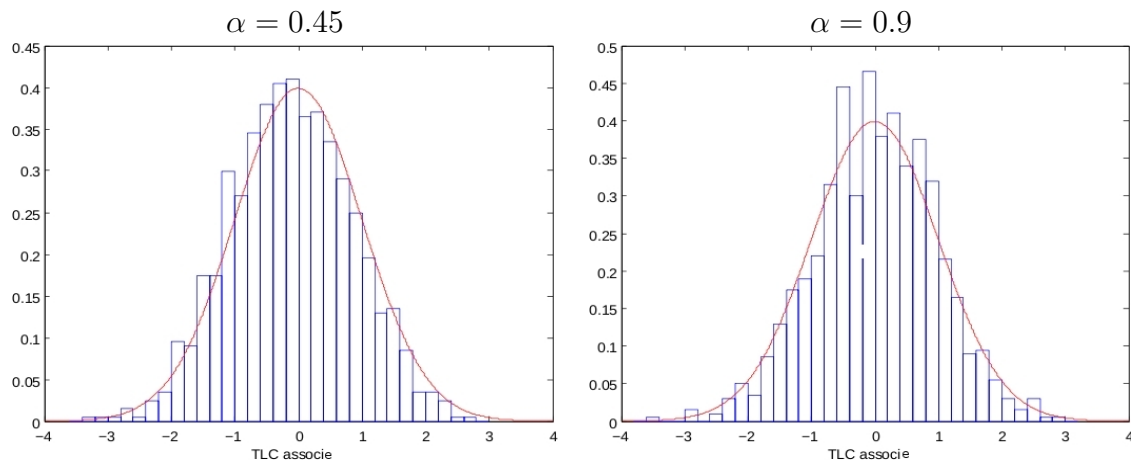


FIG. 8.7: Illustration de la normalité asymptotique de $\hat{f}_n(x'\hat{\theta}_n)$ calculés sur $\mathcal{B} = 1000$ échantillons de taille $n = 1000$ issus du modèle (M1) pour $p = 2$ avec le choix du noyau Gaussien et la largeur de fenêtre $n^{-0.45}$ ou $n^{-0.9}$

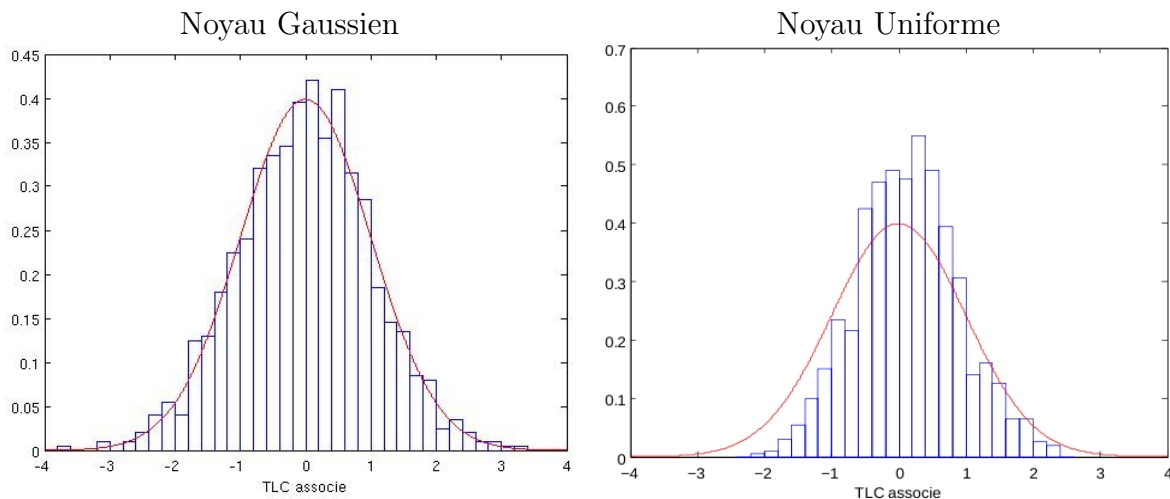


FIG. 8.8: Illustration de la normalité asymptotique de $\hat{f}_n(x'\hat{\theta}_n)$ calculés sur $\mathcal{B} = 1000$ échantillons de taille $n = 1000$ issus du modèle (M2) pour $p = 1$ avec la largeur de fenêtre $n^{-0.5}$ et le noyau Gaussien ou noyau Uniforme

8.4 Conclusion

Dans ce chapitre, nous avons étudié sur des simulations le comportement numérique de l'estimateur récursif combinant l'estimateur de Nadaraya-Watson récursif de f à l'estimateur SIR récursif $\hat{\theta}_n$. Nous avons illustré le bon comportement de l'estimateur proposé en considérant différents modèles, différentes dimensions, plusieurs valeurs pour la covariable x ainsi que différentes valeurs de σ . Nous avons montré que le choix du point cible, la variance du bruit ont de l'influence sur la qualité de l'estimateur proposé. De même la largeur de la fenêtre h_n a un effet sur la performance de l'estimateur proposé. Nous avons illustré aussi la normalité asymptotique de l'estimateur récursif proposé.

Bibliographie

- [1] Bercu, B., Chafaï, D. (2007) Modélisation stochastique et simulation. *Dunod*.
- [2] Bercu, B., Portier, B. (2008). Kernel density estimation and goodness-of-fit test in adaptive tracking. *SIAM J. Control Optim.* , **47**, 2440-2457.
- [3] Bercu, B. (1998). Central limit theorem and law of iterated logarithm for least squares algorithms in adaptive tracking. *SIAM J. Control Optim.* , **Vol. 36, No. 3**, 910-928.
- [4] Duflo, M. (1997) Random Iterative Models. *Collection mathématiques et applications, Springer*.
- [5] Dacunha-Castelle, D. et Duflo, M. (1993) Probabilité et Statistiques. *Collection mathématiques et applications pour la Maîtrise, Tome 2*.
- [6] Hall, P. (1984). Asymptotic properties of integrated square error and cross validation for kernel estimation of a regression function. *Z. Wahrsch. Verw. Gebiete*, **67**, 175-196.
- [7] Hall, P., Heyde, C.C. (1980). Martingale limit theory and its application. *Academic Press, Inc., New York*, 308 pages.
- [8] Härdle, W. and Marron, J.S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.* , **13.4**, 1465-1481.
- [9] Härdle, W. (1990) Applied nonparametric regression. *Econometric Society Monographs*.
- [10] Härdle, W. and Marron, J.S (1985) Optimal bandwidth selection in nonparametric regression function estimation. *The Annals of Statistics*, **13**, 4, 1465-1481.
- [11] Mokkadem, A., Pelletier, M. and Slaoui, Y (2009) The stochastic approximation method for the estimation of a multivariate probability density. *J. Sattist. Plann. Inference*, **139**, 2459-2478.
- [12] Mokkadem, A., Pelletier, M. and Slaoui, Y (2009) Revisiting Révész stochastic approximation method for the estimation of a regression function. *Alea* **6**, 63-114.
- [13] Nadaraya, E.A. (1964) On estimating regression. *Theory of Probability and its Applications* **10**, 186-190.
- [14] Nadaraya, E. A. (1989). *Nonparametric Estimation of Probability Densities and Regression Curves*. Kluwer, Dordrecht.
- [15] Révész, P. (1977). How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. *Math. Operationsforsch. Statist., Ser. Statistics*, **8**, 119-126.

- [16] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, **27**, 832-837.
- [17] Roussas, G. (1992). Exact rates of almost sure convergence of a recursive kernel estimate of a probability density function : Application to regression and hazard rate estimate. *J. of Nonparam. Statist.*, **3**, 171-195.
- [18] Silverman, B. W (1986). Density estimation for statistics and data analysis, Chapman & Hall.
- [19] Stone, C.J. (1977). Consistent nonparametric regression. *Ann. Statist.*, **5.4**, 595-645.

Conclusion et Perspectives

Conclusion

Nous avons développé dans cette thèse une approche qui est fondée sur une méthode introduite par Li (1991), appelée Sliced Inverse Regression (SIR). Cette méthode permet d'estimer la partie paramétrique θ d'un modèle de régression semi-paramétrique sans avoir à estimer la fonction de lien f . Nous avons proposé une approche récursive de la méthode SIR. Dans le cas particulier où l'on considère le nombre de tranches égal à 2, il est possible d'obtenir une expression analytique de l'estimateur de la direction du paramètre θ . Nous avons proposé une forme récursive pour cet estimateur. Nous avons également donné une forme récursive de l'estimateur de la matrice d'intérêt pour un nombre quelconque de tranches. Nous avons obtenu la convergence presque sûre avec la vitesse correspondante ainsi que la normalité asymptotique de notre estimateur. Les simulations ont illustré le bon comportement numérique de l'approche récursive proposée. Nos résultats ont confirmé l'avantage majeur de cette approche récursive : sa rapidité par rapport aux méthodes SIR non récursive et SIR classique.

Nous avons ensuite proposé une nouvelle approche appelée "SIRoneslice" (récursive ou non récursive) de la méthode SIR basée sur l'utilisation de l'information contenue dans une seule tranche optimale (qu'il faudra choisir parmi un nombre quelconque de tranches). Nous avons aussi proposé un critère "bootstrap naïf" pour le choix du nombre de tranches. Nous avons donné un résultat de convergence presque sûre avec la vitesse correspondante de notre estimateur. Nous avons montré les bonnes performances numériques de cette méthode SIRoneslice (récursive ou non récursive) sur des simulations, ainsi que l'avantage principal de l'utilisation de la version récursive de SIR et de SIRoneslice du point de vue des temps de calculs.

Par ailleurs, nous avons travaillé sur des données réelles de valvométrie mesurées sur les bivalves de mollusques. L'objectif est d'étudier les perturbation des rythmes qui peuvent induire une perturbation du milieu. Sur ces données, nous avons comparé le comportement numérique des trois estimateurs Nadaraya-Watson, Nadaraya-Watson récursif et Révész. Les résultats obtenus ont permis de confirmer que les estimateurs récursifs sont statistiquement meilleurs que leur version non récursive.

Finalement, nous avons proposé une méthode permettant de combiner l'estimation récursive de la fonction de lien f avec l'estimateur de Nadaraya-Watson récursif et l'estimation du paramètre θ via l'estimateur SIR récursif. Nous avons établi une Loi des Grands Nombres ponctuelle ainsi qu'un Théorème de la Limite Centrale ponctuelle. Nous avons illustré ces résultats théoriques par des simulations numériques montrant le bon comportement numérique de la méthode d'estimation proposée.

Perspectives

Dans la continuité de ce travail, plusieurs perspectives sont envisageables. Elles concernent essentiellement les méthodes récursives. Les principales perspectives de cette thèse sont les suivantes.

Application sur des vraies données pour la partie "réduction de dimension". Dans la seconde partie, nous avons travaillé sur des données réelles (mesurées sur des bivalves) dont l'objectif est d'étudier les perturbations des rythmes qui peuvent in-

duire une perturbation du milieu comme une pollution. Nous avons considéré le modèle de régression non paramétrique donné par $Y = f(X) + \varepsilon$, où Y est l'amplitude d'ouverture en millimètre ; X est le temps en heure ; f est une fonction inconnue à estimer qui exprime la valeur moyenne de l'écartement valvaire de nos bivalves en fonction du temps et ε est le terme aléatoire d'erreur de loi inconnue, indépendant de X .

Afin de mieux caractériser les effets des paramètres environnementaux sur l'activité valvaire des animaux, il serait donc intéressant de prendre en compte l'ensemble de leur contexte naturel via les mesures des paramètres du milieu (température de l'eau, vitesse du courant, O_2 , pH, salinité, fluorescence, pression hydrostatique). Dans ce cadre, la variable explicative $X \in \mathbb{R}^p$ avec $p > 1$ remplace $X \in \mathbb{R}$ dans ce modèle et une étape de réduction de dimension pourrait très certainement être utile.

SIR récursif pour “des nouvelles données arrivant par paquet”. Dans la thèse, nous avons confirmé l'avantage majeur des méthodes récursives : leur rapidité par rapport aux méthodes non récursives. L'intérêt des méthodes récursives est qu'il n'est pas nécessaire de relancer tous les calculs d'estimation à chaque fois que la base de données est complétée par des nouvelles observations. Nous pourrions envisager d'utiliser des méthodes récursives en remplaçant la nouvelle observation (x_n, y_n) arrivant par plusieurs observations nouvelles qui arrivent simultanément.

En s'inspirant des travaux de Chavent et al.(2010) qui ont proposé une approche basée sur la méthode de SIR afin d'estimer l'espace engendré par une direction EDR commune lorsque la population est stratifiée, nous pouvons proposer une méthode permettant de gérer un nouveau paquet de données arrivant dans la base de données. L'idée consiste à calculer l'estimateur $\hat{b}^{(1)}$ de la direction EDR pour les anciennes observations $\{(x_i, y_i), i = 1, \dots, n^{(1)}\}$. Ensuite, pour le nouveau paquet d'observations $\{(x_i, y_i), i = n^{(1)} + 1, \dots, n^{(2)}\}$ arrivant, nous calculons l'estimateur $\hat{b}^{(2)}$ de la direction EDR associée à ces nouvelles observations. Puis, nous construisons la matrice $\hat{B} = [\hat{b}^{(1)}, \hat{b}^{(2)}]$. Enfin, le vecteur propre associé à la plus grande valeur propre de la matrice $\hat{B}\hat{B}'$ est la direction EDR commune estimée pour la totalité des données. Il sera aussi intéressant d'étudier s'il y a une “dérive de la direction EDR” au cours du temps, c'est à dire s'il y a bien une direction EDR commune entre les nouvelles et les anciennes observations.

Extension d'autres méthodes SIR (SIR_{II} et SIR_α) à un cadre récursif. Nous avons présenté quelques extensions de la méthode SIR à la section 1.4 du chapitre 1 de la première partie de cette thèse. Il serait intéressant de développer les approches SIR_{II} et SIR_α dans le cadre récursif. Nous pouvons également envisager l'extension au cadre SIR multivarié ($y \in \mathbb{R}^q$).

Utilisation d'un autre estimateur récursif de la fonction de lien f : l'estimateur de Révész. Dans la partie 3 de cette thèse, nous avons proposé une méthode permettant de combiner l'estimation récursive de la fonction de lien f par l'estimateur de Nadaraya-Watson récursif et l'estimation du paramètre θ via l'estimateur SIR récursif. Nous pouvons envisager de remplacer l'estimateur de Nadaraya-Watson récursif par l'estimateur de Révész dans cette approche. Il serait intéressant d'étudier les propriétés asymptotiques de cet estimateur, ainsi que d'évaluer son comportement sur simulations et de le comparer à l'estimateur proposé dans la dernière partie de la thèse.