



HAL
open science

deux contributions à l'étude semi-paramétrique d'un modèle de régression

Céline Roget-Vial

► **To cite this version:**

Céline Roget-Vial. deux contributions à l'étude semi-paramétrique d'un modèle de régression. Mathématiques [math]. Université Rennes 1, 2003. Français. NNT: . tel-00008730

HAL Id: tel-00008730

<https://theses.hal.science/tel-00008730>

Submitted on 8 Mar 2005

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 de la thèse : 2923

THÈSE

Présentée

DEVANT L'UNIVERSITÉ DE RENNES I

pour obtenir

le grade de DOCTEUR DE L'UNIVERSITÉ DE RENNES I

Mention Mathématiques et Applications

par

Céline ROGET-VIAL

CREST – ENSAI, École Doctorale MATISSE

TITRE DE LA THÈSE :

*Deux contributions à l'étude semi-paramétrique
d'un modèle de régression*

Sous la direction de MICHEL DELECROIX

Soutenue le 24 octobre 2003 devant la Commission d'Examen

COMPOSITION DU JURY :

ANESTIS ANTONIADIS	Rapporteur (UJF, Grenoble)
DENIS BOSQ	Examineur (Université Paris VI)
MICHEL DELECROIX	Directeur de recherche (CREST-ENSAI)
BERNARD DELYON	Examineur (Université de Rennes 1)
JEAN DESHAYES	Examineur (Université de Rennes 1)
DOMINIQUE PICARD	Examineur (Université Paris VII)
VLADIMIR SPOKOINY	Rapporteur (Weierstrass Institute, Berlin)
ALAIN TROGNON	Invité (GENES, Paris)

Cela fait déjà quatre ans que je connais Michel Delecroix : je l'ai découvert comme enseignant et j'apprécie encore ses qualités pédagogiques. Il a été un directeur de thèse attentionné dont l'optimisme et l'enthousiasme m'ont encouragée tout au long de ces trois années de thèse. Je le remercie sincèrement de la confiance qu'il me témoigne.

Merci à Anestis Antoniadis et Vladimir Spokoiny d'avoir accepté de rapporter sur ma thèse malgré un emploi du temps déjà bien rempli, et de l'intérêt qu'ils portent à celle-ci et qui me touche tout particulièrement. Je remercie chaleureusement Dominique Picard, Denis Bosq et Bernard Delyon pour leur participation au jury et plus personnellement Jean Deshayes, qui m'a fait découvrir la Statistique, et Alain Trognon pour l'attention qu'il a toujours porté à mon travail.

Deux rencontres déterminantes ont marqué mes premiers pas dans la recherche : sans Peter Hall et Vladimir Spokoiny, ma thèse n'aurait pas été la même. Je les remercie pour leur disponibilité et les discussions enrichissantes que nous avons eues à maintes reprises.

J'ai profité de nombreuses fois de l'expérience de mes aînés ; Philippe Berthet, Jean-Yves Dauxois, Denys Pommeret, Marian Hristache, Marc Dambrine n'ont jamais hésité à donner de leur temps pour répondre à mes questions (parfois bêtes !). Je n'oublie pas Pierre Druilhet, François Coquet et Eric Matzner-Løber qui, par leur bonne humeur et leur dynamisme, ont eux aussi contribué à rendre ces trois années plus agréables. Grâce à leur enthousiasme ils m'ont tous communiqué leur goût des statistiques, qu'avant eux des enseignants tels Michel Bonneau et Jérôme Pagès avaient fait naître. Je tiens aussi à remercier Michel Pierre pour la sagesse de ses encouragements.

J'ai eu la chance de passer trois ans dans le cadre exceptionnel de l'ENSAI, merci à tous les permanents qui m'ont accueillie très chaleureusement. L'IRMAR et L'ENS m'ont permis d'utiliser leurs calculateurs ; j'ai souvent hanté les couloirs des étages d'analyse numérique et de processus stochastiques et le bâtiment A. Sauvy de l'ENS (n'est-ce pas Arnaud ?). Sur ces lieux de travail, j'ai rencontré de jeunes chercheurs (les conseils de Glenn portent leurs fruits...), avec lesquels il n'était pas toujours questions de travail ! Pour ces pauses thé ou café qui ont parfois dégénéré en plaintes collectives, merci à : Julien, Stéphanie, Hélène, Agathe, Vincent Lefieux, Solen, Gweno, Damien, Ronan, Sandrine, Benoîte, Vincent Leprince, Manuela, Glenn, Erwan B., Katell...

Merci à mes parents, Stéphane et Dedel qui ont toujours soutenu leur petite dernière ! Je tiens encore à remercier les amis et la famille de Paris, Lyon, Marseille ou d'ailleurs.

Comment remercier mon plus grand supporter, sinon en lui dédiant cette thèse ?

Table des matières

Introduction Générale	5
I Étude d'un modèle de régression par une méthode adaptative	11
Introduction	13
1 Détermination des composantes linéaires	15
1.1 Comment reconnaître une composante linéaire ?	15
1.2 Une estimation "améliorée" de g	17
1.3 L'algorithme itératif	17
1.4 Estimation de la variance du bruit	19
1.5 Choix des paramètres	20
2 Qualité des estimateurs	21
2.1 Hypothèses	21
2.2 Amélioration en une étape	23
2.3 Propriétés de l'estimateur initial	38
2.4 Qualité du l'estimateur final	40
2.5 Estimation de la variance	45
3 Exploration du modèle partiellement linéaire	50
3.1 Séparation entre composantes linéaires et non-linéaires	50
3.2 Test d'une hypothèse sur M	52
3.3 Identification de la partie non-linéaire	58
3.4 Estimation et inférence sur les composantes linéaires	60
4 Applications	62
4.1 Algorithme modifié	62
4.2 Résultats obtenus sur des données simulées	63
4.2.1 Qualité de l'estimation	64
4.2.2 Séparation et exploration du modèle	66

4.3	Un exemple sur des données réelles	68
Annexe : Propriétés de grande déviation pour des vecteurs gaussiens		71
II Estimation et test d'une direction révélatrice unique : une approche géométrique		75
Introduction		77
5	Approche géométrique de la condition H_0	81
5.1	Les cylindres	81
5.2	Un premier critère de l'hypothèse SIM	82
5.3	Critère global de l'hypothèse SIM	86
5.4	La méthode statistique : description	90
6	Hypothèses sur le modèle	93
7	Étude asymptotique de \hat{I}_S	97
7.1	Étude du processus A_S	98
7.1.1	Convergence et majoration uniforme de $A_{S,1}$	99
7.1.2	Étude du U -processus $A_{S,2}$	102
7.1.3	Vitesse de convergence de $A_{S,3}$	102
7.1.4	Ordre de grandeur de $A_{S,4}$	107
7.2	Étude du processus $B_S(g)$	108
7.2.1	Ordre de grandeur de $B_{S,1}$	109
7.2.2	Étude du U -processus $B_{S,2}$	109
7.2.3	Vitesse de convergence de $B_{S,3}$	110
7.2.4	Vitesse de convergence $B_{S,4}$	111
7.2.5	Ordre de grandeur de $B_{S,5}$	115
7.3	Propriétés asymptotiques de \hat{I}_S	115
8	Convergence de l'estimateur de l'axe du SIM	117
9	Étude du test	127
9.1	Calibration par Bootstrap	127
9.2	Propriétés du test	128
Annexe 1 : Étude de la continuité des intégrales		137
10.1	Les cylindres et leur paramétrisation	137
10.2	Continuité	138

Annexe 2 : Rappels sur les U-statistiques	143
11.1 Introduction	143
11.2 Quelques résultats sur les familles euclidiennes	144
11.2.1 Familles de fonctions euclidiennes	144
11.2.2 Ordre uniforme en probabilité	145
Annexe 3 : Distances dans un échantillon	147
12.1 Développement asymptotique de $\mathbb{E} \{D^d(\mathbf{x})\}$	147
12.1.1 Développement asymptotique de $\mathbb{E} (D_1^d)$	150
12.1.2 Développement asymptotique de $\{\mathbb{E} (D_2^d) - \mathbb{E} (D_1^d)\}$	155
12.1.3 Choix de $D^d(\cdot)$	159
12.2 Application à l'étude de $\tilde{I}_S(g)$	160
12.3 Étude du comportement asymptotique de $D(\mathbf{x})$	160
Bibliographie	166

Introduction Générale

Le travail qui suit apporte une contribution à l'étude de deux modèles de régression semi-paramétriques. Ces modèles proposent des éléments de réponse au classique phénomène du *fléau de la dimension*.

Le problème statistique de la régression, peut-être le plus usuellement étudié depuis qu'existe la statistique mathématique, est celui de l'analyse de l'influence d'un ensemble de d variables *explicatives*, noté X , $X \in \mathbb{R}^d$, sur une variable dite *expliquée*, notée Y . En considérant le couple (X, Y) comme aléatoire, le problème sera d'estimer la fonction g , appelée "espérance conditionnelle"

$$g(\mathbf{x}) = \mathbb{E}(Y|X = \mathbf{x}).$$

En effet, dans la mesure où $g(\mathbf{x})$ est la meilleure approximation L^2 de Y , l'estimation de cette fonction est considérée comme réponse au problème précédent. On dispose pour cela d'un échantillon $(X_i, Y_i)_{i=1, \dots, n}$, observé, de vecteurs aléatoires indépendants, dont la loi commune est celle de (X, Y) .

Pour estimer g à partir des (X_i, Y_i) une première solution historique a été le choix de modèles paramétriques (modèle linéaire ou linéaire généralisé, par exemple). Il suppose alors d'estimer les paramètres du modèle pour définir \hat{g} .

Lorsque le statisticien ne dispose a priori d'aucune information sur la loi des observations, et veut éviter les erreurs de spécification, il est naturel d'essayer une estimation "directe" de g par une méthode non-paramétrique. Ces méthodes (noyau, polynômes locaux, ondelettes, splines...) largement développées depuis les années soixante-dix n'ont cependant pas forcément la faveur des praticiens dans la mesure où elles souffrent d'un manque d'interprétabilité et exigent de très grandes tailles d'échantillon pour être performantes. Ce dernier point, connu sous le nom de *fléau de la dimension*, peut être simplement mis en évidence par la notion de *vitesse asymptotique minimale*, cf. *Ibragimov et Has'Minskii (1980)*. On peut résumer ce résultat ainsi : si g est k fois différentiable, un estimateur non-paramétrique de g ne peut converger vers g plus vite que $n^{\frac{-2k}{2k+d}}$, ce qui conduit à une vitesse extrêmement faible si d est grand. De nombreux auteurs se sont attaqués à ce problème dans les deux dernières décennies, en proposant des méthodes d'estimation non-paramétrique de g , sous des hypothèses diverses concernant la loi de (X, Y) , ou plus exactement la forme de g . Ces hypothèses permettent d'échapper au *fléau* évoqué ci-dessus

tout en conservant le caractère non-paramétrique de la méthode d'estimation. L'ensemble de ces méthodes est usuellement repris sous le vocable d'étude de la *réduction de dimension*. Le travail qui suit entre dans cette problématique.

Sans souci d'exhaustivité, rappelons d'abord quelques types d'hypothèses existant dans ce cadre.

1. Dans le *modèle additif*, on suppose que la fonction de régression est de la forme

$$g(\mathbf{x}) = g_1(x_1) + \cdots + g_d(x_d),$$

où les g_i sont des fonctions univariées inconnues. La régression non-paramétrique additive a été introduite par *Stone (1985)* et étudiée entre autres par *Hastie et Tibshirani (1986)* et plus récemment par *Opsomer et Ruppert (1999)*. L'hypothèse introduite est cependant difficile à tester et ce modèle ne prend pas en compte les possibles interactions entre variables explicatives.

2. Le modèle à *direction révélatrice unique*, noté SIM, suppose l'existence d'un vecteur $\varphi_0 \in \Theta = \{\theta \in \mathbb{R}^d / \|\theta\| = 1\}$ et d'une fonction g_1 tels que

$$g(\mathbf{x}) = g_1(\mathbf{x}^\top \varphi_0). \quad (1)$$

Ce modèle a été étudié en particulier par *Powell, Stock et Stoker (1989)* et est une généralisation non-paramétrique du classique modèle linéaire généralisé, cf *Nelder et Weddurburn (1972)*.

3. Dans le *modèle partiellement linéaire*, la fonction de régression a la forme

$$g(\mathbf{x}) = \theta^\top \mathbf{x}_1 + g_1(\mathbf{x}_2), \quad \text{où } \mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2). \quad (2)$$

On suppose de plus $\dim(\mathbf{x}_2) = M \ll d$ (en pratique $M = 1, 2$ ou 3 , sinon le modèle est de peu d'intérêt). Ce modèle a été introduit par *Engle et al. (1986)* et par *Hekman (1986)*. Il permet d'échapper au problème relatif à la dimension, car la fonction inconnue à estimer non-paramétriquement, g_1 , est une fonction de M variables, $M \ll d$.

4. Le modèle dit "*projection pursuit*", est une généralisation du modèle additif et du modèle précédent, on suppose ici que

$$g(\mathbf{x}) = g_1(\mathbf{x}^\top \varphi_1) + \dots + g_k(\mathbf{x}^\top \varphi_k),$$

où $k \in \mathbb{N}$ et $k \leq d$. Ce modèle a été introduit par *Friedman et Stuetzle (1981)* et *Friedman et Tukey (1974)*. Ce modèle, comme ceux définis aux points 1 et 2, ne faisant intervenir que des fonctions univariées, échappe au *fléau de la dimension*.

5. Le modèle dit à *directions révélatrices multiples*, noté MIM, qui englobe la plupart des modèles précédents admet pour fonction de régression une fonction de la forme

$$g(\mathbf{x}) = g_{k+1}(\mathbf{x}^\top \varphi_0, \dots, \mathbf{x}^\top \varphi_k),$$

où k est un entier évidemment supposé relativement petit devant la dimension, d , du modèle.

Fan et Gijbels (1996) complètent cette liste avec des modèles du type additif faisant apparaître aussi des interactions entre variables explicatives.

Nous nous sommes intéressés plus particulièrement aux modèles (2) et (1). Les contributions apportées sur chacun de ces modèles constituent les deux parties de la thèse, discutées ci-dessous

Première partie

Cette partie est consacrée à l'étude d'un modèle partiellement linéaire, pour lequel on ne connaît pas a priori la "structure". On suppose que

$$y = g(\mathbf{x}) + \varepsilon, \quad g(\mathbf{x}) = \theta^\top \mathbf{x}_1 + g_1(\mathbf{x}_2), \quad (3)$$

où $\mathbf{x}^\top = (\mathbf{x}_1^\top, \mathbf{x}_2^\top)$, $\dim(\mathbf{x}_2) = M$, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. L'objectif principal sera d'identifier les composantes apparaissant dans le modèle de façon non linéaire et constituant le vecteur \mathbf{x}_2 . On note \mathcal{J} l'ensemble des indices correspondant à cette partie de cardinal M . L'approche utilisée ici est basée sur une propriété du gradient de g .

Dans un premier chapitre, nous allons définir une *mesure de linéarité* $\mathbf{v}^* = (v_1^*, \dots, v_d^*)$ pour chacune des d composantes de X_i . Elle est construite à l'aide du gradient de la fonction de régression g , elle est nulle pour les composantes linéaires et positive pour les autres. Nous allons l'estimer en définissant un algorithme itératif, qui estime simultanément \mathbf{v}^* et $(g, \nabla g)$:

- On commence par estimer g et son gradient par la méthode des moindres carrés locaux en utilisant une largeur de fenêtre usuelle dans l'estimation fonctionnelle. Nous en déduisons un estimateur pour \mathbf{v}^* .
- L'étape suivante consiste à se servir de cette première estimation pour ajuster au mieux une fenêtre suivant chaque composante de \mathbf{x} . On obtient donc de meilleurs estimateurs de $(g, \nabla g)$ et \mathbf{v}^* . On itère les étapes précédentes.
- La dernière étape, k_n , nous donne les estimateurs de $(g, \nabla g)$ et \mathbf{v}^* .

Le deuxième chapitre donne la convergence théorique de nos estimateurs vers les vraies valeurs. L'un des résultats principaux est la convergence, à la vitesse $n^{1/2}$, de chaque \hat{v}_m , $m \notin \mathcal{J}$, vers une combinaison de variables aléatoires gaussiennes indépendantes.

Dans le chapitre 3, disposant de ces estimateurs de $(g, \nabla g)$ et \mathbf{v}^* , nous pouvons les utiliser pour explorer la structure du modèle. La première conséquence du théorème de convergence des estimateurs est une distance théorique entre composantes linéaires et non linéaire de l'ordre de $(n^{-1} \log(n))^{2/3}$. Ensuite nous développons un test de l'hypothèse $\mathcal{H}_0 = \mathcal{H}_{\mathcal{M}} : M \leq \mathcal{M}$. Au vu des résultats du chapitre 2, on ordonne les estimations \hat{v}_m par ordre décroissant et on décide de rejeter l'hypothèse, si $\hat{v}_{(\mathcal{M}+1)}$ est suffisamment positif. Le seuil sera déterminé en utilisant une méthode de Bootstrap : on rééchantillonne suivant un modèle imitant le modèle sous $\mathcal{H}_{\mathcal{M}}$. Cette procédure est à la base de l'estimation des quantités \mathcal{J} et M . On utilise là encore une procédure itérative, on teste les hypothèses

\mathcal{H}_M en faisant croître M . La première valeur de M , pour laquelle l'hypothèse n'est pas rejetée correspond à l'estimation de M . De plus les M plus grandes composantes du vecteur $\widehat{\mathbf{v}}$ correspondent alors aux composantes non linéaires. On obtient un majorant de la probabilité de mauvaise identification de la partie non linéaire. Dans un dernier point, on étudie un estimateur de la partie linéaire dont la convergence a lieu à la vitesse $n^{1/2}$. On peut alors en déduire un estimateur de g_1 en utilisant \widehat{g} et θ .

Le dernier chapitre contient une mise en oeuvre pratique de ces différents algorithmes sur des données simulées, puis sur des données réelles. Celles-ci confirment la bonne performance de l'algorithme principal et la bonne détermination des composantes linéaires et non linéaires.

Deuxième partie

Dans cette deuxième partie, on s'intéresse au modèle à direction révélatrice unique (1). Cette partie apporte deux contributions à l'étude de ce modèle : nous définissons d'abord un nouvel estimateur de l'axe $\varphi_0 \in \Theta$ du modèle lorsque l'hypothèse :

$$\mathcal{H}_0 : g(x) = g_1(\varphi_0^\top \mathbf{x})$$

est vérifiée. Puis nous construisons un test de cette hypothèse.

Il est clair que sous \mathcal{H}_0 le problème du fléau de la dimension est résolu : si l'on se place sous \mathcal{H}_0 et que l'on sait estimer φ_0 , alors on peut aussi estimer g_1 en utilisant un estimateur non-paramétrique classique *unidimensionnel*.

Notre étude se base sur une caractérisation géométrique du modèle, qui fait l'objet du chapitre 5 et peut s'énoncer de la façon suivante : sous \mathcal{H}_0 l'intégrale moyenne de g sur les cylindres \mathcal{S} d'axe θ (la définition précise sera donnée par la suite), notée $I_{\mathcal{S}}$ est invariante par changement de cylindre congruent si et seulement si θ est orthogonal à l'axe du modèle φ_0 .

On définit alors une mesure de la variabilité de $I_{\mathcal{S}}$ suivant les cylindres d'axe orthogonal à une direction quelconque, φ , notée $W(\varphi)$. Dans le cas où (1) est vérifiée, on a $W(\varphi_0) = 0$. Finalement on donne un estimateur pour chacune des quantités $I_{\mathcal{S}}$, W et φ_0 , basés sur :

$$\widehat{I}_{\mathcal{S}} = (n\|\mathcal{S}\|_d)^{-1} \sum_{i=1}^n Y_i \mathbb{I}_{(X_i \in \mathcal{S})} [\widehat{f}^{(i)}(X_i)]^{-1},$$

où $\widehat{f}^{(i)}(X_i) = \frac{1}{nb^d} \sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{X_j - X_i}{b}\right)$ et b est la largeur de fenêtre. On obtient \widehat{W} en rem-

plaçant $I_{\mathcal{S}}$ par $\widehat{I}_{\mathcal{S}}$ dans l'expression de W , alors

$$\widehat{\varphi} = \operatorname{argmin}_{\varphi \in \Theta} \widehat{W}(\varphi).$$

On teste l'hypothèse \mathcal{H}_0 à l'aide de la statistique $\widehat{W}(\widehat{\varphi})$.

Dans les chapitres 7 et 8, on établit la convergence des différentes statistiques : \widehat{I}_S , $\widehat{\varphi}$ et $\widehat{W}(\widehat{\varphi})$.

Le dernier chapitre donne les résultats relatifs au test de \mathcal{H}_0 . Les quantiles de la loi limite de $\widehat{W}(\widehat{\varphi})$ étant relativement inaccessibles, on décide de déterminer le seuil de rejet du test à l'aide d'une procédure Bootstrap. On donnera ensuite les propriétés du test : niveau asymptotique, puissance, puissance sous une suite d'hypothèses alternatives locales, c'est-à-dire

$$\lim_{n \rightarrow +\infty} \mathbb{P}(\mathcal{H}_0 \text{ est rejetée} / \mathcal{H}_0) = \alpha,$$

$$\lim_{n \rightarrow +\infty} \mathbb{P}(\mathcal{H}_0 \text{ est rejetée} / \mathcal{H}_1) = 1,$$

$$\lim_{|c| \rightarrow +\infty} \lim_{n \rightarrow +\infty} \mathbb{P}(\mathcal{H}_0 \text{ est rejetée} / \mathcal{H}_{1,n,c}) = 1,$$

où α est le niveau du test, \mathcal{H}_1 est l'hypothèse “ g n'est pas un SIM” et $\mathcal{H}_{1,n,c}$ l'hypothèse :

$$g(\mathbf{x}) = g_1(\varphi_0^\top \mathbf{x}) + c n^{-1/2} g_2(\mathbf{x}),$$

où g_2 n'est pas un SIM d'axe φ_0 et c un nombre réel.

Première partie

Étude d'un modèle de régression par une méthode adaptative

Introduction

On dispose d'un échantillon $(X_i, Y_i)_{i=1, \dots, n}$, où $X_i \in \mathbb{R}^d$ et $Y_i \in \mathbb{R}$, généré par le modèle suivant

$$y = g(\mathbf{x}) + \varepsilon, \quad g(\mathbf{x}) = \theta^\top \mathbf{x}_1 + g_1(\mathbf{x}_2), \quad (4)$$

où $\mathbf{x}^\top = (\mathbf{x}_1^\top, \mathbf{x}_2^\top)$, $\dim(\mathbf{x}_2) = M$, $\dim(\mathbf{x}_1) = d - M$. La partie non paramétrique $g_1(\cdot)$ de la fonction de régression, le paramètre θ , et la loi de bruit ε sont inconnus. Nous ne connaissons pas non plus les d_1 variables correspondant à la partie linéaire \mathbf{x}_1 . Nous proposons dans ce chapitre une méthode d'étude d'un tel modèle dit *partiellement linéaire*. Le but est donc d'identifier les régresseurs non-linéaires, d'estimer les coefficients des régresseurs linéaires et la fonction de lien g_1 , pour les régresseurs non-linéaires. La méthode développée ici permet aussi de sélectionner les régresseurs les plus "significativement" non-linéaires. Nous utiliserons ces résultats pour construire un test sur la dimension de la partie non-linéaire. On peut remarquer que la méthode est entièrement adaptative, puisque nous ne connaissons pas la structure précise du modèle. Elle offre de l'intérêt si $M \ll d$, puisqu'on pourra alors estimer g_1 en évitant le problème dit de *fléau de la dimension*.

En outre, on suppose que les données $(X_i, Y_i)_{i=1, \dots, n}$ générées par le modèle (4) vérifient :

- les X_i sont des points non aléatoires, mais on leur impose une répartition suffisamment régulière ;
- les ε_i sont des variables aléatoires indépendantes identiquement distribuées de loi normale centrée et de variance σ^2 .

Les études existantes des modèles du type (4), comme dans *Engle, Granger, Rice et Weiss (1986)* concernent généralement le problème d'estimation des parties paramétrique et non paramétrique, quand les vecteurs \mathbf{x}_1 et \mathbf{x}_2 sont spécifiés. L'ouvrage récent de *Härdle, Liang et Gao (2000)* donne de plus amples références et applications. Les différents estimateurs obtenus convergent aux vitesses classiques \sqrt{n} pour la partie linéaire θ et $\sqrt{nh_n^M}$ pour la partie non paramétrique.

Actuellement nous ne connaissons qu'un seul article concernant la sélection des variables non-linéaires dans un modèle défini par (4), *partiellement linéaire* : celui de *Chen et Chen (1991)*. Dans celui-ci, les variables "splines" sont sélectionnées en utilisant une méthode basée sur l'estimation sans biais du risque, c'est-à-dire de la somme des carrés des résidus pour lesquels ils approchent $g(\mathbf{x}_2)$ à l'aide de fonctions splines. Ils obtiennent une probabi-

lité de bonne classification des régresseurs non-linéaires tendant vers 1 quand n tend vers l'infini.

Le même type de problème est abordé dans *Härdle et Korostelev (1996)*, dans le cas particulier d'un modèle additif. On y montre que l'erreur de classification d'une variable peut être rendue "exponentiellement petite". *Shively, Kohn et Wood (1999)* se sont aussi intéressés au problème de la sélection de variables et d'estimation de fonction dans les modèles additifs; ils utilisent pour cela une approche bayésienne empirique.

Härdle, Sperlich et Spokoiny (2001) proposent enfin une méthode pour identifier les composantes linéaires dans un modèle additif en utilisant le développement de toutes les composantes additives sur les bases d'ondelettes de Haar. L'avantage de notre méthode est de ne pas avoir besoin de supposer de structure additive pour le modèle. De plus notre méthode d'identification des composantes garantit un niveau fixé d'erreur de spécification uniformément sur une classe de modèles, dont les composantes linéaires et non-linéaires sont distantes de $O((n^{-1} \log n)^{4/3})$. On peut noter aussi que cette méthode donne de bons résultats pour des échantillons de taille usuelle.

Les techniques d'estimation utilisées s'inspirent de celles apparues dans les travaux de *Hristache, Juditsky et Spokoiny (2001)* et *Hristache, Juditsky, Polzehl et Spokoiny (2001)* (que nous noterons respectivement HJS et HJPS) concernant le problème de réduction de la dimension.

Dans HJPS, on considère le problème d'estimation dans un modèle de régression à *directions révélatrices multiples*. La fonction de régression est alors de la forme $g(x) = g_1(\theta_1^\top \mathbf{x}, \dots, \theta_p^\top \mathbf{x})$, où $\theta_1, \dots, \theta_p$ sont des vecteurs de \mathbb{R}^d inconnus. Le modèle (4) apparaît comme un cas particulier de ces modèles vérifiant : $p = M + 1$, $\theta_1 = (\theta^\top, 0_M^\top)$ et pour $i = 2, \dots, p$, θ_i est un vecteur de la base canonique de \mathbb{R}^d , tel que $(\theta_2^\top \mathbf{x}, \dots, \theta_p^\top \mathbf{x}) = \mathbf{x}_2$. Il serait donc possible d'utiliser ici la méthode proposée par HJPS, mais on ne prendrait pas en compte la structure très spécifique du modèle (4). Nous proposons donc une méthode alternative, originale à notre connaissance, adaptée au modèle (4).

Cette partie s'organise de la façon suivante. La première partie sera consacrée à la caractérisation des composantes non-linéaires, à la définition de la procédure itérative d'estimation de g , de \mathbf{x}_1 et \mathbf{x}_2 . Nous étudierons ensuite la convergence de l'algorithme itératif utilisé et la qualité finale de l'estimation. La troisième partie s'attachera à l'exploration du modèle, à savoir la séparation des composantes linéaires et non-linéaires, le test visant à l'identification des composantes non-linéaires, l'estimation de θ . Le test présenté dans cette partie est dans l'esprit de celui de *Spokoiny (2001)*, qui essaie de trouver la distance minimale entre l'hypothèse nulle et l'hypothèse alternative. La dernière partie illustre les résultats obtenus sur des exemples simulés et sur des données réelles.

Chapitre 1

Détermination des composantes linéaires

L'idée de la méthode que nous allons utiliser peut se résumer par les deux remarques suivantes :

- Si la fonction de régression était correctement estimée, il serait facile de séparer les parties \mathbf{x}_1 et \mathbf{x}_2 .
- Si les vecteurs \mathbf{x}_1 et \mathbf{x}_2 étaient connus, on pourrait alors estimer plus précisément la fonction de régression.

Ceci amène à construire une procédure itérative, qui permette d'estimer g et les parties linéaire et non-linéaire. Tout d'abord on estime g non paramétriquement, ensuite on se sert de cette estimation pour approcher la structure du modèle, c'est-à-dire reconnaître \mathbf{x}_1 et \mathbf{x}_2 . On utilise alors cette estimation de la structure pour améliorer l'estimateur de la fonction de régression. Et ainsi de suite... De cette façon, progressivement, on améliore la connaissance de la structure du modèle et l'estimation de g .

1.1 Comment reconnaître une composante linéaire ?

Le modèle (4) est essentiellement décrit par ses composantes non-linéaires, notons \mathcal{J} l'ensemble des indices correspondants. On peut aussi définir cet ensemble en utilisant le gradient de la fonction de régression, g . En effet, si $g(\cdot)$ est linéaire en la $m^{\text{ème}}$ composante, alors la dérivée partielle de g en x_m , $\frac{\partial g}{\partial x_m}$, est constante (pour tout $m \in \{1, \dots, d\}$, x_m est la $m^{\text{ème}}$ composante du vecteur \mathbf{x} de \mathbb{R}^d). Pour “capturer” cette propriété, on pourrait utiliser un estimateur de la quantité $n^{-1} \sum_{i=1}^n \left(\frac{\partial g}{\partial x_m}(X_i) - n^{-1} \sum_{j=1}^n \frac{\partial g}{\partial x_m}(X_j) \right)^2$, ce qui ferait intervenir des termes quadratiques, dont la convergence est difficile à obtenir. Pour éviter ce problème et mesurer le degré de linéarité de g , nous allons donc utiliser une méthode analogue à celle de HJPS, en introduisant :

$\{\psi_{1m}, \dots, \psi_{Lm}\}$ une famille de fonctions de \mathbb{R} dans \mathbb{R} vérifiant

$$\sum_{i=1}^n \psi_{lm}(X_{i,m}) = 0, \quad n^{-1} \sum_{i=1}^n \psi_{lm}(X_{i,m}) \psi_{l'm}(X_{i,m}) = \delta_{ll'}$$

pour $l, l' = 1, \dots, L$. Si on note $\boldsymbol{\psi}_{lm} = (\psi_{lm}(X_{i,m}))_{i=1, \dots, n}$, alors les $\boldsymbol{\psi}_{lm}$ sont des vecteurs de \mathbb{R}^n . On introduit le produit scalaire sur $\mathbb{R}^n : \langle u, v \rangle = n^{-1} \sum_{i=1}^n u_i v_i$. Par les deux relations ci-dessus, on voit la famille de vecteurs de $\mathbb{R}^n \{\boldsymbol{\psi}_{lm}, l = 1, \dots, L\}$ est orthonormale, et orthogonale au vecteur $\mathbb{1} = (1, \dots, 1)^\top$, donc à tout vecteur constant. Pour construire une telle famille, on peut par exemple orthonormaliser $(1, X_{i,m}, X_{i,m}^2, \dots, X_{i,m}^L)_{i=1, \dots, n}$. Ces propriétés nous permettent d'affirmer que pour $m \notin \mathcal{J}$

$$\beta_{lm}^* = n^{-1} \sum_{i=1}^n \frac{\partial g}{\partial x_m}(X_i) \psi_{lm}(X_{i,m}) \equiv 0, \quad (1.1)$$

pour $l = 1, \dots, L$. Et la somme

$$v_m^* = \sum_{l=1}^L (\beta_{lm}^*)^2$$

peut être utilisée pour mesurer la non-linéarité de g en x_m .

Si $\widehat{\frac{\partial g}{\partial x_m}}$ est un estimateur de la dérivée partielle de g par rapport à x_m (il sera défini par la suite), nous estimons β_{lm}^* par

$$\widehat{\beta}_{lm} = n^{-1} \sum_{i=1}^n \widehat{\frac{\partial g}{\partial x_m}}(X_i) \psi_{lm}(X_{i,m}). \quad (1.2)$$

Un indicateur de la non-linéarité de g par rapport à la $m^{\text{ème}}$ composante de \mathbf{x} est alors

$$\widehat{v}_m = \widehat{\beta}_{1m}^2 + \dots + \widehat{\beta}_{Lm}^2.$$

Cette quantité nous donne des informations sur la structure du modèle et nous allons l'utiliser pour améliorer l'estimation de g et de son gradient, ∇g .

REMARQUE : CHOIX DE L

Ainsi définie, la famille de fonctions $(\psi_{l,m})_{l=1, \dots, L}$, étant une famille orthonormale pour le produit scalaire $\langle \cdot, \cdot \rangle$, elle engendre un sous-espace vectoriel, Ψ_L , de dimension L dans \mathbb{R}^n . Ce sous-espace est de plus orthogonal à l'espace engendré par $(1, \dots, 1)^\top$. La quantité v_m^* représente donc le carré de la norme de la projection du vecteur $(\frac{\partial g}{\partial x_m}(X_i))_{i=1, \dots, n}$ sur ce sous-espace Ψ_L . On perd donc de l'information en considérant cette projection, ce qui serait évité en choisissant un grand nombre de vecteurs, par exemple $L = L_n = n - 1$. En fait on peut montrer par des arguments de simulation, qu'on ne perd pas beaucoup d'information sur le caractère constant ou non de $\frac{\partial g}{\partial x_m}$ en utilisant un nombre fixé L de vecteurs. Par contre il faudra choisir L assez grand pour avoir une bonne estimation.

1.2 Une estimation “améliorée” de g

Supposons que nous disposions de $(v_m^*)_{m=1,\dots,d}$, défini au paragraphe précédent. Nous allons expliquer de quelle façon cette donnée va permettre d’améliorer la qualité de l’estimation de la fonction g .

Un estimateur naturel de g et de son gradient, ∇g , est obtenu par le critère définissant la méthode dite de “polynôme locaux”, cf. *Fan, Gijbels(1996)* :

$$\begin{aligned} \begin{pmatrix} \widehat{g}(X_i) \\ \widehat{\nabla}g(X_i) \end{pmatrix} &= \underset{a \in \mathbb{R}, c \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{j=1}^n \left[Y_j - a - c^\top (X_j - X_i) \right]^2 K \left(\frac{|X_j - X_i|^2}{b^2} \right), \\ &= \left\{ \sum_{j=1}^n \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix}^\top K \left(\frac{|X_{ij}|^2}{b^2} \right) \right\}^{-1} \sum_{j=1}^n Y_j \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} K \left(\frac{|X_{ij}|^2}{b^2} \right), \end{aligned}$$

où $X_{ij} = X_j - X_i$, $|\cdot|$ est la norme euclidienne de \mathbb{R}^d , b est la largeur de fenêtre et K est un noyau positif univarié à support sur $[0, 1]$. Ainsi seuls les points de l’échantillon se trouvant dans la boule de centre X_i et de rayon b interviennent, c’est-à-dire qu’on ne considère que des voisinages sphériques de même rayon autour de chaque point.

La difficulté de ce type d’estimateur est la sélection de b . En effet, il faut la choisir telle que la boule de centre X_i et de rayon b contienne suffisamment d’observations X_j . Si d est élevé cela peut amener à choisir b très grand, cf. *Fan et Gijbels (1996)*. Ce phénomène est appelé *le fléau de la dimension*. Or, dans notre cas, g a des propriétés de régularité anisotropes, c’est-à-dire que pour les composantes linéaires elle a un ordre de dérivation infini et pour les autres un ordre a priori plus faible. Nous allons donc utiliser cette remarque en considérant une largeur de fenêtre anisotrope adaptée à la structure du modèle contenue dans \mathcal{J} . Cette façon de procéder a été introduite, entre autres, par *Carroll, Fan, Gijbels et Wand (1997)*. On obtient alors l’estimateur suivant, que nous utiliserons désormais :

$$\begin{pmatrix} \widehat{g}(X_i) \\ \widehat{\nabla}g(X_i) \end{pmatrix} = \left\{ \sum_{j=1}^n \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix}^\top K_d(X_{ij}, \mathbf{b}) \right\}^{-1} \sum_{j=1}^n Y_j \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} K_d(X_{ij}, \mathbf{b}),$$

où $K_d(\mathbf{x}, \mathbf{b}) = K(|\mathbf{b}^{-1}\mathbf{x}|^2)$ et $\mathbf{b} = \operatorname{diag}(b_1, \dots, b_d)$ est une matrice diagonale d’entrées diagonales b_1, \dots, b_d . Connaissant v_m^* , on va choisir les valeurs de b_m de la façon suivante : b_m sera choisie petite, si v_m^* est suffisamment positif, c’est-à-dire $m \in \mathcal{J}$; sinon elle sera plus grande. Ainsi dans le cas non-linéaire en choisissant une fenêtre petite, on prend mieux en compte les variations locales de la courbe de régression.

1.3 L’algorithme itératif

L’algorithme défini ci-dessous est basé sur les raisonnements exposés dans les deux paragraphes précédents. Il va nous permettre d’obtenir simultanément des estimateurs de g ,

∇g et de \mathbf{v}^* . A partir de ces valeurs, \widehat{g} , $\widehat{\nabla g}$, $\widehat{\mathbf{v}}$, on pourra estimer le modèle qui a généré les données, cf. *chapitre 3*.

Pour définir cet algorithme nous avons besoin de quelques paramètres d'entrée : h_1 , h_{\max} , ρ_1 , ρ_{\min} que l'on choisit tels que : $h_1 < h_{\max}$, $\rho_{\min} < \rho_1$, au fur et à mesure des itérations la suite ρ décroîtra de façon géométrique de ρ_1 à ρ_{\min} avec une raison $a_\rho < 1$ et h croîtra de façon géométrique de h_1 à h_{\max} , avec une raison $a_h > 1$. Le choix de ces paramètres, ainsi que des fonctions $\{\psi_{lm}\}$, sera discuté ultérieurement, cf. *paragraphe 1.5*. L'algorithme est le suivant

1. Fixer h_1 . On pose $\rho_1 = 1$ et pour $m = 1, \dots, d$, $\widehat{v}_m^{(0)} = 0$ et $b_m^{(1)} = h_1$. Soit $k = 1$.

2. Calculer

$$b_m^{(k)} = h_k \left(1 + \rho_k^{-2} \widehat{v}_m^{(k-1)} \right)^{-1/2}, \quad m = 1, \dots, d. \quad (1.3)$$

Définir $\mathbf{b}^{(k)} = \text{diag}(b_1^{(k)}, \dots, b_d^{(k)})$.

3. Pour tout X_i calculer

$$V_i^{(k)} = \sum_{j=1}^n \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix}^\top K_d(X_{ij}, \mathbf{b}^{(k)}), \quad (1.4)$$

$$\begin{pmatrix} \widehat{g}^{(k)}(X_i) \\ \widehat{\nabla g}^{(k)}(X_i) \end{pmatrix} = \left(V_i^{(k)} \right)^{-1} \sum_{j=1}^n \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} Y_j K_d(X_{ij}, \mathbf{b}^{(k)}), \quad (1.5)$$

où $X_{ij} = X_j - X_i$.

Calculer $\widehat{\varepsilon}_i^{(k)} = Y_i - \widehat{g}^{(k)}(X_i)$.

4. Pour $m = 1, \dots, d$ et $l = 1, \dots, L$, calculer

$$\widehat{\beta}_{lm}^{(k)} = n^{-1} \sum_{i=1}^n \frac{\widehat{\partial g}^{(k)}}{\partial x_m}(X_i) \psi_{lm}(X_{i,m}), \quad \widehat{v}_m^{(k)} = \sum_{l=1}^L \left| \widehat{\beta}_{lm}^{(k)} \right|^2. \quad (1.6)$$

5. Augmenter k de 1. Alors $\rho_{k+1} = a_\rho \rho_k$, $h_{k+1} = a_h h_k$. Si $\rho_{k+1} \geq \rho_{\min}$, alors poser $k = k + 1$ et continuer en revenant à l'étape **2**; sinon terminer.

Le nombre d'itérations effectuées est noté k_n . On définit alors, en utilisant les résultats de la dernière itération, les estimateurs, que l'on utilisera dans le chapitre 3 :

- $\widehat{\beta}_{lm} = \widehat{\beta}_{lm}^{(k_n)}$ (resp. $\widehat{v}_m = \widehat{v}_m^{(k_n)}$) sera l'estimateur de β_{lm}^* (resp. v_m^*),
- $\widehat{g}(X_i)$ sera l'estimateur de $g(X_i)$, à partir duquel on peut calculer l'estimateur de ε_i :
 $\widehat{\varepsilon}_i = \widehat{\varepsilon}_i^{(k_n)} = Y_i - \widehat{g}(X_i)$.

REMARQUE : A chaque itération la valeur de $b_m^{(k)}$ est comprise entre h_k et $\rho_k h_k$ et dépend de la valeur de \widehat{v}_m obtenue à l'itération précédente, $\widehat{v}_m^{(k-1)}$. Si $m \notin \mathcal{J}$, alors $\widehat{v}_m^{(k-1)}$ est petit et donc $b_m^{(k)}$ est de l'ordre de h_k . Si $m \in \mathcal{J}$ et que la valeur de v_m^* est grande, alors celle de $\widehat{v}_m^{(k-1)}$ l'est aussi et $b_m^{(k)}$ sera alors de l'ordre de $h_k \rho_k$. Ainsi comme h_k croît et $h_k \rho_k$ décroît, on obtient bien à la dernière itération une largeur de fenêtre anisotrope.

1.4 Estimation de la variance du bruit

La variance du bruit σ^2 n'intervient pas a priori dans la définition de la procédure. Cependant nous l'utiliserons dans la procédure de rééchantillonnage utilisée pour le test sur M et il est aussi possible de l'utiliser pour définir la condition d'arrêt de l'algorithme. Rappelons que les données $(X_i, Y_i)_{i=1, \dots, n}$ ont été générées par un modèle de la forme $Y_i = g(X_i) + \varepsilon_i$, où $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. On veut estimer σ^2 à l'aide de l'algorithme défini précédemment. Dans notre cas, une façon naturelle de définir l'estimateur de la variance à chaque étape k est :

$$|\tilde{\sigma}^{(k)}|^2 = n^{-1} \sum_{i=1}^n |\hat{\varepsilon}_i^{(k)}|^2,$$

où $\hat{\varepsilon}_i^{(k)}$ est défini à l'étape 3 de l'algorithme. On peut améliorer cet estimateur – voir par exemple *Gasser, Sroka and Jennen-Steinmetz (1986)* ou *Spokoiny (2002)* – en utilisant l'écriture des résidus $\hat{\varepsilon}_i^{(k)}$ sous la forme $\hat{\varepsilon}_i^{(k)} = \sum_{j=1}^n c_{i,j}^{(k)} Y_j$ où les $c_{i,j}^{(k)}$ sont des coefficients connus. Dans notre cas, on a

$$\begin{aligned} \hat{\varepsilon}_i^{(k)} &= Y_i - \hat{g}^{(k)}(X_i), \\ &= Y_i - \begin{pmatrix} 1 \\ 0_d \end{pmatrix}^\top \left(V_i^{(k)} \right)^{-1} \sum_{j=1}^n \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} K_d(X_{ij}, \mathbf{b}^{(k)}) Y_j, \\ &= \sum_{j=1}^n c_{i,j}^{(k)} Y_j, \end{aligned}$$

avec $c_{i,j}^{(k)} = \delta_{ij} - \begin{pmatrix} 1 \\ 0_d \end{pmatrix}^\top \left(V_i^{(k)} \right)^{-1} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} K_d(X_{ij}, \mathbf{b}^{(k)})$. Ces coefficients sont aléatoires, car ils dépendent de Y_i à travers la largeur de fenêtre aléatoire $\hat{b}_m^{(k)}$. Si les Y_i formaient un bruit blanc, un estimateur de la forme $\sum_{j=1}^n c_{i,j} Y_j$, où les $c_{i,j}$ sont déterministes, vérifierait

$$\mathbb{E} \left| \sum_{j=1}^n c_{i,j} Y_j \right|^2 = \sigma^2 \sum_{j=1}^n |c_{i,j}|^2.$$

On en déduirait un estimateur de σ^2 par :

$$|\hat{\sigma}^{(k)}|^2 = \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^n |c_{i,j}|^2 \right)^{-1} \left| \sum_{j=1}^n c_{i,j} Y_j \right|^2.$$

Par analogie, l'estimateur que l'on considèrera dans la suite est

$$|\hat{\sigma}^{(k)}|^2 = \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^n |c_{i,j}^{(k)}|^2 \right)^{-1} |\hat{\varepsilon}_i^{(k)}|^2. \quad (1.7)$$

Les propriétés asymptotiques de $\widehat{\sigma}^2$ sont données dans le paragraphe 2.5.

1.5 Choix des paramètres

La qualité de l'estimation obtenue par la procédure itérative dépend fortement du choix des raisons, a_ρ et a_h , des deux suites h_k et ρ_k et des valeurs de ces suites à la dernière itération. h_k croît de h_1 à h_{\max} , simultanément ρ_k décroît de ρ_1 à ρ_{\min} .

La valeur de h_1 doit être bien choisie pour que la première estimation du gradient ne soit pas trop mauvaise. Une condition nécessaire pour obtenir cette condition est : la boule $\mathcal{B}_h(X_i) = \{x : |X_i - x| < h\}$ de centre X_i et de rayon h , doit contenir au moins $(d + 1)$ points. Ici on choisit comme h_1 initial une valeur telle que

$$\#\{i : N_h(X_i) \geq d + 1\} \geq n/2,$$

où $N_h(X_i)$ est le nombre de points appartenant à $\mathcal{B}_h(X_i)$. On commence par $h_{1,0} = (n/d)^{-d}$ avec une raison géométrique $a_{h_1} = e^{1/20}$.

En accord avec les théorèmes de la partie suivante, on choisit h_{\max} de l'ordre de 1 et ρ_{\min} le plus petit possible c'est-à-dire de l'ordre de $n^{-1/3}$.

Les voisinages que l'on considère sont elliptiques et de la forme

$$\mathcal{E}^{(k)}(X_i) = \left\{ x : |\mathbf{b}^{(k)-1}(X_i - x)|^2 \leq 1 \right\} = \left\{ x : \sum_{m=1}^d \frac{(X_{im} - x_m)^2}{b_m^{(k)2}} \leq 1 \right\}.$$

si $m \notin \mathcal{J} : b_m^{(k+1)} \sim a_h h_k = a_h b_m^{(k)}$.

si $m \in \mathcal{J} : b_m^{(k+1)} \sim a_h h_k \rho_k a_\rho = a_h a_\rho b_m^{(k)}$.

Ainsi la mesure de Lebesgue de $\mathcal{E}^{(k+1)}$ est de l'ordre de $a_h^d a_\rho^M$ fois celle de $\mathcal{E}^{(k)}$. On choisit donc $a_h^d a_\rho^M \geq 1$. Si les X_i sont aléatoires, on obtient un nombre croissant de points à l'intérieur du voisinage elliptique.

Chapitre 2

Qualité des estimateurs

Le but de ce chapitre est d'évaluer la qualité des estimations des quantités β_m^* et v_m^* , obtenues à l'aide de la procédure précédente. Dans un premier temps, on présente les hypothèses sur le modèle et sur les paramètres de l'algorithme. Ensuite nous quantifions, d'un point de vue théorique, l'amélioration entre deux itérations successives lors de la procédure itérative introduite. Finalement on donne la précision de l'estimateur final de β_m^* .

2.1 Hypothèses

Nous allons tout d'abord faire la même transformation que dans HJPS, en renormalisant la fonction de lien. Le modèle (4) n'est pas identifiable a priori : la définition de g_1 n'est pas unique. En effet, on peut aussi écrire $g(\mathbf{x}) = \theta^\top \mathbf{x}_1 + g_2(O_M^\top \mathbf{x}_2)$, où $g_2(\mathbf{z}) = g_1(O_M \mathbf{z})$, c'est-à-dire qu'on obtient la même fonction après rotation. Par contre, si on fixe g_1 de la forme $g_1(\mathbf{x}) = \tilde{g}_1(R^* \mathbf{x}_2)$, où R^* est une matrice diagonale $M \times M$ de termes diagonaux $\sqrt{v_m^*}$ ($m \in \mathcal{J}$) alors \tilde{g}_1 est unique : $\tilde{g}_1(\mathbf{z}_2) = g(\mathbf{z}_1, R^{*-1} \mathbf{z}_2) - \theta^\top \mathbf{z}_1$. On notera encore g_1 la fonction de lien non-linéaire par soucis de simplification. Ceci nous amène à considérer le modèle

$$y = g(\mathbf{x}) + \varepsilon, \quad \text{où } g(\mathbf{x}) = \mathbf{x}_1^\top \theta + g_1(\mathcal{R}^* \mathbf{x}_2). \quad (2.1)$$

Introduisons alors les hypothèses fondamentales, qui assurent la convergence des estimateurs.

Hypothèse H1. [Noyau] *Le noyau $K(\cdot)$ est continûment différentiable décroissant sur \mathbb{R}^+ avec $K(0) = 1$ et $K(x) = 0$ pour $|x| \geq 1$.*

Hypothèse H2. [Erreurs] *Les variables aléatoires ε_i dans (4) sont indépendantes et normalement distribuées, de moyenne nulle et de variance σ^2 .*

Hypothèse H3. [Paramètres de l'algorithme] Les paramètres de la procédure itérative sont choisis tels que : $\rho_1 = 1$, $\rho_{\min} = (\sigma^2 n^{-1} L \log n)^{1/3}$, $h_1 = C_0 n^{-\frac{1}{4\sqrt{d}}}$ avec des constantes $C_0 \geq 1$, $h_{\max} \geq 1$ et $a_h^d a_\rho^M \geq 1$.

Hypothèse H4. [Fonction de lien] La fonction g_1 de (2.1) est deux fois différentiable avec une dérivée seconde bornée telle qu'il existe une constante C_{g_1} et que pour tout $u, v \in \mathbb{R}^M$

$$|g_1(v) - g_1(u) - (v - u) \nabla g_1(u)| \leq C_{g_1} |u - v|^2.$$

La dernière hypothèse concerne les propriétés de répartition des variables explicatives. La réussite de la procédure est liée à la bonne estimation du gradient. Il faut qu'à chaque itération les estimateurs calculés à partir des largeurs de fenêtre $(b_1^{(k)}, \dots, b_d^{(k)})$ soient bien définis. Cette dernière hypothèse sur la bonne répartition des points X_1, \dots, X_n contient donc une condition d'inversibilité pour toutes les matrices $V_i^{(k)}$ définies par (1.4).

Introduisons quelques notations utiles par la suite : on définit pour la $k^{\text{ème}}$ itération la matrice diagonale des largeurs de fenêtre idéales et anisotropes, $\mathbf{b}^{*(k)}$, qui a pour termes diagonaux $b_m^{*(k)} = h_k (1 + \rho_k^{-2} v_m^*)^{-1/2}$. La proximité entre $\mathbf{b}^{(k)}$ et sa valeur idéale $\mathbf{b}^{*(k)}$ peut être caractérisée par les valeurs

$$U_m^{(k)} = \left(b_m^{*(k)} / b_m^{(k)} \right)^2 = (1 + \rho_k^{-2} \widehat{v}_m^{(k-1)}) / (1 + \rho_k^{-2} v_m^*), \quad (2.2)$$

pour $m = 1, \dots, d$. En effet si $\widehat{v}_m^{(k-1)} = v_m^*$, alors $U_m^{(k)} = 1$.

Pour chaque itération nous recherchons une répartition régulière des variables explicatives dans un voisinage elliptique de centre X_i défini par $\left\{ \mathbf{x} : \sum_{m=1}^d (x_m - X_{i,m})^2 / b_m^{(k)2} \leq 1 \right\}$.

On définit $Z_{ij}^{(k)} = (\mathbf{b}^{*(k)})^{-1} (X_j - X_i)$ pour $i, j = 1, \dots, n$. Ce vecteur décrit la localisation des points dans le nouveau repère centré en X_i , avec les changements d'échelle donnés par $\mathbf{b}^{*(k)}$.

Pour un vecteur $\mathbf{U} = (U_1, \dots, U_d)^\top \in \mathbb{R}^{+d}$, on notera $D_{\mathbf{U}} = \text{diag}(U_1, \dots, U_d)$. En utilisant les notations ci-dessus, on a alors $\mathbf{b} = \mathbf{b}^{*(k)} D_{\mathbf{U}^{(k)}}^{-1/2}$, où $\mathbf{U}^{(k)} = (U_1^{(k)}, \dots, U_d^{(k)})^\top$. D'après (2.2)

$$\begin{aligned} K_d(X_{i,j}, \mathbf{b}^{(k)}) &= K(X_{i,j}^\top (\mathbf{b}^{(k)})^{-2} X_{i,j}), \\ &= K(X_{i,j}^\top (\mathbf{b}^{*(k)})^{-1} D_{\mathbf{U}^{(k)}}^{1/2} D_{\mathbf{U}^{(k)}}^{1/2} (\mathbf{b}^{*(k)})^{-1} X_{i,j}), \\ &= K(Z_{ij}^\top D_{\mathbf{U}^{(k)}} Z_{ij}). \end{aligned} \quad (2.3)$$

Nous pouvons alors introduire la dernière hypothèse nécessaire à la preuve des théorèmes

de convergence. Pour cela posons, pour un vecteur \mathbf{U} quelconque

$$\begin{aligned} N_i^{(k)}(\mathbf{U}) &= \sum_{j=1}^n K\left(Z_{ij}^{(k)\top} D_{\mathbf{U}} Z_{ij}^{(k)}\right), \quad i = 1, \dots, n, \\ \mathcal{V}_i^{(k)}(\mathbf{U}) &= \sum_{j=1}^n \begin{pmatrix} 1 \\ Z_{ij}^{(k)} \end{pmatrix} \begin{pmatrix} 1 \\ Z_{ij}^{(k)} \end{pmatrix}^\top K\left(Z_{ij}^{(k)\top} D_{\mathbf{U}} Z_{ij}^{(k)}\right), \quad i = 1, \dots, n. \end{aligned}$$

L'hypothèse s'écrit alors

Hypothèse H5. [Design] *Il existe des constantes C_V , C_K , $C_{K'}$ et $\alpha \in (0, 1/2)$, telles que pour tout vecteur $\mathbf{U} \in \mathbb{R}^d$, satisfaisant $|U_m - 1| \leq \alpha$, $m = 1, \dots, d$, et pour $k \leq k_n$, la matrice $\mathcal{V}_i^{(k)}(\mathbf{U})$ est bien inversible et vérifie*

$$N_i^{(k)}(\mathbf{U}) \left\| \mathcal{V}_i^{(k)}(\mathbf{U})^{-1} \right\| \leq C_V, \quad i = 1, \dots, n.$$

Ici $\|A\|$ correspond à la norme matricelle associée à la norme vectorielle euclidienne, i.e. $\|A\| = \sup_{\lambda \neq 0} |A\lambda|/|\lambda|$.

De plus, pour $i, j = 1, \dots, n$,

$$\begin{aligned} \sum_{i=1}^n \frac{1}{N_i^{(k)}(\mathbf{U})} K\left(Z_{ij}^{(k)\top} D_{\mathbf{U}} Z_{ij}^{(k)}\right) &\leq C_K, \\ \sum_{i=1}^n \frac{1}{N_i^{(k)}(\mathbf{U})} \left| K'\left(Z_{ij}^{(k)\top} D_{\mathbf{U}} Z_{ij}^{(k)}\right) \right| &\leq C_{K'}, \\ \sum_{j=1}^n \frac{1}{N_i^{(k)}(\mathbf{U})} \left| K'\left(Z_{ij}^{(k)\top} D_{\mathbf{U}} Z_{ij}^{(k)}\right) \right| &\leq C_{K'}. \end{aligned} \tag{2.4}$$

où K' est la dérivée de K .

REMARQUE . On peut montrer que ces hypothèses sont vérifiées dans le cas où les X_i sont à densité continue strictement positive sur un ensemble de grande probabilité.

2.2 Amélioration en une étape

Cette partie consiste à quantifier l'amélioration de l'estimation à chaque itération. Pour cela, on fixe h , ρ et $\mathbf{v} = (v_1, \dots, v_d)^\top$. Cette dernière quantité peut être vue comme l'approximation de \mathbf{v}^* à l'itération précédente et les quantités h et ρ comme les valeurs correspondant à l'itération en cours.

De façon analogue au paragraphe précédent, on pose

$$b_m = h(1 + \rho^{-2}v_m)^{-1/2}, \quad m = 1, \dots, d, \tag{2.5}$$

et $\mathbf{b} = \text{diag}(b_1, \dots, b_d)$.

On considèrera dans cette partie les estimateurs $\widehat{g}(X_i)$, $\widehat{\nabla}g(X_i)$ et $\widehat{\beta}_{lm}$ donnés par (1.3) et (1.2) en utilisant la matrice de largeurs de fenêtre \mathbf{b} ci-dessus. Le but de cette section est d'évaluer l'erreur $|\widehat{\beta}_{lm} - \beta_{lm}^*|$. On introduit pour cela des notations complémentaires : $P_{\rho,m} = (1 + \rho^{-2}v_m^*)^{-1/2}$ et sa matrice diagonale associée

$$P_\rho = \text{diag}(P_{\rho,1}, \dots, P_{\rho,d}),$$

qui intervient de la façon suivante dans la mesure de proximité entre b_m et b_m^* :

$$U_m = P_{\rho,m}^2(1 + \rho^{-2}v_m) = (1 + \rho^{-2}v_m^*)^{-1}(1 + \rho^{-2}v_m),$$

de façon similaire $U_m^* = P_{\rho,m}^2(1 + \rho^{-2}v_m^*) = 1$.

On peut remarquer que les vecteurs \mathbf{v} et $\widehat{\beta}_m$ donnés par (1.2) sont en bijection avec le vecteur $\mathbf{U} = (U_1, \dots, U_d)^\top \in \mathbb{R}^d$. On fera apparaître cette dépendance en écrivant : $\mathbf{v} = \mathbf{v}(\mathbf{U})$ et $\widehat{\beta}_m = \widehat{\beta}_m(\mathbf{U})$.

La connaissance de l'erreur ponctuelle, i.e. \mathbf{U} étant fixé, ne suffit pas pour connaître l'erreur à la dernière itération. Nous avons besoin d'une majoration uniforme dans un voisinage, \mathcal{U}_α , de la valeur idéale $\mathbf{U}^* = (1, \dots, 1)^\top$. En effet, lors des itérations, la largeur de fenêtre \mathbf{b} dépend des Y_i ; elle est donc aléatoire et le résultat ponctuel ne suffit donc plus. En choisissant convenablement un voisinage \mathcal{U}_α , on pourra utiliser les résultats, uniformes sur les voisinages \mathcal{U}_α , établis dans cette partie pour en déduire l'amélioration finale des estimations, cf. *paragraphe 2.4*. Définissons ces voisinages : soit un vecteur $\alpha = (\alpha_1, \dots, \alpha_d)^\top$ de \mathbb{R}^d dont les coordonnées vérifient $\alpha_m \in (0, 1)$, on pose

$$\mathcal{U}_\alpha = \left\{ \mathbf{U} = (U_1, \dots, U_d)^\top \in \mathbb{R}^d : |U_m - 1| \leq \alpha_m, m = 1, \dots, d \right\}. \quad (2.6)$$

Soit aussi $\alpha^* = \max_{m=1, \dots, d} \alpha_m$.

On rappelle que pour $\mathbf{U} = (U_1, \dots, U_d)^\top \in \mathbb{R}^d$, on note $D_{\mathbf{U}} = \text{diag}\{U_1, \dots, U_d\}$ et que pour tous $i, j = 1, \dots, n$ $Z_{ij} = (\mathbf{b}^*)^{-1}X_{ij} = h^{-1}P_\rho^{-1}(X_j - X_i)$. Pour $i, j = 1, \dots, n$, on note

$$\begin{aligned} \widehat{K}_{ij}(\mathbf{U}) &= K(Z_{ij}^\top D_{\mathbf{U}} Z_{ij}) = K_d(X_{ij}, \mathbf{b}), \\ \mathcal{V}_i(\mathbf{U}) &= \sum_{j=1}^n \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix}^\top K_{ij}(\mathbf{U}), \\ \widehat{s}_i(\mathbf{U}) &= h^{-1} \mathcal{V}_i(\mathbf{U})^{-1} \sum_{j=1}^n \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix} Y_j K_{ij}(\mathbf{U}). \end{aligned}$$

On veut réécrire $\begin{pmatrix} \widehat{g}(X_i) \\ \widehat{\nabla}g(X_i) \end{pmatrix}$ à l'aide de ces quantités.

$$\begin{pmatrix} \widehat{g}(X_i) \\ \widehat{\nabla}g(X_i) \end{pmatrix} = V_i^{-1} \sum_{j=1}^n \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} K_d(X_{ij}, \mathbf{b}) Y_j,$$

or

$$\begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & hP_\rho \end{pmatrix} \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix},$$

donc

$$\begin{aligned} V_i &= \sum_{j=1}^n \begin{pmatrix} 1 & 0 \\ 0 & hP_\rho \end{pmatrix} \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix}^\top \begin{pmatrix} 1 & 0 \\ 0 & hP_\rho \end{pmatrix} K_{ij}(\mathbf{U}), \\ V_i &= \begin{pmatrix} 1 & 0 \\ 0 & hP_\rho \end{pmatrix} \mathcal{V}_i(\mathbf{U}) \begin{pmatrix} 1 & 0 \\ 0 & hP_\rho \end{pmatrix}, \end{aligned}$$

et

$$V_i^{-1} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & h^{-1}P_\rho^{-1} \end{pmatrix} \mathcal{V}_i^{-1}(\mathbf{U}) \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix}.$$

D'où

$$\begin{pmatrix} \widehat{g}(X_i) \\ \widehat{\nabla}g(X_i) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & h^{-1}P_\rho^{-1} \end{pmatrix} h\widehat{s}_i(\mathbf{U}),$$

finalement

$$\widehat{s}_i(\mathbf{U}) = \begin{pmatrix} h^{-1}\widehat{g}(X_i) \\ P_\rho \widehat{\nabla}g(X_i) \end{pmatrix}.$$

Nous nous intéressons à la convergence de $P_{\rho,m}\widehat{\beta}_{l,m}$ vers $P_{\rho,m}\beta_{l,m}^*$, or

$$P_{\rho,m}\widehat{\beta}_{l,m} = n^{-1} \sum_{i=1}^n \widehat{s}_{i,m}(\mathbf{U}) \psi_{lm}(X_{i,m}),$$

où $\widehat{s}_{i,m}(\mathbf{U})$ est la $(m+1)$ ^{ème} coordonnée du vecteur $\widehat{s}_i(\mathbf{U})$. Y_i étant donné par le modèle de régression (2.1), on peut décomposer $\widehat{s}_i(\mathbf{U})$ en une partie stochastique $\mathcal{S}_i(\mathbf{U})$ et une partie déterministe $s_i(\mathbf{U})$:

$$\widehat{s}_i(\mathbf{U}) = s_i(\mathbf{U}) + \mathcal{S}_i(\mathbf{U}),$$

avec

$$\begin{aligned} s_i(\mathbf{U}) &= h^{-1} \mathcal{V}_i(\mathbf{U})^{-1} \sum_{j=1}^n \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix} g(X_j) K_{ij}(\mathbf{U}), \\ \mathcal{S}_i(\mathbf{U}) &= h^{-1} \mathcal{V}_i(\mathbf{U})^{-1} \sum_{j=1}^n \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix} \varepsilon_j K_{ij}(\mathbf{U}). \end{aligned}$$

Finalement en utilisant la description précédente, la différence $P_{\rho,m}\{\widehat{\beta}_{lm}(\mathbf{U}) - \beta_{lm}^*\}$ comporte deux parties à étudier,

$$\begin{cases} P_{\rho,m}\{\mathbb{E} \widehat{\beta}_{lm}(\mathbf{U}) - \beta_{lm}^*\} &= \frac{1}{n} \sum_{i=1}^n \{s_{i,m}(\mathbf{U}) - P_{\rho,m} \nabla g_m(X_i)\} \psi_{lm}(X_{i,m}), \\ P_{\rho,m}\{\widehat{\beta}_{lm}(\mathbf{U}) - \mathbb{E} \widehat{\beta}_{lm}\} &= \frac{1}{n} \sum_{i=1}^n \mathcal{S}_{i,m}(\mathbf{U}) \psi_{lm}(X_{i,m}). \end{cases} \quad (2.7)$$

On peut remarquer que la première partie correspond au biais et que la deuxième, notée $\zeta_{lm}(\mathbf{U}) := n^{-1} \sum_{i=1}^n \mathcal{S}_{i,m}(\mathbf{U}) \psi_{lm}(X_{i,m})$, est (pour tout \mathbf{U}) une combinaison linéaire des erreurs gaussiennes ε_i . C'est donc un vecteur gaussien centré de \mathbb{R}^d .

On utilisera dans la suite la projection orthogonale \mathcal{E}_d de \mathbb{R}^{d+1} dans \mathbb{R}^d sur le plan orthogonal à la première composante de la base canonique de \mathbb{R}^{d+1} :

$$\begin{aligned} \mathcal{E}_d : \quad \mathbb{R}^{d+1} &\rightarrow \mathbb{R}^d \\ (x_0, \dots, x_d)^\top &\mapsto \mathcal{E}_d(x_0, \dots, x_d) = (x_1, \dots, x_d). \end{aligned}$$

Propriété 2.2.1. *Sous les hypothèses H1– H5, en notant $\beta_{lm}(\mathbf{U}) = \mathbb{E} \widehat{\beta}_{lm}(\mathbf{U})$, on obtient les résultats suivants.*

Pour le biais :

$$\sup_{\mathbf{U} \in \mathcal{U}_\alpha} \sum_{m=1}^d \sum_{l=1}^L |P_{\rho,m}\{\beta_{lm}(\mathbf{U}) - \beta_{lm}^*\}|^2 \leq \left(\frac{C_{g_1} C_V^{1/2}}{1 - \alpha^*} \rho^2 h \right)^2.$$

Pour la partie stochastique :

Pour tout $l = 1, \dots, L$ et $m = 1, \dots, d$, il existe une variable aléatoire gaussienne ξ_{lm} , combinaison linéaire des erreurs ε_i avec des coefficients déterministes, qui dépendent uniquement de \mathbf{v}^ , des variables explicatives $\{X_i\}_{i=1, \dots, n}$, des fonctions de la base $\psi_{lm}(\cdot)$, et du noyau K . La variable aléatoire ξ_{lm} est donc d'espérance nulle et sa variance vérifie*

$$\max_{m,l} \mathbb{E} \xi_{lm}^2 \leq 2\sigma^2 C_V^2 C_K \quad (2.8)$$

et

$$\mathbb{P} \left(\max_{m,l} \sup_{\mathbf{U} \in \mathcal{U}_\alpha} \left| P_{\rho,m}\{\widehat{\beta}_{lm}(\mathbf{U}) - \beta_{lm}(\mathbf{U})\} - \frac{\xi_{lm}}{h\sqrt{n}} \right| > \frac{\sigma \bar{\psi} C_{\alpha,n} |\alpha|}{h\sqrt{n}} \right) \leq \frac{2}{n},$$

où le maximum est évalué pour $m \in \{1, \dots, d\}$ et $l \in \{1, \dots, L\}$, $\bar{\psi} = \max_{i,l,m} |\psi_{lm}(X_{im})|$ et

$$C_{\alpha,n} = \left(\frac{\sqrt{2} C_V C_{K'}}{(1 - \alpha^*)^{3/2}} + \frac{2^{3/2} C_V^2 C_{K'} C_K}{(1 - \alpha^*)^{5/2}} \right) \left(2 + \sqrt{2 \log(ndL) + d \log(4n)} \right).$$

Afin de démontrer cette propriété nous introduisons quatre lemmes. Le premier fournit une majoration de $|Z_{ij}|^2$ qui interviendra à maintes reprises par la suite. Ensuite le lemme 2.2.2 concerne l'étude du biais et la majoration (2.9) conduit à la première relation de la proposition. Les deux lemmes suivants permettent de maîtriser la partie stochastique et sont à la base des énoncés stochastiques de la propriété 2.2.1. Une fois ces résultats intermédiaires établis, on pourra achever la preuve de la Propriété 2.2.1.

Lemme 2.2.1. *Soit $|U_m - 1| \leq \alpha_m < 1$, pour tout $m = 1, \dots, d$. Alors pour tout i, j , l'inégalité $|Z_{ij}^\top D_{\mathbf{U}} Z_{ij}| \leq 1$ implique $|Z_{ij}|^2 \leq 1/(1 - \alpha^*)$ et $1 + |Z_{ij}|^2 \leq 2/(1 - \alpha^*)$.*

DÉMONSTRATION.

$$\left| |Z_{ij}^\top D_{\mathbf{U}} Z_{ij}| - |Z_{ij}|^2 \right| = |Z_{ij}^\top (D_{\mathbf{U}} - I) Z_{ij}| \leq \alpha^* |Z_{ij}|^2,$$

par définition d'une norme matricielle et avec $\alpha^* \geq \max_m |U_m - 1|$. D'où

$$\begin{aligned} |Z_{ij}|^2 &\leq (1 - \alpha^*)^{-1} |Z_{ij}^\top D_{\mathbf{U}} Z_{ij}| \leq (1 - \alpha^*)^{-1}, \\ 1 + |Z_{ij}|^2 &\leq \frac{2 - \alpha^*}{1 - \alpha^*} \leq 2(1 - \alpha^*)^{-1}. \end{aligned}$$

■

Lemme 2.2.2. *Sous les mêmes hypothèses que la proposition 2.2.1, on a*

$$\sup_{\mathbf{U} \in \mathcal{U}_\alpha} |\mathcal{E}_d s_i(\mathbf{U}) - P_\rho \nabla g(X_i)| \leq \frac{C_{g_1} C_V^{1/2}}{1 - \alpha^*} h \rho^2, \quad i = 1, \dots, n, \quad (2.9)$$

DÉMONSTRATION. En utilisant les remarques faites en début de ce paragraphe 2.2, on écrit

$$\begin{aligned} \begin{pmatrix} h^{-1} g(X_i) \\ P_\rho \nabla g(X_i) \end{pmatrix} &= \mathcal{V}_i(\mathbf{U})^{-1} \sum_{j=1}^n \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix}^\top \begin{pmatrix} h^{-1} g(X_i) \\ P_\rho \nabla g(X_i) \end{pmatrix} K_{ij}(\mathbf{U}), \\ &= h^{-1} \mathcal{V}_i(\mathbf{U})^{-1} \sum_{j=1}^n \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix} \{g(X_i) + X_{ij}^\top \nabla g(X_i)\} K_{ij}(\mathbf{U}), \end{aligned}$$

d'où

$$\begin{aligned} s_i(\mathbf{U}) - \begin{pmatrix} h^{-1} g(X_i) \\ P_\rho \nabla g(X_i) \end{pmatrix} &= \frac{1}{h} \mathcal{V}_i(\mathbf{U})^{-1} \sum_{j=1}^n \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix} \{g(X_j) - g(X_i) - X_{ij}^\top \nabla g(X_i)\} K_{ij}(\mathbf{U}), \\ &= \frac{1}{h} \mathcal{V}_i(\mathbf{U})^{-1} \sum_{j=1}^n \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix} r_{ij} K_{ij}(\mathbf{U}), \end{aligned}$$

alors on utilise la forme du modèle (2.1), avec \mathcal{R}^* matrice diagonale $M \times M$, de coefficients diagonaux $\sqrt{v_m^*}$, $m \in \mathcal{J}$:

$$\begin{aligned} r_{ij} &= X_{j,1}^\top \theta + g_1(\mathcal{R}^* X_{j,2}) - X_{i,1}^\top \theta - g_1(\mathcal{R}^* X_{i,2}) - (X_{j,1} - X_{i,1})^\top \theta \\ &\quad - (X_{j,2} - X_{i,2})^\top \mathcal{R}^* \nabla g_1(\mathcal{R}^* X_{i,2}), \\ &= g_1(\mathcal{R}^* X_{j,2}) - g_1(\mathcal{R}^* X_{i,2}) - (\mathcal{R}^* X_{j,2} - \mathcal{R}^* X_{i,2})^\top \nabla g_1(\mathcal{R}^* X_{i,2}). \end{aligned}$$

On introduit Z_{ij} :

$$\left| \sqrt{v_m^*} X_{j,m} - \sqrt{v_m^*} X_{i,m} \right|^2 = h^2 v_m^* (1 + \rho^{-2} v_m^*)^{-1} Z_{ij,m}^2 \leq h^2 \rho^2 Z_{ij,m}^2,$$

d'où

$$|\mathcal{R}^* X_{j,2} - \mathcal{R}^* X_{i,2}|^2 \leq h^2 \rho^2 |Z_{ij}|^2.$$

En appliquant le lemme 2.2.1 et l'hypothèse H4, on obtient pour toute paire (i, j) telle que $Z_{ij}^\top D_{\mathbf{U}} Z_{ij} \leq 1$:

$$|r_{ij}| \leq C_{g_1} h^2 \rho^2 (1 - \alpha^*)^{-1}.$$

De plus, si on note

$$\mathcal{E}_d s_i(\mathbf{U}) - P_\rho \nabla g(X_i) = \mathcal{E}_d H,$$

en utilisant la relation générale

$$|\mathcal{E}_d H| \leq |H| = \frac{|H^\top H|}{|H|} \leq \sup_{\substack{\lambda \in \mathbb{R}^{d+1} \\ |\lambda|=1}} |\lambda^\top H|,$$

on peut écrire

$$\begin{aligned} |\mathcal{E}_d s_i(\mathbf{U}) - P_\rho \nabla g(X_i)| &\leq h^{-1} \sup_{\substack{\lambda \in \mathbb{R}^{d+1} \\ |\lambda|=1}} \left| \lambda^\top \mathcal{V}_i(\mathbf{U})^{-1} \sum_{j=1}^n \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix} r_{ij} K_{ij}(\mathbf{U}) \right|, \\ &\leq h^{-1} \sup_{\substack{\lambda \in \mathbb{R}^{d+1} \\ |\lambda|=1}} \left| \lambda^\top \mathcal{V}_i(\mathbf{U})^{-1} \sum_{j=1}^n \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix} K_{ij}(\mathbf{U})^{1/2} \right. \\ &\quad \left. \times r_{ij} K_{ij}(\mathbf{U})^{1/2} \right|. \end{aligned}$$

On utilise alors l'inégalité de Cauchy-Schwarz pour des vecteurs de \mathbb{R}^d et la majoration concernant les $|r_{ij}|$ que l'on vient d'établir

$$\begin{aligned}
|\mathcal{E}_d s_i(\mathbf{U}) - P_\rho \nabla g(X_i)| &\leq h^{-1} \sup_{\substack{\lambda \in \mathbb{R}^{d+1} \\ |\lambda|=1}} \left| \lambda^\top \mathcal{V}_i(\mathbf{U})^{-1} \left(\sum_{j=1}^n \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix}^\top K_{ij}(\mathbf{U}) \right) \right. \\
&\quad \left. \times \left(\sum_{j=1}^n r_{i,j}^2 K_{ij}(\mathbf{U}) \right) \mathcal{V}_i(\mathbf{U})^{-1} \lambda \right|^{1/2}, \\
&\leq h^{-1} \sup_{\substack{\lambda \in \mathbb{R}^{d+1} \\ |\lambda|=1}} \left| \lambda^\top \mathcal{V}_i(\mathbf{U})^{-1} \lambda \right|^{1/2} |N_i(\mathbf{U})|^{1/2} \frac{C_{g_1} h^2 \rho^2}{(1 - \alpha^*)}, \\
&\leq C_{g_1} h^2 \rho^2 (1 - \alpha^*)^{-1} \left[\|\mathcal{V}_i(\mathbf{U})^{-1}\| |N_i(\mathbf{U})| \right]^{1/2}, \\
&\leq C_{g_1} C_V^{1/2} h \rho^2 (1 - \alpha^*)^{-1}.
\end{aligned}$$

La dernière inégalité nous donne (2.9). ■

Par définition $\zeta_{lm}(\mathbf{U})$ est une combinaison linéaire des ε_i , c'est-à-dire qu'il existe des coefficients $c_{i,lm}(\mathbf{U})$ tels que

$$\zeta_{lm}(\mathbf{U}) = \sum_{i=1}^n c_{i,lm}(\mathbf{U}) \varepsilon_i.$$

Le lemme suivant et sa preuve fournissent une définition explicite des coefficients $c_{i,lm}(\mathbf{U})$ et des propriétés remarquables qu'ils vérifient.

Lemme 2.2.3. *Pour tout $l = 1, \dots, L$ et $m = 1, \dots, d$*

- (i) $\sum_{i=1}^n |c_{i,lm}(\mathbf{U}^*)|^2 \leq \frac{2C_V^2 C_K}{h^2 n}$;
- (ii) $\sup_{\mathbf{U} \in \mathcal{U}_\alpha} \sum_{i=1}^n |c_{i,lm}(\mathbf{U})|^2 \leq \frac{2C_V^2 C_K}{(1 - \alpha^*) h^2 n}$;
- (iii) $\sup_{\mathbf{U} \in \mathcal{U}_\alpha} \left| \frac{dc_{i,lm}(\mathbf{U})}{d\mathbf{U}} \right| \leq \frac{\kappa_\alpha}{nh}$,

$$\text{où } \kappa_\alpha = \sqrt{2}(1 - \alpha^*)^{-3/2} C_V C_K \bar{\psi} + 2^{3/2} (1 - \alpha^*)^{-5/2} C_V^2 C_K C_{K'} \bar{\psi}.$$

DÉMONSTRATION. On définit pour $i, j = 1, \dots, n$

$$v_{ij}(\mathbf{U}) = \mathcal{V}_i(\mathbf{U})^{-1} \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix}.$$

Dans la suite on notera encore $v_{ij,m}$ la $(m+1)^{\text{ème}}$ composante de ce vecteur.

Le lemme 2.2.1 et l'hypothèse H5 impliquent : si $\|D_{\mathbf{U}} - I\| \leq \alpha$ alors pour tout i, j tels que $Z_{ij}^\top D_{\mathbf{U}} Z_{ij} \leq 1$

$$\begin{aligned} |N_i(\mathbf{U}) v_{ij}(\mathbf{U})| &\leq |N_i(\mathbf{U})| \|\mathcal{V}_i(\mathbf{U})\|^{-1} \left| \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix} \right| \leq C_V (1 + |Z_{ij}|^2)^{1/2}, \\ &\leq C_V \sqrt{2} (1 - \alpha^*)^{-1/2}. \end{aligned} \quad (2.10)$$

Pour $m \in \{1, \dots, d\}$, on a

$$\begin{aligned} \zeta_{lm}(\mathbf{U}) &= \frac{1}{nh} \sum_{i=1}^n \psi_{lm}(X_{i,m}) \left(\mathcal{V}_i(\mathbf{U})^{-1} \sum_{j=1}^n \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix} K_{ij}(\mathbf{U}) \varepsilon_j \right)_{(m+1)}, \\ \zeta_{lm}(\mathbf{U}) &= \frac{1}{nh} \sum_{i=1}^n \psi_{lm}(X_{i,m}) \sum_{j=1}^n v_{ij,m}(\mathbf{U}) K_{ij}(\mathbf{U}) \varepsilon_j, \\ &= \sum_{j=1}^n \left(\frac{1}{nh} \sum_{i=1}^n \psi_{lm}(X_{i,m}) v_{ij,m}(\mathbf{U}) K_{ij}(\mathbf{U}) \right) \varepsilon_j, \\ &= \sum_{j=1}^n c_{j,lm}(\mathbf{U}) \varepsilon_j. \end{aligned}$$

On a directement, étant données les propriétés des ε_i

$$\mathbb{E} |\zeta_{lm}(\mathbf{U})|^2 = \sigma^2 \sum_{j=1}^n c_{j,lm}^2(\mathbf{U}).$$

En utilisant l'inégalité de Cauchy-Schwarz, puis l'inégalité (2.10), on peut écrire

$$\begin{aligned} \sum_{j=1}^n c_{j,lm}^2(\mathbf{U}) &= \frac{1}{n^2 h^2} \sum_{j=1}^n \left(\sum_{i=1}^n \psi_{lm}(X_{i,m}) v_{ij,m}(\mathbf{U}) K_{ij}(\mathbf{U}) \right)^2, \\ &\leq \frac{1}{n^2 h^2} \sum_{j=1}^n \left(\sum_{i=1}^n \left(\psi_{lm}(X_{i,m}) |v_{ij,m}(\mathbf{U})|^{1/2} K_{ij}(\mathbf{U})^{1/2} \right)^2 \right) \\ &\quad \times \left(\sum_{i=1}^n \left(|v_{ij,m}(\mathbf{U})|^{1/2} K_{ij}(\mathbf{U})^{1/2} \right)^2 \right), \\ &\leq \frac{1}{n^2 h^2} \sum_{j=1}^n \left(\sum_{i=1}^n \psi_{lm}^2(X_{i,m}) |v_{ij,m}(\mathbf{U})| K_{ij}(\mathbf{U}) \right) \left(\sum_{i=1}^n |v_{ij,m}(\mathbf{U})| K_{ij}(\mathbf{U}) \right), \\ &\leq \frac{2C_V^2}{(1 - \alpha^*) n^2 h^2} \sum_{j=1}^n \left(\sum_{i=1}^n \psi_{lm}^2(X_{i,m}) \frac{K_{ij}(\mathbf{U})}{N_i(\mathbf{U})} \right) \left(\sum_{i=1}^n \frac{K_{ij}(\mathbf{U})}{N_i(\mathbf{U})} \right). \end{aligned}$$

Ensuite on prend en compte l'hypothèse H5 concernant les variables explicatives et l'orthonormalité de la famille $(\psi_m)_m$:

$$\begin{aligned} \sum_{j=1}^n c_{j,lm}^2(\mathbf{U}) &\leq \frac{2C_V^2 C_K}{(1-\alpha^*)n^2 h^2} \sum_{j=1}^n \sum_{i=1}^n \psi_{lm}^2(X_{i,m}) \frac{K_{ij}(\mathbf{U})}{N_i(\mathbf{U})}, \\ &\leq \frac{2C_V^2 C_K}{(1-\alpha^*)n^2 h^2} \sum_{i=1}^n \psi_{lm}^2(X_{i,m}) = \frac{2C_V^2 C_K}{(1-\alpha^*)n h^2}. \end{aligned}$$

Cette inégalité prise dans le cas particulier où $D_{\mathbf{U}} = D_{\mathbf{U}^*} = I$ et $\alpha^* = 0$ nous permet d'obtenir

$$\sum_{j=1}^n c_{j,lm}^2(\mathbf{U}^*) \leq \frac{2C_V^2 C_K}{n h^2},$$

ainsi on a prouvé (i) et (ii) dans le lemme 2.2.3.

On veut maintenant borner la dérivée de chaque coefficient $c_{j,lm}(\mathbf{U})$ par rapport à \mathbf{U} . Celle-ci s'écrit

$$\frac{dc_{j,lm}(\mathbf{U})}{d\mathbf{U}} = \frac{1}{nh} \sum_{i=1}^n v_{ij,m}(\mathbf{U}) \psi_{lm}(X_{i,m}) \frac{dK_{ij}(\mathbf{U})}{d\mathbf{U}} + \frac{1}{nh} \sum_{i=1}^n \frac{dv_{ij,m}(\mathbf{U})}{d\mathbf{U}} K_{ij}(\mathbf{U}) \psi_{lm}(X_{i,m}).$$

Nous allons trouver une borne pour chacun des termes. Pour cela nous avons besoin dans un premier temps de borner $\left| \frac{d}{d\mathbf{U}} K_{ij}(\mathbf{U}) \right|$:

Pour tout i, j tels que $Z_{ij}^\top D_{\mathbf{U}} Z_{ij} \leq 1$, le lemme 2.2.1 implique

$$\begin{aligned} \left| \frac{d}{d\mathbf{U}} K_{ij}(\mathbf{U}) \right| &= \left| K'(Z_{ij}^\top D_{\mathbf{U}} Z_{ij}) \right| \left(\sum_{m=1}^d Z_{ij,m}^4 \right)^{1/2} \leq \left| K'(Z_{ij}^\top D_{\mathbf{U}} Z_{ij}) \right| |Z_{ij}|^2 \\ &\leq (1-\alpha^*)^{-1} \left| K'(Z_{ij}^\top D_{\mathbf{U}} Z_{ij}) \right|. \end{aligned} \quad (2.11)$$

Intéressons nous maintenant à $v_{ij,m}$: on peut écrire $v_{ij,m}$ de la façon suivante

$$v_{ij,m}(\mathbf{U}) = (1 + |Z_{ij}|^2)^{1/2} \mathbf{e}_m^\top \mathcal{V}_i(\mathbf{U})^{-1} \mathbf{o}_2,$$

en prenant $\mathbf{o}_2 = (1 + |Z_{ij}|^2)^{-1/2} \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix}$ et où \mathbf{e}_m est le $(m+1)$ ème vecteur de la base canonique de \mathbb{R}^{d+1} . Pour toute paire (i, j) telle que $Z_{ij}^\top D_{\mathbf{U}} Z_{ij} \leq 1$ et comme $\frac{dv_{ij,m}(\mathbf{U})}{d\mathbf{U}} = \left(\frac{\partial v_{ij,m}(\mathbf{U})}{\partial U_1}, \dots, \frac{\partial v_{ij,m}(\mathbf{U})}{\partial U_d} \right)$, il vient

$$\left| \frac{dv_{ij,m}(\mathbf{U})}{d\mathbf{U}} \right| \leq (1 + |Z_{ij}|^2)^{1/2} \left(\sum_{m'=1}^d \left| \frac{\partial \mathbf{e}_m^\top \mathcal{V}_i(\mathbf{U})^{-1} \mathbf{o}_2}{\partial U_{m'}} \right|^2 \right)^{1/2}.$$

Nous allons donc commencer par l'étude, pour tout vecteur unitaire de \mathbb{R}^{d+1} \mathbf{o}_1 et \mathbf{o}_2 , de la quantité $\frac{\mathbf{o}_1^\top \partial \mathcal{V}_i(\mathbf{U})^{-1} \mathbf{o}_2}{\partial U_m}$ pour $m = 1, \dots, d$. On peut écrire pour tout $m = 1, \dots, d$

$$\begin{aligned} \frac{\mathbf{o}_1^\top \partial \mathcal{V}_i(\mathbf{U})^{-1} \mathbf{o}_2}{\partial U_m} &= -\mathbf{o}_1^\top \mathcal{V}_i(\mathbf{U})^{-1} \left(\frac{\partial}{\partial U_m} \mathcal{V}_i(\mathbf{U}) \right) \mathcal{V}_i(\mathbf{U})^{-1} \mathbf{o}_2, \\ &= -\mathbf{o}_1^\top \mathcal{V}_i(\mathbf{U})^{-1} \left(\sum_{j=1}^n \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix}^\top K'(Z_{ij}^\top D_{\mathbf{U}} Z_{ij}) Z_{ij,m}^2 \right) \\ &\quad \times \mathcal{V}_i(\mathbf{U})^{-1} \mathbf{o}_2. \end{aligned}$$

Le lemme 2.2.1 et l'hypothèse H5 conduisent à

$$\begin{aligned} \left| \frac{\partial \mathbf{o}_1^\top \mathcal{V}_i(\mathbf{U})^{-1} \mathbf{o}_2}{\partial U_m} \right| &\leq \|\mathcal{V}_i(\mathbf{U})^{-1}\|^2 \sum_{j=1}^n (1 + |Z_{ij}|^2) |K'(Z_{ij}^\top D_{\mathbf{U}} Z_{ij})| Z_{ij,m}^2, \\ &\leq \frac{2C_V^2}{(1 - \alpha^*) |N_i(\mathbf{U})|^2} \sum_{j=1}^n |K'(Z_{ij}^\top D_{\mathbf{U}} Z_{ij})| Z_{ij,m}^2. \end{aligned}$$

Ce dernier résultat nous permet d'écrire

$$\begin{aligned} \left| \frac{dv_{ij,m}(\mathbf{U})}{d\mathbf{U}} \right| &\leq \frac{2^{1/2} (1 - \alpha^*)^{-1/2} 2C_V^2}{(1 - \alpha^*) |N_i(\mathbf{U})|^2} \left[\sum_{m'=1}^d \left(\sum_{j=1}^n |K'(Z_{ij}^\top D_{\mathbf{U}} Z_{ij})| Z_{ij,m'}^2 \right)^2 \right]^{1/2}, \\ &\leq \frac{2^{3/2} C_V^2}{(1 - \alpha^*)^{3/2} |N_i(\mathbf{U})|^2} \sum_{m'=1}^d \sum_{j=1}^n |K'(Z_{ij}^\top D_{\mathbf{U}} Z_{ij})| Z_{ij,m'}^2, \\ &\leq \frac{2^{3/2} C_V^2 C_{K'}}{(1 - \alpha^*)^{3/2} |N_i(\mathbf{U})|^2} \sum_{j=1}^n |K'(Z_{ij}^\top D_{\mathbf{U}} Z_{ij})| |Z_{ij}|^2, \\ &\leq \frac{2^{3/2} C_V^2 C_{K'}}{(1 - \alpha^*)^{3/2} |N_i(\mathbf{U})|^2} \frac{1}{(1 - \alpha^*)} \sum_{j=1}^n |K'(Z_{ij}^\top D_{\mathbf{U}} Z_{ij})|, \\ \left| \frac{dv_{ij,m}(\mathbf{U})}{d\mathbf{U}} \right| &\leq \frac{2^{3/2} C_V^2 C_{K'}}{(1 - \alpha^*)^{5/2} |N_i(\mathbf{U})|^2}. \end{aligned} \tag{2.12}$$

Revenons à l'étude de la dérivée de $c_{j,lm}(\mathbf{U})$: l'utilisation des inégalités (2.10), (2.11),

(2.12) et de l'hypothèse H5 conduit à

$$\begin{aligned} \left| \frac{dc_{j,lm}(\mathbf{U})}{d\mathbf{U}} \right| &\leq \frac{\sqrt{2}C_V\bar{\psi}}{nh(1-\alpha^*)^{3/2}} \sum_{i=1}^n \frac{|K'(Z_{ij}^\top D_{\mathbf{U}} Z_{ij})|}{N_i(\mathbf{U})} + \frac{2^{3/2}C_V^2 C_{K'}\bar{\psi}}{nh(1-\alpha^*)^{5/2}} \sum_{i=1}^n \frac{K_{ij}(\mathbf{U})}{N_i(\mathbf{U})}, \\ &\leq \frac{\sqrt{2}C_V C_{K'}\bar{\psi}}{nh(1-\alpha^*)^{3/2}} + \frac{2^{3/2}C_V^2 C_{K'} C_K \bar{\psi}}{nh(1-\alpha^*)^{5/2}}, \end{aligned}$$

ce qui nous donne bien l'assertion (iii) du lemme. \blacksquare

Lemme 2.2.4. *Sous les mêmes hypothèses que la proposition 2.2.1*

$$\mathbb{P} \left(\max_{l,m} \sup_{\mathbf{U} \in \mathcal{U}_\alpha} |\zeta_{lm}(\mathbf{U}) - \zeta_{lm}(\mathbf{U}^*)| > \frac{\sigma C_{\alpha,n} |\boldsymbol{\alpha}| \bar{\psi}}{h\sqrt{n}} \right) \leq 2/n, \quad (2.13)$$

$$\max_{l,m} \mathbb{E} |\zeta_{lm}(\mathbf{U}^*)|^2 \leq \frac{2\sigma^2 C_V^2 C_K}{h^2 n}. \quad (2.14)$$

où le maximum est considéré pour $l = 1, \dots, L$ et $m = 1, \dots, d$.

DÉMONSTRATION.

Nous allons utiliser le lemme 2.2.3 pour démontrer les inégalités (2.13) et (2.14).

Obtention de (2.14) : On sait que $\mathbb{E} |\zeta_{lm}(\mathbf{U})|^2 = \sigma^2 \sum_{j=1}^n c_{j,lm}^2(\mathbf{U})$, et en utilisant la condition (i) du lemme 2.2.3, on a directement

$$\mathbb{E} |\zeta_{lm}(\mathbf{U}^*)|^2 \leq \frac{2C_V^2 C_K \sigma^2}{nh^2}$$

Obtention de (2.13) : Ce résultat repose sur le lemme 4.3.3 de l'Annexe 4.3 que l'on peut appliquer connaissant la majoration (iii) du lemme 2.2.3, pour

- $\Gamma = \left\{ \gamma = (l, m), l \in \{1, \dots, L\}, m \in \{1, \dots, d\} \right\}$, donc $|\Gamma| = Ld$,
- $a_{i,\gamma}(\mathbf{U}) = \sqrt{n} c_{i,lm}(\mathbf{U})$,
- $\kappa = \frac{\kappa_\alpha}{h\sqrt{n}}$,
- $r = |\boldsymbol{\alpha}|$.

d'où

$$\mathbb{P} \left(\max_{(l,m) \in \Gamma} \sup_{|\mathbf{U} - \mathbf{U}^*| \leq r} \left| \sum_{i=1}^n \{c_{i,lm}(\mathbf{U}) - c_{i,lm}(\mathbf{U}^*)\} \varepsilon_i \right| > \sigma \kappa r t \right) \leq \frac{2}{n},$$

avec

$$\kappa = \frac{\kappa_\alpha}{h\sqrt{n}} = \left(\frac{\sqrt{2} C_V C_{K'}}{(1-\alpha^*)^{3/2}} + \frac{2^{3/2} C_V^2 C_{K'} C_K}{(1-\alpha^*)^{5/2}} \right) \frac{\bar{\psi}}{h\sqrt{n}},$$

et $t = 2 + \sqrt{2 \log(ndL) + d \log(4n)}$. On obtient donc le résultat suivant :

$$\mathbb{P}\left(\max_{m,l} \sup_{|\mathbf{U}-\mathbf{U}^*| \leq |\boldsymbol{\alpha}|} \left| \sum_{i=1}^n \{c_{i,l,m}(\mathbf{U}) - c_{i,l,m}(\mathbf{U}^*)\} \varepsilon_i \right| > \frac{\sigma \bar{\psi} C_{\alpha,n} |\boldsymbol{\alpha}|}{h \sqrt{n}}\right) \leq \frac{2}{n}. \quad (2.15)$$

Or si $\mathbf{U} \in \mathcal{U}_\alpha$, alors pour $m = 1, \dots, d$, $|U_m - 1| < \alpha_m$ et donc $|\mathbf{U} - \mathbf{U}^*| < |\boldsymbol{\alpha}|$.

Si on note γ les paires (l, m) et $Q_\gamma(\mathbf{U})$ la fonction dont on cherche à minorer la borne supérieure sur le voisinage \mathcal{U}_α , alors

$$\begin{aligned} \sup_{\mathbf{U} \in \mathcal{U}_\alpha} |Q_\gamma(\mathbf{U})| &\leq \sup_{|\mathbf{U}-\mathbf{U}^*| \leq |\boldsymbol{\alpha}|} |Q_\gamma(\mathbf{U})|, \\ \mathbb{P}\left(\max_{\gamma} \sup_{\mathbf{U} \in \mathcal{U}_\alpha} |Q_\gamma(\mathbf{U})| > \nu\right) &\leq \mathbb{P}\left(\max_{\gamma} \sup_{|\mathbf{U}-\mathbf{U}^*| \leq |\boldsymbol{\alpha}|} |Q_\gamma(\mathbf{U})| > \nu\right) \leq 2/n, \end{aligned}$$

où la dernière inégalité est due à (2.15). Finalement on obtient la majoration (2.13). ■

DÉMONSTRATION DE LA PROPRIÉTÉ 2.2.1

Dans un premier temps, on traite la partie déterministe, puis on s'intéressera à la partie stochastique.

Étude de la partie déterministe.

On obtient la première affirmation de la proposition 2.2.1 de la façon suivante :

Pour $m \leq d$ fixé, $\boldsymbol{\psi}_{lm} \in \mathbb{R}^n$, la famille $\{\boldsymbol{\psi}_{lm}, l = 1, \dots, L\}$ est une famille orthogonale de \mathbb{R}^n pour le produit scalaire défini par $\langle u, v \rangle = n^{-1} \sum_{i=1}^n u_i v_i$.

Soit $E_\psi = \text{Vect}\{\boldsymbol{\psi}_{lm} \in \mathbb{R}^n\}$, un sous-espace vectoriel de \mathbb{R}^n . On note h le vecteur de \mathbb{R}^n de coordonnées : $h_i = s_{i,m}(\mathbf{U}) - P_{\rho,m} \nabla g_m(X_i)$. Sa projection sur E_ψ , S_ψ , sera donc $S_\psi = \sum_{l=1}^L \langle h, \boldsymbol{\psi}_{lm} \rangle \boldsymbol{\psi}_{lm}$.

Par l'inégalité de Bessel $\|h\|^2 \geq \|S_\psi\|^2$ or $\|h\|^2 = n^{-1} \sum_{i=1}^n |s_{i,m}(\mathbf{U}) - P_{\rho,m} \nabla g_m(X_i)|^2$ et

$$\begin{aligned} \|S_\psi\|^2 &= \left\langle \sum_{l'=1}^L \langle h, \boldsymbol{\psi}_{lm} \rangle \boldsymbol{\psi}_{lm}, \sum_{l=1}^L \langle h, \boldsymbol{\psi}_{l'm} \rangle \boldsymbol{\psi}_{l'm} \right\rangle, \\ &= \sum_{l=1}^L \sum_{l'=1}^L \langle h, \boldsymbol{\psi}_{lm} \rangle \langle h, \boldsymbol{\psi}_{l'm} \rangle \langle \boldsymbol{\psi}_{lm}, \boldsymbol{\psi}_{l'm} \rangle, \\ &= \sum_{l=1}^L \langle h, \boldsymbol{\psi}_{lm} \rangle^2, \\ &= \sum_{l=1}^L \left(n^{-1} \sum_{i=1}^n (s_{i,m}(\mathbf{U}) - P_{\rho,m} \nabla g_m(X_i)) \boldsymbol{\psi}_{lm}(X_{i,m}) \right)^2, \\ &= \sum_{l=1}^L P_{\rho,m}^2 \left(\mathbb{E} \widehat{\beta}_{l,m} - \beta_{l,m}^* \right)^2. \end{aligned}$$

L'inégalité de Bessel se réécrit donc

$$n^{-1} \sum_{i=1}^n \left| s_{i,m}(\mathbf{U}) - P_{\rho,m} \nabla g_m(X_i) \right|^2 \geq \sum_{l=1}^L P_{\rho,m}^2 \left(\mathbb{E} \hat{\beta}_{l,m} - \beta_{l,m}^* \right)^2.$$

En outre (2.9) implique pour tout $\mathbf{U} \in \mathcal{U}_\alpha$

$$n^{-1} \sum_{i=1}^n \sum_{m=1}^d \left| s_{i,m}(\mathbf{U}) - P_{\rho,m} \nabla g_m(X_i) \right|^2 \leq \left(\frac{C_g C_V^{1/2}}{1 - \alpha^*} h \rho^2 \right)^2.$$

Finalement, pour tout $\mathbf{U} \in \mathcal{U}_\alpha$,

$$\sum_{m=1}^d \sum_{l=1}^L P_{\rho,m}^2 \left(\mathbb{E} \hat{\beta}_{l,m} - \beta_{l,m}^* \right)^2 \leq \left(\frac{C_g C_V^{1/2}}{1 - \alpha^*} h \rho^2 \right)^2,$$

on obtient donc la majoration du biais.

Étude de la partie stochastique.

On rappelle que $\zeta_{lm}(\mathbf{U}) = \sum_{i=1}^n c_{i,lm}(\mathbf{U}) \varepsilon_i$. Pour obtenir les résultats souhaités, il suffit de poser $\zeta_{lm}(\mathbf{U}^*) = \frac{\xi_{lm}}{h\sqrt{n}}$ dans les inégalités de le lemme 2.2.4. La deuxième relation est obtenue directement à partir de (2.13) en utilisant la relation de (2.7) concernant $\zeta_{lm}(\mathbf{U})$. ■

On rappelle que β_m^* est le vecteur de \mathbb{R}^L , de composantes β_{lm}^* , et que son estimateur est $\hat{\beta}_m = \hat{\beta}_m(\mathbf{U})$, de coordonnées $\hat{\beta}_{lm}(\mathbf{U})$. Le corollaire suivant donne une borne pour l'erreur $P_{\rho,m} \left(\hat{\beta}_m(\mathbf{U}) - \beta_m^* \right)$.

Corollaire 2.2.1. *Soit $z_n = (1 + 2 \log(nd) + 2 \log \log(nd))^{1/2}$ et*

$$\delta = \frac{C_{g_1} C_V^{1/2}}{1 - \alpha^*} h \rho^2 + \frac{\sqrt{2L} \sigma C_V C_K^{1/2} z_n}{h\sqrt{n}} + \frac{\sqrt{L} \sigma \bar{\psi} C_{\alpha,n} |\boldsymbol{\alpha}|}{h\sqrt{n}}. \quad (2.16)$$

Alors sous les mêmes hypothèses que la proposition 2.2.1 on a

$$\mathbb{P} \left(\max_{m=1, \dots, d} \sup_{\mathbf{U} \in \mathcal{U}_\alpha} \left| P_{\rho,m} \left(\hat{\beta}_m(\mathbf{U}) - \beta_m^* \right) \right| > \delta \right) \leq 3/n.$$

DÉMONSTRATION.

$$\begin{aligned} \max_m \sup_{\mathbf{U} \in \mathcal{U}_\alpha} \left| P_{\rho,m} \left(\hat{\beta}_m(\mathbf{U}) - \beta_m^* \right) \right| &\leq \max_m \sup_{\mathbf{U} \in \mathcal{U}_\alpha} \left| P_{\rho,m} \left(\hat{\beta}_m(\mathbf{U}) - \beta_m \right) - \frac{\xi_{l,m}}{h\sqrt{n}} \right| \\ &+ \max_m \left| \frac{\xi_{l,m}}{h\sqrt{n}} \right| + \max_m \sup_{\mathbf{U} \in \mathcal{U}_\alpha} \left| P_{\rho,m} \left(\beta_m(\mathbf{U}) - \beta_m^* \right) \right| \end{aligned}$$

On peut majorer le dernier terme en utilisant la proposition 2.2.1. En effet

$$\sup_{\mathbf{U} \in \mathcal{U}_\alpha} \sum_{m=1}^d \sum_{l=1}^L |P_{\rho,m}(\beta_{l,m}(\mathbf{U}) - \beta_{l,m}^*)|^2 \leq \left(\frac{C_{g_1} C_V^{1/2}}{1 - \alpha^*} \rho^2 h \right)^2,$$

or

$$\begin{aligned} \sup_{\mathbf{U} \in \mathcal{U}_\alpha} \sum_{m=1}^d \sum_{l=1}^L |P_{\rho,m}(\beta_{l,m}(\mathbf{U}) - \beta_{l,m}^*)|^2 &= \sup_{\mathbf{U} \in \mathcal{U}_\alpha} \sum_{m=1}^d |P_{\rho,m}(\beta_m(\mathbf{U}) - \beta_m^*)|^2, \\ &\geq \sup_{\mathbf{U} \in \mathcal{U}_\alpha} \max_{m=1, \dots, d} |P_{\rho,m}(\beta_m(\mathbf{U}) - \beta_m^*)|^2, \\ &\geq \left(\sup_{\mathbf{U} \in \mathcal{U}_\alpha} \max_{m=1, \dots, d} |P_{\rho,m}(\beta_m(\mathbf{U}) - \beta_m^*)| \right)^2, \end{aligned}$$

donc

$$\sup_{\mathbf{U} \in \mathcal{U}_\alpha} \max_{m=1, \dots, d} |P_{\rho,m}(\beta_m(\mathbf{U}) - \beta_m^*)| \leq \frac{C_{g_1} C_V^{1/2}}{1 - \alpha^*} \rho^2 h.$$

De plus, l'étude de la partie aléatoire fournit l'existence d'un ensemble aléatoire A tel que $\mathbb{P}(A) \geq 1 - 2/n$, et sur A on observe

$$\sup_{\mathbf{U} \in \mathcal{U}_\alpha} \max_{m,l} \left| P_{\rho,m}(\widehat{\beta}_{l,m}(\mathbf{U}) - \mathbb{E} \widehat{\beta}_{l,m}(\mathbf{U})) - \frac{\xi_{l,m}}{h\sqrt{n}} \right| \leq \frac{\sigma \bar{\psi} C_{\alpha,n} |\boldsymbol{\alpha}|}{h\sqrt{n}}.$$

Soit

$$\sup_{\mathbf{U} \in \mathcal{U}_\alpha} \max_m \left| P_{\rho,m}(\widehat{\beta}_m(\mathbf{U}) - \mathbb{E} \widehat{\beta}_m(\mathbf{U})) - \frac{\xi_m}{h\sqrt{n}} \right| \leq \frac{\sqrt{L} \sigma \bar{\psi} C_{\alpha,n} |\boldsymbol{\alpha}|}{h\sqrt{n}}.$$

Pour le dernier terme on utilise le lemme 4.3.2 de l'Annexe 4.3, qui donne

$$\mathbb{P}\left(|\xi_m| > z_n \sqrt{\mathbb{E} |\xi_m|^2}\right) \leq z_n e^{(z_n^2 - 1)/2} \leq 1/nd.$$

D'autre part, on sait que $\mathbb{E} |\xi_m|^2 \leq 2L\sigma^2 C_V^2 C_K$. Il existe donc un ensemble aléatoire A_2 tel que $\mathbb{P}(A_2) > 1 - 1/nd > 1 - 1/n$ et sur A_2

$$|\xi_m| \leq z_n \sqrt{\mathbb{E} |\xi_m|^2},$$

c'est-à-dire

$$|\xi_m| \leq z_n \sqrt{2L} \sigma C_V C_K^{1/2}.$$

Finalement, pour $\mathcal{A} = A \cap A_2$, on a $\mathbb{P}(\mathcal{A}^c) \leq 3/n$ et sur \mathcal{A}

$$\begin{aligned} \max_m \sup_{\mathbf{U} \in \mathcal{U}_\alpha} \left| P_{\rho,m}(\widehat{\beta}_m(\mathbf{U}) - \beta_m^*(\mathbf{U})) \right| &\leq \frac{\sqrt{L} \sigma \bar{\psi} C_{\alpha,n} |\boldsymbol{\alpha}|}{h\sqrt{n}} + \frac{z_n \sqrt{2L} \sigma C_V C_K^{1/2}}{h\sqrt{n}} \\ &\quad + \frac{C_{g_1} C_V^{1/2}}{1 - \alpha^*} \rho^2 h, \end{aligned}$$

ce qui nous donne le résultat escompté. ■

Grâce au corollaire, on obtient une borne pour l'estimation de l'erreur $P_{\rho,m}^2(\widehat{v}_m(U) - v_m^*)$.

Propriété 2.2.2. *Sous les mêmes hypothèses que la proposition 2.2.1,*

$$\mathbb{P}\left(\sup_{U \in \mathcal{U}_\alpha} |P_{\rho,m}^2(\widehat{v}_m(U) - v_m^*)| \leq \delta^2 + 2\delta\tau_m, \text{ pour tout } m = 1, \dots, d\right) \geq 1 - 3/n,$$

avec $\tau_m = \rho\sqrt{v_m^*}(\rho^2 + v_m^*)^{-1/2} \leq \min\{\rho, \sqrt{v_m^*}\}$.

Nous allons utiliser le lemme suivant

Lemme 2.2.5. *Si $P_{\rho,m}|\widehat{\beta}_m - \beta_m^*| < \delta$, alors $P_{\rho,m}^2|\widehat{v}_m - v_m^*| < \delta^2 + 2\delta\tau_m$.*

DÉMONSTRATION. On introduit le vecteur $\widehat{u}_m \in \mathbb{R}^L$ (resp. u_m^*), dont les coordonnées sont $P_{\rho,m}\widehat{\beta}_{lm}$ (resp. $P_{\rho,m}\beta_{lm}^*$).

On a alors $P_{\rho,m}^2\widehat{v}_m(U) = |\widehat{u}_m(U)|^2$ et $P_{\rho,m}^2v_m^* = |u_m^*|^2 = \tau_m^2$. En effet étant donnée la définition de τ_m , on a

$$P_{\rho,m}^2v_m^* = (\rho^2 + v_m^*)^{-1}v_m^* = \tau_m^2 \leq \min\{\rho, \sqrt{v_m^*}\}. \quad (2.17)$$

$$\begin{aligned} \left||\widehat{u}_m|^2 - |u_m^*|^2\right| &= (|\widehat{u}_m| - |u_m^*|)(|\widehat{u}_m| + |u_m^*|) \leq |\widehat{u}_m - u_m^*|(|\widehat{u}_m - u_m^*| + 2|u_m^*|), \\ \left||\widehat{u}_m|^2 - |u_m^*|^2\right| &\leq |\widehat{u}_m - u_m^*|^2 + 2|\widehat{u}_m - u_m^*| \cdot |u_m^*|, \\ &\leq 2\delta\tau_m + \delta^2, \end{aligned}$$

ce qui achève la preuve. ■

DÉMONSTRATION DE LA PROPOSITION 2.2.2

Par le corollaire 2.2.1, on connaît l'existence d'un ensemble aléatoire A , tel que $\mathbb{P}(A) \geq 1 - 3/n$ et pour tout $m = 1, \dots, d$,

$$\sup_{\mathbf{U} \in \mathcal{U}_\alpha} \|P_{\rho,m}(\widehat{\beta}_m(\mathbf{U}) - \beta_m^*(\mathbf{U}))\| \leq \delta,$$

donc en utilisant le lemme 2.2.5, pour tout $m = 1, \dots, d$

$$\sup_{\mathbf{U} \in \mathcal{U}_\alpha} \|P_{\rho,m}^2(\widehat{v}_m - v_m^*)\| < \delta^2 + 2\delta\tau_m. \quad \blacksquare$$

Ces deux propositions sont essentielles car elles fournissent l'erreur commise en estimant \widehat{v} et $\widehat{\beta}_m$ en une itération. Elles constituent la base des deux paragraphes suivants.

2.3 Propriétés de l'estimateur initial

On donne ici la qualité de l'estimation après la première itération, qui correspond en fait à l'estimation non paramétrique usuelle d'une fonction et de son gradient. On donne aussi la qualité d'approximation pour v_m^* à la première itération.

Rappelons que β_m^* est le vecteur de \mathbb{R}^L de composantes β_{lm}^* , $l = 1, \dots, L$.

Propriété 2.3.1. *On se place sous les hypothèses H1–H5.*

Pour $m = 1, \dots, d$, on obtient pour l'estimation de β_m^ à la première itération, notée $\widehat{\beta}_m^{(1)}$:*

$$\widehat{\beta}_m^{(1)} - \beta_m^* = s_m h_1 + \frac{\eta_m}{h_1 \sqrt{n}},$$

où s_m est un vecteur déterministe de \mathbb{R}^L satisfaisant

$$|s_m| \leq C_{g_1} C_V^{1/2} v_{(1)}^*,$$

avec $v_{(1)}^* = \max_{m=1, \dots, d} v_m^*$. Les η_m sont des vecteurs aléatoires gaussiens de \mathbb{R}^L de moyenne nulle et dont les composantes η_{lm} vérifient, pour $l = 1, \dots, L$,

$$\mathbb{E} |\eta_{lm}|^2 \leq 2\sigma^2 C_V^2 C_K.$$

On obtient aussi

$$\mathbb{P} \left(\max_{m=1, \dots, d} |\widehat{\beta}_m^{(1)} - \beta_m^*| > \delta_1 \right) \leq \frac{1}{n}, \quad (2.18)$$

où

$$\delta_1 = C_{g_1} C_V^{1/2} v_{(1)}^* h_1 + \frac{\sqrt{2L} \sigma C_V C_K^{1/2} z_n}{h_1 \sqrt{n}}, \quad (2.19)$$

et $z_n = (1 + 2 \log(nd) + 2 \log \log(nd))^{1/2}$.

De plus, pour l'estimateur de v_m^* à la première itération, $\widehat{v}_m^{(1)}$, $m = 1, \dots, d$, on a :

$$\mathbb{P} \left(\left| \widehat{v}_m^{(1)} - v_m^* \right| \leq \delta_1^2 + 2\delta_1 \tau_{m,1}, \forall m = 1, \dots, d \right) \geq 1 - \frac{1}{n}, \quad (2.20)$$

où $\tau_{m,1} = \sqrt{v_m^*} (1 + v_m^*)^{-1/2} \leq \min\{1, \sqrt{v_m^*}\}$.

DÉMONSTRATION. La preuve consiste en une version simplifiée de celle de la proposition 2.2.1 :

Dans notre cas $P_{\rho,m} = 1$, $\rho = 1$ et $\alpha^* = 0$. On reprend l'écriture

$$P_{\rho,m}(\widehat{\beta}_m^{(1)} - \beta_m^*) = P_{\rho,m}(\mathbb{E} \widehat{\beta}_m^{(1)} - \beta_m^*) + P_{\rho,m}(\widehat{\beta}_m^{(1)} - \mathbb{E} \widehat{\beta}_m^{(1)}), \quad (2.21)$$

$$:= s_m h_1 + \frac{\eta_m}{n\sqrt{h_1}}. \quad (2.22)$$

La partie déterministe se traite de la même façon que dans la partie 2.2. Le seul changement dans la démonstration se fait au niveau de la majoration des $|r_{i,j}|$, en effet

$$|r_{i,j}| = g_1(\mathcal{R}^* X_{j,2}) - g_1(\mathcal{R}^* X_{i,2}) - (\mathcal{R}^* X_{j,2} - \mathcal{R}^* X_{i,2})^\top \nabla g_1(\mathcal{R}^* X_{i,2}), \quad (2.23)$$

où $\mathcal{R}^* = \text{diag}(\sqrt{v_1^*}, \dots, \sqrt{v_d^*})$ et comme $\rho = 1$,

$$\left| \sqrt{v_m^*} X_{j,m} - \sqrt{v_m^*} X_{i,m} \right|^2 = h_1^2 v_m^* (1 + \rho^{-2} v_m^*)^{-1} |Z_{ij,m}|^2 \leq h_1^2 v_m^* |Z_{ij,m}|^2,$$

d'où

$$|\mathcal{R}^* X_{j,2} - \mathcal{R}^* X_{i,2}|^2 \leq h_1^2 \max_m v_m^* |Z_{ij}|^2.$$

Finalement

$$|r_{i,j}| \leq \frac{C_{g_1} h_1^2 v_{(1)}^*}{1 - \alpha^*} = C_{g_1} h_1^2 v_{(1)}^*.$$

Cette nouvelle majoration de $|r_{ij}|$, nous permet d'écrire

$$P_{\rho,m}(\mathbb{E} \widehat{\beta}_m^{(1)} - \beta_m^*) = s_m h_1, \quad (2.24)$$

avec $|s_m| \leq C_{g_1} C_V^{1/2} v_{(1)}^*$.

Pour la partie aléatoire,

$$P_{\rho,m}(\widehat{\beta}_m^{(1)} - \mathbb{E} \widehat{\beta}_m^{(1)}) = \sum_{i=1}^n c_{i,m}(\mathbf{U}) \varepsilon_i := \zeta_m. \quad (2.25)$$

Les ε_i sont des variables aléatoires indépendantes de loi $\mathcal{N}(0, \sigma^2)$, donc $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^\top$ est un vecteur gaussien et il en est de même pour ζ_m , car les coefficients sont déterministes. On pose $\eta_m = h_1 \sqrt{n} \zeta_m$ et en reprenant la démonstration de la proposition 2.2.1

$$\mathbb{E} |\zeta_{l,m}|^2 \leq \frac{\sigma^2 2 C_V^2 C_K}{n h_1^2},$$

c'est-à-dire

$$\mathbb{E} |\eta_m|^2 \leq 2L\sigma^2 C_V^2 C_K.$$

Pour montrer l'inégalité (2.18), il suffit d'appliquer directement le lemme 4.3.2 de l'annexe au vecteur gaussien η_m , comme dans la preuve du corollaire 2.2.1 ; c'est pour cette raison que l'on obtient une majoration par $1/n$ et non pas par $2/n$ comme dans la proposition 2.2.1.

Pour l'inégalité sur $\widehat{v}_m^{(1)}$, on utilise le même raisonnement que pour la proposition 2.2.2 : on emploie le lemme 2.2.5 avec $\rho = 1$. ■

2.4 Qualité du l'estimateur final

Dans le prochain théorème et dans le théorème 3.1.1 plus bas, ρ (resp. h) représente ρ_{k_n} (resp. h_{k_n}) à la dernière itération. Rappelons que h et ρ satisfont les conditions $h \geq 1$ et $\rho = (\sigma^2 n^{-1} L \log n)^{1/3}$.

Théorème 2.4.1. *On se place sous les hypothèses H1–H5. Il existe un ensemble aléatoire A avec $\mathbb{P}(A) \geq 1 - 3k_n/n$ et, pour tout $m = 1, \dots, d$, un vecteur aléatoire gaussien de moyenne nulle $\xi_m^* = (\xi_{1m}^*, \dots, \xi_{Lm}^*)^\top \in \mathbb{R}^L$, défini comme une combinaison linéaire des erreurs ε_i avec des coefficients déterministes, qui dépendent uniquement de la largeur de fenêtre "idéale" $\mathbf{b}^* = \mathbf{b}^{*(k_n)}$, des observations (X_1, \dots, X_n) , des fonctions de base $\psi_{lm}(\cdot)$ et du noyau K , tel que*

$$\mathbb{E} |\xi_{lm}^*|^2 \leq 2\sigma^2 C_V^2 C_K, \quad l = 1, \dots, L \text{ et } m = 1, \dots, d,$$

et sur A on observe

$$\begin{aligned} \max_{m=1, \dots, d} \left| P_{\rho, m}(\widehat{\beta}_m - \beta_m^*) - n^{-1/2} \xi_m^* \right| &\leq C (\sigma^2 n^{-1} L \log n)^{2/3}, \\ \max_{m=1, \dots, d} \left| P_{\rho, m}(\widehat{\beta}_m - \beta_m^*) \right| &\leq \delta_n, \\ \max_{m=1, \dots, d} \left| P_{\rho, m}^2(\widehat{v}_m - v_m^*) \right| &\leq \delta_n^2 + 2\delta_n \tau_m, \end{aligned} \quad (2.26)$$

où $C = C(d, M, C_{g_1}, C_V, C_K, C_{K'}, \bar{\psi})$, $\bar{\psi} = \max_{i, l, m} |\psi_{lm}(X_i)|$,

$$\delta_n = \sqrt{2C_V^2 C_K \sigma^2 n^{-1} L z_n^2} + C (\sigma^2 n^{-1} L \log n)^{2/3}. \quad (2.27)$$

et $\tau_m = \rho \sqrt{v_m^*} (\rho^2 + v_m^*)^{-1/2} \leq \min\{\rho, \sqrt{v_m^*}\}$. En particulier pour tout $m \notin \mathcal{J}$, on observe sur A :

$$|\widehat{\beta}_m - n^{-1/2} \xi_m^*| \leq \omega_n, \quad |\widehat{\beta}_m| \leq \delta_n, \quad \left| |\widehat{\beta}_m|^2 - n^{-1} |\xi_m^*|^2 \right| \leq \omega_n^2 + 2\omega_n \delta_n, \quad (2.28)$$

où $\omega_n = C (\sigma^2 n^{-1} L \log n)^{2/3}$.

Soient les nombres h_k et ρ_k décrits comme dans l'algorithme, pour k variant de 1 à k_n . Définissons successivement les valeurs δ_k et celles des composantes $\alpha_{k, m}$ du vecteur $\boldsymbol{\alpha}_k$ de \mathbb{R}^d . Pour $k = 1$, $\boldsymbol{\alpha}_1 = 0$ et δ_1 est donné par (2.19); pour $k = 2, \dots, k_n$

$$\begin{aligned} \delta_k &= \frac{C_{g_1} C_V^{1/2}}{(1 - \alpha_k^*)} h_k \rho_k^2 + \frac{\sqrt{2L} \sigma C_V C_K^{1/2} z_n}{h_k \sqrt{n}} + \frac{\sqrt{L} \sigma \bar{\psi} C_{\alpha_k, n} |\boldsymbol{\alpha}_k|}{h_k \sqrt{n}}, \\ \alpha_{k, m} &= \rho_k^{-2} (2\delta_{k-1} \tau_{k, m} + \delta_{k-1}^2), \quad m = 1, \dots, d, \end{aligned} \quad (2.29)$$

avec

$$\begin{aligned}\alpha_k^* &= \max_{m=1,\dots,d} \alpha_{k,m}, \\ \tau_{k,m} &= \rho_k \sqrt{v_m^*} (\rho_k^2 + v_m^*)^{-1/2} \leq \min\{\rho_k, \sqrt{v_m^*}\}, \\ \bar{\psi} &= \max_{m,l,i} |\psi_{lm}(X_{im})|, \\ z_n \text{ et } C_{\alpha_k,n} &\text{ sont définis dans le corollaire 2.2.1.}\end{aligned}$$

Pour démontrer le théorème 2.4.1, nous avons besoin des deux lemmes suivants.

Lemme 2.4.1. *Pour n suffisamment grand, les α_k satisfont $\max_{k \leq k_n} \alpha_k^* < 1/4$. De plus, pour la dernière itération k_n , on a*

$$\mu_n := \frac{C_{g_1} C_V^{1/2}}{(1 - \alpha_{k_n}^*)} h_{k_n} \rho_{k_n}^2 + \frac{\sqrt{L} \sigma \bar{\psi} C_{\alpha_{k_n},n} |\alpha_{k_n}|}{h_{k_n} \sqrt{n}} \leq C (\sigma^2 n^{-1} L \log n)^{2/3},$$

et $\delta_{k_n} \leq \delta_n$, où δ_n est défini dans (2.27) et C correspond à une constante dépendant de d , M et des constantes intervenant dans les Hypothèses H1–H5.

DÉMONSTRATION. Pour $k > 1$, établissons une borne supérieure pour $|\alpha_k|$

$$\alpha_{k,m} = \rho_k^{-2} (2\delta_{k-1} \tau_{k,m} + \delta_{k-1}^2), \quad (2.30)$$

avec $\tau_{k,m} = \rho_k \sqrt{v_m^*} (\rho_k^2 + v_m^*)^{-1/2}$.

Si $m \notin \mathcal{J}$, alors $\alpha_{k,m} = \delta_{k-1}^2 / \rho_k^2$. Si $m \in \mathcal{J}$, alors $\alpha_{k,m} \leq \frac{2\delta_{k-1}}{\rho_k} + \frac{\delta_{k-1}^2}{\rho_k^2}$.

Ainsi

$$|\alpha_k|^2 \leq (d - M) \left(\frac{\delta_{k-1}^4}{\rho_k^4} \right) + M \left(\frac{2\delta_{k-1}}{\rho_k} + \frac{\delta_{k-1}^2}{\rho_k^2} \right)^2.$$

Si

$$(d - M) \left(\frac{\delta_{k-1}^2}{\rho_k^2} \right) \leq 4M, \quad (2.31)$$

on a alors

$$|\alpha_k|^2 \leq 4M \left(\frac{\delta_{k-1}^2}{\rho_k^2} \right) + M \left(\frac{2\delta_{k-1}}{\rho_k} + \frac{\delta_{k-1}^2}{\rho_k^2} \right)^2,$$

$$|\alpha_k|^2 \leq 2M \left(\frac{2\delta_{k-1}}{\rho_k} + \frac{\delta_{k-1}^2}{\rho_k^2} \right)^2.$$

On démontre la première partie du lemme par récurrence. On remarque tout d'abord que $|\alpha_1| = 0 \leq 1/4$. On suppose ensuite $H_k : \max_{k \leq k_n} \alpha_k^* < 1/4$ et on montre que H_{k+1} est vraie :

Par l'inégalité précédente

$$|\alpha_{k+1}|^2 \leq 2M \left(\frac{2\delta_k}{a_\rho \rho_k} + \frac{\delta_k^2}{a_\rho^2 \rho_k^2} \right)^2.$$

On veut donc majorer δ_k/ρ_k . On peut écrire

$$\frac{\delta_k}{\rho_k} = \frac{C_{g_1} C_V^{1/2}}{1 - \alpha_k^*} h_k \rho_k + \frac{\sqrt{2L}\sigma C_V C_K^{1/2}}{h_k \rho_k} z_n n^{-1/2} + \frac{\sqrt{L}\sigma \bar{\psi} C_{\alpha_k, n} |\alpha_k|}{h_k \rho_k} n^{-1/2},$$

or $h_k \rho_k = h_{k-1} \rho_{k-1} a_h a_\rho < h_{k-1} \rho_{k-1}$, car $a_h a_\rho < 1$, donc la suite $(h_k \rho_k)_k$ est décroissante. De plus $h_k \rho_k < h_1 \rho_1 = C_0 n^{-1/(4\vee d)}$ et

$$\begin{aligned} n^{1/2} h_k \rho_k &> n^{1/2} h_{k_n} \rho_{k_n}, \\ &\geq C_0 n^{1/2} (n^{-1} \sigma^2 \log n L)^{1/3}, \\ &\geq C (\log n)^{1/3} n^{1/6}. \end{aligned}$$

On obtient donc la majoration suivante

$$\frac{\delta_k}{\rho_k} \leq C \frac{n^{-1/(4\vee d)}}{1 - \alpha_k^*} + \left(\sqrt{2L}\sigma C_V C_K^{1/2} z_n + \sqrt{L}\sigma \bar{\psi} C_{\alpha_k, n} |\alpha_k| \right) C' (\log n)^{-1/3} n^{-1/6}.$$

Pour n assez grand, δ_k/ρ_k est assez petit donc $\alpha_{k+1}^* \leq 1/4$, pour n assez grand. De plus l'inégalité sur δ_k/ρ_k permet d'affirmer que l'hypothèse (2.31) est bien vérifiée.

Pour la deuxième partie du lemme, i.e. majorer μ_n , on veut trouver un majorant de $|\alpha_{k_n}|$, pour cela il faut déjà en obtenir un pour δ_{k_n-1} . Il faut donc s'intéresser à chacun des trois termes de (2.29). Par ce qui précède, on a :

$$\begin{aligned} |\alpha_{k_n-1}|^2 &\leq 2M \left(\frac{2\delta_{k-2}}{\rho_{k-1}} + \frac{\delta_{k-2}^2}{\rho_{k-1}^2} \right)^2, \\ |\alpha_{k_n-1}|^2 &\leq 2M \left(\frac{4M}{d-M} + 4\sqrt{\frac{M}{d-M}} \right)^2, \end{aligned}$$

d'où $|\alpha_{k_n-1}|^2 \leq C_1(d, M)$. De plus $h_{k_n-1} \geq C_0/a_h$ et $a_\rho \rho_{k_n} \leq \rho_{\min} \leq \rho_{k_n}$, ce que l'on peut réécrire

$$\rho_{\min}/a_\rho \leq \rho_{k_n-1} \leq \rho_{\min}/a_\rho^2,$$

alors on a

$$\left\{ \begin{array}{l} \frac{C_{g_1} C_V^{1/2}}{1 - \alpha_{k_n-1}} h_{k_n-1} \rho_{k_n-1}^2 \leq C h_{\max} \rho_{\min}^2 / a_\rho^4, \\ \frac{\sqrt{2L}\sigma C_V C_K^{1/2} z_n}{h_{k_n-1} \sqrt{n}} \leq C a_h z_n n^{-1/2}, \\ \frac{\sqrt{L}\sigma \bar{\psi} C_{\alpha_{k_n-1}, n} |\alpha_{k_n-1}|}{h_{k_n-1} \sqrt{n}} \leq \sqrt{L}\sigma \bar{\psi} a_h C_{\alpha_{k_n-1}, n} C_1 n^{-1/2}. \end{array} \right.$$

Or pour n assez grand, $z_n \leq 2\sqrt{\log n}$ et $C_{\alpha_{k_n-1}, n} \leq C\sqrt{\log n}$, donc

$$\delta_{k_n-1} \leq C(\sigma^2 n^{-1} L \log n)^{1/2}.$$

On peut maintenant borner $|\alpha_{k_n}|$

$$\begin{aligned} |\alpha_{k_n}|^2 &\leq 2M \left[\frac{C_2 \sigma^2 n^{-1} L \log n}{(\sigma^2 n^{-1} L \log n)^{2/3}} + 2C \frac{(\sigma^2 n^{-1} L \log n)^{1/2}}{(\sigma^2 n^{-1} L \log n)^{1/3}} \right], \\ |\alpha_{k_n}|^2 &\leq C' (\sigma^2 n^{-1} L \log n)^{1/3}, \\ |\alpha_{k_n}| &\leq C' (\sigma^2 n^{-1} L \log n)^{1/6}. \end{aligned}$$

Comme $h_{\max} \geq h_{k_n} \geq 1$ et que ρ_{k_n} est de l'ordre de $(\sigma^2 n^{-1} L \log n)^{1/3}$ on peut majorer μ_n de la façon suivante

$$\begin{aligned} \mu_n &\leq \frac{4C_{g_1} C_V^{1/2}}{3} h_{\max} (\sigma^2 n^{-1} L \log n)^{2/3} + \frac{\sqrt{L} \sigma \bar{\Psi} C'}{C_0} (\sigma^2 n^{-1} L \log n)^{1/6} n^{-1} C_{\alpha_{k_n}, n}, \\ \mu_n &\leq C (\sigma^2 n^{-1} L \log n)^{2/3}. \end{aligned}$$

Cette dernière majoration et le fait que $h_{k_n} \geq 1$, nous permet de conclure

$$\delta_{k_n} \leq \sqrt{2C_V^2 C_K \sigma^2 n^{-1} L z_n^2} + C (\sigma^2 n^{-1} L \log n)^{2/3}. \quad (2.32)$$

■

Lemme 2.4.2. *Pour n suffisamment grand, il existe $\mathcal{A}_1 \supseteq \dots \supseteq \mathcal{A}_{k_n-1}$, des ensembles aléatoires, tels que $\mathbb{P}(\mathcal{A}_k) \geq 1 - \frac{3k}{n}$ et sur \mathcal{A}_k on a*

$$\max_{m=1, \dots, d} |P_{\rho_{k+1}, m}(\widehat{\beta}_m^{(k)} - \beta_m^*)| \leq \delta_k, \quad k = 1, \dots, k_n - 1.$$

DÉMONSTRATION. On démontre ce résultat par récurrence sur k .

Pour $k = 1$, on utilise les résultats d'exactitude à la première itération donné à la proposition 2.3.1 et le fait que

$$P_{\rho_2, m} = (1 + \rho_2^{-2} v_m^*)^{-1/2} \leq 1 = P_{\rho_1, m},$$

alors sur \mathcal{A}_1 , $\mathbb{P}(\mathcal{A}_1) \geq 1 - 1/n \geq 1 - 3/n$

$$\max_{m=1, \dots, d} |P_{\rho_2, m}(\widehat{\beta}_m^{(1)} - \beta_m^*)| \leq \max_{m=1, \dots, d} |\widehat{\beta}_m^{(1)} - \beta_m^*| \leq \delta_1.$$

Montrons que $H_k \Rightarrow H_{k+1}$: on veut alors utiliser la propriété 2.2.1, qui donne le gain à chaque étape et pour cela il faut donc montrer que $\mathbf{U}^{(k+1)} \in \mathcal{U}_{\alpha_{k+1}}$.

Pour $m \in \{1, \dots, d\}$, sur \mathcal{A}_k , tel que $\mathbb{P}(\mathcal{A}_k) \geq 1 - 3k/n$, en utilisant le lemme 2.2.5, on obtient une inégalité sur les $\widehat{v}_m^{(k)}$ et on peut écrire

$$\begin{aligned} |U_m^{(k+1)} - 1| &= |P_{\rho_{k+1}, m}^2 (1 + \rho_k^{-2} \widehat{v}_m^{(k)}) - 1|, \\ &= |P_{\rho_{k+1}, m}^2 \rho_{k+1}^{-2} (\widehat{v}_m^{(k)} - v_m^*)|, \\ &\leq \rho_{k+1}^{-2} (\delta_k^2 + 2\delta_k \tau_{k+1, m}) = \alpha_{k+1, m}. \end{aligned}$$

Pour conclure, il suffit d'utiliser le corollaire 2.2.1 et le fait que $\rho_{k+2} < \rho_{k+1}$, donc que $P_{\rho_{k+2},m} \leq P_{\rho_{k+1},m}$ sur $\mathcal{A}_k \cap A_{k+1}$. Ainsi il existe un ensemble A_{k+1} tel que $\mathbb{P}((\mathcal{A}_k \cap A_{k+1})^c) \leq 3k/n + 3/n$, on a

$$|P_{\rho_{k+2},m}(\widehat{\beta}_m^{(k+1)} - \beta_m^*)| \leq \delta_{k+1}.$$

Ceci achève la démonstration du lemme. ■

DÉMONSTRATION DU THÉORÈME 2.4.1.

D'après le lemme précédent, on sait qu'il existe un ensemble aléatoire \mathcal{A}_{k_n-1} , tel que $\mathbb{P}(\mathcal{A}_{k_n-1}) \geq 1 - \frac{3k_n-3}{n}$ et sur cet ensemble,

$$\max_{m=1,\dots,d} |P_{\rho_{k_n},m}(\widehat{\beta}_m^{(k_n-1)} - \beta_m^*)| \leq \delta_{k_n-1}.$$

Le vecteur $\mathbf{U}^{(k_n)}$ admet les coordonnées $U_m^{(k_n)} = P_{\rho_{k_n},m}^2(1 + \rho_{k_n}^{-2}\widehat{v}_m^{(k_n-1)})$ ($m = 1, \dots, d$). En utilisant le même raisonnement que dans la démonstration du lemme 2.4.2, sur \mathcal{A}_{k_n-1} on obtient $|U_m^{(k_n)} - 1| \leq \alpha_{k_n,m}$. On introduit alors le L -vecteur gaussien ξ_m provenant de la proposition 2.2.1 appliquée avec $h = h_{k_n}$ et $\rho = \rho_{k_n}$. Ainsi l'existence d'un ensemble A_{k_n} est assurée, tel que $\mathbb{P}(A_{k_n}) \geq 1 - 2/n$ et tel que sur A_{k_n} , pour tout $\mathbf{U} \in \mathcal{U}_{\alpha_{k_n}}$

$$\max_{m=1,\dots,d} \left| P_{\rho_{k_n},m}(\widehat{\beta}_m^{(k_n)}(\mathbf{U}) - \beta_m^*) - \frac{\xi_m}{h\sqrt{n}} \right| \leq \mu_n,$$

où μ_n est défini dans le lemme 2.4.1. En posant alors $\mathcal{A}_{k_n} = \mathcal{A}_{k_n-1} \cap A_{k_n}$, on a

$$\mathbb{P}(\mathcal{A}_{k_n}) \geq 1 - \frac{3k_n - 1}{n}.$$

On sait que d'après ce qui précède que sur l'ensemble \mathcal{A}_{k_n} , $\mathbf{U}^{(k_n)} \in \mathcal{U}_{\alpha_{k_n}}$, et

$$\max_{m=1,\dots,d} \left| P_{\rho_{k_n}}(\widehat{\beta}_m - \beta_m^*) - n^{-1/2}\xi_m^* \right| \leq C(\sigma^2 n^{-1} L \log n)^{2/3},$$

où $\xi_m^* = h^{-1}\xi_m$ et $\widehat{\beta}_m = \widehat{\beta}_m^{(k_n)}$.

Comme $h = h_{k_n} \geq 1$

$$\mathbb{E} |\xi_{lm}^*|^2 = h^{-2} \mathbb{E} |\xi_{lm}|^2 \leq 2\sigma^2 C_V^2 C_K.$$

On a donc établi les deux premières assertions du théorème. Pour les deux suivantes, il suffit d'utiliser le lemme 4.3.1 et le lemme 2.2.5, comme dans la démonstration du corollaire, ce qui nous donne un nouvel ensemble aléatoire $\mathcal{A}_{k_n}^*$ avec $\mathbb{P}(\mathcal{A}_{k_n}^*) \geq 1 - 1/n$. Finalement sur $A = \mathcal{A}_{k_n} \cap \mathcal{A}_{k_n}^*$, tel que $\mathbb{P}(A) \geq 1 - 3k_n/n$,

$$\max_{m=1,\dots,d} |P_{\rho,m}(\widehat{\beta}_m - \beta_m^*)| \leq \delta_n.$$

Les trois dernières inégalités (2.28) correspondent aux trois majorations (2.26) appliquées dans le cas où $m \notin \mathcal{J}$, i.e. $\beta_m^* = 0$ et $P_{\rho_m} = 1$, on obtient donc

$$\begin{aligned} |\widehat{\beta}_m - n^{-1/2}\xi_m| &\leq w_n, \\ |\widehat{\beta}_m| &\leq \delta_n, \\ \|\widehat{\beta}_m\|^2 - n^{-1}|\xi_m|^2 &\leq w_n^2 + 2\delta_n w_n. \end{aligned}$$

■

2.5 Estimation de la variance

Nous utilisons l'estimateur $\widehat{\sigma}^2$ défini dans la partie 1.4, pour lequel nous obtenons la convergence à la vitesse \sqrt{n} . Il nous faut renforcer les hypothèses

Hypothèse H6. [Design] *Pour $M \geq 1$, soit $N_{i,1} = \#\{X_j / \|\mathbf{b}^{-1}X_{i,j}\| \leq 1\}$, on suppose qu'il existe C_1 dépendant uniquement de d tel que*

$$\frac{1}{n} \sum_{i=1}^n N_{i,1} \leq C_1 n \rho^M.$$

Hypothèse H7. *Il existe une constante $C > 0$ telle que*

$$\frac{\lambda_{\max}}{\lambda_{\min}} \leq C,$$

où λ_i sont les valeurs propres de la matrice $A^\top A$ définie dans la démonstration ci-dessous, on suppose de plus que le nombre de valeurs propres nulles est fini égale à N_A .

On peut remarquer que cette hypothèse est vérifiée dans le cas où les X_i sont à densité continue, sur un ensemble de probabilité tendant vers 1, quand la taille de l'échantillon tend vers l'infini.

Théorème 2.5.1. *Sous les hypothèses H1–H7, il existe une constante C_σ dépendante des constantes intervenant dans les hypothèses précédentes telles que*

$$\mathbb{P}(\sqrt{n}|\widehat{\sigma}^2 - \sigma^2| > C_\sigma \sigma^2 \gamma) \leq 2e^{-\gamma^2/4} + e^{-\gamma\sqrt{2}/6} + 3k_n/n.$$

DÉMONSTRATION.

$$\begin{aligned} \widehat{\sigma}^2 &= \sum_{j=1}^n c_{i,j} Y_j, \\ &= \sum_{j=1}^n c_{i,j} g(X_j) + \sum_{j=1}^n c_{i,j} \varepsilon_j, \end{aligned}$$

où les $c_{i,j}$ sont donnés par

$$c_{i,j} = \delta_{ij} - \begin{pmatrix} 1 \\ 0_d \end{pmatrix}^\top (V_i)^{-1} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} K_d(X_{ij}, \mathbf{b}) = \delta_{ij} - a_{ij}.$$

Ces coefficients vérifient

$$\begin{aligned} \sum_{j=1}^n c_{i,j} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix}^\top &= - \left[\begin{pmatrix} 1 \\ 0_d \end{pmatrix}^\top V_i^{-1} \sum_{j=1}^n \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix}^\top K_d(X_{ij}, \mathbf{b}) - \begin{pmatrix} 1 \\ 0_d \end{pmatrix}^\top \right], \\ &= - \begin{pmatrix} 1 \\ 0_d \end{pmatrix}^\top + \begin{pmatrix} 1 \\ 0_d \end{pmatrix}^\top. \end{aligned}$$

On obtient donc les relations suivantes

$$\begin{cases} \sum_{j=1}^n a_{ij} = 1, \\ \sum_{j=1}^n a_{ij} X_{ij} = 0_d. \end{cases} \quad (2.33)$$

On pose $s_i^2 = \sum_{j=1}^n c_{ij}^2$ et l'estimateur de σ^2 a pour expression

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n \frac{\hat{\varepsilon}_i^2}{s_i^2} = \sum_{i=1}^n (b_i + \xi_i)^2,$$

avec

$$b_i = n^{-1/2} s_i^{-1} \sum_{j=1}^n c_{ij} g(X_j), \quad (2.34)$$

$$\xi_i = n^{-1/2} s_i^{-1} \sum_{j=1}^n c_{ij} \varepsilon_j. \quad (2.35)$$

On va adopter une écriture vectorielle :

- $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^\top$,
- $\mathbf{b} = (b_1, \dots, b_n)^\top$,
- $A = (\alpha_{i,j})_{i,j}$ matrice $n \times n$, avec $\alpha_{i,j} = n^{-1/2} s_i^{-1} c_{ij}$.

On peut donc réécrire $\hat{\sigma}^2$ de la façon suivante

$$\hat{\sigma}^2 = \mathbf{b}^\top \mathbf{b} + 2\mathbf{b}^\top A \varepsilon + \varepsilon^\top A^\top A \varepsilon.$$

Il y a donc trois termes à traiter séparément : le biais, une variable aléatoire normale, Q_1 , et une forme quadratique de variables aléatoires gaussiennes, Q_2 .

Le biais : $b^\top b$

En utilisant (2.33), on peut écrire

$$\begin{aligned} \sum_{i=1}^n b_i^2 &= \sum_{i=1}^n n^{-1} s_i^{-2} \left(\sum_{j=1}^n a_{ij} g(X_j) - g(X_i) \right)^2, \\ &= \sum_{i=1}^n n^{-1} s_i^{-2} \left(\sum_{j=1}^n a_{ij} [g(X_j) - g(X_i) - X_{ij}^\top \nabla g(X_i)] \right)^2, \\ &= \sum_{i=1}^n n^{-1} s_i^{-2} \left(\sum_{j=1}^n a_{ij} r_{ij} \right)^2. \end{aligned}$$

Alors on utilise la même méthode que dans la majoration du biais de la proposition 2.2.1, en remarquant que dans le cas paramétrique, $M = 0$, $r_{ij} = 0$.

$$\begin{aligned} \left| \sum_{j=1}^n a_{ij} r_{ij} \right| &\leq \sum_{j=1}^n |a_{ij}| |r_{ij}|, \\ &\leq \frac{C_{g_1} h^2 \rho^2}{(1 - \alpha^*)} \sum_{j \neq i} |a_{ij}|. \end{aligned}$$

En reprenant les notations de l'hypothèse H6 et en utilisant Cauchy-Schwartz, on obtient

$$\begin{aligned} \sum_{i=1}^n n^{-1} s_i^{-2} \left(\sum_{j=1}^n a_{ij} r_{ij} \right)^2 &\leq \sum_{i=1}^n n^{-1} s_i^{-2} \frac{C_{g_1} h^4 \rho^4}{(1 - \alpha^*)^2} \left(\sum_{j \neq i} |a_{ij}|^2 \right) N_{i,1}, \\ &\leq \frac{C_{g_1} h^4 \rho^4}{(1 - \alpha^*)^2} n^{-1} \sum_{i=1}^n N_{i,1}. \end{aligned}$$

Ainsi pour $M \geq 1$,

$$\sum_{i=1}^n n^{-1} s_i^{-2} \left(\sum_{j=1}^n a_{ij} r_{ij} \right)^2 \leq \frac{C_1 C_{g_1} h^4 n \rho^{4+M}}{(1 - \alpha^*)^2}.$$

La forme linéaire Q_1

Q_1 est une combinaison linéaire des variables aléatoires gaussiennes, ε_i centrées et de variance σ^2 et a pour expression

$$Q_1 = 2b^\top A\varepsilon.$$

Il est immédiat que son espérance est nulle et on veut maîtriser sa variance :

$$\begin{aligned} \mathbb{E}(Q_1^2) &= 4\mathbb{E}(b^\top A\varepsilon\varepsilon^\top A^\top b), \\ &= 4b^\top A\mathbb{E}(\varepsilon\varepsilon^\top)A^\top b = 4\sigma^2 b^\top AA^\top b, \\ &\leq 4\sigma^2 |b|^2 \lambda_A. \end{aligned}$$

où λ_A est la plus grande valeur propre de AA^\top . En utilisant le lemme 4.3.1, Q_1 étant une variable aléatoire normale :

$$\mathbb{P}(|Q_1| > \gamma(2\lambda_A)^{1/2}\sigma|b|) \leq \exp\left(-\frac{\gamma^2(2\lambda_A)\sigma^2|b|^2}{2\mathbb{E} Q_1^2}\right) \leq e^{-\gamma^2/4},$$

ce qui nous donne le résultat pour la deuxième partie. On pose $S_A^2 = 2\text{trace}(A^\top A)^2$, on a donc $\lambda_A \leq \left(\frac{S_A^2}{2}\right)^{1/2}$ et

$$\mathbb{P}(|Q_1| > \gamma(2S_A)^{1/2}\sigma|b|) \leq e^{-\gamma^2/4},$$

La forme quadratique Q_2

La forme quadratique Q_2 s'écrit

$$Q_2 = \varepsilon^\top A^\top A \varepsilon - \sigma^2 \text{trace}(A^\top A).$$

La matrice $A^\top A$ est symétrique positive, elle est donc diagonalisable à l'aide d'une matrice orthogonale, notée P . On note λ_i ses valeurs propres et on peut écrire

$$Q_2 = \sum_{i=1}^n \lambda_i \tilde{\varepsilon}_i^2 - \sigma^2,$$

où le vecteur $\tilde{\varepsilon} = P\varepsilon$ est aussi un vecteur gaussien de \mathbb{R}^n centré et de matrice de variance-covariance : $\text{diag}(\sigma^2, \dots, \sigma^2)$.

On veut maintenant majorer $\mathbb{P}(Q_2 > z)$, pour cela on utilise la remarque précédente et l'inégalité de Chebychev, pour tout $\mu > 0$:

$$\begin{aligned} \mathbb{P}(Q_2 > z) &\leq e^{-\mu z} \mathbb{E} e^{\mu Q_2}, \\ &\leq e^{-\mu z} \mathbb{E} \exp\left(\mu \sum_{i=1}^n \lambda_i (\tilde{\varepsilon}_i^2 - \sigma^2/n)\right), \\ &\leq \exp\{-\mu z - \mu\sigma^2\} \mathbb{E} \prod_{i=1}^n \exp(\mu \lambda_i \tilde{\varepsilon}_i^2). \end{aligned}$$

car $\text{trace}(A^\top A) = \sum_{i=1}^n \lambda_i = 1$. Or les variables $\tilde{\varepsilon}_i$ sont des combinaisons linéaires de variables aléatoires normales de moyenne nulle et de variance σ^2 . De plus ces variables sont indépendantes. Si $2\mu\lambda_i\sigma^2 < 1$ pour tout i , on peut écrire

$$\begin{aligned} \mathbb{P}(Q_2 > z) &\leq \exp\{-\mu z - \mu\sigma^2\} \prod_{i=1}^n \mathbb{E} \exp(\mu \lambda_i \tilde{\varepsilon}_i^2), \\ &\leq \exp\{-\mu z - \mu\sigma^2\} \prod_{i=1}^n \frac{1}{(1 - 2\mu\lambda_i\sigma^2)^{1/2}}, \\ &\leq \exp\left\{-\mu z - \mu\sigma^2 - \frac{1}{2} \sum_{i=1}^n \log(1 - 2\mu\lambda_i\sigma^2)\right\}. \end{aligned}$$

On va utiliser le fait que

$$-\log(1-u) \leq u + u^2, \quad \text{pour } u \leq 2/3.$$

Si cette condition est vérifiée, on peut alors affirmer

$$\begin{aligned} \mathbb{P}(Q_2 > z) &\leq \exp \left\{ -\mu z - \mu \sigma^2 + \frac{1}{2} \sum_{i=1}^n (2\mu \lambda_i \sigma^2 + 4\mu^2 \lambda_i^2 \sigma^4) \right\}, \\ &\leq \exp \left\{ -\mu z + 2 \left(\sum_{i=1}^n \lambda_i^2 \right) \mu^2 \sigma^4 \right\}. \end{aligned}$$

En posant $S_A^2 = 2\text{trace}[(A^\top A)^2]$, et $z = \gamma S_A \sigma^2$

$$\mathbb{P}(Q_2 > \gamma S_A \sigma^2) \leq \exp \{ -\mu \gamma S_A \sigma^2 + \mu^2 S_A^2 \sigma^4 \}.$$

Il reste alors à choisir μ sachant qu'il doit vérifier pour tout i , $2\mu \lambda_i \sigma^2 < 2/3$.

Si $\mu < 1/3\lambda_A \sigma^2$, où $\lambda_A = \max_i \lambda_i$, alors on a la dernière majoration. On doit distinguer deux cas suivant la valeur de γ :

- si $\gamma \leq \frac{2S_A}{3\lambda_A}$, on prend $\mu = \frac{\gamma}{2S_A \sigma^2}$,
- si $\gamma > \frac{2S_A}{3\lambda_A}$, on prend $\mu = \frac{1}{3\lambda_A \sigma^2}$.

donc

$$\mathbb{P}(Q_2 > \gamma S_A \sigma^2) \leq e^{-\gamma^2/4} + e^{-\gamma S_A/6\lambda_A}.$$

Or $S_A/\lambda_A > \sqrt{2}$ et par la majoration de la trace de $(A^\top A)^2$ obtenue précédemment on a

$$\mathbb{P}(Q_2 > \gamma S_A \sigma^2) \leq e^{-\gamma^2/4} + e^{-\gamma\sqrt{2}/6}.$$

Conclusion : Avec une probabilité d'au moins $1 - 2e^{-\gamma^2/4} - e^{-\gamma\sqrt{2}/6}$ on a

$$\hat{\sigma}^2 - \sigma^2 \leq \frac{C_1 C_{g_1} h^4 n \rho^{4+M}}{(1-\alpha^*)^2} + \gamma \sqrt{2} \sigma \frac{(C_1 C_{g_1} h^4 \rho^{4+M})^{1/2} S_A^{1/2}}{(1-\alpha^*)} + \gamma \sigma^2 S_A.$$

Établissons l'ordre de S_A :

$$S_A^2/2 = \sum_{i=1}^n \lambda_i^2 \leq (\sum_{i=1}^n \lambda_i) \lambda_{\max} \leq \lambda_{\max}$$

car $\sum_{i=1}^n \lambda_i = 1$. De même on peut déduire de cette égalité que $\lambda_{\min} \leq 1/(n - N_A)$. En utilisant alors l'hypothèse H7, on obtient $\lambda_{\max} \leq C/(n - N_A)$. Ceci se traduit pour S_A par

$$S_A^2 \leq 2\lambda_{\max} \leq 2C/(n - N_A).$$

Finalement on a bien $S_A = \mathcal{O}(n^{-1/2})$. Alors comme ρ est de l'ordre de $(n^{-1} \log n)^{-1/3}$, on a bien la convergence en $n^{1/2}$ de l'estimateur de la variance vers la vraie valeur. ■

Chapitre 3

Exploration du modèle partiellement linéaire

Cette partie consiste en l'exploitation des résultats obtenus par la procédure itérative, en particulier de l'estimation de la mesure de non-linéarité, v_m^* . En effet ces résultats nous assurent une bonne séparation entre régresseurs linéaires et non-linéaires, si v_m^* est suffisamment séparé de 0, pour $m \in \mathcal{J}$. Cette propriété fait l'objet de la première partie. Dans un deuxième temps, nous utiliserons ce résultat pour définir un test sur le nombre maximal de composantes non-linéaires \mathcal{M} dans le modèle. Ce test est à la base de l'estimation du nombre de composantes non-linéaires M et de l'ensemble des indices correspondants, \mathcal{J} . Dans un dernier point, nous étudierons l'estimation du vecteur θ , correspondant à la composante linéaire de la régression.

3.1 Séparation entre composantes linéaires et non-linéaires

Le théorème suivant établit la séparation entre les composantes linéaires et non-linéaires. Il est une conséquence directe du théorème 2.4.1.

Théorème 3.1.1. *Soit $u_n = \delta_n/\rho < \sqrt{2} - 1$, où $\rho = \rho_{k_n}$ et δ_n est donné par (2.27). Soit r un réel, tel que $r \geq 1$. Si pour tout $m \in \mathcal{J}$, $v_m^* > (rs_r\delta_n)^2$, avec*

$$s_r = \frac{1 + \sqrt{1 + (r^2 + 1)(1 - u_n^2 - 2u_n)}}{1 - u_n^2 - 2u_n},$$

alors sur A , défini dans le théorème 2.4.1,

$$\hat{v}_m > r^2\delta_n^2, \text{ pour } m \in \mathcal{J} \text{ et } \hat{v}_m \leq \delta_n^2 \text{ pour } m \notin \mathcal{J}.$$

DÉMONSTRATION. Le théorème 2.4.1 donne l'existence d'un ensemble aléatoire A sur lequel $\hat{v}_m \leq \delta_n^2$, pour $m \notin \mathcal{J}$. Montrons que $\hat{v}_m > r^2\delta_n^2$, pour $m \in \mathcal{J}$, sur A .

Soit $m \in \mathcal{J}$, sur A on a

$$|\hat{v}_m - v_m^*| \leq P_{\rho,m}^{-2}(\delta_n^2 + 2\delta_n\tau_m),$$

d'où

$$\widehat{v}_m \geq v_m^* - P_{\rho,m}^{-2}(\delta_n^2 + 2\delta_n\tau_m),$$

$$\text{or } \tau_m = \sqrt{v_m^*}(1 + \rho^{-2}v_m^*)^{-1/2} = P_{\rho,m}\sqrt{v_m^*},$$

$$\widehat{v}_m \geq s^2\delta_n^2 - \delta_n^2(1 + \rho^{-2}v_m^*) - 2\delta_n(1 + \rho^{-2}v_m^*)^{1/2}\sqrt{v_m^*}.$$

On pose $s^2 = v_m^*/\delta_n^2$ et $u = \delta_n/\rho$.

$$\widehat{v}_m \geq s^2\delta_n^2 - \delta_n^2 - u^2s^2\delta_n^2 - 2s\delta_n^2(1 + u^2s^2)^{1/2},$$

$$\frac{\widehat{v}_m}{\delta_n^2} \geq s^2 - 1 - u^2s^2 - 2s(1 + u^2s^2)^{1/2},$$

$$\frac{\widehat{v}_m}{\delta_n^2} \geq s^2(1 - u^2) - 1 - 2s(1 - us),$$

$$\frac{\widehat{v}_m}{\delta_n^2} \geq s^2(1 - u^2 - 2u) - 1 - 2s.$$

On définit $h(s) = s^2(1 - u^2 - 2u) - 1 - 2s$. On cherche une condition sur s pour que $h(s) > r^2$. En étudiant la fonction on trouve que si $s > s_r$ alors $\widehat{v}_m > \delta_n^2 r^2$.

On suppose dans notre cas, $r \geq 1$ et $v_m^* > (rs_r\delta_n)^2$, donc $s \geq rs_r > s_r$ et on obtient la conclusion : $\frac{\widehat{v}_m}{\delta_n^2} \geq r^2$.

Le résultat pour les composantes linéaires est donné par l'inégalité (2.28) du théorème 2.4.1.

■

Ce théorème nous assure, en prenant $r = 1$, que

- pour $m \in \mathcal{J}$, si $v_m^* \geq (s_1\delta_n)^2$, alors avec une grande probabilité $\widehat{v}_m \geq \delta_n^2$,
- pour $m \notin \mathcal{J}$, le théorème 2.4.1 donne $\widehat{v}_m < \delta_n^2$.

Si on ordonne les coordonnées de l'estimation $\widehat{\mathbf{v}}$:

$$\widehat{v}_{(1)} \geq \widehat{v}_{(2)} \geq \dots \geq \widehat{v}_{(d)},$$

les M plus grandes valeurs correspondent aux composantes non-linéaires. On a donc obtenu une condition suffisante de séparation des composantes linéaires et non-linéaires :

$$\min_{m \in \mathcal{J}} v_m^* \geq (s_1\delta_n)^2.$$

On se sert de ce résultat dans le schéma de rééchantillonnage, que l'on décrit dans le paragraphe suivant. Mais dans ce cas, on veut s'assurer d'une plus grande séparation, donc on choisit $r = s_1 > 1$.

Étant donné les ordres de grandeur de δ_n et de ρ , u_n tend vers 0 quand n tend vers l'infini. Ainsi, si n est suffisamment grand, s_r est une quantité bornée, dont la borne ne dépend que de r . Le seuil $t^* = (rs_r\delta_n)^2$, qui permet d'avoir une bonne séparation entre des composantes de nature différente, est donc de l'ordre de δ_n^2 (sur A , ensemble de grande probabilité).

3.2 Test d'une hypothèse sur M

Dans cette partie nous allons développer un test portant sur le nombre de composantes non-linéaires dans le modèle, M . Dans toute la suite on se donne $\mathcal{M} \in \{0, \dots, d\}$, quantité qui représente le nombre de composantes non-linéaires hypothétiques. Nous définissons l'ensemble des indices associés à ces composantes :

$$\widehat{\mathcal{J}}_{\mathcal{M}} = \{i / \widehat{v}_i \text{ est une des } \mathcal{M} \text{ plus grandes valeurs de } \widehat{v}_m\}.$$

On remarque immédiatement que grâce au théorème 3.1.1,

$$\text{si } M \leq \mathcal{M}, \text{ alors } \mathcal{J} \subseteq \widehat{\mathcal{J}}_{\mathcal{M}}.$$

En effet pour $m \in \mathcal{J}$, $\widehat{v}_m > r^2 \delta_n^2$ et pour $m \notin \mathcal{J}$, la deuxième relation de (2.28) du théorème 2.4.1, donne immédiatement $\widehat{v}_m^2 \leq \delta_n^2$. Ces deux inégalités assurent que les plus grandes valeurs de \widehat{v}_m correspondent à la partie non-linéaire. Ceci peut encore s'écrire $\mathcal{J} \subset \widehat{\mathcal{J}}_{\mathcal{M}}$.

Cette remarque nous pousse à ne pas tester directement $\mathcal{M} = M$, mais $M \leq \mathcal{M}$ et à utiliser les coordonnées de $\widehat{\mathbf{v}}$ rangées dans l'ordre décroissant. Pour cela on décide de rejeter $M \leq \mathcal{M}$, si $\widehat{v}_{(\mathcal{M}+1)}$ est suffisamment positif, car sous \mathcal{H}_0 l'indice $(\mathcal{M} + 1)$ correspond à une composante linéaire, donc $v_{(\mathcal{M}+1)}^* = 0$.

Si on se place dans le cas idéal : le vrai modèle vérifie $M \leq \mathcal{M}$ et $\min_{m \in \mathcal{J}} v_m^* \geq t^* = (r s_r \delta_n)^2$ où $r \geq 1$.

- Pour $m \in \mathcal{J}$, le théorème 3.1.1 implique qu'avec une forte probabilité, i.e. sur A , \widehat{v}_m est séparé de zéro d'une distance de l'ordre δ_n^2 .
- Pour $m \notin \mathcal{J}$, le théorème 2.4.1 nous donne une approximation de la distribution de \widehat{v}_m . Elle ne dépend que très peu de la fonction de régression g . En effet la variable aléatoire $|\xi_m^*|^2$ vers laquelle elle converge ne dépend de g qu'à travers la largeur de fenêtre idéale b^* .

Ainsi sous $\mathcal{H}_{\mathcal{M}}$, l'estimateur $\widehat{v}_{(\mathcal{M}+1)}$ a une distribution très peu dépendante de g . Si sa loi limite était explicite on aurait accès au seuil asymptotique. Ici on a opté pour une méthode d'obtention du seuil par bootstrap.

Le paragraphe précédent donne des propriétés remarquables pour les estimateurs \widehat{v}_m . C'est pour cette raison que nous décidons de rééchantillonner suivant un modèle qui conserve les valeurs $\widehat{v}_{(1)} \geq \widehat{v}_{(2)} \geq \widehat{v}_{(3)} \dots \geq \widehat{v}_{(\mathcal{M}+1)}$ et pour lequel \mathbf{v}^* est tout à fait connu.

La première étape définit la nouvelle fonction de régression $\widetilde{g}_{\mathcal{M}}$ qui aura exactement \mathcal{M} composantes non-linéaires. La deuxième étape explique l'obtention de l'échantillon bootstrapé. La dernière étape concerne la procédure d'obtention du seuil t_{α} , pour un niveau $\alpha < 1$ fixé, par une méthode du type Monté-Carlo.

1° Dans un premier temps, on construit une nouvelle fonction $\widetilde{g}_{\mathcal{M}}$, qui a exactement \mathcal{M} composantes non-linéaires. Celles-ci correspondent aux indices contenus dans l'ensemble

$\widehat{\mathcal{J}}_{\mathcal{M}}$. De plus les quantités $\widetilde{\beta}_{l,m}^*$ de $\widetilde{g}_{\mathcal{M}}$ correspondent aux $\widehat{\beta}_{l,m}$, c'est-à-dire

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial \widetilde{g}_{\mathcal{M}}(X_i)}{\partial x_m} \psi_{lm}(X_{i,m}) = \begin{cases} \widehat{\beta}_{lm} & \text{si } m \in \widehat{\mathcal{J}}_{\mathcal{M}}, \\ 0 & \text{sinon,} \end{cases} \quad \text{pour } l = 1, \dots, L.$$

La partie linéaire de $\widetilde{g}_{\mathcal{M}}$ est identiquement nulle. Cette définition nous assure une composante $(\mathcal{M} + 1)$ linéaire.

Par exemple, on peut construire $\widetilde{g}_{\mathcal{M}}$ comme une combinaison linéaire des $\frac{x_m^l}{l}$ pour $l = 2, \dots, L + 1$:

$$\widetilde{g}_{\mathcal{M}}(\mathbf{x}) = \sum_{m \in \widehat{\mathcal{J}}_{\mathcal{M}}} \sum_{l=2}^{L+1} c_{l-1,m} x_m^l / l, \quad (3.1)$$

où les coefficients $c_{l,m}$ sont alors définis par :

$$\Psi_m c_m = \widehat{\beta}_m, \quad m \in \widehat{\mathcal{J}}_{\mathcal{M}}, \quad (3.2)$$

avec

– Ψ_m est une matrice $L \times L$ d'élément :

$$\Psi_{m,l'l} = \frac{1}{n} \sum_{i=1}^n X_{i,m}^l \psi_{l'm}(X_{i,m}), \quad l, l' = 1, \dots, L;$$

- $c_m \in \mathbb{R}^L$ de coefficients $c_{l,m}$;
- $\widehat{\beta}_m \in \mathbb{R}^L$ de coefficients $\widehat{\beta}_{l,m}$.

2° On peut alors définir le modèle de rééchantillonnage

$$\widetilde{Y}_i = \widetilde{g}_{\mathcal{M}}(X_i) + \widehat{\sigma}_i \widetilde{\varepsilon}_i,$$

où $\widetilde{\varepsilon}_i$ sont des variables aléatoires i.i.d. de loi normale centrée réduite. Pour l'estimateur de la variance il y a deux possibilités :

- Soit prendre $\widehat{\sigma}_i^2 = \widehat{\varepsilon}_i^2$, qui correspond à un “wild bootstrap”. Cet estimateur peut être utilisé dans le cas des bruits hétérogènes.
- Soit prendre $\widehat{\sigma}_i^2 = \widehat{\sigma}^2$, explicité au paragraphe 2.5. Ce dernier cas correspond mieux au cas d'un bruit homogène.

3° L'estimation de la valeur critique se fait de la façon suivante :

1. On génère $\widetilde{Y}_i = \widetilde{g}_{\mathcal{M}}(X_i) + \widehat{\sigma}_i \widetilde{\varepsilon}_i$, pour tout $i = 1, \dots, n$, où $\widetilde{\varepsilon}_i$ provient d'une loi normale centrée réduite.

2. On utilise $\{\tilde{Y}_i, X_i : i = 1, \dots, n\}$ pour estimer $\tilde{\beta}_{lm}^*$ par l'estimateur (1.2) basé sur l'estimation du gradient avec la fenêtre correspondant à la dernière itération $\mathbf{b} = \mathbf{b}^{(k_n)}$. On note cet estimateur $\tilde{\beta}_{lm}$.
On calcule alors

$$\tilde{v}_m = \sum_{l=1}^L |\tilde{\beta}_{lm}|^2, \quad \text{pour tout } m = 1, \dots, d.$$

La statistique \tilde{T}_M correspond alors à $\tilde{v}_{(\mathcal{M}+1)}$.

3. t_α est le $(1 - \alpha)$ -quantile empirique de la distribution de \tilde{T}_M . Il est obtenu en répétant les étapes 1 et 2 plusieurs fois.

Une fois t_α déterminé, on rejette H_0 , si $\hat{v}_{(\mathcal{M}+1)} > t_\alpha$.

Le théorème suivant montre la convergence du test sous H_0 .

Théorème 3.2.1. *Supposons les hypothèses H1–H5 vérifiées et $\min_{m \in \mathcal{J}} v_m^* \geq (rs_r \delta_n)^2$ avec $r = s_1$. Si $\mathcal{M} = M$, Alors*

$$\mathbb{P}(H_{\mathcal{M}} \text{ est rejetée}) \leq \alpha + (3k_n + 1)/n + \eta_n,$$

où $\eta_n = O(n^{-1/6}(\log n)^{5/6})$.

Pour démontrer ce théorème nous avons besoin de connaître la convergence de $\tilde{v}_{(\mathcal{M}+1)}$, que fournit le lemme suivant

Lemme 3.2.1. *Sous les hypothèses du théorème 3.2.1 et sachant $(X_i, Y_i)_{i=1, \dots, n}$ il existe un ensemble aléatoire \tilde{A}^* , tel que $\mathbb{P}(\tilde{A}^*) > 1 - 1/n$ et sur lequel, pour tout $m \notin \mathcal{J}$*

$$\left| |\tilde{\beta}_m|^2 - n^{-1} |\tilde{\xi}_m^*|^2 \right| \leq \tilde{w}_n^2 + 2\tilde{w}_n \tilde{\delta}_n, \quad \text{avec } \tilde{\delta}_n = \sqrt{2C_V^2 C_K \sigma^2 n^{-1} L z_n^2} + \tilde{w}_n.$$

DÉMONSTRATION. Pour prouver cette convergence, il nous faut reprendre la démonstration de la proposition 2.2.1 :

Remarque 1. *Ici on prend pour tout $m = 1, \dots, d$*

$$U_m = \frac{1 + \rho^{-2} \hat{v}_m^{(k_n-1)}}{1 + \rho^{-2} v_m^*}.$$

Or sachant le vrai échantillon $\{(Y_i, X_i)\}_{i=1, \dots, n}$, cette quantité est déterministe, ce qui simplifiera la démonstration.

On peut aussi remarquer que cette définition, implique que $|U_m - 1| \leq \alpha_{m, k_n}$, donc que le lemme 2.2.1 peut être encore appliqué.

$$P_{\rho,m} \tilde{\beta}_{lm}(\mathbf{U}) = n^{-1} \sum_{i=1}^n \tilde{s}_{i,m}(\mathbf{U}) \psi_{lm}(X_{i,m}),$$

où $\tilde{s}_i(\mathbf{U}) = s_i(\mathbf{U}) + \zeta_i(\mathbf{U})$ avec

$$s_i(\mathbf{U}) = h^{-1} \nu_i(\mathbf{U})^{-1} \sum_{j=1}^n \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix} \tilde{g}_{\mathcal{M}}(X_j) K_{ij}(\mathbf{U}),$$

$$\zeta_i(\mathbf{U}) = \hat{\sigma} h^{-1} \nu_i(\mathbf{U})^{-1} \sum_{j=1}^n \begin{pmatrix} 1 \\ Z_{ij} \end{pmatrix} \tilde{\varepsilon}_j K_{ij}(\mathbf{U}).$$

On remarque que $s_i(\mathbf{U})$ est une quantité déterministe, sachant $\{(Y_i, X_i)\}_{i=1, \dots, n}$.
On obtient pour chaque coordonnée $m = 1, \dots, d$,

$$P_{\rho,m} \{\mathbb{E} \tilde{\beta}_{lm}(\mathbf{U}) - \hat{\beta}_{lm}\} = \frac{1}{n} \sum_{i=1}^n \left\{ s_{i,m+1}(\mathbf{U}) - P_{\rho,m} \frac{\partial \tilde{g}_{\mathcal{M}}}{\partial x_m}(X_i) \right\} \psi_{lm}(X_{i,m}),$$

$$P_{\rho,m} \{\tilde{\beta}_{lm}(\mathbf{U}) - \mathbb{E} \tilde{\beta}_{lm}(\mathbf{U})\} = \frac{1}{n} \sum_{i=1}^n \zeta_{i,m+1}(\mathbf{U}) \psi_{lm}(X_{i,m}).$$

Étude du biais

$\tilde{g}_{\mathcal{M}}$ est une fonction polynomiale, elle est donc deux fois dérivable et de dérivée seconde bornée, puisque les X_i sont à support compact (en nombre fini). On peut donc écrire

$$|\mathcal{E}_d s_i(\mathbf{U}) - P_{\rho} \nabla \tilde{g}_{\mathcal{M}}(X_i)| \leq \frac{1}{1 - \alpha^*} C C_V^{1/2} h \rho^2,$$

$$\sum_{m=1}^d \sum_{l=1}^L \left| P_{\rho,m} \{\mathbb{E} \tilde{\beta}_{lm}(\mathbf{U}) - \tilde{\beta}_{lm}^*\} \right|^2 \leq \left(\frac{C C_V^{1/2}}{1 - \alpha^*} \rho^2 h \right)^2.$$

On obtient donc le même type de majoration du biais que pour $\hat{\beta}_{lm}$.

Étude de la partie stochastique

$$\tilde{\zeta}_{l,m}(\mathbf{U}) = P_{\rho,m} \{\tilde{\beta}_{lm}(\mathbf{U}) - \mathbb{E} \tilde{\beta}_{lm}(\mathbf{U})\} = \sum_{i=1}^n \tilde{c}_{i,l,m}(\mathbf{U}) \hat{\sigma} \tilde{\varepsilon}_i,$$

$$\text{où } \tilde{c}_{i,l,m}(\mathbf{U}) = \frac{1}{nh} \sum_{j=1}^n \psi_{l,m}(X_{i,m}) v_{i,j,m}(\mathbf{U}) K_{i,j}(\mathbf{U}).$$

On remarque que les coefficients, $\tilde{c}_{i,l,m}(\cdot)$, sont les mêmes que ceux dans la proposition 2.2.1, $c_{j,l,m}(\cdot)$, car ils ne font intervenir que $\{X_i\}_{i=1, \dots, n}$, K , $\psi_{l,m}$, qui ne changent pas dans le modèle rééchantillonné.

D'autre part $\tilde{\zeta}_{lm}(\cdot)$ est une combinaison linéaire des $\tilde{\varepsilon}_i$, c'est donc une variable aléatoire de loi normale centrée et de variance $\hat{\sigma}^2 \sum_{j=1}^n |\tilde{c}_{j,l,m}(\mathbf{U})|^2$. On pose $\tilde{\zeta}_{l,m}^* = \sigma \sum_{j=1}^n \tilde{c}_{j,l,m}(\mathbf{U}^*) \tilde{\varepsilon}_i$.

On va étudier l'écart entre $\tilde{\zeta}_{l,m}(\mathbf{U})$ et $\tilde{\zeta}_{l,m}^*$, c'est-à-dire

$$\tilde{\zeta}_{l,m}(\mathbf{U}) - \tilde{\zeta}_{l,m}^* = \sum_{i=1}^n (\hat{\sigma} \tilde{c}_{i,l,m}(\mathbf{U}) - \sigma \tilde{c}_{i,l,m}(\mathbf{U}^*)) \tilde{\varepsilon}_i,$$

donc la différence $(\tilde{\zeta}_{l,m}(\mathbf{U}) - \tilde{\zeta}_{l,m}^*)$ est une variable aléatoire normale centrée et de variance $\sum_{i=1}^n (\hat{\sigma} \tilde{c}_{i,l,m}(\mathbf{U}) - \sigma \tilde{c}_{i,l,m}(\mathbf{U}^*))^2$. Nous voulons contrôler cette variance pour pouvoir utiliser le lemme 4.3.2.

$$\begin{aligned} \sum_{i=1}^n (\hat{\sigma} \tilde{c}_{i,l,m}(\mathbf{U}) - \sigma \tilde{c}_{i,l,m}(\mathbf{U}^*))^2 &= (\hat{\sigma}^2 - \sigma^2) \sum_{j=1}^n |\tilde{c}_{j,l,m}(\mathbf{U})|^2 + \\ &\quad \sigma^2 \sum_{j=1}^n (|\tilde{c}_{j,l,m}(\mathbf{U})| - |\tilde{c}_{j,l,m}(\mathbf{U}^*)|)^2. \end{aligned}$$

La première partie se majore en utilisant la vitesse de convergence de $\hat{\sigma}$ et l'inégalité (i) du lemme 2.2.3. On obtient

$$(\hat{\sigma}^2 - \sigma^2) \sum_{j=1}^n |\tilde{c}_{j,l,m}(\mathbf{U})|^2 \leq O(n^{-1/2}) \frac{2C_V^2 C_K}{(1 - \alpha_{k_n}^*) h^2 n}.$$

Pour la deuxième partie, on utilise l'inégalité (iii) du même lemme,

$$\begin{aligned} \sigma^2 \sum_{j=1}^n (|\tilde{c}_{j,l,m}(\mathbf{U})| - |\tilde{c}_{j,l,m}(\mathbf{U}^*)|)^2 &\leq \sigma^2 \sum_{j=1}^n \left(\sup_{\mathbf{U} \in \mathcal{U}_\alpha} \left| \frac{dc_{i,l,m}(\mathbf{U})}{d\mathbf{U}} \right| \right)^2 |\mathbf{U} - \mathbf{U}^*|^2, \\ &\leq \sigma^2 \left(\frac{\kappa_\alpha}{nh} \right)^2 |\boldsymbol{\alpha}_{k_n}|^2, \\ &\leq \frac{\sigma^2 \kappa_\alpha^2 |\boldsymbol{\alpha}_{k_n}|^2}{nh^2}. \end{aligned}$$

Finalement

$$\sum_{i=1}^n (\hat{\sigma} \tilde{c}_{i,l,m}(\mathbf{U}) - \sigma \tilde{c}_{i,l,m}(\mathbf{U}^*))^2 \leq O(n^{-1/2}) \frac{2C_V^2 C_K}{(1 - \alpha^*) h^2 n} + \frac{\sigma^2 \kappa_\alpha^2 |\boldsymbol{\alpha}_{k_n}|^2}{nh^2} := var.$$

On utilise alors le lemme 4.3.2, de la même façon que dans le corollaire 2.2.1, pour trouver un ensemble \tilde{A} de probabilité au moins égale à $1 - 1/n$ tel que sur \tilde{A} , pour tout $m = 1, \dots, d$

$$|\tilde{\zeta}_m(\mathbf{U}) - \tilde{\zeta}_m^*| \leq z_n \sqrt{var}.$$

Conclusion : On note $\tilde{\zeta}_{l,m}^* = h^{-1}n^{-1/2}\tilde{\xi}_{l,m}^*$ et on peut alors écrire sur l'ensemble \tilde{A}^* , vérifiant $\mathbb{P}(\tilde{A}^*) > 1 - 1/n$, on a

$$\begin{aligned} |P_{\rho,m}(\tilde{\beta}_m - \tilde{\beta}_m^*) - h^{-1}n^{-1/2}\tilde{\zeta}_m^*| &\leq z_n\sqrt{\text{var}} + \frac{CC_V^{1/2}}{1-\alpha^*}\rho^2h, \\ &\leq \tilde{C}(\sigma^2n^{-1}L\log n)^{2/3} = \tilde{w}_n, \end{aligned}$$

et

$$P_{\rho,m}(\tilde{\beta}_m - \tilde{\beta}_m^*) \leq \tilde{\delta}_n, \quad \text{avec} \quad \tilde{\delta}_n = \sqrt{2C_V^2C_K\sigma^2n^{-1}Lz_n^2} + \tilde{w}_n$$

Donc avec une probabilité $1 - 1/n$, sur \tilde{A}^* , pour $m \notin \mathcal{J}$, $\tilde{\beta}_m^* = 0$, donc

$$|\tilde{\beta}_m|^2 - n^{-1}|\tilde{\xi}_m^*|^2 \leq \tilde{w}_n^2 + 2\tilde{w}_n\tilde{\delta}_n.$$

Ainsi sur \tilde{A}^*

$$\left| n|\tilde{v}_m| - |\tilde{\xi}_m^*|^2 \right| \leq n(\tilde{w}_n^2 + 2\tilde{w}_n\tilde{\delta}_n),$$

où $\tilde{\xi}_m^*$ a la même loi que ξ_m^* , car ce sont les mêmes combinaisons linéaires de variables aléatoires indépendantes de loi normale centrée et réduite. \blacksquare

DÉMONSTRATION DU THÉORÈME 3.2.1.

Sous l'hypothèse $M = \mathcal{M}$

$$\begin{array}{c|c} v_{(1)}^* \geq \dots \geq v_{(\mathcal{M})}^* & v_{(\mathcal{M}+1)}^* \geq \dots \geq v_{(d)}^* \\ \text{composantes non-linéaires} & \text{composantes linéaires} \end{array}$$

On teste le modèle en comparant $\hat{v}_{(\mathcal{M}+1)}$ à t_α , $(1 - \alpha)$ -quantile de la loi de $\tilde{v}_{(\mathcal{M}+1)}$. Dans le vrai modèle, $\hat{v}_{(\mathcal{M}+1)}$ correspond à une composante linéaire. Le théorème 2.4.1 nous donne l'erreur faite par l'estimation sur un ensemble aléatoire A , tel que $\mathbb{P}(A) \geq 1 - 3k_n/n$. Pour montrer le résultat il suffit de montrer que

$$\mathbb{P}(\hat{\mathcal{J}} \neq \mathcal{J} \mid A) \leq \alpha + 1/n.$$

En effet

$$\begin{aligned} \mathbb{P}(\hat{\mathcal{J}} \neq \mathcal{J}) &= \mathbb{P}(\hat{\mathcal{J}} \neq \mathcal{J} \cap A) + \mathbb{P}(\hat{\mathcal{J}} \neq \mathcal{J} \cap A^c), \\ &\leq \mathbb{P}(\hat{\mathcal{J}} \neq \mathcal{J} \mid A)\mathbb{P}(A) + \mathbb{P}(A^c), \\ &\leq \mathbb{P}(\hat{\mathcal{J}} \neq \mathcal{J} \mid A) + 3k_n/n. \end{aligned}$$

On suppose que l'événement A est vérifié.

Dans le vrai modèle : grâce au théorème 2.4.1, on sait que la loi de $n\hat{v}_{(\mathcal{M}+1)}$ peut être approchée par celle de $|\xi_{(\mathcal{M}+1)}^*|^2$, car $(\mathcal{M} + 1) \notin \mathcal{J}$, avec un contrôle de cette distance :

$$\left| n\hat{v}_{(\mathcal{M}+1)} - |\xi_{(\mathcal{M}+1)}^*|^2 \right| < n(w_n^2 + 2w_n\delta_n) := \nu_n.$$

De plus $\xi_{(\mathcal{M}+1)}^*$ est une combinaison linéaire des ε_i avec des coefficients déterministes ne dépendant que de la largeur de fenêtre idéale $\mathbf{b}^{*(k_n)}$, des observations (X_1, \dots, X_n) , des vecteurs $\boldsymbol{\psi}_{l,m}$ et du noyau K .

Le lemme 3.2.1 nous donne le résultat attendu : sachant $\{(Y_i, X_i)\}_{i=1, \dots, n}$, la distribution de $n\tilde{v}_{(\mathcal{M}+1)}$ peut être approchée par celle de $|\xi_{(\mathcal{M}+1)}^*|^2$. Ainsi

$$\begin{aligned} |\mathbb{P}(n\hat{v}_{(\mathcal{M}+1)} > t|A) - \mathbb{P}(n\tilde{v}_{(\mathcal{M}+1)} > t)| &\leq 1/n + |\mathbb{P}(|\xi_{(\mathcal{M}+1)}^*|^2 > t - n(w_n^2 + 2\delta_n w_n)|A) \\ &\quad - \mathbb{P}(|\tilde{\xi}_{(\mathcal{M}+1)}^*|^2 > t - n(\tilde{w}_n^2 + 2\tilde{\delta}_n \tilde{w}_n)|\tilde{A}^*)|, \\ &\leq \eta_n + 1/n. \end{aligned}$$

L'ordre de η_n est obtenu en utilisant le théorème des accroissements finis :

$$\eta_n \leq \mathcal{O}(|\tilde{\nu}_n - \nu_n|) = \mathcal{O}\left(n^{-1/6}(\log n)^{5/6}\right),$$

car les deux variables aléatoires ont la même loi continue et $\nu_n = n(w_n^2 + w_n \delta_n) \leq C'n^{-1/6}(\log n)^{5/6}$, pour une certaine constante C' , on a une majoration du même ordre pour $n(\tilde{w}_n^2 + 2\tilde{\delta}_n \tilde{w}_n)$.

$$\begin{aligned} \mathbb{P}(\hat{\mathcal{J}} \neq \mathcal{J}) &= \mathbb{P}(\hat{v}_{(\mathcal{M}+1)} > t_\alpha | A), \\ \mathbb{P}(n\hat{v}_{(\mathcal{M}+1)} > nt_\alpha | A) &\leq \mathbb{P}(n\tilde{v}_{(\mathcal{M}+1)} > nt_\alpha) + 1/n + \eta_n, \\ \mathbb{P}(n\hat{v}_{(\mathcal{M}+1)} > nt_\alpha | A) &\leq \alpha + 1/n + \eta_n, \end{aligned}$$

On a donc démontré le résultat attendu. ■

3.3 Identification de la partie non-linéaire

Le test précédent permet de localiser la vraie valeur M par rapport à \mathcal{M} . En itérant ce test, nous allons pouvoir obtenir un estimateur du nombre de composantes non-linéaires M et des indices correspondants, \mathcal{J} . En effet la procédure est la suivante : on commence avec $\mathcal{M} = 0$. On considère le modèle avec \mathcal{M} composantes non-linéaires construit comme en (3.1) et (3.2). On teste alors l'hypothèse $M \leq \mathcal{M}$ au niveau α , comme dans la partie précédente. Si l'hypothèse $\mathcal{H}_{\mathcal{M}} : M \leq \mathcal{M}$ n'est pas rejetée c'est fini, sinon on incrémente \mathcal{M} et on recommence.

Finalement $\hat{\mathcal{M}}$ correspond au premier \mathcal{M} non rejeté et on définit alors $\hat{\mathcal{J}} = \hat{\mathcal{J}}_{\hat{\mathcal{M}}}$. Le théorème suivant donne la probabilité d'une mauvaise estimation lorsque les tests sont effectués pour un niveau $\alpha < 1$ fixé.

Théorème 3.3.1. *Sous les hypothèses H1–H5 et $\min_{m \in \mathcal{J}} v_m^* \geq (rs_r \delta_n)^2$ avec $r = s_1$, alors*

$$\mathbb{P}(\hat{\mathcal{J}} \neq \mathcal{J}) \leq \alpha + (3k_n + 3M + 1)/n + \eta_n,$$

où $\eta_n = \mathcal{O}\left(n^{-1/6}(\log n)^{5/6}\right)$.

DÉMONSTRATION. Les deux estimateurs de la structure du modèle $\widehat{\mathcal{M}}$ et $\widehat{\mathcal{J}}$ sont obtenus après avoir effectué un ou plusieurs tests du type $\mathcal{H}_{\mathcal{M}} : M \leq \mathcal{M}$. En utilisant donc le théorème 3.2.1, on va pouvoir approcher les probabilités de se tromper : $\mathbb{P}(\widehat{\mathcal{M}} \neq M)$ et $\mathbb{P}(\widehat{\mathcal{J}} \neq \mathcal{J})$.

Comme on fait l'hypothèse d'une bonne séparation,

$$\mathbb{P}(\widehat{\mathcal{J}} \neq \mathcal{J}) = \mathbb{P}(\widehat{\mathcal{M}} \neq M).$$

D'après le théorème 2.4.1, $\exists A / \mathbb{P}(A) > 1 - 3k_n/n$, il suffit donc de prouver que

$$\mathbb{P}(\widehat{\mathcal{M}} \neq M | A) \leq \alpha + 2(M + 1)/n.$$

Le théorème 3.1.1 affirme que, sachant A :

- $\widehat{v}_m < \delta_n^2$, pour $m \notin \mathcal{J}$,
- $\widehat{v}_m > (r\delta_n)^2$, pour $m \in \mathcal{J}$.

Si $\mathcal{M} \leq M$, alors $\widehat{v}_{(\mathcal{M})} > (s_1\delta_n)^2$, car il fait partie des composantes non-linéaires. De plus $\widehat{v}_{(M+1)} \leq \delta_n^2$, car M est la vraie dimension non-linéaire et l'indice $(M + 1)$ correspond à celui d'une composante linéaire.

- Si $\mathcal{M} < M$, on échantillonne à partir du modèle ayant exactement \mathcal{M} composantes non-linéaires. Dans le nouveau modèle, la vraie mesure de non-linéarité $\widetilde{v}_m^* = \widehat{v}_m$, pour $m \in \widehat{\mathcal{J}}_{\mathcal{M}}$.

On applique les propositions 2.2.1 et 2.2.2 d'amélioration en une étape, comme dans le démonstration du théorème 3.1.1, alors il existe $\widetilde{A}_{\mathcal{M}}$ tel que $\mathbb{P}(\widetilde{A}_{\mathcal{M}} | A) > 1 - 3/n$ et sur $\widetilde{A}_{\mathcal{M}}$, pour $m \notin \widehat{\mathcal{J}}_{\mathcal{M}}$, $\widetilde{v}_m \leq \delta_n^2$.

Finalement sur $\widetilde{A}_{\mathcal{M}}$, $\widetilde{v}_{(M+1)} \leq \delta_n^2$. Il en est de même pour le $(1 - \alpha)$ -quantile de $\widetilde{v}_{(M+1)}$, en effet si $\alpha > 3/n$,

$$\mathbb{P}(\widetilde{v}_{(M+1)} \leq t_\alpha) = 1 - \alpha < 1 - 3/n < \mathbb{P}(\widetilde{A}_{\mathcal{M}} | A) < \mathbb{P}(\widetilde{v}_{(M+1)} \leq \delta_n^2),$$

donc $\{\widetilde{v}_{(M+1)} \leq t_\alpha\} \subset \{\widetilde{v}_{(M+1)} \leq \delta_n^2\}$ et $t_\alpha < \delta_n^2$.

Or, dans le vrai modèle, $(M + 1)$ est encore dans \mathcal{J} , donc il correspond à une composante non-linéaire et sur A on a $\widehat{v}_{(M+1)} > (s_1\delta_n)^2$ et donc $\widehat{v}_{(M+1)} > t_\alpha$. Finalement l'hypothèse $\mathcal{H}_{\mathcal{M}}$ est rejetée.

Ceci signifie que sur $A \cap \widetilde{A}_{\mathcal{M}}$, pour tout $\mathcal{M} < M$, l'hypothèse $\mathcal{H}_{\mathcal{M}}$ est rejetée. Étant donnée l'égalité des ensembles suivants $\{\widehat{\mathcal{M}} < M\} = \{\exists \mathcal{M} < M, \mathcal{H}_{\mathcal{M}} \text{ non rejetée}\}$, on a

$$\begin{aligned} \mathbb{P}(\widehat{\mathcal{M}} < M | A) &= \mathbb{P}\left(\bigcup_{\mathcal{M} < M} \{\mathcal{H}_{\mathcal{M}} \text{ est non rejetée}\} | A\right), \\ &\leq \sum_{\mathcal{M} < M} \mathbb{P}\left(\mathcal{H}_{\mathcal{M}} \text{ est non rejetée} | A\right), \end{aligned}$$

et \mathcal{M} prend les valeurs $0, \dots, M-1$. De plus

$$\begin{aligned} \mathbb{P}\left(\mathcal{H}_{\mathcal{M}} \text{ est non rejetée} | A\right) &= 1 - \mathbb{P}\left(\mathcal{H}_{\mathcal{M}} \text{ est rejetée} | A\right), \\ &\leq 1 - \mathbb{P}\left(\tilde{A}_{\mathcal{M}} | A\right), \\ &\leq 3/n. \end{aligned}$$

d'où

$$\mathbb{P}(\widehat{\mathcal{M}} < M | A) \leq 3M/n.$$

• Si $\widehat{\mathcal{M}} > M$: $\widehat{\mathcal{M}}$ étant défini comme le premier \mathcal{M} , pour lequel l'hypothèse est non rejetée, on a l'inclusion

$$\{\widehat{\mathcal{M}} > M\} \subset \{\widehat{v}_{(M+1)} > t_{\alpha}(M)\},$$

où $t_{\alpha}(M)$ est le seuil critique évalué lors du test lié au schéma de rééchantillonnage défini avec $\mathcal{M} = M$. Par le théorème 3.2.1, on obtient

$$\mathbb{P}\{\widehat{\mathcal{M}} > M | A\} \leq \alpha + 1/n + \eta_n.$$

Finalement

$$\mathbb{P}(\widehat{\mathcal{M}} \neq M | A) \leq \alpha + (3M + 1)/n + \eta_n,$$

et on trouve le résultat annoncé dans le théorème. ■

3.4 Estimation et inférence sur les composantes linéaires

Les deux paragraphes précédents nous permettent de séparer les composantes linéaires des composantes non-linéaires. Dans cette partie nous allons nous intéresser plus précisément aux composantes linéaires, qui sont caractérisées par leur pente θ_m . On sait que pour $m \notin \mathcal{J}$, $\frac{\partial g}{\partial x_m}(\cdot) = \theta_m$. L'algorithme, défini à la partie 1.3, nous donne un estimateur du gradient de la fonction de régression, g , et nous permet de définir

$$\widehat{\theta}_m = \frac{1}{n} \sum_{i=1}^n \frac{\widehat{\partial g}}{\partial x_m}(X_i),$$

avec $\frac{\widehat{\partial g}}{\partial x_m}(X_i)$ provenant de l'estimation du gradient obtenu à la dernière itération de l'algorithme 1.3. Dans le théorème suivant nous obtenons la convergence à la vitesse $n^{1/2}$ de $\widehat{\theta}_m$ et la distance entre $(\widehat{\theta}_m - \theta_m)$ et une variable aléatoire gaussienne. Ce théorème est une conséquence du théorème 2.4.1, pour $m \notin \mathcal{J}$.

Théorème 3.4.1. *Sous les hypothèses H1–H5. Pour tout $m \notin \mathcal{J}$, il existe une variable aléatoire gaussienne de moyenne nulle, γ_m^* , qui est définie comme une combinaison*

linéaire des erreurs ε_i avec des coefficients déterministes dépendant uniquement de la largeur de fenêtre “idéale” $\mathbf{b}^* = \mathbf{b}^{*(k_n)}$, des variables explicatives (X_1, \dots, X_n) et du noyau K ; sa variance vérifie

$$\mathbb{E} |\gamma_m^*|^2 \leq 2\sigma^2 C_V^2 C_K,$$

et sur l'ensemble aléatoire A introduit au théorème 2.4.1, qui vérifie $\mathbb{P}(A) \geq 1 - 3k_n/n$, on a

$$\max_{m \notin \mathcal{J}} \left| \widehat{\theta}_m - \theta_m - n^{-1/2} \gamma_m^* \right| \leq C_1 (\sigma^2 n^{-1} L \log n)^{2/3},$$

où $C_1 = C_1(d, M, C_{g_1}, C_V, C_K, C_{K'})$.

DÉMONSTRATION. On peut comparer les deux estimateurs suivants

$$\widehat{\theta}_m = \frac{1}{n} \sum_{i=1}^n \widehat{\frac{\partial g}{\partial x_m}}(X_i) \quad \text{et} \quad \widehat{\beta}_{l,m} = \frac{1}{n} \sum_{i=1}^n \widehat{\frac{\partial g}{\partial x_m}}(X_i) \psi_{l,m}(X_i).$$

On remarque alors que $\widehat{\theta}_m$ est une version de $\widehat{\beta}_{l,m}$ pour laquelle $\psi_{l,m} \equiv 1$, pour tout $l = 1, \dots, L$. On peut donc reprendre exactement la même démonstration que celle du théorème 2.4.1. En effet, l'inégalité (2.9) donne directement la majoration concernant le biais et la partie stochastique s'étudie de la même façon en prenant $\psi_{l,m} \equiv 1$ et $L = 1$. ■

Chapitre 4

Applications

Ce chapitre a pour but d'illustrer les résultats théoriques obtenus dans les chapitres précédents. Nous allons tout d'abord présenter un "algorithme modifié", que nous utilisons lors des applications. Ensuite nous mettrons en évidence les performances de cet algorithme et de la procédure de test sur des données simulées puis sur des données réelles. Le code utilisé a été rédigé en Fortran 77; les simulations ont été effectuées sur le serveur de calcul de l'IRMAR (bi-processeur Xéon 2 GHz).

4.1 Algorithme modifié

Comme on le voit dans la description de la procédure itérative, l'estimateur $\widehat{\nabla}g(X_i)$ doit être bien défini, ce qui demande des propriétés de régularité sur les observations, X_i . Si cela n'est pas vérifié, l'estimateur du gradient correspondant va fausser toutes les estimations qui en dépendent. On contrevient à ceci en deux temps : en utilisant un paramètre de mémoire η pour la définition de l'estimateur du gradient, puis en pondérant les β_{lm} pour tenir compte de cette régularité locale. Cette dernière transformation de l'algorithme est semblable à celle faite dans HJS et HJPS.

L'algorithme est le suivant

1. Fixer h_1 . On pose $\rho_1 = 1$ et pour $m = 1, \dots, d$, $\widehat{v}_m^{(0)} = 0$ et $b_m^{(1)} = h_1$. Pour $i = 1, \dots, n$

$$V_i^{(0)} = \sum_{j=1}^n \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix}^\top, \quad S_i^{(0)} = \sum_{j=1}^n \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} Y_j, \quad (4.1)$$

où $X_{ij} = X_j - X_i$. On pose $k = 1$. **2.** Calculer

$$b_m^{(k)} = h_k \left(1 + \rho_k^{-2} \widehat{v}_m^{(k-1)} \right)^{-1/2}, \quad m = 1, \dots, d. \quad (4.2)$$

Définir $\mathbf{b}^{(k)} = \text{diag}(b_1^{(k)}, \dots, b_d^{(k)})$.

3. Pour tout X_i calculer

$$V_i^{(k)} = \eta V_i^{(k-1)} + (1 - \eta) \sum_{j=1}^n \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix}^\top K_d(X_{ij}, \mathbf{b}^{(k)}), \quad (4.3)$$

$$S_i^{(k)} = \eta S_i^{(k-1)} + (1 - \eta) \sum_{j=1}^n \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} Y_j K_d(X_{ij}, \mathbf{b}^{(k)}), \quad (4.4)$$

et

$$\begin{pmatrix} \widehat{g}^{(k)}(X_i) \\ \widehat{\nabla g}^{(k)}(X_i) \end{pmatrix} = \left(V_i^{(k)} \right)^{-1} S_i^{(k)} \quad (4.5)$$

Calculer $\widehat{\varepsilon}_i^{(k)} = Y_i - \widehat{g}^{(k)}(X_i)$. Calculer les valeurs propres de la matrice $V_i^{(k)}$, on définit w_i comme la plus petite racine carrée des valeurs propres de $V_i^{(k)}$: $w_i^2 = \lambda_{\min}(V_i^{(k)})$.

4. Pour $m = 1, \dots, d$ et $l = 1, \dots, L$, définir les fonctions $\psi_{lm}^{(k)}$, $l = 1, \dots, L$, satisfaisant les conditions

$$\sum_{i=1}^n w_i \psi_{lm}^{(k)}(X_{i,m}) = 0, \quad \left(\sum_{i=1}^n w_i \right)^{-1} \sum_{i=1}^n w_i \psi_{lm}^{(k)}(X_{i,m}) \psi_{l'm}^{(k)}(X_{i,m}) = \delta_{ll'}$$

avec les w_i obtenus ci-dessus et $l, l' = 1, \dots, L$. Pour tout $l = 1, \dots, L$ et $m = 1, \dots, d$, calculer

$$\widehat{\beta}_{lm}^{(k)} = \left(\sum_{i=1}^n w_i \right)^{-1} \sum_{i=1}^n \frac{\widehat{\partial g}^{(k)}}{\partial x_m}(X_i) \psi_{lm}^{(k)}(X_{i,m}) w_i. \quad (4.6)$$

Calculer $\widehat{v}_m^{(k)}$ par (1.6) pour $m = 1, \dots, d$. Si $\widehat{v}_m^{(k)} > 1$, alors fixer $\widehat{v}_m^{(k)} = 1$.

5. Augmenter k de 1. Soit $\rho_{k+1} = a_\rho \rho_k$, $h_{k+1} = a_h h_k$. Si $\rho_{k+1} \geq \rho_{\min}$, alors poser $k = k + 1$ et continuer avec l'étape 2; sinon terminer.

REMARQUE : Dans l'algorithme, on décide de tronquer \widehat{v}_m à 1. Cette condition permet d'empêcher une trop grande réduction du voisinage dans la direction du $m^{\text{ème}}$ régresseur. Cela peut en effet arriver si $\widehat{v}_m^{(k)}$ est trop grand.

4.2 Résultats obtenus sur des données simulées

On s'intéresse ici à la performance de notre algorithme sur des données simulées. On veut montrer de quelle façon, elle dépend de la taille n de l'échantillon, de la dimension d du modèle, du nombre M de composantes non-linéaires et de la variance σ^2 du bruit. On

pourra aussi constater l'amélioration au cours des itérations. Dans un premier temps, on s'intéressera aux résultats de la procédure itérative, c'est-à-dire de la partie estimation. On illustrera, ensuite, les résultats de bonne classification des composantes et on s'intéresse au test.

Dans toute l'étude, nous appliquerons la procédure modifiée du paragraphe précédent avec les paramètres suivants

$$\begin{aligned} \rho_1 &= 1, & \rho_{\min} &= n^{-1/3}, & a_\rho &= e^{-1/6}, \\ \eta &= 0.2 & h_{\max} &= 2\sqrt{d}, & a_h &= a_\rho^{-1/2}. \end{aligned}$$

Ainsi le nombre d'itérations k_n est de l'ordre de $\frac{\log(\rho_{\min}/\rho_1)}{\log a_\rho} = 2 \log n$. On utilise le noyau $K(|x|^2) = (1 - |x|^2)_+^2$.

La famille $\{\psi_{lm}(x_m)\}$ est obtenue à partir de la famille des polynômes en x_m de degré inférieur ou égal à L et telle que

$$\sum_{i=1}^n w_i \psi_{lm}(X_{i,m}) \psi_{l'm}(X_{i,m}) / \sum_{i=1}^n w_i = \delta_{ll'} \quad \text{et} \quad \sum_{i=1}^n w_i \psi_{lm}(X_{i,m}) = 0,$$

avec $w_i = w_i^{(k)} = \lambda_{\min}^{1/2}(V_i^{(k)})$, à la $k^{\text{ème}}$ itération.

Cette nouvelle détermination des $\{\psi_{l,m}\}$ correspond à un changement de produit scalaire, associé à la matrice diagonale $n \times n$ de coefficients diagonaux $w_i / \sum_{i=1}^n w_i$. On choisit ici $L = 6$.

Le modèle considéré ici est

$$Y_i = \theta^T X_i + g_1(X_{i,d-M+1}, \dots, X_{i,d}) + \varepsilon_i.$$

On fait varier M de 1 à 3 et θ , vecteur de \mathbb{R}^d , est pris de la forme $(1, 2, 3, 4, 0, \dots, 0)^T$. La dimension du modèle, d , est supérieure ou égale à $4 + M$. La fonction de lien, g_1 , sera définie par

- pour $M = 1$, $g_1(u) = \gamma(u) = e^u + e^{-u}$,
- pour $M = 2$, $g_1(u_1, u_2) = \gamma(u_1)\gamma(u_2)$,
- pour $M = 3$, $g_1(u_1, u_2, u_3) = \gamma(u_1)\gamma(u_2)\gamma(u_3)$.

Les erreurs, ε_i , sont des variables aléatoires indépendantes identiquement distribuées de loi $\mathcal{N}(0, \sigma^2)$, avec, sauf spécification particulière, $\sigma^2 = 0, 1$.

Les X_1, \dots, X_n proviennent de variables aléatoires telles que $X_i \sim \mathcal{N}(0, 2; 0, 8^2 I_d)$ restreinte au cube $[-1, 1]^d$.

4.2.1 Qualité de l'estimation

Les tableaux 4.1 et 4.2 résument la qualité de l'estimation dans les cas $M = 1$ et $M = 2$. On remarque que la qualité de l'estimation pour la composante linéaire s'améliore quand on augmente la dimension de d . Ces deux tableaux permettent de se rendre compte de

TAB. 4.1 – Cas $M = 1$: perte moyenne, $|\widehat{v}_m - v_m^*|$ pour le régresseur non-linéaire et $|\widehat{v}_1 - v_1^*|$ pour le premier régresseur linéaire, correspondant à la première, la quatrième et la dernière itération. Les résultats sont obtenus pour $N = 250$ simulations. L'intervalle interquartile des erreurs est donné entre parenthèses.

d	n	régresseur non-linéaire			régresseur linéaire x_1
		1 ^{ère}	4 ^{ème}	finale	finale
5	100	0.9580	0.6656	0.3069	0.0139
		(0.1865)	(0.1546)	(0.2400)	(0.0113)
5	200	0.9395	0.7711	0.2424	0.0072
		(0.1378)	(0.1300)	(0.2024)	(0.0057)
6	200	0.9432	0.7207	0.1641	0.0018
		(0.1231)	(0.1067)	(0.1766)	(0.0016)
8	200	0.9362	0.6703	0.2232	0.0006
		(0.1253)	(0.1003)	(0.1797)	(0.0005)
10	100	0.9574	0.6743	0.5822	0.0005
		(0.2064)	(0.1526)	(0.2756)	(0.0004)
10	200	0.9406	0.6777	0.3690	0.0002
		(0.1522)	(0.1202)	(0.2213)	(0.0002)
10	400	0.9348	0.7217	0.2316	0.0001
		(0.0925)	(0.0838)	(0.1399)	(0.0001)

TAB. 4.2 – Cas $M = 2$: perte moyenne, $|\widehat{v}_m - v_m^*|$ pour les régresseurs non-linéaires et $|\widehat{v}_1 - v_1^*|$ pour le premier régresseur linéaire, correspondant à la première, la quatrième et la dernière itération. Les résultats sont obtenus pour $N = 250$ simulations. L'intervalle interquartile des erreurs est donné entre parenthèses.

d	n	σ^2	1 ^{er} régresseur non-linéaire			2 ^{ème} régresseur non-linéaire			régresseur linéaire x_1
			1 ^{ère}	4 ^{ème}	finale	1 ^{ère}	4 ^{ème}	finale	finale
6	200	0.1	4.6117	3.7349	0.4763	4.6337	3.7576	0.4473	0.0081
			(0.6646)	(0.5028)	(0.5211)	(0.6257)	(0.5402)	(0.4785)	(0.0063)
8	200	0.1	4.6397	3.4423	0.4244	4.5942	3.4085	0.4058	0.0025
			(0.6683)	(0.5431)	(0.4108)	(0.6646)	(0.4621)	(0.4607)	(0.0019)
10	100	0.1	4.6338	3.1450	0.7573	4.6862	3.1642	0.7089	0.0043
			(0.8840)	(0.7307)	(0.5302)	(1.0155)	(0.7312)	(0.4938)	(0.0032)
10	200	0.1	4.5537	3.2806	0.5812	4.5904	3.2917	0.5489	0.0011
			(0.7458)	(0.5065)	(0.3404)	(0.7649)	(0.5875)	(0.4014)	(0.0010)
10	400	0.1	4.5198	3.5276	0.4457	4.5584	3.5594	0.4319	0.0004
			(0.4850)	(0.3566)	(0.3121)	(0.4168)	(0.3562)	(0.3023)	(0.0003)
10	400	0.2	4.5198	3.5284	0.4403	4.5584	3.5602	0.4325	0.0007
			(0.4850)	(0.3483)	(0.3949)	(0.4167)	(0.3642)	(0.3948)	(0.0006)
10	400	0.4	4.5198	3.5297	0.4637	4.5584	3.5615	0.4666	0.0017
			(0.4850)	(0.3316)	(0.4891)	(0.4167)	(0.3727)	(0.5029)	(0.0013)

l'amélioration de l'estimation au cours des itérations et en augmentant n , pour $d = 10$. Dans le tableau 4.2, on montre en plus que l'estimation dépend de la variance du bruit σ^2 . Pour les composantes non-linéaires, l'erreur augmente un peu avec σ . Pour les composantes linéaires l'augmentation est proportionnelle à σ .

Dans le cas non-linéaire, l'erreur est due en grande partie au biais non paramétrique,

qui disparaît au cours des itérations dans le cas linéaire. Ceci peut expliquer pourquoi l'augmentation de l'erreur est proportionnelle à σ dans le cas linéaire.

4.2.2 Séparation et exploration du modèle

La figure 4.1 suivante illustre les résultats du théorème 3.1.1, sur la séparation entre composantes linéaires et non-linéaires. On note $\mathcal{L}_N(\xi)$ la fonction de répartition empirique de la variable ξ , basée sur un échantillon de taille N . On entend par bonne séparation entre les composantes de nature différente, si $\mathcal{L}_N(\widehat{v}_m)$, pour tout $m \in \mathcal{J}$ et $1 - \mathcal{L}_N(\widehat{v}_m)$, pour $m \notin \mathcal{J}$, ont des supports disjoints. On observe sur la figure 4.1 une très bonne séparation pour $n = 100$ et une séparation presque parfaite pour $n = 200$.

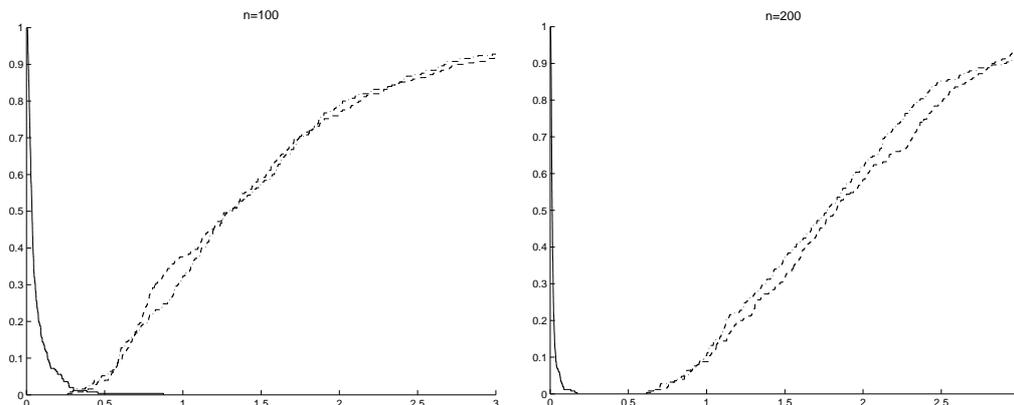


FIG. 4.1 – Cas $M = 2$, $d = 6$: $\mathcal{L}_N(\widehat{v}_6)$ et $1 - \mathcal{L}_N(\max_{m=1,\dots,4} \widehat{v}_m)$ pour $n = 100, 200$ pour $N = 250$ réalisations.

Ensuite nous illustrons les résultats du paragraphe 3.4. En effet la figure 4.2 donne des box-plots correspondant à l'estimation de $n^{1/2} \|\widehat{\theta} - \theta^*\|$ (pour la dernière itération), pour $d = 6$ et $M = 2$ et différentes tailles d'échantillon. La qualité de l'estimation s'améliore quand n croît.

Finalement nous nous intéressons au test de l'hypothèse $M \leq \mathcal{M}$ et à la classification des composantes par la méthode donnée dans le paragraphe 3.3. Pour se rendre compte de la performance de notre méthode, nous faisons varier la taille de l'échantillon, n , la dimension du modèle, d et le cardinal de \mathcal{J} , M . Le tableau 4.2.2 donne le proportion de mauvaises classifications pour les composantes non-linéaires et pour tout le modèle.

On remarque que les résultats s'améliorent encore quand d augmente. Ce lien peut s'expliquer par le fait que la distribution de la statistique de test utilisée pour classifier les composantes est de plus en plus dégénérée, quand d augmente.

Quand $M = 3$, on voit que la procédure a besoin d'un échantillon de taille minimale pour commencer à bien sélectionner les composantes non-linéaires. En effet pour $n = 100$, on obtient presque toujours $\mathcal{M} < M$. Pour $n = 200$ et $d = 7$, on a une bonne classification

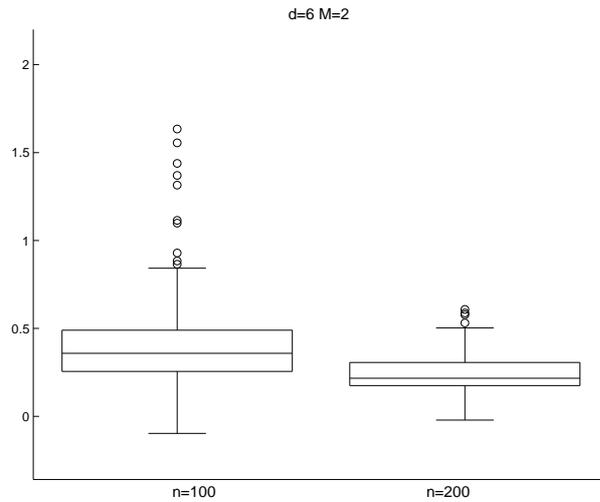


FIG. 4.2 – Cas $M = 2$: box-plots de l'estimation des erreurs $n^{1/2} \|\hat{\theta} - \theta^*\|$ de la partie linéaire à la dernière itération pour $d = 6$. Les résultats sont obtenus pour $N = 250$ réalisations.

M	d	n	1 ^{er} c.n.l.	2 ^{ème} c.n.l.	3 ^{ème} c.n.l.	$\{\mathcal{J} \neq \mathcal{J}\}$
1	5	100	0.152	–	–	0.18
		200	0.024	–	–	0.056
		400	0	–	–	0.028
1	10	200	0	–	–	0.004
		400	0	–	–	0.004
1	20	400	0	–	–	0.176
2	6	100	0.268	0.308	–	0.38
		200	0.056	0.048	–	0.1
		400	0.004	0.004	–	0.024
2	10	200	0	0	–	0.008
		400	0	0	–	0.004
2	20	400	0	0	–	0
3	7	100	0.976	0.96	0.964	0.992
		200	0.62	0.636	0.636	0.748
		400	0.076	0.06	0.072	0.1
3	10	200	0.004	0	0.004	0.004
		400	0	0	0	0

TAB. 4.3 – Proportion de mauvaises classifications pour chaque régresseur non-linéaire et pour tout le modèle. Les résultats sont obtenus avec $N = 250$ simulations et 500 itérations Bootstrap.

seulement dans 30% des cas, mais pour $d = 10$ la proportion de mauvaise classification est très faible.

Avec la dernière figure 4.3 nous avons voulu illustrer la qualité d'estimation de σ^2 par $\hat{\sigma}^2$ pour le modèle $d = 6$ et $M = 2$, avec différentes tailles d'échantillon. Ces résultats sont bien en accord avec la convergence théorique de $\hat{\sigma}^2$ à la vitesse \sqrt{n} .

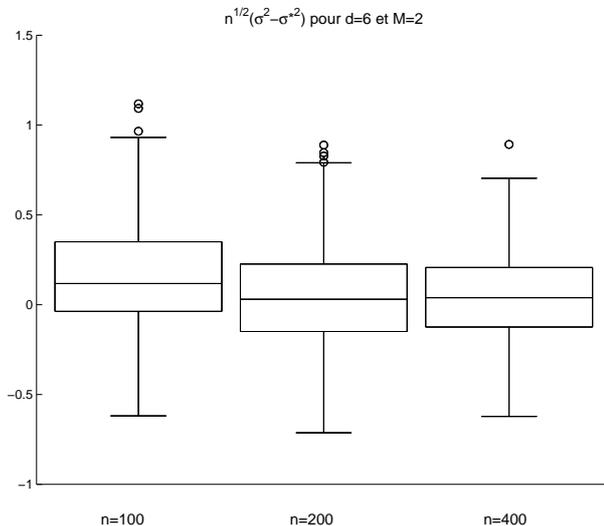


FIG. 4.3 – Box-plots de l'erreur d'estimation $n^{1/2}\|\hat{\sigma}^2 - \sigma^2\|$ pour $d = 6$, $M = 2$ et différentes tailles d'échantillon, n .

4.3 Un exemple sur des données réelles

Dans cette partie nous présentons les résultats obtenus en appliquant notre algorithme à des données réelles. Nous avons choisi d'appliquer notre méthode à un exemple déjà étudié par *Sperlich (1998)* et *Härdle, Sperlich, Spokoiny (2001)*. Il s'agit de l'étude d'un sous-échantillon d'un panel socio-économique d'Allemagne, datant de 1992. Le but est d'analyser le nombre d'heures de travail Y_i de 607 femmes travaillant et vivant avec leur partenaire. On utilise les variables suivantes : l'âge de ces femmes, X_1 , compris entre 25 et 60 ans ; leur revenu par heure, X_2 ; l'indice du type de profession (indice de prestige de Treiman), X_3 ; leur loyer ou remboursement mensuel pour leur appartement ou maison, X_4 ; le revenu net mensuel de leur conjoint, X_5 ; le nombre d'années d'études, X_6 ; le taux de chômage dans la région où elles habitent, X_7 ; le nombre d'enfant de moins de 16 ans, X_8 . Les trois dernières variables sont des variables discrètes, on ne les fera donc pas intervenir lors des tests successifs. Les estimations obtenues à la dernière itération, \hat{v}_m , avec la procédure d'estimation sont données dans le tableau 4.4. On obtient aussi une estimation de la variance du bruit $\hat{\sigma}^2 = \hat{\sigma}^{(k_n)^2} = 0.736$.

Ensuite nous identifions les composantes linéaires en commençant par $\mathcal{M} = 0$, comme décrit dans la partie 3.3. Le tableau 4.5 donne les p-valeurs $PV_{\mathcal{M}}$ pour chaque test $\mathcal{H}_{\mathcal{M}}$,

TAB. 4.4 – Estimateur \hat{v}_m de v_m^* .

\hat{v}_1	\hat{v}_2	\hat{v}_3	\hat{v}_4	\hat{v}_5	\hat{v}_6	\hat{v}_7	\hat{v}_8
0.05259	0.00729	0.00441	0.00012	0.00060	0.00142	0.00015	0.00875

elles sont obtenues après $B = 1000$ itérations Bootstrap et définies par :

$$PV_{\mathcal{M}} = \frac{1}{B} \sum_{b=1}^B 1_{\{\hat{v}_{(\mathcal{M}+1)}^{(b)} > \hat{v}_{(\mathcal{M}+1)}\}}$$

TAB. 4.5 – p-valeurs pour les tests consecutifs

\mathcal{M}	$\hat{v}_{(\mathcal{M}+1)}$	p-valeurs
0	\hat{v}_1	0.003996
1	\hat{v}_2	0.086913
2	\hat{v}_3	0.01998
3	\hat{v}_5	0.47153

Les trois premières hypothèses $\mathcal{H}_0 : M = 0$, $\mathcal{H}_1 : M \leq 1$ et $\mathcal{H}_2 : \leq 2$ sont rejetées au niveau 10% et l'hypothèse $\mathcal{H}_3 : \leq 3$ n'est pas rejetée. Ainsi pour le modèle considéré, le nombre de composantes non-linéaires est estimé par 3 et les variables non-linéaires sont X_1 , X_2 et X_3 .

La nature de nos variables coïncide avec celle obtenue par *Härdle, Sperlich et Spokoiny (2001)*, mais pour des p-valeurs assez différentes : dans notre cas X_1 , i.e. l'âge, est la variable plus non-linéaire et X_2 , i.e. le revenu par heure, est la variable la moins non-linéaire parmi les trois. Dans *Härdle, Sperlich et Spokoiny (2001)* la situation est inversée. On peut noter que leurs résultats sont obtenus pour un modèle possédant une structure additive, alors que les nôtres sont obtenus pour une situation générale, qui ne suppose pas avoir une structure additive.

Annexe : Propriétés de grande déviation pour des vecteurs gaussiens

Dans cette annexe nous avons rassemblé trois lemmes concernant les queues de distribution de vecteurs gaussiens. Ces trois propriétés sont utilisées dans la démonstration de la convergence de l'algorithme, puisque nous considérons un modèle de régression où le bruit est gaussien.

Lemme 4.3.1. *Queue d'une variable aléatoire réelle gaussienne.*

Soit ζ une variable aléatoire réelle de loi $\mathcal{N}(0, \sigma^2)$ alors

$$\mathbb{P}(|\zeta| > z\sigma) \leq e^{-z^2/2}.$$

DÉMONSTRATION. Ce résultat est une conséquence de l'inégalité suivante vraie pour $z > 1$

$$\left(\frac{1}{z} - \frac{1}{z^3}\right) \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \leq 1 - \mathbb{P}(\zeta \leq z\sigma) \leq \frac{1}{z} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}. \quad (4.7)$$

On note $F(z) = \mathbb{P}(|\zeta| \leq z\sigma)$ et $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} = F'(z)$, on a alors

$$\left(\left(\frac{1}{z} - \frac{1}{z^3}\right)f(z)\right)' = \left(-1 + \frac{3}{z^4}\right) f(z) \quad (4.8)$$

$$\left(\frac{1}{z}f(z)\right)' = \left(\frac{-1}{z^2} - 1\right) f(z) \quad (4.9)$$

$$\left(1 - F(z)\right)' = -f(z) \quad (4.10)$$

d'où

$$\left(\frac{-1}{z^2} - 1\right) f(z) \leq -f(z) \leq \left(-1 + \frac{3}{z^4}\right) f(z).$$

En intégrant entre z et $+\infty$, on obtient

$$\left(\frac{1}{z} - \frac{1}{z^3}\right) f(z) \leq 1 - F(z) \leq \frac{1}{z} f(z),$$

d'où

$$\mathbb{P}(|\zeta| > z\sigma) = 2 \times (1 - F(z)) < e^{-z^2/2}.$$

■

Lemme 4.3.2. *Queue d'un vecteur aléatoire gaussien.*

Soit ξ un vecteur gaussien de \mathbb{R}^d . Alors pour tout $z \geq 1$

$$\mathbb{P}\left(|\xi| > z\sqrt{\mathbb{E}|\xi|^2}\right) \leq z e^{-(z^2-1)/2}.$$

DÉMONSTRATION. Pour toute matrice orthonormale $d \times d$, O , le vecteur $O\xi$ est aussi gaussien et $|\xi| = |O\xi|$. Ainsi en prenant une transformation O particulière, on peut réduire le cas général à une situation où toutes les composantes ξ_i du vecteur ξ sont indépendantes. On notera $v_i^2 = \mathbb{E}\xi_i^2$ et $V^2 = \mathbb{E}|\xi|^2$. On a immédiatement $V^2 = \sum_{i=1}^d v_i^2$ et par l'inégalité de Chebyshev, pour tout $\mu > 0$, qui vérifie $2\mu v_i^2 < 1 \ \forall i \leq d$, on a

$$\mathbb{P}(|\xi| > zV) = \mathbb{P}(|\xi|^2 > z^2\mathbb{E}|\xi|^2) \leq \exp(-\mu z^2 V^2) \mathbb{E} \exp(\mu|\xi|^2).$$

L'indépendance des composantes ξ_i de ξ nous permet d'écrire

$$\begin{aligned} \mathbb{E} \exp(\mu|\xi|^2) &= \mathbb{E} \exp\left(\sum_{i=1}^d \mu\xi_i^2\right) = \prod_{i=1}^d \mathbb{E} \exp(\mu\xi_i^2), \\ &= \prod_{i=1}^d \frac{1}{\sqrt{1-2\mu v_i^2}} = \exp\left(-\frac{1}{2} \sum_{i=1}^d \log(1-2\mu v_i^2)\right), \end{aligned}$$

ainsi

$$\mathbb{P}(|\xi| > zV) \leq \exp\left(-\mu z^2 V^2 - \frac{1}{2} \sum_{i=1}^d \log(1-2\mu v_i^2)\right).$$

On applique alors cette inégalité avec $\mu = (1-z^{-2})V^{-2}/2$, on a bien $z \geq 1$, $\mu > 0$ et $2\mu v_i^2 < 1$, $\forall i \leq d$. L'inégalité de convexité $-\log(1-x) - \log(1-y) \leq -\log(1-x-y)$ vérifiée pour tout x, y avec $x+y < 1$, conduit à

$$\begin{aligned} \mathbb{P}(|\xi| > zV) &\leq \exp\left(-\mu z^2 V^2 - \frac{1}{2} \log(1-2\mu V^2)\right), \\ &\leq \exp\left(-\frac{z^2-1}{2} - \frac{1}{2} \log(z^{-2})\right) = z \exp\left(-\frac{z^2-1}{2}\right), \end{aligned}$$

c'est-à-dire le résultat attendu. ■

Lemme 4.3.3. *Soit r un nombre positif et soit Γ un ensemble fini. On considère les fonctions $a_{i,\gamma}(u)$ de $u \in \mathbb{R}^d$ à valeurs réelles, vérifiant les conditions*

$$\sup_{\gamma \in \Gamma} \sup_{|u-u^*| \leq r} \left| \frac{d}{du} a_{i,\gamma}(u) \right| \leq \kappa, \quad i = 1, \dots, n. \quad (4.11)$$

Si les ε_i sont des variables aléatoires indépendantes identiquement distribuées de loi $\mathcal{N}(0, \sigma^2)$, alors

$$\mathbb{P} \left(\sup_{\gamma \in \Gamma} \sup_{|u-u^*| \leq r} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^n \{a_{i,\gamma}(u) - a_{i,\gamma}(u^*)\} \varepsilon_i \right| > \sigma \kappa r t \right) \leq \frac{2}{n},$$

où $t = 2 + \sqrt{2 \log(n|\Gamma|) + d \log(4n)}$ et $|\Gamma|$ représente le nombre d'éléments dans Γ .

Ce lemme provient d'une modification du lemme 8 de HJS. En effet dans notre cas $a_{i,\gamma}$ est une fonction à valeurs réelles et donc fait appel au lemme 4.3.1; on veut aussi une minoration de la borne supérieure sur deux ensembles γ et u .

DÉMONSTRATION. Soit \mathcal{B}_r la boule $\{u : |u - u^*| \leq r\}$ et Σ_r un ε -réseau sur \mathcal{B}_r tel que pour tout $u \in \mathcal{B}_r$ il existe un élément u_ℓ de Σ_r tel que $|u - u_\ell| \leq \frac{r}{\sqrt{n}}$. Il est alors facile de voir qu'on peut construire un tel réseau de cardinal $N_r < (4n)^{d/2}$. Pour un élément $u_\ell \in \Sigma_r$, on note

$$\eta(u_\ell) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{a_{i,\gamma}(u_\ell) - a_{i,\gamma}(u^*)\} \varepsilon_i.$$

alors par (4.11)

$$\mathbb{E} |\eta(u_\ell)|^2 = \frac{\sigma^2}{n} \sum_{i=1}^n |a_{i,\gamma}(u_\ell) - a_{i,\gamma}(u^*)|^2 \leq \sigma^2 \kappa^2 r^2,$$

et pour tout $t \geq 1$ par le lemme 4.3.1

$$\mathbb{P}(|\eta(u_\ell)| > t \sqrt{\mathbb{E} |\eta(u_\ell)|^2}) \leq e^{-t^2/2}.$$

Alors si on prend $t = \sqrt{2 \log(n|\Gamma|) + d \log(4n)}$, alors

$$\begin{aligned} \mathbb{P} \left(\sup_{\gamma} \sup_{u_\ell \in \Sigma_r} |\eta(u_\ell)| > t \sigma \kappa r \right) &\leq \sum_{\gamma \in \Gamma} \sum_{\ell=1}^{N_r} \mathbb{P}(|\eta(u_\ell)| > t \sigma \kappa r) \\ &\leq N_r |\Gamma| \exp \left(-\log(n|\Gamma|) - \frac{d}{2} \log(4n) \right) \\ &< 1/n. \end{aligned} \quad (4.12)$$

Par construction du réseau Σ_r , pour tout $u \in B_r$, il existe un point $u_\ell(u) \in \Sigma_r$ tel que $|u - u_\ell(u)| \leq \frac{r}{\sqrt{n}}$.

Alors on a par l'inégalité de Cauchy-Schwarz et (4.11) :

$$|\eta(u) - \eta(u_\ell(u))|^2 \leq \frac{1}{n} \sum_{i=1}^n |a_{i,\gamma}(u_\ell(u)) - a_{i,\gamma}(u)|^2 \sum_{i=1}^n \varepsilon_i^2 \leq \frac{\kappa^2 r^2}{n} \sum_{i=1}^n \varepsilon_i^2.$$

Comme $\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 > 4\sigma^2\right)$ est plus petite que n^{-1} , donc

$$\mathbb{P}\left(\sup_{\gamma \in \Gamma} \sup_{u \in B_r} |\eta(u) - \eta(u_\ell(u))| > 2\kappa \sigma r\right) \leq \frac{1}{n}. \quad (4.13)$$

Par (4.12) et (4.13), on obtient

$$\begin{aligned} & \mathbb{P}\left(\sup_{\gamma \in \Gamma} \sup_{u \in B_r} |\eta(u)| > \kappa \sigma r (2 + \sqrt{2 \log(n|\Gamma|) + d \log(4n)})\right) \\ & \leq \mathbb{P}\left(\sup_{u_\ell \in \Sigma_r} |\eta(u_\ell)| > \kappa \sigma r \sqrt{2 \log(n|\Gamma|) + d \log(4n)}\right) \\ & \quad + \mathbb{P}\left(\sup_{u \in B_r} |\eta(u) - \eta(u_\ell(u))| > 2\kappa \sigma r\right) \leq \frac{2}{n} \end{aligned}$$

et le lemme est démontré. ■

Deuxième partie

Estimation et test d'une direction révélatrice unique : une approche géométrique

Introduction

Dans cette seconde partie, nous nous intéressons au modèle à direction révélatrice unique, c'est-à-dire vérifiant l'hypothèse

$$\mathcal{H}_0 : \quad \exists \varphi_0, \quad \mathbb{E} [Y_i | X_i] = \mathbb{E} [Y_i | X_i \cdot \varphi_0], \quad (4.14)$$

où “ \cdot ” désigne le produit scalaire dans \mathbb{R}^d , les (X_i, Y_i) constituent un échantillon i.i.d. de variables aléatoires à valeurs dans $\mathbb{R}^d \times \mathbb{R}$ et Y_i est de carré intégrable. Il est clair que si cette hypothèse est vérifiée, un estimateur de la fonction de régression de Y_i sur X_i sera fourni via un estimateur de Y_i sur $X_i \cdot \varphi_0$ et convergera donc à la vitesse unidimensionnelle. On rappelle que, en définissant g comme la fonction de régression de Y_i par rapport aux X_i , l'hypothèse \mathcal{H}_0 est équivalente à

$$\exists g_1, \exists \varphi_0 \quad g(\mathbf{x}) = g_1(\mathbf{x} \cdot \varphi_0) \quad \mathbb{P}_{X_i} - p.s. \quad (4.15)$$

Dans la suite, nous appellerons *fonction de lien*, la fonction g_1 et *axe* du modèle le vecteur φ_0 .

Le but de notre étude est de construire une procédure de test de l'hypothèse \mathcal{H}_0 .

A notre connaissance, le seul test existant actuellement de l'hypothèse \mathcal{H}_0 telle qu'écrivent ci-dessus, apparaît dans l'article de *Fan et Li (1996)*. Il s'appuie sur une statistique, convenablement normée, basée sur les résidus $Y_i - \hat{g}_1(X_i \cdot \hat{\varphi})$, où \hat{g}_1 (resp. $\hat{\varphi}$) est estimé non paramétriquement (resp. par la méthode ADE introduite par *Powell, Stock et Stoker en 1989*) sous l'hypothèse \mathcal{H}_0 . Les auteurs montrent que le test est convergent à un niveau fixé, mais n'étudient pas son comportement sous une suite d'alternatives du type :

$$\mathcal{H}_{1,n,c} : \quad g(\mathbf{x}) = g_1(\mathbf{x} \cdot \varphi_0) + cn^{-1/2}g_2(\mathbf{x}),$$

où g_2 ne s'écrit pas sous la forme $h(\mathbf{x} \cdot \varphi_0)$. Nous nous proposons ici de définir un test dont la puissance sous une suite d'hypothèses du type précédent tend vers 1.

D'autres tests ont été étudiés pour des hypothèses connexes à \mathcal{H}_0 . On peut citer *Härdle et Horowitz (1996)*, *Härdle, Sperlich et Spokoiny (1997)* et *Horowitz et Spokoiny (2001)*. Ces deux derniers considèrent une approche minimax du problème de test. Cependant ces tests correspondent à une hypothèse \mathcal{H}_0 paramétrique, par exemple $g(\mathbf{x}) = g_0(\mathbf{x} \cdot \varphi_0)$ avec g_0 fixée et connue.

Le test qui suit, basé sur un estimateur $\widehat{\varphi}$ de φ_0 , est construit par une méthode, à notre connaissance entièrement nouvelle, qui prend en compte une caractérisation géométrique de l'hypothèse de direction révélatrice unique.

La technique à mettre en œuvre pour la définir étant très lourde nous allons simplement, dans cette introduction, donner l'idée de cette propriété dans le cas où $d = 2$. On considère θ_0 un vecteur de direction orthogonale à φ_0 et les deux cylindres de direction θ_0 , \mathcal{S}_1 et \mathcal{S}_2 donnés sur la figure 4.4. Sous \mathcal{H}_0 , g est une fonction de $\mathbf{x} \in \mathbb{R}^2$, qui ne varie que suivant la valeur de la trace de \mathbf{x} sur φ_0 , on a donc

$$I_{\mathcal{S}_1} = \int_{\mathcal{S}_1} g(\mathbf{y}) d\lambda(\mathbf{y}) / \int_{\mathcal{S}_1} d\lambda(\mathbf{y}) = \int_{\mathcal{S}_2} g(\mathbf{y}) d\lambda(\mathbf{y}) / \int_{\mathcal{S}_2} d\lambda(\mathbf{y}) = I_{\mathcal{S}_2},$$

où λ est la mesure de Lebesgue sur \mathbb{R}^2 et en notant $I_{\mathcal{S}}$ l'intégrale moyenne de g sur un cylindre \mathcal{S} (un rectangle ici).

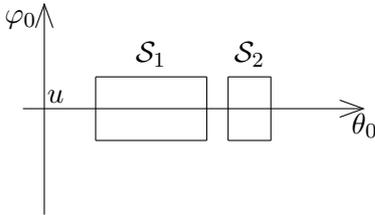


FIG. 4.4 – Cylindres de direction θ_0

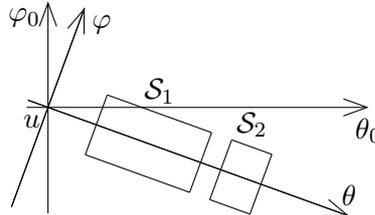


FIG. 4.5 – Cylindres de direction θ

Considérant maintenant une direction $\varphi \neq \varphi_0$, les traces sur φ_0 des cylindres d'axe $\theta \perp \varphi$ étant distinctes, les valeurs de $I_{\mathcal{S}}$ sont distinctes, cf. figure 4.5. Cette remarque est à la base de notre caractérisation géométrique. Pour capturer le fait que $I_{\mathcal{S}}$ reste constante ou non suivant la direction de l'axe des cylindres et à u fixé, on calcule la moyenne par rapport à \mathcal{S} (les centres et longueurs des cylindres varient) des intégrales $I_{\mathcal{S}}$, à φ et u fixés; on note $V(u, \varphi)$ le résultat. On définit alors une mesure de variation, notée $W(\varphi)$, des intégrales $I_{\mathcal{S}}$ correspondant à la direction φ par rapport à leur moyenne, en sommant les carrés des écarts entre les $I_{\mathcal{S}}$ et leurs moyennes $V(u, \varphi)$, d'abord sur \mathcal{S} (la largeur et le centre varient) puis sur u . La quantité $W(\varphi)$ peut servir à caractériser φ_0 et \mathcal{H}_0 , puisque $W(\varphi_0)$ est nulle sous \mathcal{H}_0 .

Clairement si nous savons estimer $\int_{\mathcal{S}} g d\lambda$ pour tout cylindre \mathcal{S} , nous disposerons d'un estimateur $\widehat{W}(\varphi)$, donc d'un estimateur de φ_0 en en prenant l'Argmin. L'estimateur de $I_{\mathcal{S}}$ peut résulter de la remarque suivante

$$\int_{\mathcal{S}} g d\lambda = \mathbb{E} [Y_i \mathbb{I}_{\{X_i \in \mathcal{S}\}} / f(X_i)],$$

où f est la densité de X_i . La définition de l'estimateur de $\int_{\mathcal{S}} g d\lambda$ dépendra donc essentiellement de l'estimateur choisi pour la quantité $1/f(X_i)$.

Dans le cas général le problème nécessite la définition soignée d'un grand nombre de quantités. Le travail correspondant sera ainsi divisé en cinq chapitres.

Le premier chapitre consiste en la formalisation et la généralisation de la propriété donnée dans l'exemple précédent. On y exprime un critère pour que \mathcal{H}_0 soit vérifiée, basé sur la caractérisation suivante, qui sera précisée par la suite : *si on considère des "cylindres" de même axe et de bases congruentes, les intégrales moyennes de la fonction de régression sur ces cylindres sont invariantes par changement de cylindre si et seulement si l'axe de ceux-ci est orthogonal à l'axe φ_0* . Le travail est alors séparé en quatre sous-paragraphes : le premier définit la notion de *cylindres axiaux*; dans le deuxième on montre que \mathcal{H}_0 est équivalente à la propriété donnée ci-dessus; on en déduit ensuite une condition globale qui s'écrit pour une certaine statistique W , " $W(\varphi_0) = 0$ "; dans le dernier point, on définit les statistiques permettant d'utiliser la propriété précédente : estimation de l'intégrale moyenne de g sur un cylindre \mathcal{S} , $I_{\mathcal{S}}$ et estimateur de W .

Les trois chapitres suivants donnent les résultats de convergence de $\widehat{I}_{\mathcal{S}}$, estimateur de $I_{\mathcal{S}}$, et ceux concernant $\widehat{\varphi}$ et $\widehat{W}(\widehat{\varphi})$, qui sont respectivement les estimateurs de φ_0 et $W(\varphi_0)$. Le dernier chapitre est consacré à la convergence du test sous \mathcal{H}_0 et sous différentes alternatives.

Chapitre 5

Approche géométrique de la condition H_0

5.1 Les cylindres

Soit \mathcal{Q} un sous-ensemble compact de \mathbb{R}^d convexe et $\Theta = \{\theta \in \mathbb{R}^d, \|\theta\| = 1\}$. Nous allons définir un outil essentiel pour la suite : les “cylindres”.

Définition 5.1.1. Soit $\theta \in \Theta$, $\theta \perp \varphi$

Soit $\mathcal{R}_{\theta\mathbf{x}}$ une sphère, de rayon r fixé, de centre \mathbf{x} , non dégénérée, appartenant à l’hyperplan orthogonal à θ passant par \mathbf{x} .

On définit alors le cylindre $\mathcal{S} = \mathcal{S}(s, \mathcal{R}_{\theta\mathbf{x}})$ comme étant le produit direct de $\mathcal{R}_{\theta\mathbf{x}}$ et d’un segment de longueur s orthogonal à $\mathcal{R}_{\theta\mathbf{x}}$, qui a son centre au centre, \mathbf{x} , de $\mathcal{R}_{\theta\mathbf{x}}$. On dira que s et $\mathcal{R}_{\theta\mathbf{x}}$ sont respectivement la “largeur” et la “base” de $\mathcal{S}(s, \mathcal{R}_{\theta\mathbf{x}})$.

Plus précisément, si on complète (θ, φ) en une base orthonormale de \mathbb{R}^d , on a $(\varphi_i)_{i=1, \dots, d}$ avec $\varphi_1 = \theta$ et $\varphi_2 = \varphi$.

$$\mathcal{S}(s, \mathcal{R}_{\theta\mathbf{x}}) = \left\{ \mathbf{y} \in \mathbb{R}^d / \mathbf{y} \cdot \theta \in [\mathbf{x} \cdot \theta - s/2, \mathbf{x} \cdot \theta + s/2], \sum_{i=2}^d (\mathbf{y} \cdot \varphi_i - \mathbf{x} \cdot \varphi_i)^2 \leq r^2 \right\}$$

Les figures 5.1 illustrent la définition appliquée aux cas $d = 2$ et $d = 3$.

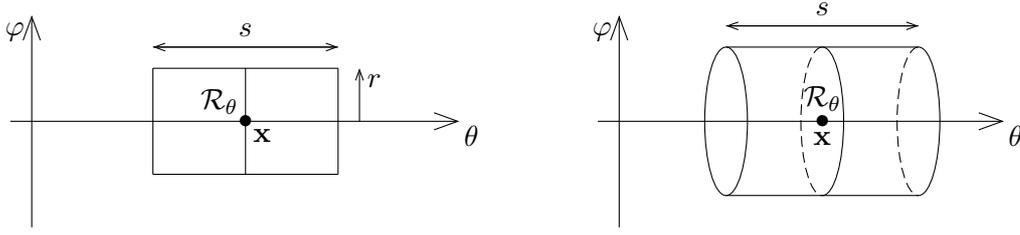


FIG. 5.1 – Cylindres en dimensions 2 et 3.

Définition 5.1.2. On dira que deux cylindres $\mathcal{S}(s, \mathcal{R}_{\theta\mathbf{x}})$ et $\mathcal{S}(s', \mathcal{R}_{\theta\mathbf{x}'})$ sont axiaux si

- leurs bases sont congruentes, i.e. elles ont la même projection sur un plan orthogonal au vecteur θ ;
- ils ont le même axe de direction θ ;
- leurs centres sont sur un même segment $\alpha = \alpha_\theta$ parallèle à la direction θ .

Les figures 4.4 et 4.5 illustrent cette définition. Plus précisément :

Si $\mathcal{S}(s, \mathcal{R}_{\theta\mathbf{x}})$ et $\mathcal{S}(s', \mathcal{R}_{\theta\mathbf{x}'})$ sont des cylindres axiaux, alors en utilisant les notations de la définition : \mathbf{x} et \mathbf{x}' sont tels que $(\mathbf{x} - \mathbf{x}')$ est parallèle à θ et pour tout $v \in \mathbb{R}$

$$\left\{ \mathbf{y} \in \mathbb{R}^d / \mathbf{y} \cdot \varphi_1 = v, \sum_{i=2}^d (\mathbf{y} \cdot \varphi_i - \mathbf{x} \cdot \varphi_i)^2 \leq r^2 \right\} = \left\{ \mathbf{y} \in \mathbb{R}^d / \mathbf{y} \cdot \varphi_1 = v, \sum_{i=2}^d (\mathbf{y} \cdot \varphi_i - \mathbf{x}' \cdot \varphi_i)^2 \leq r^2 \right\}$$

5.2 Un premier critère de l'hypothèse SIM

Dans toute la suite \mathcal{Q} est un ensemble de \mathbb{R}^d , compact et convexe. Soit g une fonction définie sur \mathcal{Q} .

Pour un cylindre $\mathcal{S} \subseteq \mathcal{Q}$, on définit

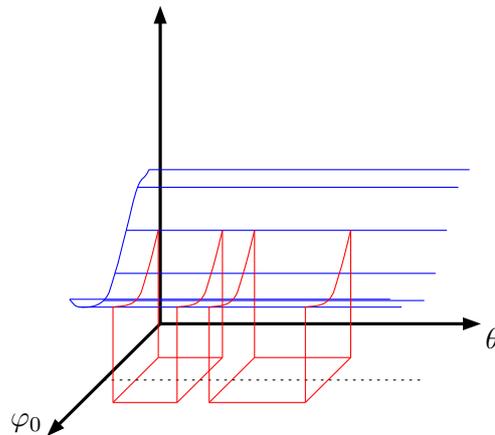
$$I_{\mathcal{S}}(g) = \left\{ \int_{\mathcal{S}} g(\mathbf{x}) d\mathbf{x} \right\} / \left(\int_{\mathcal{S}} d\mathbf{x} \right), \quad (5.1)$$

intégrale moyenne de g sur \mathcal{S} . Ainsi la propriété géométrique s'écrit de la façon suivante :

Propriété 5.2.1 (Caractérisation géométrique des SIM). La condition portant sur la fonction de régression :

$$g(\mathbf{x}) = g_1(\mathbf{x} \cdot \varphi_0) \text{ pour tout } \mathbf{x} \in \mathcal{Q}, \quad (5.2)$$

où g_1 est une fonction d'une variable réelle, est équivalente à l'existence d'un vecteur φ_0 tel que pour tout vecteur $\theta \perp \varphi_0$, tout segment α_θ de direction θ et $r > 0$, $I_{\mathcal{S}(s, \mathcal{R}_{\theta\mathbf{x}})}(g)$ ne varie pas avec les couples $(s, \mathcal{R}_{\theta\mathbf{x}})$ tels que $s > 0$, $\mathcal{R}_{\theta\mathbf{x}}$ est une boule dans le plan orthogonal à θ de rayon r et centrée en \mathbf{x} sur α_θ et $\mathcal{S}(s, \mathcal{R}_{\theta\mathbf{x}}) \subseteq \mathcal{Q}$.

FIG. 5.2 – Fonction de régression à direction révélatrice $\varphi_0 \in \mathbb{R}^2$.

DÉMONSTRATION.

Sens direct.

Soit $\theta \perp \varphi_0$, α_θ un segment de direction θ , repéré par un point terminal \mathbf{v} .

Soit $r > 0$ fixé. Considérons deux cylindres axiaux $\mathcal{S}(s, \mathcal{R}_{\theta\mathbf{x}})$ et $\mathcal{S}(s', \mathcal{R}_{\theta\mathbf{x}'})$. On veut faire un changement de variable dans l'intégrale multiple (5.1) : on complète (θ, φ_0) en une base de \mathbb{R}^d : $(\varphi_i)_{i=1, \dots, d}$ avec $\varphi_1 = \theta$ et $\varphi_2 = \varphi_0$. On pose $w_i = \mathbf{y} \cdot \varphi_i$, d'où $\mathbf{y} = \sum_{i=1}^n w_i \varphi_i$.

Il faut alors reparamétriser les cylindres : soit \mathcal{P} le sous-espace engendré par les vecteurs $(\varphi_i)_{i=2, \dots, d}$ et passant par l'extrémité \mathbf{v} de α_θ

$$\begin{aligned} \mathcal{S}(s, \mathcal{R}_\theta) &= \left\{ \mathbf{y} / \mathbf{y} \cdot \theta \in [\mathbf{x} \cdot \theta - s/2, \mathbf{x} \cdot \theta + s/2], (\mathbf{y} \cdot \varphi_i)_{i=2, \dots, d} \in Proj_{\mathcal{P}}(\mathcal{R}_\theta) \right\} \\ &= \left\{ \mathbf{y} / \mathbf{y} \cdot \theta \in [\mathbf{x} \cdot \theta - s/2, \mathbf{x} \cdot \theta + s/2], \sum_{i=2}^d (\mathbf{y} \cdot \varphi_i - \mathbf{v} \cdot \varphi_i)^2 \leq r^2 \right\} \\ &= \left\{ \mathbf{w} / w_1 \in [\mathbf{x} \cdot \theta - s/2, \mathbf{x} \cdot \theta + s/2], \sum_{i=2}^d (w_i - \mathbf{v} \cdot \varphi_i)^2 \leq r^2 \right\} \end{aligned}$$

De même pour $\mathcal{S}(s', \mathcal{R}_{\theta\mathbf{x}'})$ on obtient :

$$\mathcal{S}(s', \mathcal{R}_{\theta\mathbf{x}'}) = \left\{ \mathbf{w} / w_1 \in [\mathbf{x}' \cdot \theta - s'/2, \mathbf{x}' \cdot \theta + s'/2], \sum_{i=2}^d (w_i - \mathbf{v} \cdot \varphi_i)^2 \leq r^2 \right\}$$

En effet ces deux cylindres sont congruents et ne diffèrent donc que par l'emplacement de leur centre sur α_θ et par la largeur du cylindre, information qui est contenue dans w_1 .

Ainsi

$$I_S(g) = \frac{\int_{\Theta^{d-1}(\tilde{\mathbf{v}}, r)} g_1(w_2) d\mathbf{w}^{(1)} \int_{\mathbf{x} \cdot \theta - s/2}^{\mathbf{x} \cdot \theta + s/2} dw_1}{\int_{\Theta^{d-1}(\tilde{\mathbf{v}}, r)} d\mathbf{w}^{(1)} \int_{\mathbf{x} \cdot \theta - s/2}^{\mathbf{x} \cdot \theta + s/2} dw_1},$$

$$I_S(g) = \frac{\int_{\Theta^{d-1}(\tilde{\mathbf{v}}, r)} g_1(w_2) d\mathbf{w}^{(1)}}{\int_{\Theta^{d-1}(\tilde{\mathbf{v}}, r)} d\mathbf{w}^{(1)}},$$

où $\tilde{\mathbf{v}} = (\mathbf{v} \cdot \varphi_2, \dots, \mathbf{v} \cdot \varphi_d)$, $\mathbf{w}^{(1)} = (w_2, \dots, w_d)$ et $\Theta^{d-1}(\tilde{\mathbf{v}}, r)$ est la sphère en dimension $(d-1)$, de centre $\tilde{\mathbf{v}}$ et de rayon r . Finalement

$$I_{S(s, \mathcal{R}_{\theta \mathbf{x}})} = I_{S(s', \mathcal{R}_{\theta \mathbf{x}'})}.$$

Réciproque.

On montre la réciproque par l'absurde. Un modèle n'est pas un SIM s'il vérifie

$$\forall \varphi \in \Theta, \quad \exists \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{Q} \text{ tel que } \mathbf{x}_1 \cdot \varphi = \mathbf{x}_2 \cdot \varphi \text{ et } g(\mathbf{x}_1) \neq g(\mathbf{x}_2).$$

Montrons que l'on peut supposer que \mathbf{x}_1 et \mathbf{x}_2 ne sont pas sur le bord de \mathcal{Q} :

Si $\mathbf{x}_1 \in Fr(\mathcal{Q})$:

$$g(\mathbf{x}_1) \neq g(\mathbf{x}_2) \Rightarrow \exists V_o \in \mathcal{V}(g(\mathbf{x}_1)) \text{ et } U_o \in \mathcal{V}(g(\mathbf{x}_2)) \text{ tels que } V_o \cap U_o = \emptyset$$

Or g est continue en \mathbf{x}_1 , donc

$$\exists \mathcal{B}(\mathbf{x}_1, r_o) = W_o^1 \in \mathcal{V}(\mathbf{x}_1) \text{ telle que } g(W_o^1) \subset V_o, \quad (5.3)$$

alors l'hyperplan $\mathcal{H}_o = \{\mathbf{x}, \mathbf{x} \cdot \varphi = \mathbf{x}_1 \cdot \varphi\}$ est d'intersection non vide avec W_o^1 et on peut choisir $\mathbf{x}'_1 \in W_o^1$ qui n'est pas sur la frontière de \mathcal{Q} et tel que $\mathbf{x}'_1 \cdot \varphi = \mathbf{x}_1 \cdot \varphi$.

Or si \mathcal{H}_o est tangent au bord en \mathbf{x}_1 , alors $\mathbf{x}'_1 \notin \mathcal{Q}$, mais dans ce cas \mathbf{x}_2 , qui appartient aussi à \mathcal{H}_o , n'est plus dans \mathcal{Q} . On aboutit à une contradiction, ce cas n'est donc pas possible.

On peut donc toujours se ramener à $\mathbf{x}_1 \in Int(\mathcal{Q})$.

On peut faire de même pour \mathbf{x}_2 .

On suppose maintenant que \mathbf{x}_1 et \mathbf{x}_2 sont dans l'intérieur de \mathcal{Q} , on peut alors trouver $\mathcal{B}(\mathbf{x}_1, r_o^1) \subset W_o^1 \cap \mathcal{Q}$ et $\mathcal{B}(\mathbf{x}_2, r_o^2) \subset W_o^2 \cap \mathcal{Q}$, où W_o^i est défini comme dans (5.3), tels que $g(\mathcal{B}(\mathbf{x}_1, r_o^1)) \cap g(\mathcal{B}(\mathbf{x}_2, r_o^2)) = \emptyset$.

De plus comme g est à valeurs réelles et est continue, on peut supposer par exemple que

$$g(y) > g(z), \quad \forall y \in \mathcal{B}(\mathbf{x}_1, r_o^1), \quad \forall z \in \mathcal{B}(\mathbf{x}_2, r_o^2) \quad (5.4)$$

Pour $i = 1, 2$, si s et r sont suffisamment petits, il existe un cylindre $\mathcal{S}_i = \mathcal{S}(s, \mathcal{R}_{\theta \mathbf{x}_i})$ tel que $\mathcal{S}_i \subset \mathcal{B}(\mathbf{x}_i, r^i)$, où θ est la direction de l'axe passant par \mathbf{x}_1 et \mathbf{x}_2 . On remarque que \mathcal{S}_2 est obtenu à partir de \mathcal{S}_1 par la translation, qui transforme \mathbf{x}_1 en \mathbf{x}_2 . De plus \mathbf{x}_1 et \mathbf{x}_2 sont bien sur un même axe de direction $\theta \perp \varphi$, car $(\mathbf{x}_1 - \mathbf{x}_2) \cdot \varphi = 0$.

D'où $I_{\mathcal{S}_2}(g) = \int_{\mathcal{S}_1} g(t(\mathbf{y})) d\mathbf{y} / \int_{\mathcal{S}_1} d\mathbf{y}$ et donc en utilisant (5.4)

$$I_{\mathcal{S}_1}(g) > I_{\mathcal{S}_2}(g).$$

Finalement on a trouvé deux intégrales moyennes de g sur des cylindres axiaux de valeurs distinctes. La caractérisation par $I_{\mathcal{S}}$ n'est donc pas vérifiée. Ce qui achève la preuve. ■

La démonstration précédente laisse penser qu'il faut pouvoir prendre des cylindres aussi petit que l'on veut pour détecter la forme du modèle. En fait ce n'est ni nécessaire, ni souhaitable.

En effet pour détecter le fait qu'un modèle ne soit pas un SIM il faut que la taille des cylindres ne soit pas trop petite, sinon les variations de l'intégrale $\int_{\mathcal{S}} g d\lambda$ vont être absorbées par le volume de $\int_{\mathcal{S}} d\lambda$. D'autre part pour déterminer la valeur de φ on utilise ces intégrales sur des cylindres dont la taille doit être minorée, cf. *chapitre 8*.

En fait on va travailler avec des cylindres relativement grands dans le sens où leur taille ne convergera pas vers 0 quand n tend vers l'infini. Le rayon r ne jouera donc pas le rôle de la largeur de fenêtre traditionnelle.

Pour expliquer pourquoi on peut se fixer la taille de cylindre, on développe les arguments pour le cas $d = 2$. Pour celui-ci les bases des cylindres sont des segments et les cylindres sont des rectangles.

Pour des raisons techniques qui apparaîtront par la suite, afin d'éviter les effets de bord, on considère les valeurs de $g(\mathbf{x})$ uniquement pour $\mathbf{x} \in \mathcal{Q}_\delta$, où $\delta > 0$ étant fixé \mathcal{Q}_δ est défini par

$$\mathcal{Q}_\delta = \{\mathbf{x} \in \mathcal{Q} / d(\mathbf{x}, Fr(\mathcal{Q})) > \delta\}. \quad (5.5)$$

Pour $\delta > 0$ fixé, on pourra choisir l , $s_1 > 0$, $s_2 > s_1$ tels que pour tout $\theta \in \Theta$

$$\mathcal{Q}_\delta \subset \bigcup_{s \in [s_1, s_2], r \geq l} \mathcal{S}(s, \mathcal{R}_{\theta \mathbf{x}}^r) \subset \mathcal{Q},$$

où $\mathcal{R}_{\theta \mathbf{x}}^r$ est la sphère des rayon r de centre \mathbf{x} appartenant à l'hyperplan orthogonal à θ et passant par \mathbf{x} .

Comme \mathcal{Q} est compact, on le recouvre par une union finie de ces cylindres $\mathcal{S}(s, \mathcal{R}_{\theta \mathbf{x}}^r)$. On retire de l'union ceux qui recouvrent uniquement $\mathcal{Q} \setminus \mathcal{Q}_\delta$. On obtient alors

$$\mathcal{Q}_\delta \subset \bigcup \mathcal{S}(s, \mathcal{R}_{\theta \mathbf{x}}^r) \subset \mathcal{Q}$$

Le choix de r , s_1 et s_2 sera donné dans le paragraphe 5.3.

On va montrer que pour détecter pour tout $\varphi \in \Theta$, l'existence de \mathbf{x}_1 et \mathbf{x}_2 dans \mathcal{Q}_δ tel que $\mathbf{x}_1 \cdot \varphi = \mathbf{x}_2 \cdot \varphi$ et $g(\mathbf{x}_1) \neq g(\mathbf{x}_2)$, il suffit de prendre des bases de rayon compris entre $l/2$ et l .

On considère les cylindres d'axe $\theta \perp \varphi$, de largeur $s \in [s_1, s_2]$ et de base $\mathcal{R}_{\theta\mathbf{x}}$ de rayon l , contenant \mathbf{x}_1 et \mathbf{x}_2 .

Si les intégrales moyennes sur deux cylindres de même trace sur φ sont différentes, alors on a fini. Sinon toutes les intégrales ayant la même trace sont égales. Comme le centre de chaque cylindre varie continûment sur un segment, on peut considérer des cylindres infinitésimaux, i.e. $\mathcal{S}(0, \mathcal{R}_{\theta\mathbf{x}}) = \mathcal{R}_{\theta\mathbf{x}}$.

Comme $g(\mathbf{x}_1) \neq g(\mathbf{x}_2)$ il existe $\mathcal{R}'_{\theta\mathbf{x}_1}$ et $\mathcal{R}'_{\theta\mathbf{x}_2}$ tel que $\int_{\mathcal{R}'_{\theta\mathbf{x}_1}} g d\lambda \neq \int_{\mathcal{R}'_{\theta\mathbf{x}_2}} g d\lambda$

On choisit alors la base $\mathcal{R}_{\theta\mathbf{x}_1}$ tel que $\mathcal{R}_{\theta\mathbf{x}_1} \setminus \mathcal{R}'_{\theta\mathbf{x}_1}$ soit encore un segment (cf. figure 5.3).

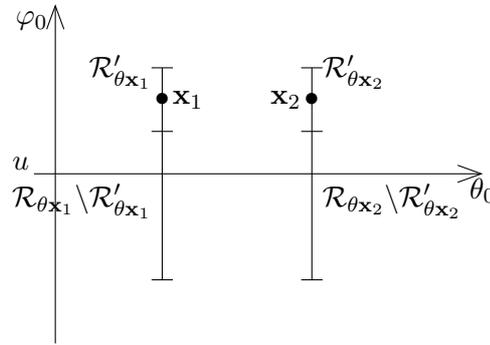


FIG. 5.3 – Choix de $r \in [l/2, l]$

Ainsi

$$\int_{\mathcal{R}_{\theta\mathbf{x}_1} \setminus \mathcal{R}'_{\theta\mathbf{x}_1}} g d\lambda \neq \int_{\mathcal{R}_{\theta\mathbf{x}_2} \setminus \mathcal{R}'_{\theta\mathbf{x}_2}} g d\lambda$$

Finalement on note $\mathcal{R}''_{\theta\mathbf{x}_i}$ celui des deux segments $\mathcal{R}_{\theta\mathbf{x}_i}$ et $\mathcal{R}_{\theta\mathbf{x}_i} \setminus \mathcal{R}'_{\theta\mathbf{x}_i}$ qui a une longueur plus grande que $l/2$. Toutes les intégrales sur les cylindres congruents aux cylindres $\mathcal{R}''_{\theta\mathbf{x}_i}$ ne sont donc pas égales. Ainsi toutes les intégrales sur $\mathcal{S}(s, \mathcal{R}''_{\theta\mathbf{x}})$, où $\mathcal{R}''_{\theta\mathbf{x}}$ est de rayon au moins $l/2$ et de centre \mathbf{x} , ne sont pas égales.

On a donc montré qu'il n'est pas nécessaire de prendre r convergeant 0.

Dans la suite on prendra des cylindres dont les bases auront toutes le même rayon $r > 0$.

5.3 Critère global de l'hypothèse SIM

Notre but est maintenant de mesurer la variabilité de $I_{\mathcal{S}}$. Il nous faut capturer les variations de $I_{\mathcal{S}}$ lorsque la direction des cylindres change. On sait que la valeur de $I_{\mathcal{S}}$ est constante sur tous les cylindres dont les axes sont contenus dans un même hyperplan fixé orthogonal à φ_0 . On va donc considérer l'intégrale des carrés des écarts de $I_{\mathcal{S}}$ à sa moyenne suivant

toutes les directions possibles et tous les hyperplans possibles.

On paramétrise alors la famille des cylindres de la façon suivante (*cf. figure 5.4 et 5.5*)

- Toutes les bases $\mathcal{R}_{\theta\mathbf{x}}$ sont de même rayon égal à $r > 0$.
- La largeur s des cylindres variera entre s_1 et s_2 , où $0 < s_1 < s_2 < +\infty$. Le choix de r, s_1, s_2 sera expliqué dans le paragraphe suivant.
- Le centre \mathbf{x} de la base $\mathcal{R}_{\theta\mathbf{x}}$ appartient à un segment $\alpha_\theta \subseteq \mathcal{Q}$, parallèle à θ . On pose $\mathcal{T}(\alpha_\theta)$ l'ensemble des couples (s, \mathbf{x}) tel que $s \in [s_1, s_2]$ et $\mathbf{x} \in \alpha_\theta$.
- Pour localiser les segments α_θ , on les indexe par le vecteur $\mathbf{v} \in \mathcal{Q}$ qui repère un de leurs points terminaux : $\alpha_\theta = \alpha_\theta(\mathbf{v})$. Nous serons obligés de limiter les segments utilisés de telle sorte que, inclus dans \mathcal{Q} , les points terminaux n'appartiennent pas à la frontière de \mathcal{Q} , pour les raisons exposées au paragraphe suivant.
- Pour chaque axe $\varphi \in \Theta$, on peut définir $\mathcal{V}(\varphi)$, l'ensemble des \mathbf{v} , comme $\bigcup_{u \in \mathcal{U}(\varphi)} \mathcal{V}(u, \varphi)$, où $\mathcal{U}(\varphi) \subset \mathbb{R}$ et $\mathcal{V}(u, \varphi) = \{\mathbf{v} \in \mathcal{V}(\varphi) : \mathbf{v} \cdot \varphi = u\}$, les $\mathcal{V}(u, \varphi)$ étant choisis de telle sorte que les extrémités \mathbf{v} forment un ensemble connexe.

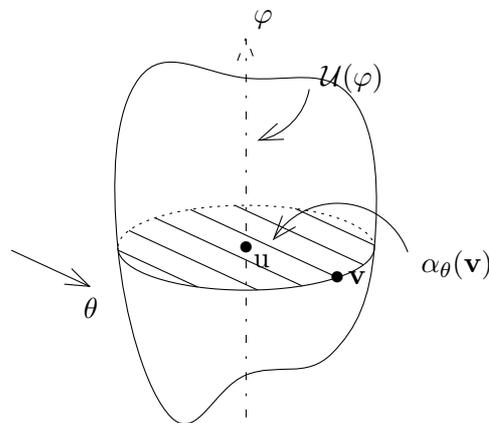


FIG. 5.4 – Paramétrisation de \mathcal{Q}_δ en dimension 3.

On choisit r, s_1 et s_2 de telle sorte que les cylindres recouvrent \mathcal{Q}_δ , défini par (5.5), c'est-à-dire $0 < r < \delta$ et $s_2 < 2(\delta^2 - r^2)^{1/2}$. En effet on choisit s_2 tel que les cylindres admissibles restent dans \mathcal{Q} (centre dans \mathcal{Q}_δ). Pour cela on impose qu'ils soient inclus dans des boules de rayon δ , à $r < \delta$ fixé, *cf. figure 5.6*. Il faut donc avoir l'inégalité $s_2/2 \leq \sqrt{\delta^2 - r^2}$.

Dans la suite, les cylindres \mathcal{S} considérés seront ceux qui recouvrent \mathcal{Q}_δ , dont les centres sont contenus dans \mathcal{Q}_δ . Ils seront donc entièrement contenus dans \mathcal{Q} , par les choix précédents. Grâce à cette définition on va pouvoir passer d'une caractérisation locale à une propriété globale.

On pose $\Theta(\varphi) = \{\theta \in \Theta : \theta \perp \varphi\}$.

Si g est un SIM d'axe φ , le critère de la proposition 5.2.1 se réécrit :

$\forall u \in \mathcal{U}(\varphi), \forall \theta \in \Theta(\varphi)$ et $\forall (s, \mathbf{x}) \in \mathcal{T}(\alpha_\theta(\mathbf{v}))$, où $\mathbf{v} \in \mathcal{V}(u, \theta)$, $I_{\mathcal{S}(s, \mathcal{R}_{\theta\mathbf{x}})}(g)$ reste constante.

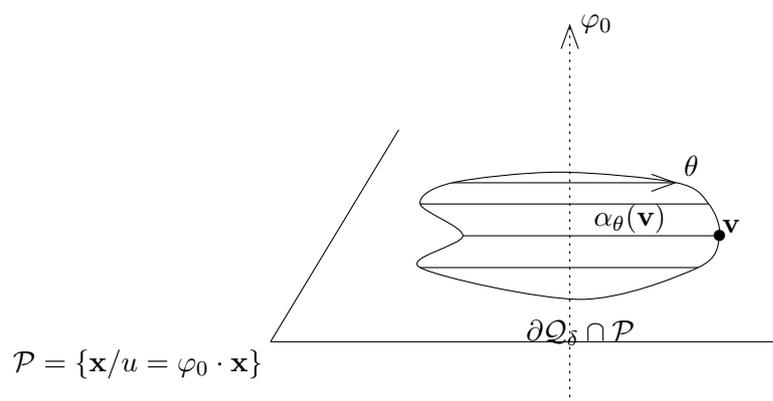
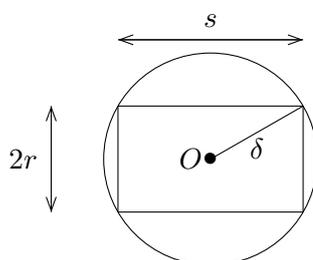


FIG. 5.5 – Paramétrisation des cylindres en dimensions 3.

FIG. 5.6 – Choix de s_2

On définit alors la valeur moyenne de I_S sur les directions θ

$$V(u, \varphi) = \frac{\int_{\mathcal{V}(u, \varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_\theta(\mathbf{v})\}} I_{S(s, \mathcal{R}_{\theta\mathbf{x}})}(g) ds d\mathbf{x}}{\int_{\mathcal{V}(u, \varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_\theta(\mathbf{v})\}} ds d\mathbf{x}},$$

qui est constant, égal à toutes les intégrales moyennes $I_{S(s, \mathcal{R}_{\theta\mathbf{x}})}(g)$, si φ est l'axe du SIM. Pour mesurer la variation de I_S , on considère la somme de ses écarts à $V(u, \varphi)$ au carré sur tous les cylindres, *i.e.*

$$W(\varphi) = \int_{\mathcal{U}(\varphi)} du \int_{\mathcal{V}(u, \varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_\theta(\mathbf{v})\}} \{I_{S(s, \mathcal{R}_{\theta\mathbf{x}})}(g) - V(u, \varphi)\}^2 ds d\mathbf{x}. \quad (5.6)$$

La caractérisation est alors la suivante :

Théorème 5.3.1 (Propriété globale des SIM). *Si la condition (5.2) est vérifiée sur \mathcal{Q} alors $\exists \varphi_0$, $W(\varphi_0) = 0$.*

Si $\exists \varphi_0$, $W(\varphi_0) = 0$ alors le modèle est à direction révélatrice unique sur \mathcal{Q}' , sous-ensemble de \mathcal{Q} recouvert par les cylindres, i.e.

$$\mathcal{Q}' = \bigcup_{\phi \in \Theta} \bigcup_{u \in \mathcal{U}(\varphi)} \bigcup_{\mathbf{v} \in \mathcal{V}(u, \varphi)} \bigcup_{\theta \in \Theta(\varphi)} \bigcup_{(s, \mathbf{x}) \in \mathcal{T}\{\alpha_\theta(\mathbf{v})\}} \mathcal{S}(s, \mathcal{R}_{\theta \mathbf{x}}).$$

DÉMONSTRATION.

Sens direct.

D'après ce qui précède si on a un SIM d'axe φ_0 , $W(\varphi_0) = 0$.

Réciproque.

Si $W(\varphi_0) = 0$, comme W est l'intégrale d'une fonction continue, voir Annexe 9.2, et positive, alors

$$\forall u, (s, \mathbf{x}), \mathbf{v}, \theta \quad I_{\mathcal{S}(s, \mathcal{R}_{\theta \mathbf{x}})}(g) = V(u, \varphi_0) \quad (5.7)$$

De plus $(s, \mathbf{x}; \theta) \mapsto I_{\mathcal{S}(s, \mathcal{R}_{\theta \mathbf{x}})}(g)$ est une fonction continue sur un compact, cf. Annexe 9.2, elle est donc bornée et atteint ses bornes : Soit u_0 fixé

$$\exists (s_0, \mathbf{x}_0), \mathbf{v}_0, \theta_0, \quad I_{\mathcal{S}(s_0, \mathcal{R}_{\theta_0 \mathbf{x}_0})}(g) = \inf \{ I_{\mathcal{S}(s, \mathcal{R}_{\theta \mathbf{x}})}(g), (s, \mathbf{x}), \mathbf{v}, \theta \}$$

Ainsi pour u_0 , (5.7) se réécrit

$$\int_{\mathcal{V}(u_0, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta \int_{\mathcal{T}\{\alpha_\theta(\mathbf{v})\}} \left\{ I_{\mathcal{S}(s, \mathcal{R}_{\theta \mathbf{x}})}(g) - I_{\mathcal{S}(s_0, \mathcal{R}_{\theta_0 \mathbf{x}_0})}(g) \right\} ds d\mathbf{x} = 0.$$

Ceci est l'intégrale d'une fonction positive continue donc

$$\forall (s, \mathbf{x}), \mathbf{v}, \theta, \quad I_{\mathcal{S}(s, \mathcal{R}_{\theta \mathbf{x}})}(g) = I_{\mathcal{S}(s_0, \mathcal{R}_{\theta_0 \mathbf{x}_0})}(g).$$

Comme on peut montrer ce résultat pour tous u_0 , on obtient la propriété (5.2.1) du paragraphe précédent pour tout les cylindres d'axe orthogonal à φ_0 recouvrant \mathcal{Q}_δ . Finalement g est un SIM sur \mathcal{Q}' . ■

Cette dernière propriété nous incite à définir l'hypothèse nulle comme étant

$$\mathcal{H}_0 : g \text{ est un SIM sur } \mathcal{Q}'. \quad (5.8)$$

Ainsi on ne teste pas le modèle SIM sur \mathcal{Q} tout-entier, mais sur un sous-ensemble \mathcal{Q}' . Par la suite on définit φ_0 , que \mathcal{H}_0 soit vérifiée ou non, par

$$\varphi_0 = \underset{\varphi \in \Theta}{\operatorname{argmin}} W(\varphi),$$

il est clair que sous \mathcal{H}_0 l'axe du SIM réalise le minimum de $W(\varphi)$. De plus \mathcal{H}_0 est équivalente à $W(\varphi_0) = 0$. Ainsi la quantité $W(\varphi)$ mesure la variation de $I_{\mathcal{S}}(g)$ et nous allons donc l'utiliser dans les deux problèmes : estimation et test.

Il reste donc à estimer la valeur de $W(\varphi)$ à partir d'un échantillon i.i.d. et à en déduire un test de \mathcal{H}_0 . Ceci fait l'objet du paragraphe suivant.

5.4 La méthode statistique : description

On considère le modèle général

$$Y_i = g(X_i) + \sigma(X_i)\varepsilon_i,$$

où g est une fonction définie sur un ensemble \mathcal{Q} de \mathbb{R}^d et à valeurs dans \mathbb{R} ; les erreurs ε_i sont des variables aléatoires indépendantes et identiquement distribuées de moyenne nulle et de variance unité; la fonction σ est bornée et positive; $\mathcal{X} = \{X_1, \dots, X_n\}$ est un échantillon aléatoire de vecteurs de \mathbb{R}^d indépendants des erreurs et de densité f sur \mathcal{Q} .

On observe les paires $\mathcal{Z} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$.

Compte tenu du paragraphe précédent, il nous faut maintenant estimer les quantités caractéristiques du modèle : $I_S(g)$, V , W .

Si on dispose d'un estimateur de $I_S(g)$, noté $\widehat{I}_S(g)$, on définit alors la mesure empirique de variation de $\widehat{I}_S(g)$ pour différents choix de base $\mathcal{R}_{\theta\mathbf{x}}$, dont l'axe θ est orthogonal à φ , de la façon suivante

$$\widehat{W}(\varphi) = \int_{\mathcal{U}(\varphi)} du \int_{\mathcal{V}(u,\varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_\theta(\mathbf{v})\}} \{\widehat{I}_{S(s,\mathcal{R}_{\theta\mathbf{x}})}(g) - \widehat{V}(u,\varphi)\}^2 ds d\mathbf{x}. \quad (5.9)$$

où \widehat{V} est l'estimateur de la valeur constante V :

$$\widehat{V}(u,\varphi) = \frac{\int_{\mathcal{V}(u,\varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_\theta(\mathbf{v})\}} \widehat{I}_{S(s,\mathcal{R}_{\theta\mathbf{x}})}(g) ds d\mathbf{x}}{\int_{\mathcal{V}(u,\varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_\theta(\mathbf{v})\}} ds d\mathbf{x}}. \quad (5.10)$$

On obtient finalement un nouvel estimateur de l'axe du SIM :

$$\widehat{\varphi} = \operatorname{argmin}_{\varphi \in \Theta} \widehat{W}(\varphi).$$

On peut alors tester l'hypothèse \mathcal{H}_0 , définie en (5.8), en rejetant \mathcal{H}_0 si $\widehat{W}(\widehat{\varphi})$ est plus grand qu'un seuil, qui sera déterminé à l'aide d'une procédure bootstrap.

Il nous reste alors à définir un estimateur de $I_S(g)$. Plusieurs méthodes s'imposent à nous suivant l'estimateur choisi pour $1/f(X_i)$.

- Notre première idée a été de faire intervenir la distance au $j^{\text{ème}}$ plus proche voisin dans l'échantillon :

On note $D_j(\mathbf{x})$ la distance de $\mathbf{x} \in \mathbb{R}^d$ à son $j^{\text{ème}}$ plus proche voisin $X_i \in \mathcal{X}$, pour lequel $X_i \neq \mathbf{x}$, et $D^d(\cdot)$ représente une combinaison linéaire des $D_j^d(\cdot)$, choisie telle que, uniformément par rapport à \mathbf{x}

$$E\{D^d(\mathbf{x})\} = \{n v_d f(\mathbf{x})\}^{-1} + o(n^{-3/2}), \quad (5.11)$$

où $v_d = \pi^{d/2}/\Gamma(1 + \frac{1}{2}d)$ est le volume de la boule unité en dimension d . L'annexe 11.2.2 fournit les combinaisons linéaires donnant un tel développement. On peut alors vérifier que l'estimateur

$$\tilde{I}_{\mathcal{S}}(g) = \|\mathcal{S}\|_d^{-1} \sum_{i=1}^n Y_i \mathbb{I}_{(X_i \in \mathcal{S})} D^d(X_i), \quad (5.12)$$

où $\|\mathcal{S}\|_d$ désigne le volume de \mathcal{S} , est asymptotiquement sans biais pour $v_d^{-1}I_{\mathcal{S}}(g)$, cf. *paragraphe 12.2*. En revanche dans notre cas nous avons besoin d'une convergence uniforme de la statistique $\tilde{I}_{\mathcal{S}}$. Or $D^d(X_i)$ n'est pas un estimateur consistant de $1/f(X_i)$, car on a

$$\lim_{n \rightarrow \infty} \mathcal{L}\left(nD^d(\mathbf{x})\right) = \mathcal{L}\left(\frac{Z(\mathbf{x})}{f(\mathbf{x})v_d}\right),$$

où $Z(\mathbf{x})$ est une variable aléatoire de loi exponentielle, cf. *Annexe 12.3*, et ceci pose problème pour obtenir la convergence à la vitesse $n^{1/2}$.

• Pour palier ce problème, une autre possibilité serait de faire intervenir un estimateur de $\mathbb{E}[D^d(X_i)|X_i]$. On peut par exemple choisir de remplacer $D^d(X_i)$ dans l'expression de $\tilde{I}_{\mathcal{S}}$ par

$$\sum_{j=1}^n D^d(X_i) \mathbb{I}_{\|X_i - X_j\| \leq \eta} / \sum_{j=1}^n \mathbb{I}_{\|X_i - X_j\| \leq \eta}$$

η représente alors une largeur de fenêtre.

• Une alternative pour conserver la distance $D^d(X_i)$ est de tronquer la statistique et d'utiliser les résultats démontrés par *Bickel et Breiman (1983)*.

Dans les cas précédents la loi limite sera sûrement la somme de deux processus gaussiens indépendants et donc le test en découlant sera moins puissant que celui présenté ici.

• Nous proposons de prendre comme estimateur de $f(X_i)$:

$$\hat{f}^{(i)}(X_i) = \frac{1}{nb^d} \sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{X_j - X_i}{b}\right),$$

où K est un noyau produit à support compact et b la largeur de fenêtre de la forme $n^{-\kappa}$ avec $\kappa < 1$. L'estimateur que l'on choisit pour $I_{\mathcal{S}}(g)$ est alors

$$\hat{I}_{\mathcal{S}}(g) = (n\|\mathcal{S}\|_d)^{-1} \sum_{i=1}^n Y_i \mathbb{I}_{(X_i \in \mathcal{S})} [\hat{f}^{(i)}(X_i)]^{-1}.$$

Dans un premier temps nous allons présenter les hypothèses liées au modèle et à la technique d'estimation de f . Dans les chapitres suivants nous étudierons successivement les propriétés de convergence de $\hat{I}_{\mathcal{S}}$, de l'estimateur de l'axe du SIM, $\hat{\varphi}$ et de la statistique

de test, $\widehat{W}(\widehat{\varphi})$. Finalement nous expliquerons en quoi consiste la calibration du test par bootstrap et les propriétés de celui-ci.

Chapitre 6

Hypothèses sur le modèle

On suppose que les données (X_i, Y_i) , $1 \leq i \leq n$ sont générées par le modèle

$$Y_i = g(X_i) + \sigma(X_i)\varepsilon_i.$$

Les hypothèses suivantes concernent le modèle ci-dessus :

Hypothèse H1. [Variables explicatives] (X_1, \dots, X_n) sont des variables aléatoires indépendantes et identiquement distribuées de densité f de support \mathcal{Q} .

Hypothèse H2. [Support des X_i] \mathcal{Q} est un sous-ensemble de \mathbb{R}^d non-dégénéré, convexe, compact et dont la frontière admet une tangente en tout point.

Hypothèse H3. [Densité des X_i] La densité f est séparée de 0 sur \mathcal{Q} , i.e.

$$\exists m > 0 \quad \text{tel que} \quad \forall \mathbf{x} \in \mathcal{Q}, \quad f(\mathbf{x}) \geq m > 0.$$

Hypothèse H4. [Erreurs] Les ε_i sont des variables aléatoires indépendantes identiquement distribuées de loi $\mathcal{N}(0, 1)$. Elles sont aussi indépendantes des X_i .

Hypothèse H5. [Variance] La fonction σ est continue, positive et bornée sur \mathcal{Q} par M_σ .

Hypothèse H6. [Régularité de f et g] La fonction de densité f admet $\tau + 1 \geq d + 2$ dérivées uniformément bornées sur \mathcal{Q} ; la fonction de régression g est continûment dérivable sur \mathbb{R} .

Pour le théorème 8.0.2, on suppose de plus

Hypothèse H7. [Modèle à direction révélatrice unique] La fonction g est à direction révélatrice sur \mathcal{Q} , i.e. il existe un unique vecteur $\varphi_0 \in \Theta = \{\theta \in \mathbb{R}^d / \|\theta\| = 1\}$ et une unique fonction univariée g_1 telle que $g(\mathbf{x}) = g_1(\varphi_0 \cdot \mathbf{x})$, pour tout $\mathbf{x} \in \mathcal{Q}$.

Hypothèse H8. φ_0 est le minimum de W , voir (5.6), et est atteint de façon quadratique, i.e. $\exists \eta > 0$ et une constante, c_3 , indépendante de φ , tels que

$$|W(\varphi) - W(\varphi_0)| > c_3 \|\varphi - \varphi_0\|^2, \quad \forall \varphi / \|\varphi - \varphi_0\| \leq \eta$$

Détaillons les hypothèses concernant la paramétrisation des “cylindres”, expliquée au paragraphe 5.3. Rappelons que $\delta > 0$ est un réel quelconque.

Hypothèse H9. *Toutes les bases des cylindres, \mathcal{R}_θ , ont le même rayon $0 < r < \delta$.*

Hypothèse H10. *À θ fixé, chaque segment $\alpha_\theta = \alpha_\theta(\mathbf{v})$ est dans \mathcal{Q} et est choisi aussi grand que possible pour avoir pour tout centre \mathbf{x} de cylindre :*

$$\mathbf{x} \in \alpha_\theta \cap \mathcal{Q}_\delta.$$

De plus $0 < s_1 < s_2 < 2\sqrt{\delta^2 - r^2}$.

Ainsi les deux hypothèses précédentes s’interprètent de la façon suivante : en prenant r , s_1 et s_2 suffisamment petits relativement à δ , on a : pour tout $\mathbf{x} \in \alpha_\theta$, $\mathcal{S}(s_2, \mathcal{R}_{\theta\mathbf{x}}) \cap \mathcal{Q}_\delta \neq \emptyset$, $\mathbf{x} \in \mathcal{Q}_\delta$ et pour tout $\theta \in \Theta$ et tout $\mathbf{x} \in \alpha_\theta$, $\mathcal{S}(s_2, \mathcal{R}_{\theta\mathbf{x}}) \subseteq \mathcal{Q}$.

Hypothèse H11. *Pour tout u et tout φ $\mathcal{V}(u, \varphi)$ et $\mathcal{U}(\varphi)$ sont choisis aussi grands que possible, tels que l’hypothèse H10 soit vérifiée.*

Hypothèse H12. *La mesure de Lebesgue de $\mathcal{V}(u, \varphi)$ est séparée de 0 uniformément pour $u \in \mathcal{U}(\varphi)$ et $\varphi \in \Theta$.*

On donne alors les hypothèses concernant l’estimation de la densité f par la méthode des noyaux

Hypothèse H13. [Noyau] *Le noyau K est une fonction de \mathbb{R}^d dans \mathbb{R} ayant la forme suivante :*

$$\forall \mathbf{x} = (x_1, \dots, x_d), \quad K(\mathbf{x}) = \prod_{i=1}^d K_1(x_i),$$

où K_1 est un noyau à support compact \mathbb{K} d’ordre $\tau > d$.

Hypothèse H14. [Largeur de fenêtre] *La largeur de fenêtre b est choisie telle que $b = n^{-\kappa}$ et*

$$\frac{1}{2\tau-1} < \kappa < \frac{1}{2d}.$$

Pour définir les distributions asymptotiques de $\widehat{\varphi}$ et $\widehat{W}(\widehat{\varphi})$ sous l’hypothèse de SIM, on utilise un processus gaussien $\zeta(\mathcal{S})$ de moyenne nulle, indexé par les ensembles \mathcal{S} et de covariance définie par

$$\text{cov}\{\zeta(\mathcal{S}_1), \zeta(\mathcal{S}_2)\} = \int_{\mathcal{S}_1 \cap \mathcal{S}_2} \sigma^2(\mathbf{x}) f^{-1}(\mathbf{x}) d\mathbf{x}.$$

On définit

$$\Delta(s, u, \mathbf{x}, \theta, \varphi) = I_{\mathcal{S}(s, \mathcal{R}_{\theta\mathbf{x}})}(g) - V(u, \varphi). \quad (6.1)$$

Propriété 6.0.1. *Il existe une fonction β telle que*

$$\Delta(s, u, \mathbf{x}, \theta, \varphi) = (\varphi - \varphi_0)^T \beta(s, u, \mathbf{x}, \theta_0, \varphi_0) + \mathcal{O}(\|\varphi - \varphi_0\|), \quad (6.2)$$

quand $\varphi \rightarrow \varphi_0$ suivant Θ .

DÉMONSTRATION.

Soit $\theta \in \Theta(\varphi)$, on pose $\theta_0 = \theta_0(\theta, \varphi)$ défini par :

$$\theta_0 \perp \varphi_0 \text{ et } \theta_0 \in \text{Vect}(\theta, \varphi)$$

alors $\theta - \theta_0 = R(\varphi - \varphi_0)$, où R est une rotation qui transforme φ_0 en θ_0 . Comme l'application $\varphi \mapsto \Delta(s, u, \mathbf{x}, \theta, \varphi)$ est de classe \mathcal{C}^1 , on peut écrire

$$\Delta(s, u, \mathbf{x}, \theta, \varphi) = \Delta(s, u, \mathbf{x}, \theta_0, \varphi_0) + (\varphi - \varphi_0)^T \frac{\partial \Delta(s, u, \mathbf{x}, \theta_0, \varphi_0)}{\partial \varphi} + \mathcal{O}(\|\varphi - \varphi_0\|),$$

car

$$\frac{\partial \Delta(s, u, \mathbf{x}, \theta_0, \varphi_0)}{\partial \varphi} = \left. \frac{\partial \Delta(s, u, \mathbf{x}, \theta_0 + R(\varphi - \varphi_0), \varphi)}{\partial \varphi} \right|_{\varphi = \varphi_0}$$

que l'on note dans la suite $\beta(s, u, \mathbf{x}, \theta_0, \varphi_0)$. En utilisant la caractérisation locale des SIM donnée dans la proposition 5.2.1, on obtient la relation désirée :

$$\Delta(s, u, \mathbf{x}, \theta, \varphi) = (\varphi - \varphi_0)^T \beta(s, u, \mathbf{x}, \theta_0, \varphi_0) + \mathcal{O}(\|\varphi - \varphi_0\|),$$

lorsque $\varphi \rightarrow \varphi_0$ suivant Θ . ■

On définit alors des nouvelles quantités qui interviennent dans l'expression des limites de $\widehat{\varphi}$ et \widehat{W} . Soit ω un vecteur de \mathbb{R}^d , on pose

$$M_\xi(u, \varphi_0) = \frac{\int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} \xi(s, u, \mathbf{x}, \theta_0, \varphi_0) ds d\mathbf{x}}{\int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} ds d\mathbf{x}}$$

où

$$\xi(s, u, \mathbf{x}, \theta_0, \varphi_0) = \|\mathcal{S}(s, \mathcal{R}_{\theta_0 \mathbf{x}})\|_d^{-1} \zeta\{\mathcal{S}(s, \mathcal{R}_{\theta_0 \mathbf{x}})\}, \quad (6.3)$$

et on définit

$$\begin{aligned} \overline{W}(\omega) = & \int_{\mathcal{U}(\varphi_0)} du \int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} \left[\|\mathcal{S}(s, \mathcal{R}_{\theta_0 \mathbf{x}})\|_d^{-1} \zeta\{\mathcal{S}(s, \mathcal{R}_{\theta_0 \mathbf{x}})\} \right. \\ & \left. + \omega^T \beta(s, u, \mathbf{x}, \theta_0, \varphi_0) - M_\xi(u, \varphi_0) \right]^2 ds d\mathbf{x}. \end{aligned} \quad (6.4)$$

On remarque que la fonction \overline{W} est strictement convexe sur \mathbb{R}^d . Elle admet donc bien un unique minimum. On pose alors

$$\Omega = \operatorname{argmin}_{\omega \in \mathbb{R}^d} \overline{W}(\omega). \quad (6.5)$$

De plus comme

$$\sup_{u \in \mathcal{U}(\varphi_0)} \sup_{\mathbf{v} \in \mathcal{V}(u, \varphi_0)} \sup_{\theta_0 \in \Theta(\varphi_0)} \sup_{(s, \mathbf{x}) \in \mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} |\zeta\{\mathcal{S}(s, \mathcal{R}_{\theta_0 \mathbf{x}})\}| = \mathcal{O}_{\mathbb{P}}(1),$$

alors

$$\Omega = \mathcal{O}_{\mathbb{P}}(1) \quad (6.6)$$

Chapitre 7

Étude asymptotique de $\widehat{I}_{\mathcal{S}}$

Dans toute cette partie et dans un souci de clareté, nous noterons \mathcal{S} le cylindres paramétrés par (θ, \mathbf{x}, s) et notés auparavant $\mathcal{S}(s, \mathcal{R}_{\theta\mathbf{x}})$. De même on omettra la dépendance en g pour $I_{\mathcal{S}}$ quand il n'y a pas de confusion possible.

Dans cette partie on s'intéresse aux processus $\widehat{I}_{\mathcal{S}}$ et $I_{\mathcal{S}}$ indexés par les cylindres, chacun de ces termes étant défini par

$$\widehat{I}_{\mathcal{S}}(g) = n^{-1} \|\mathcal{S}\|_d^{-1} \sum_{i=1}^n Y_i \mathbb{I}_{(X_i \in \mathcal{S})} [\widehat{f}^{(i)}(X_i)]^{-1}, \quad I_{\mathcal{S}}(g) = \|\mathcal{S}\|_d^{-1} \left\{ \int_{\mathcal{S}} g(\mathbf{x}) d\mathbf{x} \right\}, \quad (7.1)$$

$$\text{avec } \widehat{f}^{(i)}(X_i) = \frac{1}{nb^d} \sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{X_j - X_i}{b}\right) \text{ et } \|\mathcal{S}\|_d = \int_{\mathcal{S}} d\mathbf{x}.$$

On note \mathcal{C} l'ensemble des cylindres recouvrant \mathcal{Q}_{δ} , strictement inclus dans \mathcal{Q} ; \mathcal{S} varie donc dans \mathcal{C} . Nous voulons obtenir la convergence faible de $(\widehat{I}_{\mathcal{S}} - I_{\mathcal{S}})$ à la vitesse \sqrt{n} et l'ordre uniforme de cette quantité. Pour cela on décompose $\|\mathcal{S}\|_d \widehat{I}_{\mathcal{S}}$ en deux processus $A_{\mathcal{S}}$ et $B_{\mathcal{S}}(g)$, définis par

$$A_{\mathcal{S}} = n^{-1} \sum_{i=1}^n \sigma(X_i) \mathbb{I}_{(X_i \in \mathcal{S})} [\widehat{f}^{(i)}(X_i)]^{-1} \varepsilon_i, \quad B_{\mathcal{S}}(g) = n^{-1} \sum_{i=1}^n g(X_i) \mathbb{I}_{(X_i \in \mathcal{S})} [\widehat{f}^{(i)}(X_i)]^{-1},$$

et $\|\mathcal{S}\|_d \widehat{I}_{\mathcal{S}}(g) = A_{\mathcal{S}} + B_{\mathcal{S}}(g)$. De plus, on notera

$$\zeta_n(\mathcal{S}) = n^{1/2} \|\mathcal{S}\|_d (\widehat{I}_{\mathcal{S}}(g) - I_{\mathcal{S}}(g)).$$

Dans un premier temps, nous allons étudier les processus $A_{\mathcal{S}}$ et $B_{\mathcal{S}}(g)$, puis nous en déduirons la convergence du processus ζ_n .

7.1 Étude du processus A_S

Le but de cette partie est de montrer la convergence faible de ce processus et la majoration uniforme suivante :

$$\sup_{\mathcal{C}} |A_S| = \mathcal{O}_{\mathbb{P}}(n^{-1/2}). \quad (7.2)$$

Pour étudier A_S , nous allons décomposer ce processus en quatre parties provenant du développement suivant :

$$\frac{1}{\widehat{f}^{(i)}(X_i)} = \frac{1}{f(X_i)} - \frac{\widehat{f}^{(i)}(X_i) - f(X_i)}{f^2(X_i)} + \mathcal{O}_{\mathbb{P}}\left(\frac{(\widehat{f}^{(i)}(X_i) - f(X_i))^2}{f^3(X_i)}\right). \quad (7.3)$$

Ainsi on peut décomposer A_S de la façon suivante :

$$A_S = A_{S,1} - A_{S,2} - A_{S,3} + \mathcal{O}_{\mathbb{P}}(A_{S,4}),$$

avec

$$A_{S,1} = n^{-1} \sum_{i=1}^n \frac{\sigma(X_i)}{f(X_i)} \mathbb{I}_{(X_i \in \mathcal{S})} \varepsilon_i, \quad (7.4)$$

$$A_{S,2} = \frac{1}{n(n-1)b^d} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \left\{ \gamma_S^A(X_i) K\left(\frac{X_j - X_i}{b}\right) \varepsilon_i \right. \\ \left. - \mathbb{E} \left[\gamma_S^A(X_i) K\left(\frac{X_j - X_i}{b}\right) \varepsilon_i \middle| X_i, \varepsilon_i \right] \right\}, \quad (7.5)$$

$$A_{S,3} = \frac{1}{n(n-1)b^d} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathbb{E} \left[\gamma_S^A(X_i) K\left(\frac{X_j - X_i}{b}\right) \varepsilon_i \middle| X_i, \varepsilon_i \right] \\ - \frac{1}{n} \sum_{i=1}^n \gamma_S^A(X_i) f(X_i) \varepsilon_i, \quad (7.6)$$

$$A_{S,4} = n^{-1} \sum_{i=1}^n (\widehat{f}^{(i)}(X_i) - f(X_i))^2 |\varepsilon_i|, \quad (7.7)$$

où $\gamma_S^A(X_i) = \frac{\sigma(X_i)}{f^2(X_i)} \mathbb{I}_{(X_i \in \mathcal{S})}$.

On montre que le premier terme est de l'ordre de $n^{-1/2}$ et que les suivants sont négligeables devant celui-ci.

7.1.1 Convergence et majoration uniforme de $A_{\mathcal{S},1}$

On remarque tout d'abord que pour \mathcal{S} fixé, $A_{\mathcal{S},1}$ est une somme de variables aléatoires indépendantes que nous pouvons écrire

$$A_{\mathcal{S},1} = n^{-1} \sum_{i=1}^n \frac{\sigma(X_i)}{f(X_i)} \varepsilon_i \mathbb{I}_{\mathcal{S} \times \mathbb{R}}(X_i, \varepsilon_i) = n^{-1} \sum_{i=1}^n h_{\mathcal{S}}(X_i, \varepsilon_i),$$

où $h_{\mathcal{S}}(\mathbf{x}, y) = \frac{\sigma(\mathbf{x})}{f(\mathbf{x})} y \mathbb{I}_{\mathcal{S} \times \mathbb{R}}(\mathbf{x}, y)$. Le terme $A_{\mathcal{S},1}$ apparaît alors comme un processus empirique sur la classe de fonctions

$$\mathcal{H}_1 = \left\{ h_{\mathcal{S}}; h_{\mathcal{S}}(\mathbf{x}, y) = \frac{\sigma(\mathbf{x})}{f(\mathbf{x})} y \mathbb{I}_{\mathcal{S} \times \mathbb{R}}(\mathbf{x}, y), \mathcal{S} \in \mathcal{C} \right\}.$$

Si on montre que cette classe de fonctions est euclidienne, alors en utilisant des inégalités maximales de *Sherman (1994)* et un théorème central limite pour les processus empiriques, on obtiendra les résultats souhaités.

\mathcal{H}_1 classe de fonctions euclidienne :

Lemme 7.1.1. *Soit \mathcal{H}_1 la classe de fonctions définie ci-dessus, alors \mathcal{H}_1 est une classe de fonctions euclidienne.*

DÉMONSTRATION.

• \mathcal{C} est une classe d'ensembles de *Vapnik - Chervonenkis*. On dira une VC-classe. Pour obtenir ce résultat on introduit l'ensemble des ellipses sur \mathbb{R}^d :

$$\mathcal{E} = \left\{ (\mathbf{x} - \mathbf{a})^T \Sigma^{-1} (\mathbf{x} - \mathbf{a}) \leq 1, \Sigma \text{ matrice symétrique définitive positive, } \mathbf{a} \in \mathbb{R}^d \right\}$$

\mathcal{E} est une VC-classe, car en posant Si on pose

$$g_{\Sigma, \mathbf{a}}(\mathbf{x}) = -(\mathbf{x} - \mathbf{a})^T \Sigma^{-1} (\mathbf{x} - \mathbf{a}) + 1 = -\mathbf{x}^T \Sigma^{-1} \mathbf{x} + 2\mathbf{a}^T \Sigma^{-1} \mathbf{x} - \mathbf{a}^T \Sigma^{-1} \mathbf{a} + 1,$$

alors $\mathcal{G} = \{g_{\Sigma, \mathbf{a}}(\mathbf{x}), \Sigma \text{ matrice symétrique définitive positive, } \mathbf{a} \in \mathbb{R}^d\}$ est l'espace des polynômes à d indéterminées de degré inférieur ou égal à deux et à coefficients donnés par une matrice symétrique définie positive et par un vecteur de \mathbb{R}^d . La dimension de \mathcal{G} est donc $d(d+1)/2 + d$.

Or on peut réécrire \mathcal{E} de la façon suivante

$$\mathcal{E} = \left\{ \mathbf{x} : g_{\Sigma, \mathbf{a}}(\mathbf{x}) \geq 0 \right\}, \text{ pour } g_{\Sigma, \mathbf{a}}(\mathbf{x}) \in \mathcal{G}.$$

On sait alors, cf. *Devroye et Lugosi (2000)*, que \mathcal{E} est une VC-classe de VC-dimension inférieure ou égale à $\frac{d(d+3)}{2}$. Utilisons cette famille pour montrer que \mathcal{C} est aussi une VC-classe.

Chaque cylindre \mathcal{S} peut être obtenu à partir d'un cylindre particulier \mathcal{S}_0 : soit \mathcal{S}_0 le cylindre de centre O , d'axe Ox_1 , de rayon r et de longueur 1. Si on note

- D_s la dilatation d'axe Ox_1 et de rapport $s \in [s_1, s_2]$;
 - R_θ la rotation de centre O , qui transforme l'axe Ox_1 en θ ;
 - $t_{\mathbf{x}}$ la translation, qui envoie O au centre \mathbf{x} de \mathcal{S} ;
- et on pose $T_{r,\theta,s} = t_{\mathbf{x}} \circ R_\theta \circ D_s$, alors $T_{r,\theta,s}(\mathcal{S}_0) = \mathcal{S}$.

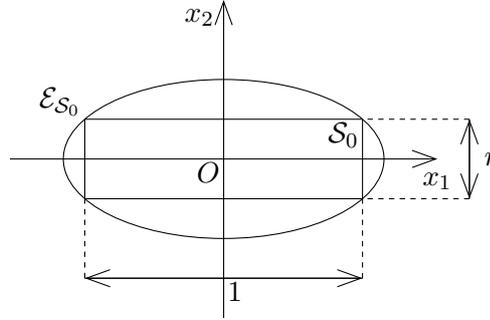


FIG. 7.1 – Cylindre \mathcal{S}_0 en dimensions 2 et son ellipse $\mathcal{E}_{\mathcal{S}_0}$.

Pour ce cylindre \mathcal{S}_0 , on peut considérer une ellipse $\mathcal{E}_{\mathcal{S}_0}$, qui lui est circonscrite et donnée par la fonction $g(\mathbf{x}) = 1 - \mathbf{x}^T \Sigma^{-1} \mathbf{x}$, avec $\Sigma = \text{diag}(1/2, 2r^2, \dots, 2r^2)$.

La transformation $T_{r,\theta,s}$ envoie donc $\mathcal{E}_{\mathcal{S}_0}$ sur $\mathcal{E}_{\mathcal{S}}$, ellipse circonscrite à \mathcal{S} .

Ceci nous permet de définir l'application $\Psi : \mathcal{C} \rightarrow \mathcal{E}$, telle que $\Psi(\mathcal{S}) = \mathcal{E}_{\mathcal{S}}$. L'image $\mathcal{E}_{\mathcal{C}}$ de cette application est un sous ensemble \mathcal{E} qui est une VC-classe et donc $\mathcal{E}_{\mathcal{C}}$ en est une. Or on peut écrire $\mathcal{C} = \Psi^{-1}(\mathcal{E}_{\mathcal{C}})$, donc \mathcal{C} est une VC-classe.

- $\mathcal{C} \times \mathbb{R}$ est une VC-classe. En considérant la projection p de \mathbb{R}^{d+1} dans \mathbb{R}^d , orthogonale à x_{d+1} , alors

$$\mathcal{C} \times \mathbb{R} = p^{-1}(\mathcal{C}),$$

et par les propriétés des VC-classes, cf. Pollard (1984), on peut en déduire que $\mathcal{C} \times \mathbb{R}$ est une VC-classe.

- \mathcal{H}_1 est une classe de fonctions euclidienne. Rappelons que

$$\mathcal{H}_1 = \left\{ h_{\mathcal{S}} / h_{\mathcal{S}}(\mathbf{x}, y) = \frac{\sigma(\mathbf{x})}{f(\mathbf{x})} y \mathbb{I}_{\mathcal{S} \times \mathbb{R}}(\mathbf{x}, y), \mathcal{S} \in \mathcal{C} \right\}$$

alors en utilisant des propriétés des VC-classes, cf. Annexe 10.2, on en déduit que \mathcal{H}_1 est une classe euclidienne. ■

Majoration uniforme de $A_{\mathcal{S},1}$:

Pour cela, on utilise les résultats de Sherman (1994) sur la majoration uniforme des U -processus. On remarque que \mathcal{H}_1 est une classe de fonctions euclidienne :

- elle est dégénérée, car pour tout $h_{\mathcal{S}} \in \mathcal{H}_1$, $\mathbb{E} h_{\mathcal{S}}(X, \varepsilon) = 0$ par indépendance des variables et par centrage de ε ;

– elle admet une enveloppe de carré intégrable, car la fonction σ étant bornée par M_σ

$$\sup_{h_S \in \mathcal{H}_1} |h_S(\mathbf{x}, y)| \leq \frac{M_\sigma |y|}{f(\mathbf{x})} \mathbb{I}_{\mathcal{Q} \times \mathbb{R}}(\mathbf{x}, y) = H(\mathbf{x}, y)$$

et

$$\mathbb{E} [H(X, \varepsilon)^2] \leq M_\sigma \left(\int_{\mathcal{Q}_\delta} \frac{d\mathbf{x}}{f(\mathbf{x})} \right) \mathbb{E} [\varepsilon^2] < \infty.$$

On a alors en utilisant le corollaire 4(ii) de *Sherman (1994)*

$$\sup_{\mathcal{C}} |A_{S,1}| = \mathcal{O}_{\mathbb{P}}(n^{-1/2}). \quad (7.8)$$

Convergence faible de $A_{S,1}$:

Propriété 7.1.1. *Sous les hypothèses du chapitre 6*

$$n^{1/2} A_{,1} \xrightarrow{w} \zeta(\cdot),$$

où $\zeta(\cdot)$ est un processus gaussien indexé par les cylindres de \mathcal{C} d'espérance nulle et de covariance

$$\text{cov}(\zeta(\mathcal{S}_1), \zeta(\mathcal{S}_2)) = \int_{\mathcal{S}_1 \cap \mathcal{S}_2} \frac{\sigma^2(\mathbf{x})}{f(\mathbf{x})} d\mathbf{x}.$$

DÉMONSTRATION. En utilisant les théorèmes classiques de convergence dans le cas des processus empiriques pour des classes de fonctions euclidiennes d'enveloppe deux fois différentiable, cf. *Van der Vaart et Wellner (1996)*, p141, on a

$$n^{1/2}(A_{,1} - \mathbb{E}(h.(X, \varepsilon))) \xrightarrow{w} G(\cdot),$$

où G est donné par la convergence des marginales et ici $\mathbb{E}(h.(X, \varepsilon)) = 0$. Pour k fixé et un k -uplet $(\mathcal{S}_1, \dots, \mathcal{S}_k)$ de cylindres, un théorème central limite multidimensionnel fournit

$$n^{1/2}(A_{\mathcal{S}_1,1}, \dots, A_{\mathcal{S}_k,1}) \xrightarrow{\mathcal{L}} \mathcal{N}_k(0, \Sigma),$$

où les covariances sont données pour $i, j \in \{1, \dots, k\}$:

$$\begin{aligned} \Sigma_{i,j} &= \mathbb{E} [\sigma(X)^2 / f(X)^2 \mathbb{I}_{\mathcal{S}_i \cap \mathcal{S}_j}(X) \varepsilon^2] \\ &= \int_{\mathcal{S}_i \cap \mathcal{S}_j} \frac{\sigma^2(\mathbf{x})}{f(\mathbf{x})} d\mathbf{x}, \end{aligned}$$

ce qui achève la démonstration. ■

7.1.2 Étude du U -processus $A_{\mathcal{S},2}$

Propriété 7.1.2. *Sous les hypothèses du chapitre 6,*

$$\sup_{\mathcal{C}} |A_{\mathcal{S},2}| = \mathcal{O}_{\mathbb{P}}(n^{-1/2}). \quad (7.9)$$

DÉMONSTRATION. On rappelle que

$$\begin{aligned} b^d A_{\mathcal{S},2} &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \left\{ \gamma_{\mathcal{S}}^A(X_i) K\left(\frac{X_j - X_i}{b}\right) \varepsilon_i \right. \\ &\quad \left. - \mathbb{E} \left[\gamma_{\mathcal{S}}^A(X_i) K\left(\frac{X_j - X_i}{b}\right) \varepsilon_i \middle| X_i, \varepsilon_i \right] \right\}, \\ &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \left\{ h_{\mathcal{S},b}^A(Z_i, Z_j) - \mathbb{E} [h_{\mathcal{S},b}^A(Z_i, Z_j) | Z_i] \right\}, \end{aligned}$$

où $Z_i = (X_i, \varepsilon_i)$ et pour $\mathbf{z}_i = (\mathbf{x}_i, \varepsilon_i)$ on a $h_{\mathcal{S},b}^A(\mathbf{z}_1, \mathbf{z}_2) = \gamma_{\mathcal{S}}^A(\mathbf{x}_1) \varepsilon_1 K\left(\frac{\mathbf{x}_2 - \mathbf{x}_1}{b}\right)$. On définit alors

$$\mathcal{H}_2 = \left\{ \tilde{h}_{\mathcal{S},b}^A / \tilde{h}_{\mathcal{S},b}^A(\cdot, \cdot) = h_{\mathcal{S},b}^A(\cdot, \cdot) - \mathbb{E} [h_{\mathcal{S},b}^A(\cdot, Z)] \right\}, \quad \mathcal{S} \in \mathcal{C}, b \in]0, 1].$$

Nous sommes donc en présence d'un U -processus d'ordre 2, dégénéré, défini pour la famille de fonctions \mathcal{H}_2 . En utilisant les lemmes de *Nolan et Pollard (1987)* et de *Pakes et Pollard (1989)*, on peut montrer que cette famille de fonctions est une famille euclidienne (cf. *Annexe 10.2*). Le corollaire 4(ii) de *Sherman (1994)* nous fournit alors

$$\sup_{\mathcal{S},b} |b^d A_{\mathcal{S},2}| = \mathcal{O}_{\mathbb{P}}(n^{-1})$$

c'est-à-dire étant donné que $b = n^{-\kappa}$ et $d\kappa < 1/2$

$$\sup_{\mathcal{C}} |A_{\mathcal{S},2}| = \mathcal{O}_{\mathbb{P}}(n^{-1}b^{-d}) = \mathcal{O}_{\mathbb{P}}(n^{-1/2}).$$

■

7.1.3 Vitesse de convergence de $A_{\mathcal{S},3}$

Propriété 7.1.3. *Sous les hypothèses du chapitre 6*

$$\sup_{\mathcal{C}} |A_{\mathcal{S},3}| = \mathcal{O}_{\mathbb{P}}(n^{-1/2}). \quad (7.10)$$

La démarche utilisée ici est différente, elle fait intervenir un réseau de cylindre. Elle sera utilisée à plusieurs reprises et se décompose de la façon suivante :

– il faut d’une part obtenir l’ordre suivant

$$\sup_{\mathcal{C}} \mathbb{P} \left\{ |A_{\mathcal{S},3}| > n^{-(1/2)-\eta_0} \right\} = \mathcal{O}(n^{-C_1}),$$

ceci fera l’objet du lemme 7.1.2;

– on utilise ensuite une technique mise en œuvre dans *Stone (1984)* : on introduit un réseau de cylindres de cardinal fini, d’ordre polynomiale en n , qui recouvre “assez bien” l’ensemble tout entier, selon un critère qui sera défini plus bas. Ce choix fait l’objet de lemme 7.1.3 et pour celui-ci on obtient

$$\sup_{\mathcal{C}} |A_{\mathcal{S},3} - A_{\mathcal{S}',3}| = \mathcal{O}_{\mathbb{P}}(n^{-1/2}),$$

où \mathcal{S}' est l’élément du réseau associé à \mathcal{S} , cf. *lemme 7.1.5*

En combinant ces deux résultats, on obtient l’ordre de $A_{\mathcal{S},3}$ uniformément sur \mathcal{C} .

Lemme 7.1.2. *Pour tout $C_1 > 0$, il existe $\eta_0 > 0$ indépendant de C_1 .*

$$\sup_{\mathcal{C}} \mathbb{P} \left\{ |A_{\mathcal{S},3}| > n^{-(1/2)-\eta_0} \right\} = \mathcal{O}(n^{-C_1}). \quad (7.11)$$

DÉMONSTRATION. On rappelle que

$$A_{\mathcal{S},3} = n^{-1} \sum_{i=1}^n \gamma_{\mathcal{S}}^A(X_i) \{ \mu_b(X_i) - f(X_i) \} \varepsilon_i := n^{-1} \sum_{i=1}^n Z_i \varepsilon_i,$$

où $\mu_b(X_i) = \mathbb{E} \left[b^{-d} K \left(\frac{X - X_i}{b} \right) | X_i \right]$. On peut alors démontrer en utilisant l’ordre du noyau et la régularité de la densité f , que

$$\mu_b(\mathbf{x}) = f(\mathbf{x}) + \mathcal{O}(b^\tau), \quad \text{uniformément en } \mathbf{x}. \quad (7.12)$$

L’outil principal est l’inégalité sur les martingales de Rosenthal, cf. *Hall et Heyde (1980)*, p.23. En effet $n A_{\mathcal{S},3}$, en tant que somme de variables aléatoires indépendantes centrées, est une martingale par rapport à la filtration $\mathcal{F}_n = \{(X_i, \varepsilon_i), i = 1, \dots, n\}$. Soit k un entier strictement positif, il existe alors une constante C qui dépend uniquement de k telle que

$$\begin{aligned} \mathbb{E} |n A_{\mathcal{S},3}|^k &\leq C_2 \left(\mathbb{E} \left[\sum_{i=1}^n \mathbb{E} (Z_i^2 \varepsilon_i^2 | \mathcal{F}_{i-1}) \right]^{k/2} + \sum_{i=1}^n \mathbb{E} (|Z_i|^k |\varepsilon_i|^k) \right), \\ &\leq C_2 \left(\mathbb{E} \left[\sum_{i=1}^n \mathbb{E} (Z_i^2) \right]^{k/2} + \sum_{i=1}^n \mathbb{E} (|Z_i|^k) M_k \right), \\ &\leq C_2 \left([n \mathbb{E} (Z_i^2)]^{k/2} + n M_k \mathbb{E} (|Z_i|^k) \right). \end{aligned}$$

Or par définition de Z_i et en utilisant le développement (7.12), on peut écrire pour tout $l \in \mathbb{N}$

$$\begin{aligned} \mathbb{E} |Z_i|^l &= \int_{\mathcal{S}} \frac{\sigma(\mathbf{x})^l}{f(\mathbf{x})^{2l}} (\mu_b(\mathbf{x}) - f(\mathbf{x}))^l f(\mathbf{x}) d\mathbf{x}, \\ &\leq \mathcal{O}(b^{l\tau}), \end{aligned}$$

indépendamment de \mathcal{S} . Finalement

$$\mathbb{E} |n A_{\mathcal{S},3}|^k \leq \mathcal{O}(b^\tau n^{1/2})^k + \mathcal{O}(b^\tau n^{1/k})^k.$$

Alors en utilisant l'inégalité de Markov

$$\begin{aligned} \mathbb{P}(|A_{\mathcal{S},3}| > n^{-1/2-\eta_0}) &\leq \frac{n^{-2k} \mathbb{E} |n A_{\mathcal{S},3}|^{2k}}{(n^{-1/2-\eta_0})^{2k}} = \mathcal{O}(b^\tau n^{-1/2+1/2+\eta_0})^{2k}, \\ &= \mathcal{O}(n^{-2k\varepsilon}), \end{aligned}$$

l'existence de $\varepsilon > 0$ est assurée par le choix d'un η_0 tel que $0 < \eta_0 < \tau\kappa$. Finalement

$$\sup_{\mathcal{C}} \mathbb{P}\{|A_{\mathcal{S},3}| > n^{-(1/2)-\eta_0}\} = \mathcal{O}(n^{-C_1}), \quad \forall C_1 > 0 \text{ et } \eta_0 \text{ indépendant de } C_1 \quad (7.13)$$

■

Lemme 7.1.3. *Il existe une sous famille \mathcal{C}' de la famille des cylindres \mathcal{C} , de cardinal d'ordre polynomial, i.e. $|\mathcal{C}'| = \mathcal{O}(n^{C_2})$ avec C_2 assez grand pour que*

$$\sup_{\mathcal{C}} \sum_{i=1}^n \mathbb{I}_{X_i \in \mathcal{S} \Delta \mathcal{S}'} = \mathcal{O}_{\mathbb{P}}(1) \quad (7.14)$$

DÉMONSTRATION. On introduit tout d'abord quelques notations : soit \mathcal{T} un ensemble de \mathbb{R}^d , on note $d(\mathbf{x}, \mathcal{T}) = \inf_{\mathbf{y} \in \mathcal{T}} |\mathbf{x} - \mathbf{y}|$, la distance du point $\mathbf{x} \in \mathbb{R}^d$ à l'ensemble \mathcal{T} et $\partial\mathcal{T}$ la frontière de l'ensemble \mathcal{T} . On définit un voisinage autour de cette frontière de la façon suivante $\mathcal{T}^\dagger = \{\mathbf{x} \in \mathbb{R}^d / d(\mathbf{x}, \partial\mathcal{T}) \leq n^{-2}\}$.

On choisit \mathcal{C}' assez fine telle que, pour tout $\mathcal{S} \in \mathcal{C}$, il existe un cylindre $\mathcal{S}' \in \mathcal{C}'$ qui vérifie

$$\mathcal{S} \Delta \mathcal{S}' \subset (\mathcal{S}')^\dagger, \quad (\text{cf. figure 7.2}).$$

Ainsi on choisit C_2 assez grand pour avoir $|\mathcal{C}'| = \mathcal{O}(n^{C_2})$ et la propriété ci-dessus. Dans ce cas pour tout $(\mathcal{S}, \mathcal{S}')$ ainsi associés, on a

$$\sum_{i=1}^n \mathbb{I}_{X_i \in \mathcal{S} \Delta \mathcal{S}'} \leq \sum_{i=1}^n \mathbb{I}_{X_i \in (\mathcal{S}')^\dagger},$$

donc

$$\sup_{\mathcal{C}} \sum_{i=1}^n \mathbb{I}_{X_i \in \mathcal{S} \Delta \mathcal{S}'} \leq \sup_{\mathcal{C}'} \sum_{i=1}^n \mathbb{I}_{X_i \in (\mathcal{S}')^\dagger}.$$

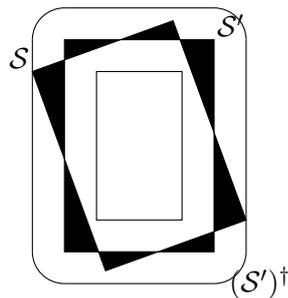


FIG. 7.2 – Construction du réseau dans \mathbb{R}^2 (l'aire grisée représente $S\Delta S'$).

Montrons que $\sup_{C'} \sum_{i=1}^n \mathbb{I}_{X_i \in (S')^\dagger} = \mathcal{O}_{\mathbb{P}}(1)$.

Soit $N(S') = \sum_{i=1}^n \mathbb{I}_{X_i \in (S')^\dagger}$, c'est-à-dire le nombre de X_i dans $(S')^\dagger$. Par définition, $N(S')$ est une variable aléatoire de loi $\mathcal{B}(n, q(S'))$, où $q(S') = \mathbb{P}(X \in (S')^\dagger) = \int_{(S')^\dagger} f(\mathbf{x}) d\mathbf{x}$. Or on sait que le volume de $(S')^\dagger$ est de l'ordre n^{-2} :

$$\text{Vol}((S')^\dagger) \leq \text{Vol}(\mathcal{S}(s + n^{-2}, r + n^{-2})) - \text{Vol}(\mathcal{S}(s - n^{-2}, r - n^{-2})) \leq \mathcal{O}(n^{-2}),$$

donc $q(S') = \mathcal{O}(n^{-2})$, uniformément en S' , puisque $s \in [s_1, s_2]$. Soit k un entier fixé dans $\{1, \dots, n\}$,

$$\begin{aligned} \mathbb{P}(N(S') \geq k) &= \sum_{p=k}^n C_n^p q(S')^p [1 - q(S')]^{n-p}, \\ &= \mathcal{O}(n^{-2k} C_n^k) = \mathcal{O}(n^{-k}), \quad \text{uniformément sur les } S'. \end{aligned}$$

Si $k > C_2$, ce qui est possible car n tend vers l'infini et C_2 est fixé, alors

$$\mathbb{P}(\sup_{C'} N(S') \geq k) \leq \mathcal{O}(n^{C_2 - k}) = \mathcal{o}(1)$$

Conclusion. Il existe k tel que

$$\mathbb{P}\left(\sup_C \sum_{i=1}^n \mathbb{I}_{X_i \in S\Delta S'} \geq k\right) \leq \mathbb{P}\left(\sup_{C'} \sum_{i=1}^n \mathbb{I}_{X_i \in (S')^\dagger} \geq k\right) = \mathcal{o}(1),$$

ce qui achève la preuve. ■

Nous pouvons alors établir les deux résultats qui serviront directement à l'obtention de (7.10).

Lemme 7.1.4. *Soit \mathcal{C}' la sous-famille de \mathcal{C} , définie dans le lemme 7.1.3. Alors il existe $C > 0$*

$$\mathbb{P}\left\{\sup_{\mathcal{C}'} |A_{\mathcal{S},3}| > n^{-(1/2)-\eta}\right\} = \mathcal{O}(n^{-C}). \quad (7.15)$$

DÉMONSTRATION. On rappelle que \mathcal{C}' est un sous-ensemble de \mathcal{C} , dont le cardinal est un $\mathcal{O}(n^{C_2})$. En utilisant (7.11),

$$\begin{aligned} \mathbb{P}\left\{\sup_{\mathcal{C}'} |A_{\mathcal{S},3}| > n^{-(1/2)-\eta_0}\right\} &\leq \sum_{\mathcal{S}' \in \mathcal{C}'} \mathbb{P}\left\{|A_{\mathcal{S}',3}| > n^{-(1/2)-\eta_0}\right\}, \\ &\leq \mathcal{O}(n^{C_2} n^{-C_1}), \end{aligned}$$

ceci étant vrai pour tout $C_1 > 0$, on obtient bien (7.15). ■

Lemme 7.1.5. *Soit \mathcal{C}' défini dans le lemme 7.1.3, alors*

$$\sup_{\mathcal{C}} |A_{\mathcal{S},3} - A_{\mathcal{S}',3}| = \mathcal{O}_{\mathbb{P}}(n^{-1/2}). \quad (7.16)$$

où \mathcal{S}' est l'élément de \mathcal{C}' associé à \mathcal{S} , défini par le lemme 7.1.3.

DÉMONSTRATION. Soit $\mathcal{S} \in \mathcal{C}$ alors il existe $\mathcal{S}' \in \mathcal{C}'$ tel que $\mathcal{S} \Delta \mathcal{S}' \subset (\mathcal{S})^\dagger$, alors

$$\begin{aligned} |A_{\mathcal{S},3} - A_{\mathcal{S}',3}| &\leq n^{-1} \sum_{i=1}^n \left| \frac{\sigma(X_i)}{f^2(X_i)} \right| |\mu_b(X_i) - f(X_i)| |\mathbb{I}_{X_i \in \mathcal{S}} - \mathbb{I}_{X_i \in \mathcal{S}'}| |\varepsilon_i|, \\ &\leq n^{-1} \sup_{\mathbf{x}} \left| \frac{\sigma(\mathbf{x})}{f^2(\mathbf{x})} \right| \sup_i |\mu_b(X_i) - f(X_i)| \sup_i |\varepsilon_i| \sum_{i=1}^n \mathbb{I}_{X_i \in \mathcal{S} \Delta \mathcal{S}'}. \end{aligned}$$

Évaluons l'ordre de chacun de ces termes :

- σ étant bornée, $\sup_{\mathbf{x}} |\sigma(\mathbf{x})| < \infty$, et $\sup_{\mathbf{x}} f(\mathbf{x}) \geq m > 0$, le premier terme est borné.
- Les ε_i suivant une loi $\mathcal{N}(0, 1)$, on peut montrer :

$$\sup_i |\varepsilon_i| = \mathcal{O}_{\mathbb{P}}(\log(n)^{1/2}).$$

En effet dans ce cas, on a

$$\mathbb{P}(\sup_i |\varepsilon_i| > A_n) \leq n \mathbb{P}(|\varepsilon_i| > A_n) \leq n \exp\left(-\frac{A_n^2}{2}\right).$$

Si on pose $A_n = M(\log(n))^{1/2}$, alors

$$\mathbb{P}(\sup_i |\varepsilon_i| > A_n) \leq n \exp\left(-\frac{M^2}{2} \log n\right) \leq n^{1-M^2/2}$$

On choisit alors M tel que $n^{1-M^2/2} \rightarrow 0$. D'où

$$\sup_i |\varepsilon_i| = \mathcal{O}_{\mathbb{P}}(\log(n)^{1/2})$$

– D'après le développement (7.12) de $\mu_b(\cdot)$

$$\sup_i |\mu_b(X_i) - f(X_i)| = \mathcal{O}_{\mathbb{P}}(n^{-\tau\kappa}).$$

Ainsi pour la quantité qui nous intéresse, on peut écrire

$$\sup_{\mathcal{C}} |A_{\mathcal{S},3} - A_{\mathcal{S}',3}| = \mathcal{O}_{\mathbb{P}}(n^{-1-\tau\kappa} \log(n)^{1/2}) \sup_{\mathcal{C}} \sum_{i=1}^n \mathbb{I}_{X_i \in \mathcal{S} \Delta \mathcal{S}'}$$

Finalement en utilisant le lemme 7.1.3, on a

$$\sup_{\mathcal{C}} |A_{\mathcal{S},3} - A_{\mathcal{S}',3}| = \mathcal{O}_{\mathbb{P}}(n^{-1-\tau\kappa} \log(n)^{1/2}) = \mathcal{O}_{\mathbb{P}}(n^{-1/2}).$$

■

DÉMONSTRATION DE LA PROPOSITION 7.1.3 On veut montrer pour $\varepsilon > 0$ fixé,

$$\mathbb{P}(n^{1/2} \sup_{\mathcal{C}} |A_{\mathcal{S},3}| > \varepsilon) \rightarrow 0.$$

Décomposons $|A_{\mathcal{S},3}|$ en introduisant le cylindre \mathcal{S}' associé à \mathcal{S}

$$\begin{aligned} |A_{\mathcal{S},3}| &= |A_{\mathcal{S},3} - A_{\mathcal{S}',3}| + |A_{\mathcal{S}',3}|, \\ &\leq |A_{\mathcal{S},3} - A_{\mathcal{S}',3}| + \sup_{\mathcal{C}'} |A_{\mathcal{S}',3}|, \\ &\leq \sup_{\mathcal{C}} |A_{\mathcal{S},3} - A_{\mathcal{S}',3}| + \sup_{\mathcal{C}'} |A_{\mathcal{S}',3}|. \end{aligned}$$

Pour $\varepsilon > 0$, on peut écrire

$$\begin{aligned} \mathbb{P}(n^{1/2} \sup_{\mathcal{C}} |A_{\mathcal{S},3}| > \varepsilon) &\leq \mathbb{P}(n^{1/2} \sup_{\mathcal{C}} |A_{\mathcal{S},3} - A_{\mathcal{S}',3}| + n^{1/2} \sup_{\mathcal{C}'} |A_{\mathcal{S}',3}| > \varepsilon), \\ &\leq \mathbb{P}(n^{1/2} \sup_{\mathcal{C}} |A_{\mathcal{S},3} - A_{\mathcal{S}',3}| > \varepsilon/2) + \mathbb{P}(n^{1/2} \sup_{\mathcal{C}'} |A_{\mathcal{S}',3}| > \varepsilon/2). \end{aligned}$$

Soit $\varepsilon' > 0$ fixé, le lemme 7.1.5 nous fournit l'existence de n_1 tel que pour $n \geq n_1$, le premier terme de droite est inférieur par ε' . Le lemme 7.1.4 et le fait que $(\exists n_2, n \geq n_2 \Rightarrow n^{-C} < \varepsilon')$ nous donne le résultat pour le deuxième terme. Finalement pour $n \geq \sup(n_1, n_2)$

$$\mathbb{P}(n^{1/2} \sup_{\mathcal{C}'} |A_{\mathcal{S},3}| > \varepsilon) \leq 2\varepsilon',$$

ceci achève la démonstration.

■

7.1.4 Ordre de grandeur de $A_{\mathcal{S},4}$

Propriété 7.1.4. *Sous les hypothèses du chapitre 6*

$$|A_{\mathcal{S},4}| = \mathcal{O}(n^{-1/2}).$$

On remarque que $A_{\mathcal{S},4}$ est en fait indépendant de \mathcal{S} .

DÉMONSTRATION. $A_{\mathcal{S},4}$ est défini par : $A_{\mathcal{S},4} = n^{-1} \sum_{i=1}^n (\widehat{f}^{(i)}(X_i) - f(X_i))^2 |\varepsilon_i|$. Évaluons $\mathbb{P}(|A_{\mathcal{S},4}| > n^{-1/2}\epsilon)$, pour $\epsilon > 0$:

$$\begin{aligned} \mathbb{P}(|A_{\mathcal{S},4}| > n^{-1/2}\epsilon) &\leq \frac{\mathbb{E} |A_{\mathcal{S},4}|}{n^{-1/2}\epsilon} \\ &\leq C n^{1/2} \mathbb{E} (\widehat{f}^{(i)}(X_i) - f(X_i))^2 \\ &\leq n^{1/2} \mathcal{O} \left(\frac{1}{n b^d} + b^{2\tau} \right) = \mathcal{O}(n^{-1/2+d\kappa}). \end{aligned}$$

Or $\kappa d < 1/2$, la dernière inégalité nous fournit donc le résultat. ■

7.2 Étude du processus $B_{\mathcal{S}}(g)$

Pour étudier $B_{\mathcal{S}}(g)$ donné par

$$B_{\mathcal{S}}(g) = n^{-1} \sum_{i=1}^n g(X_i) \mathbb{I}_{(X_i \in \mathcal{S})} [\widehat{f}^{(i)}(X_i)]^{-1},$$

nous allons utiliser le développement de $\widehat{f}^{(i)}(X_i)$ donné en (7.3). Nous faisons alors apparaître des termes plus complexes que dans le cas du processus $A_{\mathcal{S}}$, ceci est dû à l'absence de la variable aléatoire ε_i . Nous obtenons donc

$$B_{\mathcal{S}}(g) - \int_{\mathcal{S}} g(\mathbf{x}) d\mathbf{x} = B_{\mathcal{S},1} - B_{\mathcal{S},2} - B_{\mathcal{S},3} - B_{\mathcal{S},4} + \mathcal{O}(B_{\mathcal{S},5}),$$

avec

$$B_{\mathcal{S},1} = \frac{1}{b^d} \mathbb{E} [h_{\mathcal{S},b}^B(X_1, X_2)] - \int_{\mathcal{S}} g(\mathbf{x}) d\mathbf{x}$$

$$\begin{aligned} B_{\mathcal{S},2} &= \frac{1}{n(n-1)b^d} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \{h_{\mathcal{S},b}^B(X_i, X_j) - \mathbb{E} [h_{\mathcal{S},b}^B(X_i, X_j) | X_i] \\ &\quad - \mathbb{E} [h_{\mathcal{S},b}^B(X_i, X_j) | X_j] + \mathbb{E} [h_{\mathcal{S},b}^B(X_i, X_j)]\} \end{aligned}$$

$$B_{\mathcal{S},3} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \left\{ \frac{1}{b^d} \mathbb{E} [h_{\mathcal{S},b}^B(X_i, X_j) | X_i] - \gamma_{\mathcal{S}}^B(X_i) f(X_i) \right\}$$

$$B_{\mathcal{S},4} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \left\{ \frac{1}{b^d} \mathbb{E} [h_{\mathcal{S},b}^B(X_i, X_j) | X_j] - \gamma_{\mathcal{S}}^B(X_j) f(X_j) \right\}$$

$$B_{S,5} = n^{-1} \sum_{i=1}^n (\widehat{f}^{(i)}(X_i) - f(X_i))^2$$

où $h_{S,b}^B(\mathbf{x}_1, \mathbf{x}_2) = \gamma_S^B(\mathbf{x}_1) K\left(\frac{\mathbf{x}_2 - \mathbf{x}_1}{b}\right)$ et $\gamma_S^B(\mathbf{x}) = \frac{g(\mathbf{x})}{f^2(\mathbf{x})} \mathbb{I}_{(\mathbf{x} \in S)}$.

Ce processus se comporte différemment de A_S , car les termes d'ordre $n^{-1/2}$ s'éliminent. B_S est donc négligeable devant $n^{-1/2}$. Nous allons étudier chacun de ces cinq termes successivement.

7.2.1 Ordre de grandeur de $B_{S,1}$

Propriété 7.2.1. *Sous les hypothèses du chapitre 6*

$$\sup_{\mathcal{C}} |B_{S,1}| = \mathcal{O}_{\mathbb{P}}(n^{-1/2}). \quad (7.17)$$

DÉMONSTRATION.

$$\begin{aligned} B_{S,1} &= \mathbb{E} \left[\frac{g(X_1)}{f(X_1)^2} \mathbb{I}_{(X_1 \in S)} \frac{1}{b^d} K\left(\frac{X_2 - X_1}{b}\right) \right] - \int_S g(\mathbf{x}) \, d\mathbf{x} \\ &= \int_S \frac{g(\mathbf{x})}{f^2(\mathbf{x})} \int \frac{1}{b^d} K\left(\frac{\mathbf{y} - \mathbf{x}}{b}\right) f(\mathbf{y}) f(\mathbf{x}) \, d\mathbf{y} \, d\mathbf{x} - \int_S g(\mathbf{x}) \, d\mathbf{x} \\ &= \int_S \frac{g(\mathbf{x})}{f(\mathbf{x})} \int K(\mathbf{u}) f(\mathbf{x} + b\mathbf{u}) \, d\mathbf{u} \, d\mathbf{x} - \int_S g(\mathbf{x}) \, d\mathbf{x} \\ &= \int_S \frac{g(\mathbf{x})}{f(\mathbf{x})} \int K(\mathbf{u}) (f(\mathbf{x} + b\mathbf{u}) - f(\mathbf{x})) \, d\mathbf{u} \, d\mathbf{x} \\ &= \mathcal{O}(b^\tau), \end{aligned}$$

car K est le produit de noyaux d'ordre τ et que f est suffisamment régulière. De plus cet ordre est uniforme par rapport à \mathcal{S} , ceci est dû à la régularité des fonctions f et g sur le compact \mathcal{Q} . De plus l'hypothèse sur la largeur de fenêtre ($\kappa\tau > 1/2$) fournit le résultat. ■

7.2.2 Étude du U -processus $B_{S,2}$

Propriété 7.2.2. *Sous les hypothèses du chapitre 6*

$$\sup_{\mathcal{C}} |B_{S,2}| = \mathcal{O}_{\mathbb{P}}(n^{-1/2}). \quad (7.18)$$

DÉMONSTRATION. On a construit $b^d B_{S,2}$ de sorte que $\{b^d B_{S,2}, \mathcal{S} \in \mathcal{C}, b \in]0, 1]\}$ soit un

U -processus indexé par (\mathcal{S}, b) , d'ordre 2, dégénéré. En effet

$$\begin{aligned} b^d B_{\mathcal{S},2} &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \{h_{\mathcal{S},b}^B(X_i, X_j) - \mathbb{E}[h_{\mathcal{S},b}^B(X_i, X_j)|X_i] \\ &\quad - \mathbb{E}[h_{\mathcal{S},b}^B(X_i, X_j)|X_j] + \mathbb{E}[h_{\mathcal{S},b}^B(X_i, X_j)]\} \\ &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \widetilde{h}_{\mathcal{S},b}^B(X_i, X_j), \end{aligned}$$

où

$$\widetilde{h}_{\mathcal{S},b}^B(\mathbf{x}_1, \mathbf{x}_2) = h_{\mathcal{S},b}^B(\mathbf{x}_1, \mathbf{x}_2) - \mathbb{E}[h_{\mathcal{S},b}^B(\mathbf{x}_1, X)] - \mathbb{E}[h_{\mathcal{S},b}^B(X, \mathbf{x}_2)] + \mathbb{E}[h_{\mathcal{S},b}^B(X, Y)]$$

(X et Y ont la même loi que les X_i et sont indépendants) et ces fonctions varient dans l'ensemble

$$\mathcal{H}_3 = \left\{ \widetilde{h}_{\mathcal{S},b}^B, \mathcal{S} \in \mathcal{C}, b \in]0, 1] \right\}.$$

En utilisant les lemmes de *Nolan et Pollard (1987)* et de *Pakes et Pollard (1989)*, on peut montrer que cette famille de fonction est une famille euclidienne (cf. *Annexe 10.2*). Le corollaire 4(ii) de *Sherman (1994)* nous fournit alors

$$\sup_{\mathcal{S},b} |b^d B_{\mathcal{S},2}| = \mathcal{O}_{\mathbb{P}}(n^{-1}).$$

Comme $b = n^{-\kappa}$ et $\kappa d < 1/2$,

$$\sup_{\mathcal{C}} |B_{\mathcal{S},2}| = \mathcal{O}_{\mathbb{P}}(n^{-1/2} n^{-1/2+\kappa d}) = \mathcal{O}_{\mathbb{P}}(n^{-1/2}).$$

On obtient le résultat escompté. ■

7.2.3 Vitesse de convergence de $B_{\mathcal{S},3}$

Propriété 7.2.3. *Sous les hypothèses du chapitre 6*

$$\sup_{\mathcal{C}} |B_{\mathcal{S},3}| = \mathcal{O}_{\mathbb{P}}(n^{-1/2}). \quad (7.19)$$

DÉMONSTRATION. On peut réécrire $B_{\mathcal{S},3}$ de la façon suivante

$$\begin{aligned} B_{\mathcal{S},3} &= \frac{1}{n} \sum_{i=1}^n \left\{ \mathbb{E} \left[\frac{1}{b^d} h_{\mathcal{S},b}^B(X_i, X) | X_i \right] - \gamma_{\mathcal{S}}^B(X_i) f(X_i) \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \gamma_{\mathcal{S}}^B(X_i) \left(\mathbb{E} \left[\frac{1}{b^d} K\left(\frac{X_i - X}{b}\right) | X_i \right] - f(X_i) \right) \\ &= \frac{1}{n} \sum_{i=1}^n \gamma_{\mathcal{S}}^B(X_i) (\mu_b(X_i) - f(X_i)). \end{aligned}$$

On traite ce terme de la même façon que $A_{\mathcal{S},3}$, c'est-à-dire en faisant intervenir un réseau de cylindres. Il nous faut tout d'abord établir

$$\exists \eta_0, \forall C > 0, \text{ indépendant de } \eta_0, \quad \sup_{\mathcal{C}} \mathbb{P}(|B_{\mathcal{S},3}| > n^{-1/2-\eta_0}) = \mathcal{O}(n^{-C}) \quad (7.20)$$

Pour cela on montre qu'il existe $\varepsilon > 0$ tel que pour tout $k \in \mathbb{N}$

$$\mathbb{E} (|B_{\mathcal{S},3}|^{2k}) = \mathcal{O}(n^{-1/2-\varepsilon})^{2k}$$

Or pour tout $l \in \mathbb{N}$ et uniformément en \mathcal{S}

$$\mathbb{E} \left[\gamma_{\mathcal{S}}^B(X_i)^l (\mu_b(X_i) - f(X_i))^l \right] = \mathcal{O}(b^{l\tau})$$

car

$$\mu_b(X_i) - f(X_i) = \mathcal{O}(b^\tau)$$

uniformément en X_i (cf. développement (7.12) et la preuve pour $A_{\mathcal{S},3}$). Ainsi

$$\begin{aligned} \mathbb{E} \left[(B_{\mathcal{S},3})^{2k} \right] &= \frac{1}{n^{2k}} \sum_{i_1, \dots, i_{2k}} \mathbb{E} \left[\gamma_{\mathcal{S}}^B(X_{i_1}) (\mu_b(X_{i_1}) - f(X_{i_1})) \times \right. \\ &\quad \left. \dots \times \gamma_{\mathcal{S}}^B(X_{i_{2k}}) (\mu_b(X_{i_{2k}}) - f(X_{i_{2k}})) \right] \\ &= \mathcal{O}(b^{2k\tau}) \end{aligned}$$

Finalement

$$\begin{aligned} \mathbb{P}(|B_{\mathcal{S},3}| > n^{-1/2-\eta_0}) &\leq \frac{\mathbb{E} (|B_{\mathcal{S},3}|^{2k})}{(n^{-1/2-\eta_0})^{2k}} \\ &\leq \mathcal{O}(n^{-\kappa\tau+1/2+\eta_0})^{2k}, \end{aligned}$$

uniformément en $\mathcal{S} \in \mathcal{C}$. Or $\kappa\tau - 1/2 > \tau/2$, donc on peut trouver η_0 tel que (7.20) soit vérifiée.

On peut remarquer que pour démontrer le lemme 7.1.2, on aurait aussi pu utiliser cette technique.

Pour conclure on peut alors utiliser exactement la même démarche que pour $A_{\mathcal{S},3}$, le seul changement se situe dans le lemme 7.1.5. En effet dans notre cas $\sigma(X_i)\varepsilon_i$ est remplacé par $g(X_i)$, mais comme cette quantité est bornée sur \mathcal{Q} le résultat est toujours valable. Ainsi on obtient le résultat souhaité. \blacksquare

7.2.4 Vitesse de convergence $B_{\mathcal{S},4}$

Propriété 7.2.4. *Sous les hypothèses du chapitre 6*

$$\sup_{\mathcal{C}} |B_{\mathcal{S},4}| = \mathcal{O}_{\mathbb{P}}(n^{-1/2}). \quad (7.21)$$

On peut utiliser la même méthode que précédemment, mais ici l'ordre est un peu plus difficile à obtenir car la largeur de fenêtre intervient aussi dans l'indicatrice sur l'ensemble \mathcal{S} . En effet

$$\begin{aligned}
B_{\mathcal{S},4} &= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{b^d} \mathbb{E} [h_{\mathcal{S},b}^B(X, X_i) | X_i] - \gamma_{\mathcal{S}}^B(X_i) f(X_i) \right\} \\
&= \frac{1}{n} \sum_{i=1}^n \left\{ \int_{\mathcal{S}} \frac{g(\mathbf{x})}{f(\mathbf{x})} \frac{1}{b^d} K\left(\frac{X_i - \mathbf{x}}{b}\right) d\mathbf{x} - \gamma_{\mathcal{S}}^B(X_i) f(X_i) \right\} \\
&= \frac{1}{n} \sum_{i=1}^n \left\{ \int \frac{g(\mathbf{x} - b\mathbf{u})}{f(\mathbf{x} - b\mathbf{u})} K(\mathbf{u}) \mathbb{I}_{(X_i - b\mathbf{u})} d\mathbf{u} - \frac{g(X_i)}{f(X_i)} \mathbb{I}_{(X_i \in \mathcal{S})} \right\} \\
&= \frac{1}{n} \sum_{i=1}^n \left\{ \int [\psi_{\mathcal{S}}(X_i - b\mathbf{u}) - \psi_{\mathcal{S}}(X_i)] K(\mathbf{u}) d\mathbf{u} \right\} \\
&:= \frac{1}{n} \sum_{i=1}^n Z_i(\mathcal{S}, b),
\end{aligned}$$

où $\psi_{\mathcal{S}}(\mathbf{x}) = \frac{g(\mathbf{x})}{f(\mathbf{x})} \mathbb{I}_{(\mathbf{x} \in \mathcal{S})}$.

Première partie : On montre qu'il existe $\varepsilon > 0$ tel que pour tout $k \in \mathbb{N}^*$

$$\mathbb{E} \left[(n^{1/2} B_{\mathcal{S},4})^{2k} \right] = \mathcal{O}(n^{-2k\varepsilon}). \quad (7.22)$$

On remarque que

$$\begin{aligned}
\mathbb{E} [Z_i(\mathcal{S}, b)] &= \mathbb{E} \left[\frac{1}{b^d} h_{\mathcal{S},b}^B(X_1, X_2) \right] - \mathbb{E} [\gamma_{\mathcal{S}}^B(X) f(X)] \\
&= B_{\mathcal{S},1} = \mathcal{O}(b^\tau),
\end{aligned}$$

uniformément en \mathcal{S} et

$$\begin{aligned}
\mathbb{E} [Z_i(\mathcal{S}, b)^k] &= \int \left\{ \int \left[\frac{g(\mathbf{x} - b\mathbf{u})}{f(\mathbf{x} - b\mathbf{u})} \mathbb{I}_{(\mathbf{x} - b\mathbf{u} \in \mathcal{S})} - \frac{g(\mathbf{x})}{f(\mathbf{x})} \mathbb{I}_{(\mathbf{x} \in \mathcal{S})} \right] K(\mathbf{u}) d\mathbf{u} \right\}^k f(\mathbf{x}) d\mathbf{x} \\
&= \int \left\{ \frac{g(\mathbf{x})}{f(\mathbf{x})} \int [\mathbb{I}_{(\mathbf{x} - b\mathbf{u} \in \mathcal{S})} - \mathbb{I}_{(\mathbf{x} \in \mathcal{S})}] K(\mathbf{u}) d\mathbf{u} + \mathcal{O}(b) \right\}^k f(\mathbf{x}) d\mathbf{x} \\
&= \mathcal{O}(b)
\end{aligned}$$

car $\frac{g(\mathbf{x} - b\mathbf{u})}{f(\mathbf{x} - b\mathbf{u})} = \frac{g(\mathbf{x})}{f(\mathbf{x})} + \mathcal{O}(b)$, uniforme en \mathbf{x} et \mathbf{u} . De plus

$$\begin{aligned}
\mathbb{E} (n^{1/2} B_{\mathcal{S},4})^{2k} &= \frac{1}{n^k} \sum_{i_1, \dots, i_{2k}} \mathbb{E} [Z_{i_1}(\mathcal{S}, b) \times \dots \times Z_{i_{2k}}(\mathcal{S}, b)] \\
&= \frac{1}{n^k} \sum_{i_1 \neq \dots \neq i_{2k}} \mathbb{E} [Z_{i_1}(\mathcal{S}, b)^{2k}] + \dots + \frac{1}{n^k} \sum_i \mathbb{E} [Z_i(\mathcal{S}, b)^{2k}] \\
&= n^k b^{2k\tau} + n^{k-1} b^{(2k-1)\tau+1} + \dots + \frac{b}{n^{k-1}} \\
&= \mathcal{O}(b^k),
\end{aligned}$$

car $\kappa > 1/(2\tau - 1)$, c'est-à-dire $nb^{2\tau} = \mathcal{O}(b)$. On obtient donc bien le résultat (7.22) et en utilisant les inégalités de Markov à tout ordre on peut conclure :

$$\exists \eta_0, \forall C > 0, \text{ indépendant de } \eta_0, \quad \sup_C \mathbb{P}(|B_{S,4}| > n^{-1/2-\eta_0}) = \mathcal{O}(n^{-C}). \quad (7.23)$$

Deuxième partie : De la même façon que pour $A_{S,3}$ et $B_{S,3}$, on introduit un réseau de cylindres de cardinal n^A , mais cette fois la distance au réseau est n^{-a} où $a > 1$. On note alors et $\mathcal{I}_n(\mathcal{S})$ l'élément du réseau associé à \mathcal{S} et $\mathcal{I}_n(\mathcal{S})^\dagger$ l'ensemble des points de \mathbb{R}^d qui se trouvent à une distance de n^{-a} de la frontière de $\mathcal{I}_n(\mathcal{S})$. On remarque alors que

$$\begin{aligned} n^{1/2}|B_{S,4} - B_{\mathcal{I}_n(\mathcal{S}),4}| &\leq n^{-1/2} \sum_{i=1}^n \left| \int [\psi_S(X_i - b\mathbf{u}) - \psi_S(X_i)] K(\mathbf{u}) d\mathbf{u} \right. \\ &\quad \left. - \int [\psi_{\mathcal{I}_n(\mathcal{S})}(X_i - b\mathbf{u}) - \psi_{\mathcal{I}_n(\mathcal{S})}(X_i)] K(\mathbf{u}) d\mathbf{u} \right| \\ &\leq n^{-1/2} \sum_{i=1}^n \int |\psi_S(X_i - b\mathbf{u}) - \psi_{\mathcal{I}_n(\mathcal{S})}(X_i - b\mathbf{u})| K(\mathbf{u}) d\mathbf{u} \\ &\quad + n^{-1/2} \sum_{i=1}^n \int |\psi_S(X_i) - \psi_{\mathcal{I}_n(\mathcal{S})}(X_i)| K(\mathbf{u}) d\mathbf{u} \\ &\leq n^{-1/2} C \sum_{i=1}^n \int \mathbb{I}_{(X_i - b\mathbf{u} \in \mathcal{S} \Delta \mathcal{I}_n(\mathcal{S}))} K(\mathbf{u}) d\mathbf{u} \\ &\quad + n^{-1/2} C \sum_{i=1}^n \int \mathbb{I}_{(X_i \in \mathcal{S} \Delta \mathcal{I}_n(\mathcal{S}))} K(\mathbf{u}) d\mathbf{u} \\ &\leq n^{-1/2} C \sum_{i=1}^n \int \mathbb{I}_{(X_i - b\mathbf{u} \in \mathcal{I}_n(\mathcal{S})^\dagger)} K(\mathbf{u}) d\mathbf{u} \\ &\quad + n^{-1/2} C \sum_{i=1}^n \mathbb{I}_{(X_i \in \mathcal{I}_n(\mathcal{S})^\dagger)}. \end{aligned}$$

Ainsi

$$\begin{aligned} &\mathbb{P} \left(n^{1/2} \sup_C |B_{S,4} - B_{\mathcal{I}_n(\mathcal{S}),4}| \geq n^{-\eta} \right) \\ &\leq \mathbb{P} \left(n^{-1/2} \sup_{C'} \sum_{i=1}^n \int \mathbb{I}_{(X_i - b\mathbf{u} \in \mathcal{I}_n(\mathcal{S})^\dagger)} K(\mathbf{u}) d\mathbf{u} + n^{-1/2} \sup_{C'} \sum_{i=1}^n \mathbb{I}_{(X_i \in \mathcal{I}_n(\mathcal{S})^\dagger)} \geq n^{-\eta}/C \right) \\ &\leq n^A \mathbb{P} \left(\sum_{i=1}^n \int \mathbb{I}_{(X_i - b\mathbf{u} \in \mathcal{I}_n(\mathcal{S})^\dagger)} K(\mathbf{u}) d\mathbf{u} \geq \frac{n^{1/2-\eta}}{2C} \right) + n^A \mathbb{P} \left(\sum_{i=1}^n \mathbb{I}_{(X_i \in \mathcal{I}_n(\mathcal{S})^\dagger)} \geq \frac{n^{1/2-\eta}}{2C} \right). \end{aligned}$$

Étudions ces deux probabilités :

• $\mathbb{P}\left(\sum_{i=1}^n \mathbb{I}_{(X_i \in \mathcal{I}_n(\mathcal{S})^\dagger)} \geq C_1 n^{1/2-\eta}\right)$ est la probabilité pour une variable aléatoire de Bernoulli $\mathcal{B}(n, p_n)$ de dépasser $k_n = C_1 n^{1/2-\eta}$. Or $p_n = \mathbb{P}(X_i \in \mathcal{I}_n(\mathcal{S})^\dagger) = \mathcal{O}(n^{-a})$. Ainsi

$$\mathbb{P}\left(\sum_{i=1}^n \mathbb{I}_{(X_i \in \mathcal{I}_n(\mathcal{S})^\dagger)} \geq C_1 n^{1/2-\eta}\right) = \mathcal{O}(n^{1-a})^{[C_1 n^{1/2-\eta}]}$$

Or $\eta < 1/2$ et $a > 1$, finalement

$$\mathbb{P}\left(\sum_{i=1}^n \mathbb{I}_{(X_i \in \mathcal{I}_n(\mathcal{S})^\dagger)} \geq C_1 n^{1/2-\eta}\right) = \mathcal{O}(n^{-C'}), \quad \forall C' > 0$$

• Ici on s'intéresse à la variable aléatoire $\int \mathbb{I}_{(X_i - b\mathbf{u} \in \mathcal{I}_n(\mathcal{S})^\dagger)} K(\mathbf{u}) d\mathbf{u}$. Or

$$X_i - b\mathbf{u} \in \mathcal{I}_n(\mathcal{S})^\dagger \Rightarrow X_i \in \mathcal{I}_n(\mathcal{S})^{\dagger\dagger}$$

où $\mathcal{I}_n(\mathcal{S})^{\dagger\dagger}$ est l'ensemble des points à la distance inférieure ou égale à $(b \sup_{\mathbf{u} \in \mathbb{K}} \|\mathbf{u}\|)$ du bord de $\mathcal{I}_n(\mathcal{S})$. Donc K étant un noyau à support compact

$$\mathbb{P}\left(\sum_{i=1}^n \int \mathbb{I}_{(X_i - b\mathbf{u} \in \mathcal{I}_n(\mathcal{S})^\dagger)} K(\mathbf{u}) d\mathbf{u} \geq C_1 n^{1/2-\eta}\right) \leq \mathbb{P}\left(\sum_{i=1}^n \mathbb{I}_{(X_i \in \mathcal{I}_n(\mathcal{S})^{\dagger\dagger})} \geq C_1 n^{1/2-\eta}\right).$$

Il nous faut donc quantifier la probabilité d'une variable de Bernoulli $\mathcal{B}(n, q_n)$ de dépasser $k_n = C_1 n^{1/2-\eta}$. Or $q_n = \mathbb{P}(X_i - b\mathbf{u} \in \mathcal{I}_n(\mathcal{S})^{\dagger\dagger}) = \mathcal{O}(n^{-a})$. Ainsi

$$\mathbb{P}\left(\sum_{i=1}^n \mathbb{I}_{(X_i \in \mathcal{I}_n(\mathcal{S})^{\dagger\dagger})} \geq C_1 n^{1/2-\eta}\right) = \mathcal{O}(n^{1-a})^{[C_1 n^{1/2-\eta}]}$$

Or $\eta < 1/2$ et $a > 1$, donc

$$\mathbb{P}\left(\sum_{i=1}^n \int \mathbb{I}_{(X_i - b\mathbf{u} \in \mathcal{I}_n(\mathcal{S})^\dagger)} K(\mathbf{u}) d\mathbf{u} \geq C_1 n^{1/2-\eta}\right) = \mathcal{O}(n^{-C'}), \quad \forall C' > 0.$$

• Finalement en combinant ces résultats on obtient

$$\sup_{\mathcal{C}} |B_{\mathcal{S},4} - B_{\mathcal{I}_n(\mathcal{S}),4}| = \mathcal{O}(n^{-1/2}). \quad (7.24)$$

Les ordres uniformes donnés par (7.23) et (7.24) nous permettent d'obtenir (7.21) de la même façon que dans le cas de $A_{\mathcal{S},3}$.

7.2.5 Ordre de grandeur de $B_{\mathcal{S},5}$

Propriété 7.2.5. *Sous les hypothèses du chapitre 6*

$$|B_{\mathcal{S},5}| = \mathcal{O}(n^{-1/2}).$$

$B_{\mathcal{S},5}$ est indépendant de \mathcal{S} et ce résultat se démontre de la même façon que pour $A_{\mathcal{S},4}$.

DÉMONSTRATION. $B_{\mathcal{S},5}$ est donné par

$$B_{\mathcal{S},5} = n^{-1} \sum_{i=1}^n (\widehat{f}^{(i)}(X_i) - f(X_i))^2.$$

Évaluons $\mathbb{P}(|B_{\mathcal{S},5}| > n^{-1/2}\epsilon)$, pour $\epsilon > 0$:

$$\begin{aligned} \mathbb{P}(|B_{\mathcal{S},5}| > n^{-1/2}\epsilon) &\leq \frac{\mathbb{E} |B_{\mathcal{S},5}|}{n^{-1/2}\epsilon} \\ &\leq C n^{1/2} \mathbb{E} (\widehat{f}^{(i)}(X_i) - f(X_i))^2 \\ &\leq n^{1/2} \mathcal{O}\left(\frac{1}{nb^d} + b^{2\tau}\right) = \mathcal{O}(n^{-1/2+d\kappa}). \end{aligned}$$

Comme $\kappa d < 1/2$, le majorant tend donc vers 0 quand $n \rightarrow \infty$, ce qui permet de conclure.

■

7.3 Propriétés asymptotiques de $\widehat{I}_{\mathcal{S}}$

Les différents résultats des parties précédentes nous permettent d'établir que $\widehat{I}_{\mathcal{S}}$ est un estimateur asymptotiquement sans biais de $I_{\mathcal{S}}$ et qu'il converge uniformément en probabilité vers $I_{\mathcal{S}}$. Nous en déduisons aussi la convergence du processus $\zeta_n(\cdot)$ indexé sur les cylindres \mathcal{S} vers un processus gaussien.

Théorème 7.3.1. *On se place dans le cadre des hypothèses du chapitre 6*

Le processus $\zeta_n(\mathcal{S}) = n^{1/2} \left\{ A_{\mathcal{S}} + B_{\mathcal{S}} - \int_{\mathcal{S}} g \right\}$ converge faiblement vers un processus gaussien $\zeta(\cdot)$ indexé par les cylindres, centré et de covariance donnée pour des cylindres \mathcal{S}_1 et \mathcal{S}_2 distincts par

$$\text{cov}\{\zeta(\mathcal{S}_1), \zeta(\mathcal{S}_2)\} = \int_{\mathcal{S}_1 \cap \mathcal{S}_2} \sigma^2(\mathbf{x}) f^{-1}(\mathbf{x}) d\mathbf{x}.$$

DÉMONSTRATION.

• D'après la partie 7.1, on peut décomposer le processus $A_{\mathcal{S}}$

$$A_{\mathcal{S}} = A_{\mathcal{S},1} + A_{\mathcal{S},2} + A_{\mathcal{S},3} + \mathcal{O}(A_{\mathcal{S},4})$$

avec

- $n^{1/2}A_{\cdot,1} \xrightarrow{w} \zeta(\cdot)$ en tant que processus indexé par les cylindres de \mathcal{C} , cf. propriété 7.1.1.
- les propriétés 7.1.2, 7.1.3 et 7.1.4 assurent que $\sup_{\mathcal{C}}|A_{\mathcal{S},j}| = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$, pour $j = 2, 3, 4$.

Finalement

$$n^{1/2}A. \xrightarrow{w} \zeta(\cdot), \quad (7.25)$$

où $\zeta(\cdot)$ est un processus gaussien indexé par les cylindres de \mathcal{C} .

De plus en utilisant (7.8) on obtient l'ordre uniforme pour $A_{\mathcal{S}}$:

$$\sup_{\mathcal{C}}|A_{\mathcal{S}}| = \mathcal{O}_{\mathbb{P}}(n^{-1/2}). \quad (7.26)$$

- La partie 7.2 traite du processus $(B_{\mathcal{S}}(g) - \int_{\mathcal{S}} g)$, que l'on décompose lui aussi en cinq processus $B_{\mathcal{S},j}$, pour $j = 1, \dots, 5$. Les propriétés 7.2.1, 7.2.2, 7.2.3, 7.2.4 et 7.2.5 s'intéressent à chacun de ces termes et établissent que

$$\sup_{\mathcal{C}}|B_{\mathcal{S},j}(g)| = \mathcal{O}_{\mathbb{P}}(n^{1/2}) \quad \text{pour } j = 1, \dots, 5.$$

On en déduit donc que

$$\sup_{\mathcal{C}} \left| B_{\mathcal{S}}(g) - \int_{\mathcal{S}} g \right| = \mathcal{O}_{\mathbb{P}}(n^{1/2}). \quad (7.27)$$

- En rassemblant les deux dernières conclusions (7.27) et (7.25), on peut affirmer que $\zeta_n(\cdot)$ converge faiblement vers $\zeta(\cdot)$. ■

Chapitre 8

Convergence de l'estimateur de l'axe du SIM

Dans cette partie nous nous intéressons à la convergence de l'estimateur de l'axe φ_0 du SIM et de la statistique de test $\widehat{W}(\widehat{\varphi})$. Pour cela on conserve les notations introduites dans le chapitre 6 et on énonce le théorème suivant

Théorème 8.0.2. *Sous les hypothèses H1–H12, on obtient*

$$n^{1/2}(\widehat{\varphi} - \varphi_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \Omega$$

et

$$n \widehat{W}(\widehat{\varphi}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \overline{W}(\Omega).$$

où \overline{W} (resp. Ω) est défini par (6.4) (resp. (6.5)).

DÉMONSTRATION. Cette démonstration comporte neuf étapes successives, qui utilisent les résultats du chapitre précédent.

Par la suite et dans un esprit de simplification on notera $\mathcal{S} = \mathcal{S}(s, \mathcal{R}_{\theta_{\mathbf{x}}})$, $\mathcal{S}_0 = \mathcal{S}(s, \mathcal{R}_{\theta_0 \mathbf{x}})$

Étape 1- Montrons que $\sup_{\mathcal{C}} |\widehat{I}_{\mathcal{S}}(g) - I_{\mathcal{S}}(g)| = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$:

$$\begin{aligned} |\widehat{I}_{\mathcal{S}}(g) - I_{\mathcal{S}}(g)| &= \left| \|\mathcal{S}\|_d^{-1} (A_{\mathcal{S}} + B_{\mathcal{S}}(g)) - I_{\mathcal{S}}(g) \right|, \\ &\leq \|\mathcal{S}\|_d^{-1} \left[|A_{\mathcal{S}}| + \left| B_{\mathcal{S}}(g) - \int_{\mathcal{S}} g(\mathbf{x}) d\mathbf{x} \right| \right], \end{aligned}$$

car $I_{\mathcal{S}}(g) = \|\mathcal{S}\|_d^{-1} \int_{\mathcal{S}} g(\mathbf{x}) d\mathbf{x}$.

$$\begin{aligned} \sup_{\mathcal{C}} |\widehat{I}_{\mathcal{S}}(g) - I_{\mathcal{S}}(g)| &\leq \sup_{\mathcal{C}} \|\mathcal{S}\|_d^{-1} \left[|A_{\mathcal{S}}| + \left| B_{\mathcal{S}}(g) - \int_{\mathcal{S}} g(\mathbf{x}) \, d\mathbf{x} \right| \right], \\ \sup_{\mathcal{C}} |\widehat{I}_{\mathcal{S}}(g) - I_{\mathcal{S}}(g)| &\leq v^{-1} \left[\sup_{\mathcal{C}} |A_{\mathcal{S}}| + \sup_{\mathcal{C}} \left| B_{\mathcal{S}}(g) - \int_{\mathcal{S}} g(\mathbf{x}) \, d\mathbf{x} \right| \right], \\ \sup_{\mathcal{C}} |\widehat{I}_{\mathcal{S}}(g) - I_{\mathcal{S}}(g)| &\leq v^{-1} \left[\mathcal{O}_{\mathbb{P}}(n^{-1/2}) + \mathcal{O}(n^{-1/2}) \right], \end{aligned}$$

où $v = \inf_{\mathcal{S}} \|\mathcal{S}\|_d^{-1}$ et pour obtenir cette dernière inégalité nous utilisons les ordres de grandeur donnés par (7.26) (7.27). Finalement

$$\sup_{\mathcal{C}} |\widehat{I}_{\mathcal{S}}(g) - I_{\mathcal{S}}(g)| = \mathcal{O}_{\mathbb{P}}(n^{-1/2}). \quad (8.1)$$

D'autre part

$$\left| \widehat{V}(u, \varphi) - V(u, \varphi) \right| \leq \frac{\int_{\mathcal{V}(u, \varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_{\theta}(\mathbf{v})\}} |\widehat{I}_{\mathcal{S}}(g) - I_{\mathcal{S}}(g)| \, ds \, d\mathbf{x}}{\int_{\mathcal{V}(u, \varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_{\theta}(\mathbf{v})\}} ds \, d\mathbf{x}}.$$

On obtient donc

$$\sup_{u, \varphi} \left| \widehat{V}(u, \varphi) - V(u, \varphi) \right| \leq \mathcal{O} \left(\sup_{\mathcal{C}} |\widehat{I}_{\mathcal{S}}(g) - I_{\mathcal{S}}(g)| \right) = \mathcal{O}_{\mathbb{P}}(n^{-1/2}). \quad (8.2)$$

■

On définit maintenant $\widehat{\Delta}$ comme la quantité Δ introduite par l'équation (6.1) dans laquelle on remplace toutes les quantités I par \widehat{I} .

Étape 2- Montrons que $\widehat{\Delta}(s, u, \mathbf{x}, \theta, \varphi_0) = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$ uniformément en $(s, \mathbf{x}) \in \mathcal{T}\{\alpha_{\theta}(\mathbf{v})\}$ et pour tout $\theta \in \Theta(\varphi_0)$, $\mathbf{v} \in \mathcal{V}(u, \varphi_0)$ et $u \in \mathcal{U}(\varphi_0)$.

On sait que $\widehat{\Delta}(s, u, \mathbf{x}, \theta, \varphi) - \Delta(s, u, \mathbf{x}, \theta, \varphi) = \widehat{I}_{\mathcal{S}}(g) - I_{\mathcal{S}}(g) - (\widehat{V}(u, \varphi) - V(u, \varphi))$. De plus si φ_0 est l'axe du SIM, alors pour $\theta \perp \varphi_0$,

$$I_{\mathcal{S}} = cste, \quad \forall \mathbf{x} \in \alpha_{\theta}(\mathbf{v}) \quad , \quad s \in [s_1, s_2] \quad \text{tels que } \mathcal{S} \subset \mathcal{Q}.$$

Finalement pour u fixé, $V(u, \varphi_0) = I_{\mathcal{S}}$, $\forall (s, \mathbf{x}, \theta)$. D'où pour tout $(s, \mathbf{x}) \in \mathcal{T}(\alpha_{\theta}(\mathbf{v}))$, $\theta \in \Theta(\varphi_0)$, $\mathbf{v} \in \mathcal{V}(u, \varphi_0)$ et $u \in \mathcal{U}(\varphi_0)$, on a

$$\Delta(s, u, \mathbf{x}, \theta, \varphi_0) = I_{\mathcal{S}} - V(u, \varphi_0) = 0.$$

On peut donc écrire pour $\varphi = \varphi_0$:

$$\widehat{\Delta}(s, u, \mathbf{x}, \theta, \varphi_0) - 0 = \widehat{I}_{\mathcal{S}}(g) - I_{\mathcal{S}}(g) - (\widehat{V}(u, \varphi_0) - V(u, \varphi_0)).$$

En utilisant les deux relations obtenues à l'étape précédente, (8.1) et (8.2), on peut en déduire

$$|\widehat{\Delta}(s, u, \mathbf{x}, \theta, \varphi_0)| = \mathcal{O}_{\mathbb{P}}(n^{-1/2}),$$

uniformément en $(s, \mathbf{x}) \in \mathcal{T}(\alpha_{\theta}(\mathbf{v}))$, $\theta \in \Theta(\varphi_0)$, $\mathbf{v} \in \mathcal{V}(u, \varphi_0)$ et $u \in \mathcal{U}(\varphi_0)$. ■

Étape 3- Montrons que $\inf_{\theta \in \Theta} \widehat{W}(\theta) = \mathcal{O}_{\mathbb{P}}(n^{-1})$.

Par définition de \widehat{W} , on a

$$\widehat{W}(\varphi) = \int_{\mathcal{U}(\varphi)} du \int_{\mathcal{V}(u, \varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_{\theta}(\mathbf{v})\}} \widehat{\Delta}(s, u, \mathbf{x}, \theta, \varphi)^2 ds d\mathbf{x}.$$

D'après le paragraphe précédent, on sait que pour ε fixé,

$$\exists A > 0 / \mathbb{P} \left(\frac{|\widehat{\Delta}(s, u, \mathbf{x}, \theta, \varphi_0)|}{n^{-1/2}} \leq A \right) \geq 1 - \varepsilon,$$

où A est un majorant indépendant de $(u, \theta, s, \mathbf{x})$, choisis tels que $(s, \mathbf{x}) \in \mathcal{T}(\alpha_{\theta}(\mathbf{v}))$, $\theta \in \Theta(\varphi_0)$, $\mathbf{v} \in \mathcal{V}(u, \varphi_0)$ et $u \in \mathcal{U}(\varphi_0)$. Comme les ensembles sur lesquels portent les intégrales sont des compacts, on a l'inclusion suivante, où c est une constante bien choisie,

$$\left\{ \frac{|\widehat{\Delta}(s, u, \mathbf{x}, \theta, \varphi_0)|}{n^{-1/2}} \leq A \right\} \subset \left\{ \frac{|\widehat{W}(\varphi_0)|}{n^{-1}} \leq cA^2 \right\},$$

d'où $\mathbb{P} \left(\frac{|\widehat{W}(\varphi_0)|}{n^{-1}} \leq cA^2 \right) \geq 1 - \varepsilon$, i.e. $\widehat{W}(\varphi_0) = \mathcal{O}_{\mathbb{P}}(n^{-1})$. De plus $\inf_{\theta \in \Theta} \widehat{W}(\theta) \leq \widehat{W}(\varphi_0)$, on obtient le résultat attendu

$$\inf_{\theta \in \Theta} \widehat{W}(\theta) = \mathcal{O}_{\mathbb{P}}(n^{-1}).$$

■

Étape 4- Montrons que $\widehat{W}(\varphi) \xrightarrow{\mathbb{P}} W(\varphi)$ uniformément en $\varphi \in \Theta$.

$$\begin{aligned}
\left| \widehat{W}(\varphi)^{1/2} - W(\varphi)^{1/2} \right| &\leq \left[\int_{\mathcal{U}(\varphi)} du \int_{\mathcal{V}(u,\varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_\theta(\mathbf{v})\}} \left[(\widehat{I}_S(g) - \widehat{V}(u,\varphi)) \right. \right. \\
&\quad \left. \left. - (I_S(g) - V(u,\varphi)) \right]^2 ds d\mathbf{x} \right]^{1/2}, \\
&\leq \left[\int_{\mathcal{U}(\varphi)} du \int_{\mathcal{V}(u,\varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_\theta(\mathbf{v})\}} 2(\widehat{I}_S(g) - I_S(g))^2 \right. \\
&\quad \left. + 2(\widehat{V}(u,\varphi) - V(u,\varphi))^2 ds d\mathbf{x} \right]^{1/2}, \\
&\leq C \left[\sup_{\mathcal{C}} |\widehat{I}_S(g) - I_S(g)|^2 + \sup_{u,\varphi} |\widehat{V}(u,\varphi) - V(u,\varphi)|^2 \right]^{1/2}.
\end{aligned}$$

En utilisant alors les majorations uniformes de ces deux termes données en (8.1) et (8.2), on obtient uniformément en $\varphi \in \Theta$

$$\widehat{W}(\varphi)^{1/2} - W(\varphi)^{1/2} = \mathcal{O}_{\mathbb{P}}(n^{-1/2}).$$

Donc uniformément en $\varphi \in \Theta$,

$$\widehat{W}(\varphi)^{1/2} \xrightarrow{\mathbb{P}} W(\varphi)^{1/2}, \quad \text{d'où} \quad \widehat{W}(\varphi) \xrightarrow{\mathbb{P}} W(\varphi).$$

■

On peut alors obtenir des résultats de convergence et d'ordre de grandeur pour $\widehat{\varphi}$.

Étape 5- Montrons que $\widehat{\varphi}$ converge en probabilité vers φ_0 .

Propriété 8.0.1. On définit $\widehat{\varphi} = \operatorname{argmin}_{\varphi \in \Theta} \widehat{W}(\varphi)$. Alors

$$\widehat{\varphi} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \varphi_0.$$

DÉMONSTRATION. Les 4 propriétés suivantes sont vérifiées, d'après les hypothèses du théorème 8.0.2 et les étapes précédentes :

1. l'espace des paramètres Θ est un compact de \mathbb{R}^d ;
2. $\widehat{W}^{1/2}(\varphi, y)$ est une fonction continue en φ , pour tout y , et mesurable en y , pour tout $\varphi \in \Theta$;

3. $\widehat{W}(\varphi)$ converge vers $W(\varphi)$, fonction déterministe, en probabilité uniformément pour $\varphi \in \Theta$, quand $n \rightarrow +\infty$;
4. $W(\varphi)$ admet un unique minimum φ_0 .

On peut donc adapter le théorème 4.1.1 de *Amemiya (1985)* pour démontrer la convergence en probabilité voulue : soit N un voisinage ouvert de φ_0 dans \mathbb{R}^d . Alors $N' = \Theta \cap N^c$ est un compact donc W , continue, admet un minimum sur N' , qu'elle atteint.

Soit $\varepsilon = \min_{\varphi \in N'} W(\varphi) - W(\varphi_0)$ et $A_n = \left\{ |\widehat{W}(\varphi) - W(\varphi)| < \varepsilon/2, \forall \varphi \in \Theta \right\}$.

Alors sur A_n ,

$$\widehat{W}(\widehat{\varphi}) > W(\widehat{\varphi}) - \varepsilon/2, \quad (8.3)$$

$$W(\varphi_0) > \widehat{W}(\varphi_0) - \varepsilon/2. \quad (8.4)$$

Or $\widehat{W}(\widehat{\varphi}) < \widehat{W}(\varphi_0)$ et l'inégalité (8.4) impliquent

$$W(\varphi_0) > \widehat{W}(\widehat{\varphi}) - \varepsilon/2.$$

En additionnant cette dernière inégalité avec (8.3)

$$W(\varphi_0) > W(\widehat{\varphi}) - \varepsilon.$$

En utilisant la définition de ε , on obtient

$$\min_{\varphi \in N'} W(\varphi) > W(\widehat{\varphi}),$$

d'où $\widehat{\varphi}$ est dans N . On a donc $\mathbb{P}(\widehat{\varphi} \in N) \geq \mathbb{P}(A_n)$ et comme $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = 1$, on obtient

$$\widehat{\varphi} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \varphi_0.$$

■

Étape 6- Montrons que $\|\widehat{\varphi} - \varphi_0\| = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$.

Uniformément par rapport à $\varphi \in \Theta$, on a

$$\widehat{W}(\varphi)^{1/2} - W(\varphi)^{1/2} = \mathcal{O}_{\mathbb{P}}(n^{-1/2}),$$

or W atteint son minimum, 0, en φ_0 de façon quadratique, i.e. pour $\eta > 0$ assez petit et c_3 indépendant de φ ,

$$W(\varphi)^{1/2} \geq c_3 \|\varphi - \varphi_0\|, \quad \forall \varphi / \|\varphi - \varphi_0\| \leq \eta.$$

En utilisant cette propriété et celle de l'étape **3-**, on obtient

$$\mathcal{O}_{\mathbb{P}}(n^{-1/2}) = \widehat{W}(\widehat{\varphi})^{1/2}$$

et

$$\widehat{W}(\varphi)^{1/2} + \mathcal{O}_{\mathbb{P}}(n^{-1/2}) \geq c_3 \|\varphi - \varphi_0\|, \quad \forall \varphi / \|\varphi - \varphi_0\| \leq \eta.$$

On peut alors écrire pour $\varepsilon > 0$, qu'il existe B , indépendant de φ tel que $\mathbb{P}(B) \geq 1 - \varepsilon$ et sur B , il existe M tel que

$$\widehat{W}(\widehat{\varphi})^{1/2} \leq Mn^{-1/2}, \quad \widehat{W}(\widehat{\varphi})^{1/2} + Mn^{-1/2} \geq c_3 \|\widehat{\varphi} - \varphi_0\|.$$

La convergence de $\widehat{\varphi}$ vers φ_0 en probabilité nous permet d'écrire pour $A_n = \{\|\widehat{\varphi} - \varphi_0\| < \eta\}$, $\mathbb{P}(A_n) \rightarrow 1$. Sur $A_n \cap B$, en utilisant les deux inégalités ci-dessus, on obtient

$$\|\widehat{\varphi} - \varphi_0\| \leq Cn^{-1/2}$$

or $\mathbb{P}(A_n \cap B) \geq 1 - 2\varepsilon$, pour un n assez grand. Finalement

$$\|\widehat{\varphi} - \varphi_0\| = \mathcal{O}_{\mathbb{P}}(n^{-1/2}).$$

■

Ce dernier résultat implique que pour obtenir des résultats de convergence pour $\widehat{\varphi}$ et $\widehat{W}(\widehat{\varphi})$ il suffit de traiter le cas $\|\widehat{\varphi} - \varphi_0\| \leq Cn^{-1/2}$, où $C > 0$ est fixé, mais arbitrairement grand.

Étape 7- Développement de $\widehat{\Delta}(s, u, \mathbf{x}, \theta, \varphi)$ en φ_0 .

On rappelle les notations utilisées

$$\zeta_n(\mathcal{S}) = n^{1/2} \left\{ A_{\mathcal{S}} + B_{\mathcal{S}}(g) - \int_{\mathcal{S}} g \right\}$$

avec $\|\mathcal{S}\|_d \widehat{I}_{\mathcal{S}}(g) = A_{\mathcal{S}} + B_{\mathcal{S}}(g)$. D'où

$$\begin{aligned} n^{1/2} \widehat{\Delta}(s, u, \mathbf{x}, \theta, \varphi) - n^{1/2} \Delta(s, u, \mathbf{x}, \theta, \varphi) &= n^{1/2} \left(\widehat{I}_{\mathcal{S}}(g) - I_{\mathcal{S}}(g) \right) - n^{1/2} \left(\widehat{V}(u, \varphi) - V(u, \varphi) \right) \\ &= \|\mathcal{S}\|_d^{-1} \zeta_n\{\mathcal{S}\} - \frac{\int_{\mathcal{V}(u, \varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_{\theta}(\mathbf{v})\}} \|\mathcal{S}\|_d^{-1} \zeta_n\{\mathcal{S}\} ds d\mathbf{x}}{\int_{\mathcal{V}(u, \varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_{\theta}(\mathbf{v})\}} ds d\mathbf{x}}. \end{aligned}$$

En utilisant des arguments semblables à ceux du paragraphe 7, on peut démontrer que uniformément sur les cylindres vérifiant $\|\varphi - \varphi_0\| \leq Cn^{-1/2}$

$$|A_{\mathcal{S}} - A_{\mathcal{S}_0}| = \mathcal{O}_{\mathbb{P}}(n^{-1/2}), \quad \text{et} \quad \left| B_{\mathcal{S}}(g) - B_{\mathcal{S}_0}(g) - \int_{\mathcal{S}} g - \int_{\mathcal{S}_0} g \right| = \mathcal{O}_{\mathbb{P}}(n^{-1/2}).$$

Et comme la différence entre les intégrales converge uniformément vers 0 quand $\varphi \rightarrow \varphi_0$, on peut en déduire que

$$n^{1/2} \widehat{\Delta}(s, u, \mathbf{x}, \theta, \varphi) = n^{1/2} \Delta(s, u, \mathbf{x}, \theta, \varphi) + \|\mathcal{S}_0\|_d^{-1} \zeta_n\{\mathcal{S}_0\} \\ - \frac{\int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} \|\mathcal{S}_0\|_d^{-1} \zeta_n\{\mathcal{S}_0\} ds d\mathbf{x}}{\int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} ds d\mathbf{x}} + \mathcal{O}_{\mathbb{P}}(1).$$

Grâce au développement de Δ en φ_0 donné en (6.2), on a

$$n^{1/2} \widehat{\Delta}(s, u, \mathbf{x}, \theta, \varphi) = n^{1/2} (\varphi - \varphi_0)^T \beta(s, u, \mathbf{x}, \theta_0, \varphi_0) + \|\mathcal{S}_0\|_d^{-1} \zeta_n(\mathcal{S}_0) \\ - \frac{\int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} \|\mathcal{S}_0\|_d^{-1} \zeta_n(\mathcal{S}_0) ds d\mathbf{x}}{\int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} ds d\mathbf{x}} + \mathcal{O}_{\mathbb{P}}(1), \quad (8.5)$$

uniformément sur $\{\varphi / \|\varphi - \varphi_0\| \leq Cn^{-1/2}\}$. ■

Étape 8- Développement de $\widehat{W}(\varphi)$ en φ_0 .

On pose $\omega = n^{1/2}(\varphi - \varphi_0)$. Pour $\|\omega\| \leq C$, on peut écrire, car la différence entre les intégrales tend vers 0

$$\left| \left[n \widehat{W}(\varphi) \right]^{1/2} - \overline{W}(\omega)^{1/2} \right| \\ \leq \mathcal{O}(\|\omega\|) + \left[\int_{\mathcal{U}(\varphi_0)} du \int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} \left(n^{1/2} \widehat{\Delta}(s, u, \mathbf{x}, \theta, \varphi) \right. \right. \\ \left. \left. - \xi(s, u, \mathbf{x}, \theta_0, \varphi_0) - \omega^T \beta(s, u, \mathbf{x}, \theta_0, \varphi_0) + M_{\xi}(u, \varphi_0) \right)^2 ds d\mathbf{x} \right]^{1/2}.$$

On définit ξ_n (resp. M_{ξ_n}) la quantité ξ (resp. M_{ξ}) dans laquelle on remplace le processus limite, ζ , par ζ_n . On obtient en utilisant le développement (8.5)

$$\begin{aligned}
& \left| \left[n\widehat{W}(\varphi) \right]^{1/2} - \overline{W}(\omega)^{1/2} \right| \\
& \leq \mathcal{O}(\|\omega\|) + \left[\int_{\mathcal{U}(\varphi_0)} du \int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} \left(\xi_n(s, u, \mathbf{x}, \theta_0, \varphi_0) \right. \right. \\
& \quad \left. \left. - M_{\xi_n}(u, \varphi_0, w) - \xi(s, u, \mathbf{x}, \theta_0, \varphi_0) + M_{\xi}(u, \varphi_0) + \mathcal{O}_{\mathbb{P}}(1) \right)^2 ds d\mathbf{x} \right]^{1/2}. \quad (8.6)
\end{aligned}$$

Dans le paragraphe 7, on a obtenu la convergence faible de $\zeta_n(\mathcal{S})$ vers $\zeta(\mathcal{S})$, au sens des processus indexés par les cylindres de \mathcal{C} . Grace au théorème d'extension de Kolmogorov, on a alors $\zeta_n - \tilde{\zeta} \xrightarrow{\mathbb{P}} 0$ (cf. Billingsley (1968) p.207). La variable aléatoire $\tilde{\zeta}$ dépend de n , mais a la même loi que ζ . Or $\tilde{\zeta}$ et ζ_n ont un moment fini d'ordre $(2 + \eta)$, pour un $\eta > 0$, uniformément bornés donc $\{|\zeta_n - \tilde{\zeta}|^{2+\eta}\}_{n \in \mathbb{N}}$ est équi-intégrable et comme $\zeta_n - \tilde{\zeta} \xrightarrow{\mathbb{P}} 0$.

On a $\zeta_n - \tilde{\zeta} \xrightarrow{L^{2+\eta}} 0$ pour un $\eta > 0$. Donc $\zeta_n - \tilde{\zeta} \xrightarrow{L^2} 0$. Alors les deux différences de (8.6) faisant intervenir $(\zeta_n - \tilde{\zeta})$, tendent vers 0.

Étant donné que $\|\omega\| < C$, pour $C > 0$ fixé arbitrairement grand, on obtient donc :

$$n\widehat{W}(\varphi) = \widetilde{W}(n^{1/2}(\varphi - \varphi_0)) + \mathcal{O}_{\mathbb{P}}(1),$$

uniformément sur $\{\varphi/\|\varphi - \varphi_0\| \leq Cn^{-1/2}\}$, où \widetilde{W} dépend aussi de n mais a la même loi que \overline{W} défini par (6.4). \blacksquare

Étape 9- Convergence en probabilité de $n^{1/2}(\widehat{\varphi} - \varphi_0)$ vers $\Omega = \underset{\omega \in \mathbb{R}^d}{\operatorname{argmin}} \overline{W}(\omega)$ et de $n\widehat{W}(\widehat{\varphi})$ vers $\overline{W}(\Omega)$.

Convergence de $n^{1/2}(\widehat{\varphi} - \varphi_0)$

Par le théorème d'extension de Kolmogorov, la distribution de $\widehat{\varphi}$ qui minimise \widehat{W} parmi les φ vérifiant $\|\varphi - \varphi_0\| < Cn^{-1/2}$, est la même que la distribution de φ qui minimise $(\overline{W}(n^{1/2}\|\varphi - \varphi_0\|) + Z_{n,1}(\varphi))$, où \overline{W} peut être choisi comme un processus stochastique fixe, i.e. avec une construction ne dépendant pas de n et

$$\sup_{\varphi: \|\varphi - \varphi_0\| \leq n^{-1/2}C} |Z_{n,1}(\varphi)| \xrightarrow{\mathbb{P}} 0.$$

On pose alors $\psi = n^{1/2}(\varphi - \varphi_0)$ et $Z_{n,2}(\psi) = Z_{n,1}(\varphi)$. Alors on cherche la distribution de $\psi = \psi_n$ qui minimise $\overline{W}(\psi) + Z_{n,2}(\psi)$, pour ψ tel que $\|\psi\| < C$ et nous allons montrer

que $\psi_n \xrightarrow{\mathbb{P}} \psi(C)$, où $\psi(C) = \operatorname{argmin}_{\|\psi\| < C} \overline{W}(\psi)$, qui est uniquement défini car \overline{W} est strictement convexe sur $\{\|\psi\| < C\}$

Par définition de ψ_n et pour tout $\delta > 0$

$$\begin{aligned} 1 &= \mathbb{P}(\overline{W}(\psi_n) + Z_{n,2}(\psi_n) \leq \overline{W}(\psi(C)) + Z_{n,2}(\psi(C))) \\ 1 &= \mathbb{P}(\overline{W}(\psi_n) \leq \overline{W}(\psi(C)) + \delta) + o(1) \\ 1 &= \mathbb{P}(0 < \overline{W}(\psi_n) - \overline{W}(\psi(C)) \leq \delta) + o(1) \end{aligned}$$

car $\mathbb{P}(Z_{n,2}(\psi_n) < -\delta/2 \text{ ou } Z_{n,2}(\psi(C)) > \delta/2) = o(1)$.

Or $\psi(C)$ est le minimum de \overline{W} qui est continue et strictement convexe on en déduit donc que $\psi_n \xrightarrow{\mathbb{P}} \psi(C)$. Or $\psi_n = n^{1/2}(\varphi - \varphi_0)$, où φ a la même loi que $\widehat{\varphi}$ et $\psi(C) = \Omega$ pour C assez grand. On a donc la convergence souhaitée pour $n^{1/2}(\widehat{\varphi} - \varphi_0)$.

Convergence de $n\widehat{W}(\widehat{\varphi})$

Soit $\epsilon > 0$. La continuité de \overline{W} permet d'écrire :

$$\exists \eta \text{ tel que } |\omega - \omega'| < \eta \Rightarrow |\overline{W}(\omega) - \overline{W}(\omega')| < \epsilon/2.$$

On note alors

$$A_n = \left\{ \left| n\widehat{W}(\varphi) - \overline{W}(n^{1/2}(\varphi - \varphi_0)) \right| < \epsilon/2, \quad \forall \varphi / \|\varphi - \varphi_0\| < C_0 n^{-1/2} \right\},$$

et

$$B_n = \{\|\widehat{\varphi} - \varphi_0\| < n^{1/2}C_0\}, \quad C_n = \left\{ \left\| n^{1/2}(\widehat{\varphi} - \varphi_0) - \Omega \right\| < \eta \right\}.$$

Sur $A_n \cap B_n \cap C_n$:

$$\begin{aligned} \left| n\widehat{W}(\widehat{\varphi}) - \overline{W}(\Omega) \right| &\leq \left| n\widehat{W}(\widehat{\varphi}) - \overline{W}(n^{1/2}(\widehat{\varphi} - \varphi_0)) \right| + \left| \overline{W}(n^{1/2}(\widehat{\varphi} - \varphi_0)) - \overline{W}(\Omega) \right|, \\ &< \epsilon. \end{aligned}$$

Donc $\mathbb{P}(A_n \cap B_n \cap C_n) \leq \mathbb{P}(|n\widehat{W}(\widehat{\varphi}) - \overline{W}(\Omega)| < \epsilon)$, or $\mathbb{P}(A_n \cap B_n \cap C_n) \geq 1 - [\mathbb{P}(A_n^c) + \mathbb{P}(B_n^c) + \mathbb{P}(C_n^c)]$. Or $\mathbb{P}(A_n^c)$ et $\mathbb{P}(C_n^c)$ tendent vers 0 et $\mathbb{P}(B_n) > 1 - \epsilon_0$. De plus $\mathbb{P}(A_n) \rightarrow 1$ quel que soit le choix de C_0 , ainsi $\mathbb{P}(B_n) > 1 - \epsilon_0$ pour tout ϵ_0 en adaptant le choix de C_0 . Finalement

$$\mathbb{P}(|n\widehat{W}(\widehat{\varphi}) - \overline{W}(\Omega)| < \epsilon) > 1 - \epsilon_0, \quad \forall \epsilon_0 > 0.$$

On a donc en particulier

$$n\widehat{W}(\widehat{\varphi}) \xrightarrow{\mathcal{L}} \overline{W}(\Omega).$$

■

Chapitre 9

Étude du test

Dans ce chapitre nous étudions plus particulièrement le test du modèle SIM. Pour cela nous présentons dans un premier temps la calibration du test par bootstrap. Dans un deuxième temps nous étudierons la convergence du test et sa puissance.

9.1 Calibration par Bootstrap

On suppose que σ est constante et on considère le modèle

$$Y_i = g(X_i) + \varepsilon_i,$$

où $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.

Lors de la mise en œuvre de l'algorithme, nous aurons besoin de deux estimateurs de la fonction de régression g et d'un estimateur de la densité f . Ces choix peuvent se faire de la façon suivante :

- Soit $\tilde{g}(\mathbf{x})$ un estimateur non paramétrique classique de la régression, qui ne prend pas en compte l'hypothèse de SIM.
- Soit $\hat{g}_1(\hat{\varphi} \cdot \mathbf{x})$ un estimateur dans le modèle à direction révélatrice unique. Pour l'estimateur de l'axe de la régression, on peut utiliser soit celui proposé dans *Powell, Stock et Stoker (1989)*, soit celui proposé dans *Härdle, Hall et Ichimura (1993)*. Mais pour avoir une démarche cohérente, nous considérons que $\hat{\varphi}$ est l'estimateur de l'axe de la régression proposé dans le chapitre précédent. Comme $\hat{\varphi}$ étant un M-estimateur, il converge à la vitesse \sqrt{n} vers φ_0 , vecteur qui minimise $W(\varphi)$, même si g n'est plus un SIM. De même $\hat{g}_1(t)$ peut être toujours défini comme l'estimateur de $g_1(t) = \mathbb{E}[g(X)|\varphi_0 \cdot X = t]$. Par exemple \hat{g}_1 peut être construit comme estimateur localement linéaire sur l'échantillon $\{(Y_i, \hat{\varphi} \cdot X_i), i = 1, \dots, n\}$. Cet estimateur peut être construit tel que \hat{g}'_1 converge vers g'_1 , il suffit de choisir un bon paramètre de fenêtre.
- Soit \hat{f} estimateur de la densité f de la variable aléatoire X_i . Nous allons rééchantillonner les X_i , car nous sommes dans un modèle dit de "corrélation à erreurs indépendantes", cf. *Hall (1992) p.170*.

Le déroulement de la procédure bootstrap est le suivant.

1° On calcule les erreurs standards $\tilde{\varepsilon}_i = Y_i - \tilde{g}(X_i)$, sans tenir compte du modèle sous-jacent. Puis on les centre

$$\hat{\varepsilon}_i = \tilde{\varepsilon}_i - N^{-1} \sum_{i=1}^N \tilde{\varepsilon}_i$$

pour $i = 1, \dots, N$, où $N < n$, car on ne calcule pas $\hat{\varepsilon}_i$ quand X_i est trop proche du bord de \mathcal{Q} , cf. *Hall (1992) p.230*. On rééchantillonne à partir des $(\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_N)$ par remplacement, c'est-à-dire en utilisant la fonction de répartition empirique des $\hat{\varepsilon}_i$, et on obtient $(\varepsilon_1^*, \dots, \varepsilon_n^*)$.

2° Les variables explicatives sont rééchantillonnées suivant la loi de densité \hat{f} . On obtient alors un échantillon $\mathcal{X}^* = \{X_1^*, \dots, X_n^*\}$.

3° On pose $Y_i^* = \hat{g}_1(\hat{\varphi} \cdot X_i^*) + \varepsilon_i^*$ pour $i = 1, \dots, n$.

Alors on dispose d'un échantillon bootstrap $\mathcal{Z}^* = \{(X_i^*, Y_i^*), \text{ pour } i = 1, \dots, n\}$. En opérant ainsi on garde l'indépendance des X_i^* et ε_i^* .

4° On calcule les quantités bootstrap \hat{V}^* et \hat{W}^* correspondant respectivement à \hat{V} et \hat{W} à partir de l'échantillon bootstrap \mathcal{Z}^* .

5° On définit $\hat{\varphi}^* = \underset{\varphi \in \Theta}{\operatorname{argmin}} \hat{W}^*(\varphi)$ et on calcule la distribution bootstrap de $\hat{W}^*(\hat{\varphi}^*)$ conditionnellement à \mathcal{Z} .

6° Pour un niveau donné π (par exemple 0.05) on note \hat{t}_π la valeur de t telle que

$$\mathbb{P}(\hat{W}^*(\hat{\varphi}^*) \leq t | \mathcal{Z}) = 1 - \pi.$$

Le test consiste alors à rejeter l'hypothèse nulle, i.e. g est un SIM sur \mathcal{Q}_δ , au niveau π si $\hat{W}(\hat{\varphi}) > \hat{t}_\pi$.

9.2 Propriétés du test

Pour étudier les propriétés de test, des hypothèses supplémentaires sont requises. Tout d'abord nous remplaçons l'hypothèse H5 par

Hypothèse H15. *La fonction σ est constante.*

Pour assurer la convergence de la distribution de la statistique bootstrapée, nous avons besoin que le nombre des X_i proches du bord de \mathcal{Q} ne soit pas trop grand :

Hypothèse H16. *L'entier N , introduit dans l'algorithme bootstrap, satisfait*

$$\exists 0 < \eta < 1, \quad N \geq (1 - \eta)n.$$

Comme pour la convergence de $W(\varphi)$, nous avons besoin d'hypothèses de régularité sur g et f , mais aussi sur leurs estimateurs :

Hypothèse H17. Les estimateurs \widehat{f} , \widetilde{g} et \widehat{g}_1 sont construits tels que

$$\sup |\widehat{f}^{(i)} - f^{(i)}| \xrightarrow{\mathbb{P}} 0, \quad \sup |\widetilde{g} - g| \xrightarrow{\mathbb{P}} 0$$

$$\sup |\widehat{g}_1^{(j)} - g_1^{(j)}| \xrightarrow{\mathbb{P}} 0, \quad \sup |\widehat{f}^{(\tau+1)}| = \mathcal{O}_{\mathbb{P}}(1),$$

où $1 \leq i \leq \tau$, $j = 1, 2$, les première, deuxième et quatrième bornes supérieures sont prises sur \mathcal{Q}' et la troisième sur l'ensemble $\{\varphi_0 \cdot \mathbf{x} : \mathbf{x} \in \mathcal{Q}\}$.

On rappelle que

$$\mathcal{Q}' = \bigcup_{\varphi \in \Theta} \bigcup_{u \in \mathcal{U}(\varphi)} \bigcup_{\mathbf{v} \in \mathcal{V}(u, \varphi)} \bigcup_{\theta \in \Theta(\varphi)} \bigcup_{(s, \mathbf{x}) \in \mathcal{T}\{\alpha_\theta(\mathbf{v})\}} \mathcal{S}(s, \mathcal{R}_{\theta \mathbf{x}})$$

i.e. \mathcal{Q}' correspond au sous ensemble de \mathcal{Q} recouvert par les cylindres.

Dans le théorème suivant, la partie **(I)** concerne le niveau asymptotique du test bootstrap sous l'hypothèse \mathcal{H}_0 . La partie **(II)** fournit la convergence vers 1, sous l'alternative où g fixée n'est pas un SIM, de la probabilité de rejeter \mathcal{H}_0 . La partie **(III)** énonce le fait que le test est capable de distinguer une alternative distante de $n^{-1/2}$ de l'hypothèse nulle. En utilisant un argument similaire, on montrerait que si l'hypothèse alternative est du même type excepté le terme $cn^{-1/2}$, que l'on remplace par $\delta(n)$ où $\delta(n) \rightarrow 0$ et $n^{1/2}\delta(n) \rightarrow \infty$, alors la probabilité de rejeter \mathcal{H}_0 convergerait vers 1. Dans la partie **(III)** du théorème, pour laquelle g dépend de n , nous ne considérons plus la condition H6, qui est relative à g .

Théorème 9.2.1. On suppose que les hypothèses H1–H4, H6 et H8–H17 sont vérifiées, on a

- **(I)** Sous l'hypothèse \mathcal{H}_0 de SIM sur \mathcal{Q}' , pour $0 < \pi < 1$, la probabilité que $\widehat{W}(\widehat{\varphi}) > \widehat{t}_\pi$ converge vers π , quand $n \rightarrow \infty$.
- **(II)** Si g est fixée et n'est pas un SIM sur \mathcal{Q}_δ , alors pour tout $0 < \pi < 1$ la probabilité que $\widehat{W}(\widehat{\varphi}) > \widehat{t}_\pi$ converge vers 1, quand $n \rightarrow \infty$.
- **(III)** Si $g = \gamma_1 + cn^{-1/2}\gamma_2$ sur \mathcal{Q} , où (i) $\gamma_1(\mathbf{x}) = g_1(\varphi_0 \cdot \mathbf{x})$ est un SIM fixé sur \mathcal{Q} , de représentation unique sur \mathcal{Q}_δ , (ii) γ_2 n'est pas un SIM avec le même axe φ_0 sur \mathcal{Q}_δ , et (iii) g_1 et γ_2 ont des dérivées continues, alors on a

$$\lim_{|c| \rightarrow \infty} \liminf_{n \rightarrow \infty} P(\mathcal{H}_0 \text{ est rejetée} \mid g = \gamma_1 + cn^{-1/2}\gamma_2) = 1. \quad (9.1)$$

On peut remarquer que l'on ne teste pas directement sur \mathcal{Q} tout entier, mais sur \mathcal{Q}_δ . Ceci est dû au critère global de SIM donné au paragraphe 5.3.

DÉMONSTRATION. Les lois limites de $n^{1/2}(\widehat{\varphi} - \varphi_0)$ et $n\widehat{W}(\widehat{\varphi})$, obtenues dans le chapitre 8 sont complètement déterminées par f et ∇g . Ces démonstrations demandent que les

dérivées $2k(d)$ de f soient uniformément bornées et que ∇g soit continue. En utilisant des arguments similaires et en tenant compte des hypothèses de convergence sur les estimateurs de la densité f et de la régression g , les lois bootstraps limites de $n^{1/2}(\widehat{\varphi}^* - \widehat{\varphi})$ et $n\widehat{W}^*(\widehat{\varphi}^*)$ sont les lois de Ω et $\overline{W}(\Omega)$, respectivement, définies dans le chapitre 8, en remplaçant g_1 par γ_1 .

• *Partie (I)*

Par définition des seuils critiques t_π et \widehat{t}_π , on a

$$\mathbb{P}\left(\overline{W}(\Omega) < t_\pi\right) = 1 - \pi = \mathbb{P}\left(\widehat{W}^*(\widehat{\varphi}^*) < \widehat{t}_\pi \mid \mathcal{Z}\right).$$

La convergence en loi de la statistique bootstrap donne

$$\mathbb{P}\left(n\widehat{W}^*(\widehat{\varphi}^*) < z \mid \mathcal{Z}\right) = \mathbb{P}\left(\overline{W}(\Omega) < z\right) + o_{\mathbb{P}}(1),$$

cette égalité est encore vraie pour $z = n\widehat{t}_\pi$, en utilisant l'inverse de la fonction de répartition de $\overline{W}(\Omega)$ et sa régularité on obtient

$$n\widehat{t}_\pi = t_\pi + o_{\mathbb{P}}(1).$$

On peut maintenant en déduire le niveau asymptotique du test. Soit $\varepsilon > 0$, en utilisant la convergence en loi de $n\widehat{W}(\widehat{\varphi})$ et en posant $A_{n,\varepsilon} = \{|nt_\pi - t_\pi| < \varepsilon\}$, on a

$$\begin{aligned} \mathbb{P}\left(\widehat{W}(\widehat{\varphi}) > \widehat{t}_\pi\right) &= \mathbb{P}\left(\overline{W}(\Omega) > n\widehat{t}_\pi\right) + o(1) \\ &= \mathbb{P}\left(\overline{W}(\Omega) > t_\pi - \varepsilon \cap A_{n,\varepsilon}\right) + \mathbb{P}\left(\overline{W}(\Omega) > n\widehat{t}_\pi \cap A_{n,\varepsilon}^c\right) + o(1). \end{aligned}$$

Or $\mathbb{P}(A_{n,\varepsilon}) \xrightarrow[n \rightarrow \infty]{} 1$ et $\mathbb{P}(\overline{W}(\Omega) > t_\pi - \varepsilon) \xrightarrow[\varepsilon \rightarrow 0]{} \pi$. Le deuxième terme converge donc vers 0 et le premier vers π . Finalement

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\widehat{W}(\widehat{\varphi}) > \widehat{t}_\pi\right) = \pi.$$

• *Partie (II)*

Sous l'hypothèse où g est fixée et n'est pas un SIM sur \mathcal{Q}_δ , on rééchantillonne quand même suivant un modèle SIM d'axe $\widehat{\varphi} = \underset{\varphi \in \Theta}{\operatorname{argmin}} \widehat{W}(\varphi)$, qui tend vers un $\varphi_0 = \underset{\varphi \in \Theta}{\operatorname{argmin}} W(\varphi)$. Ainsi

on a encore la convergence en loi de $n\widehat{W}^*(\widehat{\varphi}^*)$ vers $\overline{W}(\Omega)$. La relation entre les seuils critiques théoriques et bootstrap est donc toujours valable, i.e. $\widehat{t}_\pi = o_{\mathbb{P}}(1)$. On s'intéresse maintenant à la convergence de $\widehat{W}(\widehat{\varphi})$: en reprenant la démonstration du chapitre 8, on obtient la convergence en probabilité de $\widehat{W}(\widehat{\varphi})$ vers $W(\varphi_0)$ et $W(\varphi_0) > 0$. En effet les étapes **2-**, **3-** et **6-** changent et font intervenir les quantités $\Delta(s, u, \mathbf{x}, \theta\varphi_0) \neq 0$ et $W(\varphi_0) > 0$. Soit $\eta > 0$ alors $\mathbb{P}(\widehat{t}_\pi > \eta) \xrightarrow[n \rightarrow \infty]{} 0$.

$$\begin{aligned} \mathbb{P}\left(\widehat{W}(\widehat{\varphi}) > \widehat{t}_\pi\right) &= \mathbb{P}\left(W(\varphi_0) > \widehat{t}_\pi\right) + o(1), \\ &= \mathbb{P}\left(\{W(\varphi_0) > \eta\} \cap \{\widehat{t}_\pi > \eta\}\right) + \mathbb{P}\left(\{W(\varphi_0) > \widehat{t}_\pi\} \cap \{\widehat{t}_\pi \leq \eta\}\right) + o(1). \end{aligned}$$

Les remarques précédentes impliquent que le deuxième terme tend vers 0 et le premier tend vers 1 lorsque l'on fait tendre tout d'abord n vers $+\infty$, puis η vers 0. Finalement

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\widehat{W}(\widehat{\varphi}) > \widehat{t}_\pi\right) = 1.$$

• **Partie (III)**

On suppose que le modèle qui engendre les données est de la forme $g = \gamma_1 + cn^{-1/2}\gamma_2$ sur \mathcal{Q} , avec les hypothèses (i),(ii) et (iii). Cette démonstration fait intervenir des résultats intermédiaires de convergence, dans un souci de clarté nous avons préféré les renvoyer à la fin de ce chapitre dans le lemme 9.2.1.

Le lemme 9.2.1 et les résultats internes à sa démonstration, donnent la convergence en probabilité de $\widehat{W}(\varphi)$ vers $W(\varphi)$, ce dernier ne faisant intervenir que la partie SIM du modèle : $\gamma_1(\varphi_0 \cdot \mathbf{x})$. On a aussi $\widehat{\varphi} \xrightarrow{\mathbb{P}} \varphi_0$. Ainsi lorsqu'on rééchantillonne suivant le modèle à direction révélatrice $\widehat{\varphi}$, on obtient encore

$$\mathbb{P}\left(n\widehat{W}^*(\widehat{\varphi}^*) < z | \mathcal{L}\right) \rightarrow \mathbb{P}\left(\overline{W}(\Omega) < z\right), \quad \text{pour tout } z,$$

où \overline{W} est obtenu par la formule (6.4) en remplaçant g_1 par γ_1 . Ainsi on a $n\widehat{t}_\pi \xrightarrow{\mathbb{P}} t_\pi$, où t_π est le $(1 - \pi)$ -quantile de $\overline{W}(\Omega)$.

On veut montrer que

$$\lim_{|c| \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}\left(\widehat{W}(\widehat{\varphi}) > \widehat{t}_\pi\right) = 1.$$

Comme $n\widehat{t}_\pi \xrightarrow{\mathbb{P}} t_\pi$, si on montre que

$$\lim_{|c| \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}\left(n\widehat{W}(\widehat{\varphi}) > C'\right) = 1, \quad \text{pour tout } C' > 0,$$

alors on a le résultat. En utilisant la convergence en loi de $n\widehat{W}(\widehat{\varphi})$, cf. lemme 9.2.1, il suffit de montrer que

$$\lim_{|c| \rightarrow \infty} \mathbb{P}\left(\overline{W}_1(\Omega_{1,c}, c) > C'\right) = 1. \quad (9.2)$$

Dans l'expression de $\overline{W}_1(\omega, c)$ interviennent le processus limite gaussien $\zeta(\mathcal{S}_0)$ et la quantité $\omega^T \beta + cI_{\mathcal{S}}(\gamma_2)$. On peut écrire

$$\begin{aligned} \overline{W}_1^{1/2}(\omega, c) \geq & \left| \left(\int_{\mathcal{U}(\varphi_0)} du \int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} \left(\zeta(\mathcal{S}_0) - M_\zeta(u, \varphi_0) \right)^2 ds d\mathbf{x} \right)^{1/2} \right. \\ & \left. - \left(\int_{\mathcal{U}(\varphi_0)} du \int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} \left(\omega^T \beta + c(I_{\mathcal{S}_0}(\gamma_2) - V_{\gamma_2}(u, \varphi_0)) \right)^2 ds d\mathbf{x} \right)^{1/2} \right| \end{aligned}$$

On note $W_2(\omega, c)$ le deuxième terme de droite, c'est-à-dire

$$W_2(\omega, c) = \int_{\mathcal{U}(\varphi_0)} du \int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} \left(\omega^{\text{T}} \beta(s, u, \mathbf{x}, \theta_0, \varphi_0) + c(I_{\mathcal{S}_0}(\gamma_2) - V_{\gamma_2}(u, \varphi_0)) \right)^2 ds d\mathbf{x}.$$

Or

$$\sup_{u \in \mathcal{U}(\varphi_0)} \sup_{\mathbf{v} \in \mathcal{V}(u, \varphi_0)} \sup_{\theta_0 \in \Theta(\varphi_0)} \sup_{(s, \mathbf{x}) \in \mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} |\zeta\{\mathcal{S}(s, \mathcal{R}_{\theta_0 \mathbf{x}})\}| = \mathcal{O}_{\mathbb{P}}(1),$$

On remarque donc que dans la minoration de \overline{W}_1 le premier terme de droite est borné en probabilité. Finalement si $W_2(\omega, c)^{1/2}$ vérifie

$$\lim_{|c| \rightarrow \infty} \inf_{\omega} W_2(\omega, c) = \infty, \quad (9.3)$$

on obtient le résultat (9.2).

La convergence (9.3) a lieu puisque γ_2 n'est pas un SIM d'axe φ_0 , donc indépendamment de ω , $I_{\mathcal{S}_0}(\gamma_2) - V_{\gamma_2}(u, \varphi_0) \neq 0$. ■

Lemme 9.2.1. *Sous les hypothèses du théorème précédent et de la partie (III)*

$$\widehat{\varphi} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \varphi_0$$

et

$$n \widehat{W}(\widehat{\varphi}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \overline{W}_1(\Omega_{1,c}, c),$$

où

$$\begin{aligned} \overline{W}_1(\omega, c) = & \int_{\mathcal{U}(\varphi_0)} du \int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} \left(\xi(s, u, \mathbf{x}, \theta_0, \varphi_0) \right. \\ & \left. + \omega^{\text{T}} \beta(s, u, \mathbf{x}, \theta_0, \varphi_0) + cI_{\mathcal{S}_0}(\gamma_2) - M_{\xi}(u, \varphi_0) - cV_{\gamma_2}(u, \varphi_0) \right)^2 ds d\mathbf{x} \end{aligned}$$

et $\Omega_{1,c} = \operatorname{argmin}_{\omega \in \mathbb{R}^d} \overline{W}_1(\omega, c)$.

DÉMONSTRATION. On reprend la démonstration du chapitre 8. Dans cette partie, pour plus de clarté, on notera $V_g(u, \varphi)$ à la place de $V(u, \varphi)$ pour bien préciser la dépendance en g de cette quantité.

Étape 1- Montrons que $\sup_c |\widehat{I}_{\mathcal{S}}(g) - I_{\mathcal{S}}(\gamma_1)| = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$: en reprenant les mêmes notations, on a ici

$$\widehat{I}_{\mathcal{S}}(g) = \widehat{I}_{\mathcal{S}}(\gamma_1) + cn^{-1/2} B_{\mathcal{S}}(\gamma_2)$$

$$\begin{aligned} \sup_{\mathcal{C}} |\widehat{I}_{\mathcal{S}}(g) - I_{\mathcal{S}}(\gamma_1)| &\leq \sup_{\mathcal{C}} |\widehat{I}_{\mathcal{S}}(\gamma_1) - I_{\mathcal{S}}(\gamma_1)| + cn^{-1/2} |B_{\mathcal{S}}(\gamma_2)| \\ &= \mathcal{O}_{\mathbb{P}}(n^{-1/2}), \end{aligned}$$

car la convergence de $\widehat{I}_{\mathcal{S}}(\gamma_1)$ est toujours valable et que la convergence de $B_{\mathcal{S}}(\gamma_2)$ se démontre de la même façon que dans la partie 7.2. De plus on a

$$\sup_{u, \varphi} \left| \widehat{V}_g(u, \varphi) - V_{\gamma_1}(u, \varphi) \right| \leq \sup_u \sup_{\mathcal{C}} |\widehat{I}_{\mathcal{S}}(g) - I_{\mathcal{S}}(\gamma_1)| = \mathcal{O}_{\mathbb{P}}(n^{-1/2}). \quad (9.4)$$

Étape 2-, **Étape 3-** et **Étape 4-** restent inchangées étant donné que notre définition de Δ ne dépend que de la partie SIM du modèle. Cette quantité s'annule donc en φ_0 . On obtient la convergence suivante, uniformément en $\varphi \in \Theta$

$$\widehat{W}(\varphi) \xrightarrow{\mathbb{P}} W(\varphi),$$

où

$$W(\varphi) = \int_{\mathcal{U}(\varphi)} du \int_{\mathcal{V}(u, \varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_{\theta}(\mathbf{v})\}} \{I_{\mathcal{S}(s, \mathcal{R}_{\theta\mathbf{x}})}(\gamma_1) - V_{\gamma_1}(u, \varphi)\}^2 ds d\mathbf{x}.$$

De plus φ_0 étant toujours atteint de façon quadratique, les **Étapes 5-** et **Étapes 6-** restent inchangées, on obtient donc la convergence

$$\widehat{\varphi} \xrightarrow{\mathbb{P}} \varphi_0.$$

Étape 7- Développement de $\widehat{\Delta}(s, u, \mathbf{x}, \theta, \varphi)$ en φ_0 . On note

$$\widehat{V}_{B, \gamma_2}(u, \varphi) = \frac{\int_{\mathcal{V}(u, \varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_{\theta}(\mathbf{v})\}} B_{\mathcal{S}(s, \mathcal{R}_{\theta\mathbf{x}})}(\gamma_2) ds d\mathbf{x}}{\int_{\mathcal{V}(u, \varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_{\theta}(\mathbf{v})\}} ds d\mathbf{x}}.$$

On peut alors écrire

$$n^{1/2} \widehat{\Delta} = n^{1/2} \left[\widehat{I}_{\mathcal{S}}(\gamma_1) - \widehat{V}_{\gamma_1}(u, \varphi) \right] + c \left[\|\mathcal{S}\|^{-1} B_{\mathcal{S}}(\gamma_2) - \widehat{V}_{B, \gamma_2}(u, \varphi) \right]$$

En utilisant la partie 7.2, on obtient le développement de $B_{\mathcal{S}}(\gamma_2)$

$$\begin{aligned} B_{\mathcal{S}}(\gamma_2) &= (B_{\mathcal{S}}(\gamma_2) - I_{\mathcal{S}}(\gamma_2)) + I_{\mathcal{S}}(\gamma_2) \\ &= I_{\mathcal{S}}(\gamma_2) + \mathcal{O}_{\mathbb{P}}(n^{-1/2}) \end{aligned}$$

On peut donc écrire

$$\begin{aligned}
& n^{1/2} \widehat{\Delta}(s, u, \mathbf{x}, \theta, \varphi) - n^{1/2} \Delta(s, u, \mathbf{x}, \theta, \varphi) \\
&= n^{1/2} \left(\widehat{I}_{\mathcal{S}}(\gamma_1) - I_{\mathcal{S}}(\gamma_1) \right) - n^{1/2} \left(\widehat{V}_{\gamma_1}(u, \varphi) - V_{\gamma_1}(u, \varphi) \right) \\
&\quad + c \left(B_{\mathcal{S}}(\gamma_2) - \widehat{V}_{B, \gamma_2}(u, \varphi) \right) \\
&= \|\mathcal{S}\|_d^{-1} \zeta_n \{\mathcal{S}\} - \frac{\int_{\mathcal{V}(u, \varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_{\theta}(\mathbf{v})\}} \|\mathcal{S}\|_d^{-1} \zeta_n \{\mathcal{S}\} ds d\mathbf{x}}{\int_{\mathcal{V}(u, \varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}\{\alpha_{\theta}(\mathbf{v})\}} ds d\mathbf{x}} \\
&\quad + c \left(B_{\mathcal{S}}(\gamma_2) - \widehat{V}_{B, \gamma_2}(u, \varphi) \right).
\end{aligned}$$

Finalement pour φ tel que $\|\varphi_0 - \varphi\| \leq Cn^{-1/2}$, en utilisant les mêmes arguments que dans le chapitre 8

$$\begin{aligned}
& n^{1/2} \widehat{\Delta}(s, u, \mathbf{x}, \theta, \varphi) \\
&= n^{1/2} \Delta(s, u, \mathbf{x}, \theta, \varphi) + \|\mathcal{S}_0\|_d^{-1} \zeta_n \{\mathcal{S}_0\} + cI_{\mathcal{S}}(\gamma_2) \\
&\quad - \frac{\int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} \|\mathcal{S}_0\|_d^{-1} \zeta_n \{\mathcal{S}_0\} + cI_{\mathcal{S}}(\gamma_2) ds d\mathbf{x}}{\int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} ds d\mathbf{x}} + \mathcal{O}_{\mathbb{P}}(1).
\end{aligned}$$

Grâce au développement de Δ en φ_0 donné en (6.2) on a uniformément sur l'ensemble $\{\varphi / \|\varphi - \varphi_0\| \leq Cn^{-1/2}\}$

$$\begin{aligned}
& n^{1/2} \widehat{\Delta}(s, u, \mathbf{x}, \theta, \varphi) = n^{1/2} (\varphi - \varphi_0)^{\top} \beta(s, u, \mathbf{x}, \theta_0, \varphi_0) + \|\mathcal{S}_0\|_d^{-1} \zeta_n(\mathcal{S}_0) + cI_{\mathcal{S}}(\gamma_2) \\
&\quad - \frac{\int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} \|\mathcal{S}_0\|_d^{-1} \zeta_n(\mathcal{S}_0) + cI_{\mathcal{S}}(\gamma_2) ds d\mathbf{x}}{\int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} ds d\mathbf{x}} + \mathcal{O}_{\mathbb{P}}(1).
\end{aligned}$$

On voit alors que dans la variable aléatoire limite la partie $\zeta(\mathcal{S})$ sera remplacée par $\zeta(\mathcal{S}) + cI_{\mathcal{S}}(\gamma_2)$. On définit donc

$$\begin{aligned}
\overline{W}_1(\omega, c) &= \int_{\mathcal{U}(\varphi_0)} du \int_{\mathcal{V}(u, \varphi_0)} d\mathbf{v} \int_{\Theta(\varphi_0)} d\theta_0 \int_{\mathcal{T}\{\alpha_{\theta_0}(\mathbf{v})\}} \left(\xi(s, u, \mathbf{x}, \theta_0, \varphi_0) \right. \\
&\quad \left. + \omega^{\top} \beta(s, u, \mathbf{x}, \theta_0, \varphi_0) + cI_{\mathcal{S}_0}(\gamma_2) - M_{\xi}(u, \varphi_0) - cV_{\gamma_2}(u, \varphi_0) \right)^2 ds d\mathbf{x}
\end{aligned}$$

et $\Omega_{1,c} = \operatorname{argmin}_{\omega \in \mathbb{R}^d} \overline{W}_1(\omega, c)$.

On peut alors reprendre la démonstration des étapes **8-** et **9-**. On conclut que

$$n\widehat{W}(\widehat{\varphi}) \xrightarrow{\mathcal{L}} \overline{W}_1(\Omega_{1,c}, c).$$

■

Annexe 1 : Étude de la continuité des intégrales

Cette partie est consacrée à l'étude des propriétés de régularité des fonctions

$$\begin{aligned}(\theta, \mathbf{v}, s, \mathbf{x}) &\mapsto I_{\mathcal{S}(s, \mathcal{R}_{\theta\mathbf{x}})}, \\(u, \varphi) &\mapsto V(u, \varphi), \\ \varphi &\mapsto W(\varphi).\end{aligned}$$

Dans un premier temps nous allons faire quelques remarques sur les cylindres et les domaines de variation des paramètres $(\theta, \mathbf{v}, s, \mathbf{x}, u, \varphi)$.

10.1 Les cylindres et leur paramétrisation

Soit $\mathcal{Q} \subset \mathbb{R}^d$ un ensemble convexe, compact, dont le bord $\partial\mathcal{Q}$ est de classe \mathcal{C}^1 . Pour $\delta > 0$ quelconque fixé, on définit

$$\mathcal{Q}_\delta = \{x \in \mathcal{Q}, d(x, \partial\mathcal{Q}) > \delta\}.$$

La paramétrisation des cylindres qui recouvrent \mathcal{Q}_δ est la suivante : fixons $\varphi \in \Theta$, une direction privilégiée. Soit $\theta \in \Theta(\varphi)$, i.e. $\theta \in \Theta$ et θ est un vecteur orthogonal à φ . Alors les cylindres sont choisis tels que

- Le rayon de leur base est $r > 0$. Ce rayon est le même pour tous les cylindres et est choisi tel que $r < \delta$.
- leur largeur s varie entre s_1 et s_2 , où $0 < s_1 < s_2 < 2\sqrt{\delta^2 - r^2}$.
- Le centre des cylindres \mathbf{x} est dans \mathcal{Q}_δ .

Un tel cylindre est noté $\mathcal{S}(s, \mathcal{R}_{\theta\mathbf{x}})$.

L'idée est de considérer un repère de base ayant son origine sur φ et on note $u = \varphi \cdot \mathbf{x}$, la cote sur l'axe φ . De plus on localise alors les centres des cylindres \mathbf{x} sur les segments de cote u , de direction θ et d'extrémité \mathbf{v} , cf. figure 5.5, i.e. l'ensemble des $\alpha_\theta(\mathbf{v})$.

On peut alors démontrer que les ensembles de variation des paramètres sont compacts.

Propriété 10.1.1. – $\Theta(\varphi_0) = \{\theta, \theta \perp \varphi_0 \text{ et } \|\theta\| = 1\}$ est compact.

– Soit $\varphi, \theta \in \Theta(\varphi)$ et \mathbf{v} fixés. $\mathcal{T}(\alpha_\theta(\mathbf{v})) = \{(s, \mathbf{x}) / s \in [s_1, s_2] \text{ et } \mathbf{x} \in \alpha_\theta(\mathbf{v})\}$ est un compact.

– Soit φ et u fixés, $\mathcal{V}(u, \varphi)$ est un compact.

– Soit φ , $\mathcal{U}(\varphi)$ est compact.

– $\mathcal{V}(\varphi) = \bigcup_{u \in \mathcal{U}(\varphi)} \mathcal{V}(u, \varphi)$ est compact.

DÉMONSTRATION.

• Pour φ_0 fixé, $\Theta(\varphi_0) \subset \Theta$ compact.

Soit $\{\theta_n\}_n \in \Theta(\varphi_0) \subset \Theta$ telle que $\theta_n \rightarrow \theta$. Montrons que $\theta \in \Theta(\varphi_0)$:

$\|\theta\| = 1$, car Θ est compact. Pour tout n , $\theta_n \cdot \varphi_0 = 0$, donc par continuité du produit scalaire, d'où $\theta \cdot \varphi_0 = 0$. Donc $\theta \in \Theta(\varphi_0)$, qui est donc un sous-ensemble fermé de Θ compact de \mathbb{R}^d . Finalement $\Theta(\varphi_0)$ est compact

• $\mathcal{T}(\alpha_\theta(\mathbf{v})) = [s_1, s_2] \times \alpha_\theta(\mathbf{v})$, or $\alpha_\theta(\mathbf{v})$ est un segment donc un compact. $\mathcal{T}(\alpha_\theta(\mathbf{v}))$ est un compact comme produit direct de deux compacts.

• $\mathcal{V}(u, \varphi) = \{\mathbf{v} / \varphi \cdot \mathbf{v} = u\}$. Soit \mathbf{v}_n une suite de $\mathcal{V}(u, \varphi) \subset \mathcal{Q}$. Or \mathcal{Q} est compact donc il existe une sous-suite $\mathbf{v}_{\phi(n)} \rightarrow \mathbf{v} \in \mathcal{Q}$ et pour tout n , $\mathbf{v}_{\phi(n)} \cdot \varphi = u$. Donc par continuité du produit scalaire $\mathbf{v} \cdot \varphi = u$, c'est-à-dire que $\mathbf{v} \in \mathcal{V}(u, \varphi)$. Finalement $\mathcal{V}(u, \varphi)$ est compact.

• $\mathcal{U}(\varphi)$ est la trace du compact \mathcal{Q} sur l'axe φ , c'est donc un segment.

• $\mathcal{V}(\varphi)$ représente un sous-ensemble fermé du bord de \mathcal{Q}_δ qui est compact, c'est donc un ensemble compact de \mathbb{R}^d . ■

10.2 Continuité

Régularité de $I_{\mathcal{S}}$

La fonction I est définie comme suit :

$$I : \mathcal{D}(u, \varphi) \rightarrow \mathbb{R} \\ (\theta, \mathbf{v}, (s, \mathbf{x})) \mapsto I_{\mathcal{S}(s, \mathcal{R}_{\theta\mathbf{x}})}(g),$$

avec $\mathcal{D}(u, \varphi) = \{(\theta, \mathbf{v}, (s, \mathbf{x})) / \theta \in \Theta(\varphi), \mathbf{v} \in \mathcal{V}(u, \varphi) \text{ et } (s, \mathbf{x}) \in \mathcal{T}(\alpha_\theta(\mathbf{v}))\}$.

Chaque cylindre \mathcal{S} peut être obtenu à partir d'un cylindre particulier \mathcal{S}_0 :

Soit \mathcal{S}_0 le cylindre de centre O , d'axe Ox_1 , de rayon r et de longueur 1. Si on note

– D_s la dilatation d'axe Ox_1 et de rapport $s \in [s_1, s_2]$;

– R_θ la rotation de centre O , qui transforme l'axe Ox_1 en θ ;

– $t_{\mathbf{x}}$ la translation, qui envoie O au centre \mathbf{x} de \mathcal{S} ;

et on pose $T_{r,\theta,s} = t_{\mathbf{x}} \circ R_{\theta} \circ D_s$, alors $T_{r,\theta,s}(\mathcal{S}_0) = \mathcal{S}$.

Le Jacobien de $T_{r,\theta,s}$ vaut s d'où, en faisant le changement de variable adéquate,

$$I_{\mathcal{S}(s,\mathcal{R}_{\theta\mathbf{x}})}(g) = s \int_{\mathcal{C}_0} g(T_{r,\theta,s}(y)) dy.$$

Donc la fonction I est de classe \mathcal{C}^1 , car g l'est.

Continuité de V

Par définition,

$$V : \mathcal{D} \rightarrow \mathbb{R} \\ (u, \varphi) \mapsto V(u, \varphi) = \int_{\mathcal{V}(u,\varphi)} d\mathbf{v} \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}_{\theta}(\mathbf{v})} I_{\mathcal{S}(s,\mathcal{R}_{\theta\mathbf{x}})}(g) ds d\mathbf{x},$$

où $\mathcal{D} = \{(u, \varphi) / \varphi \in \Theta \text{ et } u \in \mathcal{U}(\varphi)\}$. On a montré que la fonction $(\theta, \mathbf{v}, s, \mathbf{x}) \mapsto I_{\mathcal{S}(s,\mathcal{R}_{\theta\mathbf{x}})}(g)$ est continue.

Remarque 2. pour montrer que $t \mapsto \int_{g(t)}^1 f(x) dx$ est continue en t_0 , on étudie la quantité suivante

$$\Delta = \left| \int_{g(t)}^1 f(x) dx - \int_{g(t_0)}^1 f(x) dx \right| \leq \int_{g(t)}^{g(t_0)} |f(x)| dx.$$

Si g est continue en t_0 et f bornée par M , alors

$$\Delta \leq M|g(t) - g(t_0)| < \varepsilon.$$

Donc ici on va montrer que les domaines sur lesquels on intègre sont continus par rapport à leur paramètre : dans notre cas $\Delta \leq \lambda^d(K \cup K_0 \setminus K \cap K_0)$, où λ est la mesure de Lebesgue. Donc il faut montrer que la mesure de Lebesgue de cet ensemble tend vers 0 quand $K \rightarrow K_0$. On utilise en fait souvent une majoration de la mesure de Lebesgue

$$\lambda^d(K \cup K_0 \setminus K \cap K_0) \leq v_d \sup_{x \in K} \inf_{y \in K_0} d(x, y) = v_d \sup_{x \in K} d(x, K_0),$$

avec v_d le volume de la sphère en dimension d . ■

À u, φ fixés :

Pour $(\theta, \mathbf{v}) \mapsto \mathcal{T}_{\theta}(\mathbf{v})$ à θ fixé, on prouve la continuité par rapport à \mathbf{v} , cf. figure 10.1. On veut montrer que $\sup_{\mathbf{x} \in \alpha_{\theta}(\mathbf{v})} d(\mathbf{x}, \alpha_{\theta}(\mathbf{v}_0)) \rightarrow 0$ quand $\mathbf{v} \rightarrow \mathbf{v}_0$. On remarque que $\mathbf{v} \rightarrow \mathbf{v}_0$ équivaut à dire que $h \rightarrow 0$. La borne supérieure est obtenue sur le bord à une des extrémités sinon il est égal à h . On considère alors le repère $(M_0(\mathbf{v}_0), \theta', \theta)$, où $\theta' \perp \theta$ et dans le plan

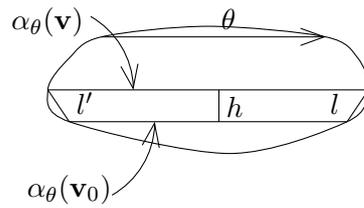


FIG. 10.1 – Continuité en $\alpha_\theta(\mathbf{v})$ par rapport à \mathbf{v} .

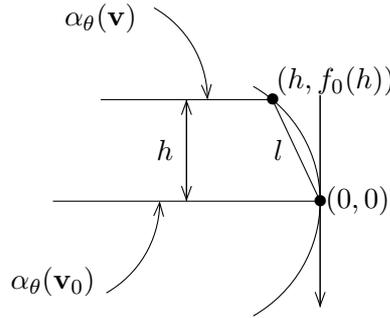


FIG. 10.2 – Continuité de $\alpha_\theta(\mathbf{v})$ par rapport à \mathbf{v}

orthogonal à φ contenant θ , cf. figure 10.2. Dans ce repère M_0 a pour coordonnées $(0, 0)$, on applique le théorème des fonction implicites pour trouver l'équation du bord : $y = f_0(\mathbf{x})$, avec $f_0 \in \mathcal{C}^1$. De plus f'_0 est quelconque et $l = \sqrt{h^2 + f_0(h)^2}$. Or le développement de Taylor entre 0 et h de f_0 est

$$f_0(h) = 0 + h \int_0^1 (1+t)f'_0(th) dt$$

Ainsi $l = h\sqrt{1 + (\int_0^1 (1+t)f'_0(th) dt)^2} = \mathcal{O}(h)$. Finalement $l \xrightarrow{h \rightarrow 0} 0$.

On peut faire de même en considérant l'autre extrémité du segment $\alpha_\theta(\mathbf{v}_0)$.

À \mathbf{v} fixé, on montre que $\theta \mapsto \alpha_\theta(\mathbf{v})$ est continue. Pour cela on étudie la distance entre

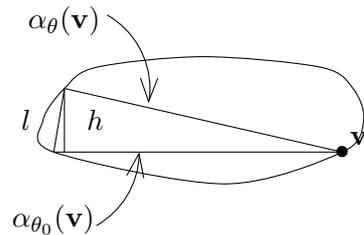


FIG. 10.3 – Continuité en $\alpha_\theta(\mathbf{v})$ par rapport à θ .

$\alpha_\theta(\mathbf{v})$ et $\alpha_{\theta_0}(\mathbf{v})$ quand $\theta \rightarrow \theta_0$. On remarque que le $\sup_{\mathbf{x} \in \alpha_\theta(\mathbf{v})} d(\mathbf{x}, \alpha_{\theta_0}(\mathbf{v}))$ est l , on fait

alors un changement de repère dans lequel la courbe est le graphe d'une fonction que l'on note encore f_0 . On obtient aussi $l = \mathcal{O}(h)$. Or ici $h = \|\alpha_\theta(\mathbf{v})\| \sin(\theta, \theta_0) \leq \tilde{\delta} \sin(\theta, \theta_0)$, où $\tilde{\delta} = \text{diam}(\mathcal{Q})$.

Finalement $l = \mathcal{O}(\theta, \theta_0)$, i.e. $l \xrightarrow{\theta \rightarrow \theta_0} 0$.

Il reste alors à montrer la continuité quand $(\mathbf{v}, \theta) \rightarrow (\mathbf{v}_0, \theta_0)$ de l'application

$$(\theta, \mathbf{v}) \mapsto \int_{\mathcal{T}_\theta(\mathbf{v})} I_S(g) ds d\mathbf{x}.$$

Pour cela il suffit de combiner les deux figures 10.2 et 10.3. On peut alors majorer

$$\sup_{\mathbf{x} \in \alpha_\theta(\mathbf{v})} d(\mathbf{x}, \alpha_{\theta_0}(\mathbf{v}_0)) \leq \sup_{\mathbf{x} \in \alpha_\theta(\mathbf{v})} d(\mathbf{x}, \alpha_{\theta_0}(\mathbf{v})) + \sup_{\mathbf{x} \in \alpha_{\theta_0}(\mathbf{v})} d(\mathbf{x}, \alpha_{\theta_0}(\mathbf{v}_0)),$$

les deux quantités de droite sont celles étudiées précédemment et tendent donc vers 0, quand $(\mathbf{v}, \theta) \rightarrow (\mathbf{v}_0, \theta_0)$.

On s'intéresse alors à la continuité de $(\varphi, \mathbf{v}) \mapsto \int_{\Theta(\varphi)} d\theta \int_{\mathcal{T}_\theta(\mathbf{v})} I_S(g) ds d\mathbf{x}$. Par définition de $\Theta(\varphi)$, i.e. $\Theta(\varphi) = \{\theta \in \Theta / \theta \cdot \varphi = 0\}$, cet ensemble est continu en φ par continuité du produit scalaire.

On considère $\Theta(\varphi)$ et $\Theta(\varphi_0)$. Soit $\mathbf{x} \in \Theta(\varphi_0)$ et $d(\mathbf{x}, \Theta(\varphi)) = |\mathbf{x} \cdot \varphi|$, or $\mathbf{x} \cdot \varphi_0 = 0$ et $\|\mathbf{x}\| = 1$.

$$\sup_{\mathbf{x} \in \Theta(\varphi_0)} |\mathbf{x} \cdot (\varphi_0 - \varphi)| < |\varphi - \varphi_0| < \varepsilon.$$

Cette dernière égalité nous donne la continuité voulue.

Il nous reste alors à montrer la continuité de $\mathcal{V}(u, \varphi)$. Or $\mathcal{V}(u, \varphi)$ est l'intersection du bord de $f(\mathbf{x}) = 0$ et du plan $\varphi \cdot \mathbf{x} = u$. On procède de même que pour les ensembles précédents :
 – à u fixé, faire varier φ : cf. figure 10.4

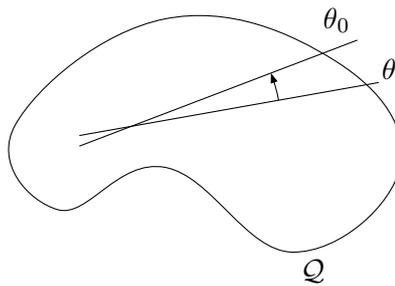


FIG. 10.4 – Continuité de $\mathcal{V}(u, \varphi)$ en dimensions 3.

– pour un même φ faire varier u : cf. figure 10.5

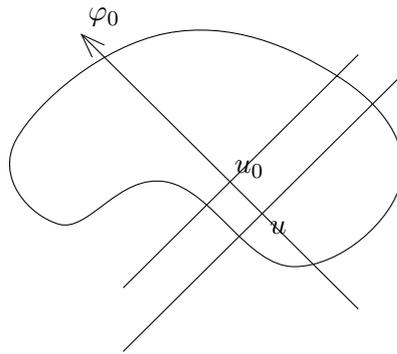


FIG. 10.5 – Continuité de $\mathcal{V}(u, \varphi)$ en dimensions 3.

La plus grande distance entre deux ensembles est encore atteinte au bord. Pour conclure il faut faire varier (u, φ) . On obtient alors la continuité de $(u, \varphi) \mapsto V(u, \varphi)$.

L'Étude de la continuité de $W(\varphi)$ se fait de même. il ne reste plus qu'à obtenir la continuité de $\mathcal{U}(\varphi)$ par rapport à φ . Or cet ensemble est la projection de \mathcal{Q} sur l'axe φ , il est donc continu en φ .

Annexe 2 : Rappels sur les U -statistiques

11.1 Introduction

Dans ce paragraphe nous rappelons les définitions de base concernant les U -statistiques, les U -processus et les classes de fonctions euclidiennes, ainsi que certains résultats obtenus par *Sherman (1994)*. En effet ces processus et classes de fonctions interviennent à plusieurs reprises dans le chapitre 7.

Soit Z_1, \dots, Z_n des observations indépendantes d'une loi \mathbb{P} sur un ensemble \mathcal{X} . Soit k un entier et \mathcal{F} une classe de fonctions sur \mathcal{X}^k , on définit :

$$U_n^k f = (n)_k^{-1} \sum_{\mathbf{i}_k} f(Z_{i_1}, \dots, Z_{i_n})$$

où $(n)_k = n(n-1) \cdots (n-k+1)$ et $\mathbf{i}_k = (i_1, \dots, i_n)$ ensemble de k indices parmi les $(n)_k$ k -uplets ordonnés d'entiers distincts de $\{1, \dots, n\}$. Pour $k = 1$, $U_n^1 = \mathbb{P}_n$ est la mesure empirique.

Définition 11.1.1. $U_n^k f$ est appelée U -statistique d'ordre k . $\{U_n^k f, f \in \mathcal{F}\}$ est appelé U -processus d'ordre k indexé par \mathcal{F} .

Si de plus, pour tout $f \in \mathcal{F}$,

$$\mathbb{E} f(s_1, \dots, s_{i-1}, Z_i, s_{i+1}, \dots, s_k) = 0 \text{ pour } i = 1, \dots, k,$$

alors \mathcal{F} est appelée classe de fonctions de \mathcal{X}^k \mathbb{P} -dégénérée et $U_n^k f$ (resp. $\{U_n^k f, f \in \mathcal{F}\}$) est appelée U -statistique (resp. U -processus) d'ordre k \mathbb{P} -dégénérée.

Définition 11.1.2. On dit qu'une fonction positive F définie sur \mathcal{X}^k est une enveloppe de \mathcal{F} si

$$\sup_{f \in \mathcal{F}} |f(\cdot)| \leq F(\cdot)$$

Définition 11.1.3. Soit \mathcal{F} une classe de fonctions sur \mathcal{X} à valeurs réelles. On dit que \mathcal{F} est **euclidienne** pour l'enveloppe F s'il existe A et V deux constantes positives avec la propriété suivante : si μ est une mesure pour laquelle $\mu F^2 < +\infty$, alors

$$D(x, d\mu, \mathcal{F}) \leq Ax^{-V}, \quad 0 < x \leq 1,$$

où

- pour $\varepsilon > 0$, $D(\varepsilon, d, T)$ est le plus grand nombre D pour lequel $\exists m_1, \dots, m_D$ des points de T tels que $d(m_i, m_j) > \varepsilon$, pour $i \neq j$;
- pour $f, g \in \mathcal{F}$, $d\mu(f, g) = \left(\frac{\mu|f-g|^2}{\mu F^2} \right)^{1/2}$.

Interprétation :

Ce nombre $D(\varepsilon, d, T)$ nous donne pour le pseudo-espace métrique (T, d) la grosseur de T selon la métrique d .

On peut remarquer que

- A et V sont des constantes indépendantes de μ ;
- on dit que \mathcal{F} est euclidienne (A, V) pour l'enveloppe F .

On trouvera des exemples de telles familles dans les ouvrages de *Pollard (1984)* et *Devroye et Lugosi (2000)*.

Les premiers exemples de telles familles sont les familles d'indicatrices sur des ensembles qui forment une VC-classe.

11.2 Quelques résultats sur les familles euclidiennes

Dans ce paragraphe nous rappelons quelques résultats utiles sur les familles euclidiennes. Ces résultats sont de deux types : les premiers donnent des exemples et des propriétés pour déterminer qu'une famille de fonctions est euclidienne. Les seconds sont des théorèmes sur les taux de convergence utilisés dans les preuves.

11.2.1 Familles de fonctions euclidiennes

Les exemples et propriétés ci-dessous se trouvent dans les articles de *Pakes et Pollard (1989)* et *Nolan et Pollard (1987)* et permettent de montrer que certaines familles de fonctions sont euclidiennes sans revenir à la définition 11.1.3.

Propriété 11.2.1. *conditions d'euclidiennité*

1. (cf. *Nolan et Pollard (1987)*, Corollaire 21) Soit \mathcal{F} une classe de fonctions de $\mathcal{X} \times \mathcal{X}$ uniformément bornée euclidienne. Alors pour toute mesure ν de \mathcal{X} , la classe $\{\int f(x, \cdot) d\nu(x), f \in \mathcal{F}\}$ est euclidienne pour une enveloppe constante.

2. (cf. Nolan et Pollard (1987), Lemme 22(ii)) Soit $g(\cdot)$ une fonction réelle à variations bornées sur \mathbb{R} . La classe de toutes les fonctions sur \mathbb{R}^d de la forme $z \mapsto g(\alpha \cdot z + \beta)$, avec α parcourant \mathbb{R}^d et β parcourant \mathbb{R} , est une famille euclidienne pour une enveloppe constante (αz est le produit scalaire entre z et α dans \mathbb{R}^d).
3. (cf. Pakes et Pollard (1989), Lemme 2.14) Si \mathcal{F} est une famille euclidienne pour l'enveloppe F et \mathcal{G} est une famille euclidienne pour l'enveloppe G . Alors $\{fg : f \in \mathcal{F}, g \in \mathcal{G}\}$ est une famille euclidienne pour une enveloppe FG .

Exemple 11.2.1. Soit K une fonction à variations bornées. Si on pose $\mathcal{F} = \{f_{b,y}(x) = K(\frac{y-x}{b}), b > 0, y \in \mathbb{R}\}$, alors \mathcal{F} est une famille euclidienne pour l'enveloppe $F = \sup |K|$.

Exemple 11.2.2. Dans notre cas on a la famille $\mathcal{G} = \{K(\frac{y-x}{b}), b \in]0, 1]\}$. K étant un noyau produit on peut en déduire en utilisant l'exemple précédent et le lemme 2.14, que \mathcal{G} est une famille euclidienne. En utilisant à nouveau le lemme 2.14 on en déduit que \mathcal{H}_2 et \mathcal{H}_3 sont des familles euclidiennes.

De plus amples exemples et conditions sont répertoriés dans les livres et articles déjà cités et dans l'ouvrage de Van der Vaart et Wellner (1996)

11.2.2 Ordre uniforme en probabilité

On rappelle la vitesse de convergence des U -processus utilisée dans les preuves du chapitre 7.

Propriété 11.2.2. [cf. Sherman (1994), Corollaire 4(ii)] Soit \mathcal{F} une classe de fonctions sur \mathcal{X}^k , $k \leq 1$, \mathbb{P} -dégénérée et $\mathbb{P}^k = \mathbb{P} \oplus \dots \oplus \mathbb{P}$. Si \mathcal{F} est euclidienne pour une enveloppe F deux fois intégrable, i.e. $\int F^2 d\mathbb{P}^k < \infty$. Alors

$$\sup_{f \in \mathcal{F}} |n^{k/2} U_n^k f| = \mathcal{O}_{\mathbb{P}}(1)$$

Annexe 3 : Distances dans un échantillon

On note, pour $j \in \{1, \dots, n\}$, $D_j(\mathbf{x})$ la distance de $\mathbf{x} \in \mathbb{R}^d$ à son $j^{\text{ème}}$ plus proche voisin dans l'échantillon \mathcal{X} .

Dans cette partie nous allons étudier tout d'abord le développement asymptotique de $\mathbb{E}(D_j^d(\mathbf{x}))$. On pourra alors en déduire une combinaison linéaire des $D_j(\cdot)$, définissant $D^d(\cdot)$ et admettant le développement (12.5). Dans un deuxième temps on s'intéressera au comportement asymptotique de la variable aléatoire $D^d(\mathbf{x})$.

12.1 Développement asymptotique de $\mathbb{E}\{D^d(\mathbf{x})\}$

Le théorème suivante nous donne la combinaison linéaire définissant $D^d(\cdot)$, dans quelques cas particuliers.

Théorème 12.1.1. *Soit \mathcal{X} un échantillon de vecteurs aléatoires de \mathbb{R}^d , de densité f sur un compact \mathcal{Q} . On suppose f séparée de zéro.*

- Si $2 \leq d \leq 3$ et si f a ses deux premières dérivées bornées alors, en prenant $D^d(\cdot) = D_1^d(\cdot)$, on a, uniformément par rapport à \mathbf{x}

$$\mathbb{E}[D^d(\mathbf{x})] = \{nv_d f(\mathbf{x})\}^{-1} + \mathcal{O}(n^{-3/2}). \quad (12.5)$$

- Si $4 \leq d \leq 7$ et si f a ses quatre premières dérivées bornées dans un voisinage de \mathbf{x} , alors on pose

$$D^d = \frac{\Gamma(3 + 2d^{-1})D_1^d - \Gamma(2 + 2d^{-1})D_2^d}{\Gamma(3 + 2d^{-1}) - 2\Gamma(2 + 2d^{-1})}$$

et (12.5) est vérifiée.

L'étude asymptotique de $\mathbb{E}(D^d(\cdot))$ revient donc à celle de $\mathbb{E}(D_1^d(\cdot))$ et $\mathbb{E}(D_2^d(\cdot))$ et fera donc l'objet des deux parties suivantes.

Remarque 3. *Le théorème ne traite pas le cas $d \geq 8$. Pour obtenir le résultat dans ce cas il faudra utiliser les mêmes techniques que celles utilisées ici et faire intervenir des distances D_j , pour $j \geq 3$.*

Pour la démonstration du théorème précédent nous avons besoin de quelques remarques préliminaires :

- On peut tout d’abord se ramener au cas $\mathbf{x} = 0$. On pose dans la suite $\mathcal{B}(t) = \mathcal{B}(0, t)$, boule de centre 0 et de rayon t . On note $D_j = D_j(0)$.
- Soit X une variable aléatoire positive à support compact, \mathcal{Q} , i.e. $X(\omega) \in [0, \delta]$, où $\delta = \text{diam}\mathcal{Q}$. Si g est une fonction dérivable vérifiant $g(0) = 0$, alors

$$\begin{aligned} \mathbb{E} [g(X)] &= \int g(X(w))d\mathbb{P}(w) = \int \int_0^{+\infty} g'(t)\mathbb{I}_{\{X(w)>t\}}dt d\mathbb{P}(w), \\ &= \int_0^{+\infty} g'(t)\mathbb{P}\{X(w) > t\}dt, \\ &= \int_0^{\delta} g'(t)\mathbb{P}\{X(w) > t\}dt. \end{aligned}$$

- D’après la définition de D_1 , on peut écrire pour tout $t \in [0, \delta]$

$$\begin{aligned} \mathbb{P}\{D_1 > t\} &= \mathbb{P}\{\text{aucun des points } X_i \text{ est dans } \mathcal{B}(t)\}, \\ &= \left(1 - \int_{\mathcal{B}(t)} f(u)du\right)^n, \\ &= \left(1 - t^d \int_{\mathcal{B}(1)} f(tz)dz\right)^n. \end{aligned}$$

Il va donc falloir déterminer dans un premier temps le développement asymptotique de $t^d \int_{\mathcal{B}(1)} f(tz)dz$, quand $t \rightarrow 0$.

Lemme 12.1.1. *Sous les mêmes hypothèses que pour la proposition : quand $t \rightarrow 0$*

$$t^d \int_{\mathcal{B}(1)} f(tz) dz = v_d f(0) t^d [1 + u_d t^2] + \mathcal{O}(t^{4+d}),$$

avec v_d , le volume de la boule unité de \mathbb{R}^d et on note $z^{(j)}$ la $j^{\text{ème}}$ composante du vecteur $z \in \mathbb{R}^d$ et

$$u_d = \frac{\sum_j f^{(j,j)}(0)}{2d v_d f(0)} \int_{\mathcal{B}(1)} \|z\|^2 dz, \quad \text{avec} \quad f^{(j,j)}(0) = (\partial/\partial z^{(j)})^2 f(z) \Big|_{z=0}.$$

DÉMONSTRATION.

f admet le développement suivant en 0

$$\begin{aligned} f(tz) &= f(0) + t \sum_{j=1}^d \frac{\partial f}{\partial z^{(j)}}(0) z^{(j)} + \frac{t^2}{2} \sum_{i=1}^d \sum_{j=1}^d f^{(j,i)}(0) z^{(j)} z^{(i)} \\ &\quad + \frac{t^3}{3!} \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d f^{(j,i,k)}(0) z^{(j)} z^{(i)} z^{(k)} + \mathcal{O}(t^4). \end{aligned}$$

On remarque que ce développement est valable uniformément en z , car la dérivée 4^{ème} de f est bornée et on considère ici $z \in \mathcal{B}(1)$, compact.

La symétrie de rotation de la sphère de centre 0 a pour conséquence que l'intégrale sur cette sphère d'une fonction impaire est nulle, ce qui conduit à

$$\int_{\mathcal{B}(1)} f(tz) dz = \int_{\mathcal{B}(1)} \left\{ f(0) + \frac{1}{2} \sum_{j=1}^d f^{(j,j)}(0) (z^{(j)})^2 t^2 \right\} dz + \mathcal{O}(t^4).$$

Explicitons chacun de ces deux termes

- $\int_{\mathcal{B}(1)} f(0) dz = f(0) v_d$;

- Comme la sphère est invariante par rotation, les intégrales, $\int_{\mathcal{B}(1)} (z^{(j)})^2 dz$, sont donc les mêmes dans toutes les directions j . Ainsi

$$\frac{t^2}{2} \sum_{j=1}^d f^{(j,j)}(0) \int_{\mathcal{B}(1)} (z^{(j)})^2 dz = \frac{t^2}{2} \sum_{j=1}^d f^{(j,j)}(0) \left(\int_{\mathcal{B}(1)} (u^{(1)})^2 du \right).$$

On obtient alors

$$t^d \int_{\mathcal{B}(1)} f(tz) dz = v_d f(0) t^d \left[1 + \frac{1}{2f(0)v_d} \sum_{j=1}^d f^{(j,j)}(0) \left(\int_{\mathcal{B}(1)} (u^{(1)})^2 du \right) t^2 \right] + \mathcal{O}(t^{4+d}).$$

On note u_d le deuxième terme de l'égalité, que l'on peut encore écrire

$$u_d = \frac{1}{2df(0)v_d} \left(\sum_{j=1}^d f^{(j,j)}(0) \right) \int_{\mathcal{B}(1)} \|u\|^2 du.$$

Finalement quand t tends vers 0

$$t^d \int_{\mathcal{B}(1)} f(tz) dz = v_d f(0) t^d [1 + u_d t^2] + \mathcal{O}(t^{4+d})$$

■

On définit dans toute la suite

$$h(t) = \log \left(1 - t^d \int_{\mathcal{B}(1)} f(tz) dz \right),$$

fonction qui intervient dans le développement asymptotique des $\mathbb{E}(D_j^d)$. En utilisant le lemme 12.1.1, on en déduit le développement de h au voisinage de 0

$$h(t) = -v_d f(0) t^d [1 + u_d t^2] + \mathcal{O}(t^{4+d})$$

12.1.1 Développement asymptotique de $\mathbb{E}(D_1^d)$

Le théorème suivant donne le développement asymptotique de $\mathbb{E}(D_1^d)$, dont nous avons besoin pour prouver le théorème 12.1.1.

Théorème 12.1.2. *Sous les mêmes hypothèses que le théorème 12.1.1, nous avons*

$$\mathbb{E}(D_1^d) = (f(0)v_d n)^{-1} \left(1 - \Gamma(2 + 2/d)n^{-2/d}w_d \right) + \mathcal{O}(n^{-1-4/d}), \quad (12.6)$$

où $w_d = u_d/(f(0)v_d)^{2/d}$ et u_d est défini dans le lemme 12.1.1.

L'obtention de ce résultat nécessite trois étapes de démonstration : trouver un équivalent, le deuxième terme du développement puis l'ordre du reste. Les outils principaux sont la méthode de Laplace et le théorème de convergence dominée.

DÉMONSTRATION. **1°** *Obtention d'un équivalent de $\mathbb{E}(D_1^d)$, quand $n \rightarrow +\infty$*

$$\begin{aligned} \mathbb{E}(D_1^d) &= \int_0^\delta dt^{d-1} \mathbb{P}(D_1 > t) dt \\ &= \int_0^\delta dt^{d-1} \left(1 - t^d \int_{\mathcal{B}(1)} f(tu) du \right)^n dt \\ &= \int_0^\delta g(t) e^{nh(t)} dt \end{aligned}$$

avec $g(t) = dt^{d-1}$ et on rappelle que $h(t) = \log \left(1 - t^d \int_{\mathcal{B}(1)} f(tu) du \right)$. On peut alors utiliser la méthode de Laplace pour trouver un équivalent de cette intégrale. En effet $\int_0^\delta |g(t)| e^{h(t)} dt$ est une intégrale convergente, h est décroissante et admet le développement suivant au voisinage de 0 :

$$h(t) = -f(0)v_d t^d + \mathcal{O}(t^d)$$

Finalement, quand $n \rightarrow \infty$

$$\mathbb{E}(D_1^d) \sim (f(0)v_d n)^{-1}$$

2° *Obtention du deuxième terme du développement*

Pour cela on étudie la différence :

$$\Delta_n = \mathbb{E}(D_1^d) - (f(0)v_d n)^{-1}$$

Celle-ci contient deux quantités : l'une qui utilise le développement de h au voisinage de 0 et qui va donner un équivalent ; la deuxième sera négligeable devant cet équivalent.

• 1^{ère} étape : développement au voisinage de 0.

Soit $\varepsilon > 0$, il existe δ_0 tel que pour $t \in [0, \delta_0]$

$$(1 - \varepsilon) \left(-f(0)v_d t^d - t^{d+2}u_d v_d f(0) \right) \leq h(t) \leq (1 + \varepsilon) \left(-f(0)v_d t^d - t^{d+2}u_d v_d f(0) \right).$$

On note

$$\Delta_{n,\delta_0} = \int_0^{\delta_0} g(t) e^{nh(t)} - \frac{1}{nv_d f(0)},$$

alors

$$\begin{aligned} \int_0^{\delta_0} g(t) e^{n(1-\varepsilon) \left(-f(0)v_d t^d - t^{d+2}u_d v_d f(0) \right)} dt - \frac{1}{nv_d f(0)} &\leq \Delta_{n,\delta_0} \\ &\leq \int_0^{\delta_0} g(t) e^{n(1+\varepsilon) \left(-f(0)v_d t^d - t^{d+2}u_d v_d f(0) \right)} dt - \frac{1}{nv_d f(0)}. \end{aligned} \quad (12.7)$$

On considère tout d'abord la partie droite de cette inégalité.

On remarque qu'il existe n_0 , tel que pour $n > n_0$: $\frac{1 - e^{-na(1-\varepsilon)\delta_0^d}}{na(1-\varepsilon)} - \frac{1}{na} > 0$. En utilisant cette inégalité avec $a = v_d f(0)$, on obtient

$$\int_0^{\delta_0} g(t) e^{n(1-\varepsilon) \left(-f(0)v_d t^d - t^{d+2}u_d v_d f(0) \right)} dt - \frac{1 - e^{-nf(0)v_d(1-\varepsilon)\delta_0^d}}{nf(0)v_d(1-\varepsilon)} \leq \Delta_{n,\delta_0}$$

Étudions maintenant cette borne inférieure, $\underline{\Delta}_{n,\delta_0}$

$$\begin{aligned} \underline{\Delta}_{n,\delta_0} &= \int_0^{\delta_0} g(t) e^{n(1-\varepsilon) \left(-f(0)v_d t^d - t^{d+2}u_d v_d f(0) \right)} dt - \frac{1 - e^{nv_d f(0)(1-\varepsilon)\delta_0^d}}{nf(0)v_d(1-\varepsilon)} \\ \underline{\Delta}_{n,\delta_0} &= \int_0^{\delta_0} dt^{d-1} e^{n(1-\varepsilon) \left(-f(0)v_d t^d \right)} \left(e^{-n(1-\varepsilon)t^{d+2}u_d v_d f(0)} - 1 \right) dt \\ \underline{\Delta}_{n,\delta_0} &= \int_0^{nv_d f(0)(1-\varepsilon)\delta_0} \left(e^{-n(1-\varepsilon)\frac{u^{1+2/d}}{(nv_d f(0)(1-\varepsilon))^{1+2/d}} u_d v_d f(0)} - 1 \right) \frac{e^{-u} du}{nv_d f(0)(1-\varepsilon)} \\ n^{1+2/d} \underline{\Delta}_{n,\delta_0} &= \int_0^{nf(0)v_d(1-\varepsilon)\delta_0} \frac{n^{2/d} e^{-u}}{v_d f(0)(1-\varepsilon)} f_n(u) du \\ n^{1+2/d} \underline{\Delta}_{n,\delta_0} &= \int_{\mathbb{R}^+} g_n(u) du \end{aligned}$$

Or, pour tout n , $g_n(u) = \frac{n^{2/d} e^{-u}}{v_d f(0)(1-\varepsilon)} f_n(u) \mathbb{I}_{[0, nv_d f(0)(1-\varepsilon)\delta_0]}(u)$ est une fonction mesurable. De plus

$$g_n(u) \sim - \frac{e^{-u} u^{1+2/d}}{(v_d f(0))^{1+2/d} (1-\varepsilon)^{1+2/d}} u_d$$

et

$$|g_n(u)| \leq e^{-u} u^{1+2/d} u_d / (v_d f(0))^{1+2/d} (1 - \varepsilon)^{1+2/d}.$$

Le théorème de convergence dominée donne alors :

$$n^{1+2/d} \underline{\Delta}_{n, \delta_0} \sim - \frac{\Gamma(2 + 2/d) u_d}{(v_d f(0))^{1+2/d} (1 - \varepsilon)^{1+2/d}}$$

c'est-à-dire : il existe n_1 , tel que pour tout $n \geq \sup(n_0, n_1)$ on a

$$\frac{-\Gamma(2 + 2/d) n^{-1-2/d} u_d}{(v_d f(0))^{1+2/d} (1 - \varepsilon)^{2/d}} \leq \Delta_{n, \delta_0} \quad (12.8)$$

On fait de même pour la partie gauche de l'inégalité 12.7

– Comme on a encore $\frac{1 - e^{-na(1+\varepsilon)\delta_0^d}}{na(1+\varepsilon)} - \frac{1}{na} < 0$, donc

$$\Delta_{n, \delta_0} \leq \int_0^{\delta_0} g(t) e^{n(1+\varepsilon) \left(-f(0)v_d t^d - t^{d+2} u_d v_d f(0) \right)} dt - \frac{1 - e^{-nv_d f(0)(1+\varepsilon)\delta^d}}{nv_d f(0)(1+\varepsilon)}.$$

– L'étude de ce majorant, $\overline{\Delta}_{n, \delta_0}$ se fait de façon analogue à celle de $\underline{\Delta}_{n, \delta_0}$, donne

$$\overline{\Delta}_{n, \delta_0} \sim - \frac{\Gamma(2 + 2/d) u_d}{(v_d f(0))^{1+2/d} (1 + \varepsilon)^{1+2/d}}$$

On obtient l'existence de n_2 , tel que pour $n \geq n_2$

$$\Delta_{n, \delta_0} \leq \frac{-\Gamma(2 + 2/d) n^{-1-2/d} u_d}{(v_d f(0))^{1+2/d} (1 + \varepsilon)^{2/d}} \quad (12.9)$$

Finalement en utilisant les inégalités 12.8 et 12.9 on obtient

$$-(1 - \varepsilon)^{-2/d} \frac{\Gamma(2 + 2/d) n^{-1-2/d} u_d}{(v_d f(0))^{1+2/d}} \leq \Delta_{n, \delta_0} \leq -(1 + \varepsilon)^{-2/d} \frac{\Gamma(2 + 2/d) n^{-1-2/d} u_d}{(v_d f(0))^{1+2/d}}$$

• 2^{ème} étape : Étude de $\Delta_n - \Delta_{n, \delta_0} = \int_{\delta_0}^{\delta} g(t) e^{nh(t)} dt$.

Posons $\mu = -h(\delta_0) > 0$. Comme h est décroissante, on a $h(x) + \mu \leq 0$ pour tout $x \geq \delta_0$, donc pour tout $x \geq \delta_0$, pour tout $n > 1$,

$$nh(x) = (n - 1)h(x) + h(x) \leq -(n - 1)\mu + h(x)$$

ce qui entraîne la convergence de $\int_{\delta_0}^{\delta} |g(x)| e^{nh(x)} dx$ et l'inégalité

$$\int_{\delta_0}^{\delta} |g(x)| e^{nh(x)} dx \leq e^{-(n-1)\mu} \int_{\delta_0}^{\delta} |g(x)| e^{nh(x)} dx$$

Or $\mu > 0$, d'où l'existence d'un entier n_3 , tel que pour tout $n \geq n_3$

$$\int_{\delta_0}^{\delta} |g(x)| e^{nh(x)} dx \leq \varepsilon \frac{\Gamma(2 + 2/d) n^{-1-2/d} u_d}{(v_d f(0))^{1+2/d}}$$

• *Conclusion* : pour $n \geq \sup(n_0, n_1, n_2, n_3)$, on obtient un encadrement, dont chaque borne est le produit d'une fonction en ε , qui tend vers 1 quand $\varepsilon \rightarrow 0$ et de la constante $\frac{-\Gamma(2+2/d)n^{-1-2/d}u_d}{(v_d f(0))^{1+2/d}}$. On obtient donc

$$\Delta_n = \int_0^{\delta} g(t) e^{nh(t)} - \frac{1}{nf(0)v_d} \sim -\frac{\Gamma(2 + 2/d) n^{-1-2/d} u_d}{(v_d f(0))^{1+2/d}} \quad (12.10)$$

3° *Ordre du reste*

On veut obtenir l'ordre du reste, i.e.

$$\Gamma_n = \int_0^{\delta_0} g(t) e^{nh(t)} - \frac{1}{nf(0)v_d} + \frac{\Gamma(2 + 2/d) n^{-1-2/d} u_d}{(v_d f(0))^{1+2/d}} = \mathcal{O}(n^{-1-4/d}).$$

Pour cela on utilise la même démarche que dans le point précédent

• *1^{ère} étape* : On utilise le développement de h au voisinage de 0 il existe δ_0 , pour tout $x \in [0, \delta_0]$, il existe M tel que

$$\begin{aligned} -Mx^{d+4} + a(x) &\leq h(x) \leq Mx^{d+4} + a(x), \\ e^{-nMx^{d+4}+na(x)} &\leq e^{nh(x)} \leq e^{nMx^{d+4}+na(x)} \end{aligned}$$

avec $a(x) = -v_d f(0)x^d - v_d f(0)u_d x^{d+2}$. On obtient alors l'encadrement

$$\begin{aligned} \int_0^{\delta_0} g(x) e^{-nMx^{d+4}+na(x)} dx - \left(\frac{1}{nv_d f(0)} - \frac{\Gamma(2 + 2/d) n^{-1-2/d} u_d}{(v_d f(0))^{1+2/d}} \right) &\leq \Gamma_{n, \delta_0} \leq \\ \int_0^{\delta_0} g(x) e^{nMx^{d+4}+na(x)} dx - \left(\frac{1}{nv_d f(0)} - \frac{\Gamma(2 + 2/d) n^{-1-2/d} u_d}{(v_d f(0))^{1+2/d}} \right) & \end{aligned}$$

En utilisant le lemme 12.1.2, l'étude du minorant revient à l'étude de la quantité

$$\begin{aligned} \Gamma_{n, \delta_0} &= \int_0^{\delta_0} g(x) e^{na(x)-Mnx^{d+4}} dx - \int_0^{\delta_0} g(x) e^{na(x)} dx \\ &= \int_0^{\delta_0} g(x) e^{na(x)} \left(e^{-Mnx^{d+4}} - 1 \right) dx \\ &= \int_0^{nv_d f(0)\delta_0} e^{na\left(\left(\frac{u}{nv_d f(0)}\right)^{1/d}\right)} \left(e^{-\frac{Mnu^{1+4/d}}{(nf(0)v_d)^{1+4/d}} - 1} \right) \frac{du}{nf(0)v_d} \end{aligned}$$

en effectuant le changement de variable $u = nf(0)v_dx^d$. On note ensuite

$$n^{1+4/d}\Gamma_{n,\delta_0} = \int_{\mathbb{R}^+} t_n(u)du,$$

où $(t_n)_n$ est une suite de fonctions mesurables vérifiant $t_n(u) \sim_{+\infty} -\frac{Mu^{1+4/d}}{(f(0)v_d)^{1+4/d}}\mathbb{I}_{[0,\infty]}(u)$ et majorée par $\frac{Mu^{1+4/d}}{(f(0)v_d)^{1+4/d}}\mathbb{I}_{[0,\infty]}(u)$, fonction intégrable. Le théorème de convergence dominée nous permet de conclure que le minorant est d'ordre $\mathcal{O}(n^{-1-4/d})$.

On fait de même avec le majorant de Γ_{n,δ_0} et on obtient

$$\Gamma_{n,\delta_0} = \mathcal{O}(n^{-1-4/d}).$$

• 2^{ème} étape : On montre que $\Gamma_n - \Gamma_{n,\delta_0}$ est de l'ordre de $\mathcal{O}(n^{-1-4/d})$, avec le même raisonnement que pour la 2^{ème} étape du deuxième point.

• Conclusion on obtient

$$\Gamma_n = \mathcal{O}(n^{-1-4/d}),$$

ce qui correspond au résultat attendu du théorème 12.1.2 ■

Lemme 12.1.2. Soit $a(x) = -v_df(0)x^d - v_df(0)u_dx^{d+2}$, alors

$$\int_0^{\delta_0} g(x)e^{na(x)} dx = \frac{1}{nv_df(0)} \left(1 - \frac{w_d\Gamma(2+2/d)}{n^{2/d}}\right) + \mathcal{O}(n^{-1-4/d})$$

où $w_d = \frac{u_d}{(f(0)v_d)^{2/d}}$.

DÉMONSTRATION. 1^o *Obtention de l'équivalent* On utilise la méthode de Laplace pour obtenir

$$\int_0^{\delta_0} g(x)e^{na(x)} dx \sim \frac{1}{nv_df(0)}$$

2^o *Deuxième terme du développement asymptotique.* On étudie la différence

$$D_n = \int_0^{\delta_0} g(x)e^{na(x)} dx - \frac{1}{nv_df(0)},$$

que l'on peut réécrire

$$D_n = \frac{1}{nv_df(0)} \int_0^{nv_df(0)\delta_0} e^{-u} \left(e^{-\frac{u_d}{(v_df(0))^{2/d}}n^{-2/d}u^{1+2/d}} - 1 \right) du + \frac{e^{-nv_df(0)\delta_0^d}}{nv_df(0)}$$

$$n^{1+2/d}D_n = \int_0^\infty a_n(u)du + \frac{e^{-nv_df(0)\delta_0^d}}{nv_df(0)}n^{1+2/d}$$

– quand $n \rightarrow \infty$, le deuxième terme tend vers 0 ;

- pour traiter le premier terme, nous faisons appel au théorème de convergence dominée, car les fonction a_n sont mesurables et vérifient $a_n(u) \sim -\frac{u_d u^{1+2/d}}{(v_d f(0))^{1+2/d}} e^{-u} \mathbb{I}_{[0,\infty]}$ et $|a_n(u)| \leq \frac{u_d u^{1+2/d}}{(v_d f(0))^{1+2/d}} e^{-u} \mathbb{I}_{[0,\infty]}$, qui est dans L^1 . On a donc

$$n^{1+2/d} D_n \sim -\frac{u_d \Gamma\left(2 + \frac{2}{d}\right)}{(v_d f(0))^{1+2/d}}$$

3° Ordre du reste On pose

$$E_n = \int_0^{\delta_0} g(x) e^{na(x)} dx - \frac{1}{n v_d f(0)} \left(1 - \frac{w_d \Gamma(2 + 2/d)}{n^{2/d}}\right),$$

que l'on peut réécrire

$$\begin{aligned} n^{1+4/d} E_n &= \frac{n^{4/d}}{v_d f(0)} \int_0^{n f(0) v_d \delta_0} e^{-u} \left(e^{-\frac{u_d n^{-2/d} u^{1+2/d}}{(v_d f(0))^{2/d}}} - 1 + \frac{u_d n^{-2/d} u^{1+2/d}}{(v_d f(0))^{2/d}} \right) du \\ &\quad - \frac{n^{4/d} e^{-n v_d f(0) \delta_0^d}}{v_d f(0)} + \int_{n f(0) v_d \delta_0}^{\infty} \frac{u_d n^{2/d} u^{1+2/d}}{(v_d f(0))^{1+2/d}} e^{-u} du \end{aligned}$$

- Le premier terme s'étudie en utilisant le développement de la fonction exponentielle un terme plus loin ;
- Le deuxième terme tend vers 0 ;
- Le troisième terme tend vers 0, car l'intégrale est inférieure à

$$n^{2/d} e^{-n f(0) v_d \delta_0 / 2} \int_{n f(0) v_d \delta_0}^{\infty} \frac{u_d u^{1+2/d}}{(v_d f(0))^{1+2/d}} e^{-u/2} du$$

et l'intégrale converge par le théorème de convergence dominée.

4° Conclusion

$$\int_0^{\delta_0} g(x) e^{na(x)} dx = \frac{1}{n v_d f(0)} \left(1 - w_d \Gamma\left(2 + \frac{2}{d}\right) n^{-2/d}\right) + \mathcal{O}\left(n^{-1-4/d}\right)$$

12.1.2 Développement asymptotique de $\{\mathbb{E}(D_2^d) - \mathbb{E}(D_1^d)\}$

Nous n'allons pas étudier directement $\mathbb{E}(D_2^d)$, mais la différence $\{\mathbb{E}(D_2^d) - \mathbb{E}(D_1^d)\}$. En effet cette dernière nous permet d'identifier exactement le nombre de points de l'échantillon \mathcal{X} , contenu dans un voisinage de la forme $\mathcal{B}(t)$. De plus comme nous disposons du développement de $\mathbb{E}(D_1^d)$, par le paragraphe précédent, nous pourrions retrouver celui de $\mathbb{E}(D_2^d)$.

Propriété 12.1.1. *Sous les conditions énoncées dans le théorème 12.1.1, on a*

$$\{\mathbb{E}(D_2^d) - \mathbb{E}(D_1^d)\} = \{v_d f(0)n\}^{-1} \left[\Gamma(2) + w_d(\Gamma(2 + 2/d) - \Gamma(3 + 2/d))n^{-2/d} \right. \\ \left. + \mathcal{O}(n^{\eta-4/d}) \right] \quad (12.11)$$

où $w_d = u_d/(f(0)v_d)^{2/d}$.

DÉMONSTRATION. Cette différence se réécrit

$$\begin{aligned} \{\mathbb{E}(D_2^d) - \mathbb{E}(D_1^d)\} &= \int_0^{+\infty} dt^{d-1} (\mathbb{P}(D_2 > t) - \mathbb{P}(D_1 > t)) \mathbb{I}_{\{t \leq \delta\}} dt, \\ &= \int_0^{+\infty} t^{d-1} \mathbb{P}(\{D_2 > t\} \cap \{D_1 < t\}) \mathbb{I}_{\{t \leq \delta\}} dt, \\ &= \int_0^{+\infty} dt^{d-1} P\{\text{un et un seul } X_i \text{ est dans } \mathcal{O}(t)\} \mathbb{I}_{\{t \leq \delta\}} dt, \\ &= n \int_0^{+\infty} dt^{d-1} \left\{ \int_{\mathcal{B}(t)} f(z) dz \right\} \left\{ 1 - \int_{\mathcal{B}(t)} f(z) dz \right\}^{n-1} \mathbb{I}_{\{t \leq \delta\}} dt, \\ &= n \int_0^{+\infty} dt^{d-1} \left\{ \int_{\mathcal{B}(1)} t^d f(tz) dz \right\} \left\{ 1 - \int_{\mathcal{B}(1)} t^d f(tz) dz \right\}^{n-1} \mathbb{I}_{\{t \leq \delta\}} dt. \end{aligned}$$

c'est-à-dire

$$\{\mathbb{E}(D_2^d) - \mathbb{E}(D_1^d)\} = n \int_0^{+\infty} g(t) e^{mh(t)} dt \quad (12.12)$$

où $m = n - 1$, $g(t) = dt^{2d-1} \int_{\mathcal{B}(1)} f(tu) du \mathbb{I}_{\{t \leq \delta\}}$ et $h(t) = \log \left(1 - t^d \int_{\mathcal{B}(1)} f(tz) dz \right)$.

1° Obtention d'un équivalent On utilise la méthode de Laplace pour trouver un équivalent. En effet l'intégrale 12.12 est convergente, h est décroissante et au voisinage de 0, on a :

$$\begin{aligned} g(t) &= v_d dt^{2d-1} f(0) + u_d v_d f(0) t^{2d+1} + \mathcal{O}(t^{2d+4}) \\ h(t) &= -v_d f(0) t^d - u_d v_d f(0) t^{d+2} + \mathcal{O}(t^{d+4}) \end{aligned}$$

On obtient donc

$$n \int_0^{+\infty} g(t) e^{mh(t)} dt \sim n \frac{v_d f(0) \Gamma(2)}{(n v_d f(0))^2} = \frac{1}{n v_d f(0)}$$

2° Deuxième terme Pour cela, il faut considérer la différence entre l'intégrale et l'équivalent trouvé ci-dessus. Comme dans la démonstration du développement de $\mathbb{E}(D_1^d)$, le terme qui donne l'équivalent est celui qui correspond au développement de h en 0. Ainsi en encadrant h obtient un encadrement du reste. Finalement le lemme 12.1.3 nous donne l'équivalent.

3° Ordre du reste En appliquant encore une fois le lemme 12.1.3 à la différence Γ_n , on trouve l'ordre du reste.

4° Conclusion On obtient bien le résultat

$$\{\mathbb{E}(D_2^d) - \mathbb{E}(D_1^d)\} = \{v_d f(0)n\}^{-1} \left[\Gamma(2) + w_d(\Gamma(2 + 2/d) - \Gamma(3 + 2/d))n^{-2/p} + \mathcal{O}(n^{-1-4/d}) \right] \quad (12.13)$$

Lemme 12.1.3. Soit $\delta_0 \in [0, \delta]$,

$$n \int_0^{\delta_0} g(x) e^{na(x)} dx = \frac{1}{nf(0)v_d} \left[1 + \left(\Gamma(2 + 2/d) - \Gamma(3 + 2/d) \right) w_d n^{-2/d} \right] + \mathcal{O}(n^{-1-4/d}), \quad (12.14)$$

où $a(x) = -v_d f(0)x^d - u_d v_d f(0)x^{d+2}$.

DÉMONSTRATION. 1° Équivalent On utilise la méthode de Laplace pour obtenir :

$$n \int_0^{\delta_0} g(x) e^{na(x)} dx \sim \frac{1}{nf(0)v_d} \quad (12.15)$$

2° Deuxième terme Pour l'obtenir nous devons introduire deux termes intermédiaires

$$\begin{aligned} n \int_0^{\delta_0} g(x) e^{na(x)} dx - \frac{1}{nf(0)v_d} &= \left(n \int_0^{\delta_0} g(x) e^{na(x)} dx - n \int_0^{\delta_0} g(x) e^{nv_d f(0)x^d} dx \right) \\ &+ \left(n \int_0^{\delta_0} g(x) e^{nv_d f(0)x^d} dx - n \int_0^{\delta_0} v_d dx^{2d-1} f(0) e^{nv_d f(0)x^d} dx \right) \\ &+ \left(n \int_0^{\delta_0} v_d dx^{2d-1} f(0) e^{nv_d f(0)x^d} dx - \frac{1}{nf(0)v_d} \right) \end{aligned}$$

Étudions séparément chacun de ces termes :

- Le premier terme se réécrit, en utilisant le changement de variable $u = nv_d f(0)x^d$

$$\begin{aligned} &n \int_0^{\delta_0} g(x) e^{na(x)} dx - n \int_0^{\delta_0} g(x) e^{-nv_d f(0)x^d} dx \\ &= n \int_0^{\delta_0} g(x) e^{-nv_d f(0)x^d} \left(e^{-nv_d f(0)u_d x^{d+2}} - 1 \right) dx \\ &= n \int_0^{nv_d f(0)\delta^d} \frac{u}{(nv_d f(0))^2} \left(\int_{\mathcal{B}(1)} f \left(\frac{u^{1/d} v}{(nf(0)v_d)^{1/d}} \right) dv \right) e^{-u} \left(e^{-\frac{nv_d f(0)u_d u^{1+2/d}}{(nv_d f(0))^{1+2/d}} - 1} \right) du \\ &= \int_{\mathbb{R}^+} f_n(u) du \end{aligned}$$

f_n est une fonction mesurable et quand $n \rightarrow \infty$

$$f_n(u) \sim -ne^{-u} \frac{u}{(nv_d f(0))^2} f(0) v_d \frac{u d u^{1+2/d}}{(nv_d f(0))^{2/d}} = -e^{-u} \frac{u^{2+2/d}}{(v_d f(0))^{1+2/d}} n^{-1-2/d}.$$

La suite de fonctions, $(n^{2+2/d} f_n)_n$, vérifie les hypothèses du théorème de convergence dominée, on obtient donc

$$n \int_0^{\delta_0} g(x) e^{na(x)} dx - n \int_0^{\delta_0} g(x) e^{-nv_d f(0)x^d} dx \sim -\frac{\Gamma(3+2/d)}{(v_d f(0))^{1+2/d}} n^{-1-2/d}$$

• Le deuxième terme

$$\begin{aligned} & n \int_0^{\delta_0} g(x) e^{nv_d f(0)x^d} dx - \int_0^{\delta_0} n v_d x^{2d-1} f(0) e^{nv_d f(0)x^d} dx \\ &= \int_0^{\delta_0} dx^{2d-1} \left(\int_{\mathcal{B}(1)} f(tz) dz - v_d f(0) \right) e^{nv_d f(0)x^d} dx \\ &= \int_0^{nv_d f(0)\delta_0^d} \frac{z}{nv_d f(0)} \left(\int_{\mathcal{B}(1)} f \left(\frac{z^{1/d} u}{(nf(0)v_d)^{1/d}} \right) du - v_d f(0) \right) e^{-z} \frac{dz}{nv_d f(0)} \\ &= \int_{\mathbb{R}} g_n(u) du \end{aligned}$$

$(g_n)_n$ est une suite de fonctions mesurables, vérifiant

$$g_n(z) \sim \frac{z}{n(v_d f(0))^2} e^{-z} \frac{u_d v_d f(0) z^{2/d}}{(nv_d f(0))^{2/d}} = \frac{u_d z^{1+2/d} e^{-z}}{n^{1+2/d} (v_d f(0))^{1+2/d}}$$

En appliquant le théorème de convergence dominée à la suite de fonctions $(n^{1+2/d} g_n)_n$ on obtient

$$\int_{\mathbb{R}} g_n(u) du \sim \frac{\Gamma(2+2/d)}{(v_d f(0))^{1+2/d}} n^{1+2/d}$$

• Pour la dernier terme, on écrit

$$\begin{aligned} & n \int_0^{\delta_0} v_d dx^{2d-1} f(0) e^{nv_d f(0)x^d} dx - \frac{1}{nf(0)v_d} = \int_0^{nv_d f(0)\delta_0^d} e^{-u} \frac{u}{nf(0)v_d} du - \frac{1}{nf(0)v_d} \\ &= \frac{1}{nf(0)v_d} \int_{nv_d f(0)\delta_0^d}^{\infty} u e^{-u} du \\ &= \frac{-e^{-nv_d f(0)\delta_0^d}}{nf(0)v_d} (nv_d f(0)\delta_0^d + 1) = o(n^{-1-2/d}) \end{aligned}$$

• Finalement avec les trois termes, on obtient le deuxième terme du développement (12.14).

3° Conclusion L'ordre du reste du développement est obtenu comme dans le lemme 12.1.2 et est essentiellement dû au développement de l'exponentielle à l'ordre supérieur. Ceci termine la démonstration du lemme. ■

12.1.3 Choix de $D^d(\cdot)$

Dans cette partie nous allons finalement expliquer pourquoi le théorème 12.1.1 propose de tels combinaisons linéaires pour $D^d(\cdot)$ à l'aides des deux paragraphes précédents.

- Si $d = 2$ ou $d = 3$, le développement trouve dans le premier paragraphe

$$\mathbb{E} (D_1^d) = (nv_d f(0))^{-1} \left(1 - \Gamma(2 + 2/d)n^{-2/d}w_d \right) + \mathcal{O}(n^{-1-4/d})$$

où $w_d = u_d/(f(0)v_d)^{2/d}$. Comme $d = 2$ ou 3 , on a $n^{-1-2/d} = \mathcal{O}(n^{-3/2})$, ainsi

$$\mathbb{E} (D^d) = (nv_d f(0))^{-1} + \mathcal{O}(n^{-3/2}).$$

On pose donc dans ce cas $D^d = D_1^d$.

- Pour $4 \leq d \leq 7$, soit $D^d(\cdot)$ définit par

$$D^d = \frac{\Gamma(3 + 2d^{-1})D_1^d - \Gamma(2 + 2d^{-1})D_2^d}{\Gamma(3 + 2d^{-1}) - 2\Gamma(2 + 2d^{-1})}$$

En utilisant les deux paragraphes précédents, on peut écrire

$$\begin{aligned} & \mathbb{E} (D^d) \\ &= \frac{\Gamma(3 + 2d^{-1}) - \Gamma(2 + 2d^{-1})}{\Gamma(3 + 2d^{-1}) - 2\Gamma(2 + 2d^{-1})} \mathbb{E} (D_1^d) - \frac{\Gamma(2 + 2d^{-1})}{\Gamma(3 + 2d^{-1}) - 2\Gamma(2 + 2d^{-1})} \mathbb{E} (D_2^d - D_1^d) \\ &= \frac{\Gamma(3 + 2d^{-1}) - \Gamma(2 + 2d^{-1})}{\Gamma(3 + 2d^{-1}) - 2\Gamma(2 + 2d^{-1})} (nv_d f(0))^{-1} \left(1 - \Gamma(2 + 2/d)n^{-2/d}w_d \right) \\ &\quad - \frac{\Gamma(2 + 2d^{-1})}{\Gamma(3 + 2d^{-1}) - 2\Gamma(2 + 2d^{-1})} \{nv_d f(0)\}^{-1} \left[\Gamma(2) + w_d(\Gamma(2 + 2/d) - \Gamma(3 + 2/d))n^{-2/d} \right] \\ &\quad + \mathcal{O}(n^{-1-4/d}) \end{aligned}$$

Finalement

$$\mathbb{E} (D^d) = \{nv_d f(0)\}^{-1} + \mathcal{O}(n^{-1-4/d}) = \{nv_d f(0)\}^{-1} + \mathcal{O}(n^{-3/2}).$$

Remarque 4. Pour $d \geq 8$, on peut faire de même, mais on voit bien qu'on a besoin d'un terme supplémentaire dans ces deux développements. Il faudra aussi calculer le développement de $\mathbb{E} (D_3^d)$ et pour cela on utilisera les mêmes méthodes, c'est-à-dire on choisira de calculer $\mathbb{E} (D_3^d) - \mathbb{E} (D_2^d)$. Les développements ainsi obtenus seraient de l'ordre de $\mathcal{O}(n^{-1-6/d})$ et nous fourniraient la combinaison linéaire à utiliser pour $d = 8, \dots, 11$. Il faudra faire intervenir D_4 , si $d \geq 12$.

12.2 Application à l'étude de $\tilde{I}_S(g)$

Propriété 12.2.1. Soit $\rho = v_d^{-1}$.

$\tilde{I}_S(g) = \|\mathcal{S}\|_d^{-1} \sum_{i=1}^n Y_i \mathbb{I}_{(X_i \in \mathcal{S})} D^d(X_i)$ est un estimateur asymptotiquement sans biais de $\rho I_S(g)$.

DÉMONSTRATION. On se sert du développement asymptotique (5.11) pour étudier \tilde{I}_S , donné par (5.12). Montrons que $\tilde{I}_S(g)$ est asymptotiquement sans biais :

$$\begin{aligned} \mathbb{E} \left(\tilde{I}_S(g) \right) &= \|\mathcal{S}\|_d^{-1} \sum_{i=1}^n \mathbb{E} \left[Y_i \mathbb{I}_{X_i \in \mathcal{S}} D^d(X_i) \right], \\ &= \|\mathcal{S}\|_d^{-1} \sum_{i=1}^n \mathbb{E} \left[\mathbb{E} [Y_i \mathbb{I}_{X_i \in \mathcal{S}} D^d(X_i) | X_i] \right], \\ &= \|\mathcal{S}\|_d^{-1} \sum_{i=1}^n \left[\int_{\mathcal{S}} \mathbb{E} [Y_i D^d(X_i) | X_i = \mathbf{x}] f(\mathbf{x}) d\mathbf{x} \right], \\ &= \|\mathcal{S}\|_d^{-1} n \int_{\mathcal{S}} g(\mathbf{x}) \mathbb{E} (D^d(\mathbf{x})) f(\mathbf{x}) d\mathbf{x}, \end{aligned}$$

car $Y_i = g(X_i) + \varepsilon_i$.

Alors en utilisant le développement asymptotique (5.11), on peut écrire

$$\begin{aligned} \mathbb{E} (\tilde{I}_S(g)) &= \|\mathcal{S}\|_d^{-1} n \int_{\mathcal{S}} g(\mathbf{x}) \frac{1}{n f(\mathbf{x}) v_d} f(\mathbf{x}) d\mathbf{x} + o(n^{-1/2}), \\ \mathbb{E} (\tilde{I}_S(g)) &= \rho I_S(g) + o(n^{-1/2}). \end{aligned}$$

■

12.3 Étude du comportement asymptotique de $D(\mathbf{x})$

Examinons tout d'abord la convergence en loi de $D^d(\mathbf{x})$.

$$\begin{aligned} \mathbb{P} \left(D(\mathbf{x}) > n^{-1/d} u \right) &= \left(1 - \mathbb{P} \left(\|X_j - \mathbf{x}\| \leq n^{-1/d} u \right) \right)^n \\ &= \left(1 - \int_{\mathcal{B}(\mathbf{x}, n^{-1/d} u)} f(z) dz \right)^n \end{aligned}$$

avec le même développement que dans le paragraphe 12.5, en supposant que la dérivée seconde est bornée, on obtient :

$$\begin{aligned} \mathbb{P} \left(D(\mathbf{x}) > n^{-1/d} u \right) &= \left(1 - f(\mathbf{x}) v_d n^{-1} u^d (1 + o(n^{-1})) \right)^n \\ &= \exp \left[-f(\mathbf{x}) v_d u^d (1 + o(1)) \right] \\ &\rightarrow \exp(-f(\mathbf{x}) v_d u^d) \end{aligned}$$

ce que l'on réécrit

$$\begin{aligned}\mathbb{P}\left(D(\mathbf{x})(f(\mathbf{x})v_d n)^{1/d} > u\right) &\rightarrow \exp(-u^d) \\ \mathbb{P}\left(D^d(\mathbf{x})f(\mathbf{x})v_d n > u\right) &\rightarrow \exp(-u) = 1 - F_{Z(\mathbf{x})}(u)\end{aligned}$$

où $Z(\mathbf{x}) \sim \mathcal{E}(1)$. Donc

$$\lim_{n \rightarrow \infty} \mathcal{L}\left(nD^d(\mathbf{x})\right) = \mathcal{L}\left(\frac{Z(\mathbf{x})}{f(\mathbf{x})v_d}\right).$$

Bibliographie

- [1] T. AMEMIYA. *Advanced econometrics*. Harvard Univ. Pr., Cambridge 1985.
- [2] P. BICKEL, L. BREIMAN. Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. *Annals of Probability* **11** (1983) 185–214.
- [3] P. BILLINGSLEY. *Convergence of probability measures*. John Wiley & Sons Inc., New York 1968.
- [4] M. BONNEU, M. DELECROIX, M. HRISTACHE. Efficient semiparametric estimation in a class of single-index model. Document de travail 9447, CREST, Centre de Recherche en Economie et Statistique 1994.
- [5] D. BOSQ, J.-P. LECOUTRE. *Théorie de l'estimation fonctionnelle*. Ecole Nationale de la statistique et de l'Administration Economique et Centre d'Etudes des Programmes Economiques. Economica, Paris 1987.
- [6] R. CARROLL, J. FAN, I. GIJBELS, M. P. WAND. Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92**, no.438 ((1997)) 477–489.
- [7] H. CHEN, K. W. CHEN. Selection of the splined variables and convergence rates in a partial linear model. *Canadian J. Statist.* **19** (1991) 323–339.
- [8] M. DELECROIX, M. HRISTACHE. M-estimateurs semi-paramétriques dans les modèles à direction révélatrice unique. *Bull. Belg. Math. Soc.* **6** (1999) 161–185.
- [9] M. DELECROIX, M. HRISTACHE, V. PATILÉA. Optimal smoothing in semiparametric index approximation of regression functions. Document de travail 9952, CREST, Centre de Recherche en Economie et Statistique 1999.
- [10] L. DEVROYE, G. LUGOSI. *Combinatorial methods in density estimation*. Springer series in statistics. Springer-Verlag, New York, Inc. 2000.
- [11] R. ENGLE, C. GRANGER, J. RICE, A. WEISS. Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Ass.* **81** (1986) 310–320.
- [12] R. EUBANK, J. HART. Testing goodness-of-fit in regression via order selection criteria. *Ann. Statist.* **20**, no. 3 (1992) 1424–1425.
- [13] R. EUBANK, C. SPIEGELMAN. Testing the goodness of fit of a linear model via nonparametric regression techniques. *J. Amer. Stat. Ass.* **85** (1990) 387–392.

-
- [14] J. FAN. Test of significance based on wavelet thresholding and neyman's truncation. *J. Amer. Statist. Ass.* **91** (1996) 674–688.
- [15] J. FAN, I. GIJBELS. *Local polynomial modelling and its applications*. Chapman and Hall, London 1996.
- [16] Y. FAN, Q. LI. Consistent model specification tests : omitted variables and semiparametric functional forms. *Econometrica* **64** (1996) 865–890.
- [17] J. FRIEDMAN, W. STUETZLE. Projection pursuit regression. *Journal of the American Statistical Association* **76** (1981) 817–823.
- [18] J. FRIEDMAN, J. TUKEY. A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers* **C-23** (1974) 881–890.
- [19] P. HALL. Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of multivariate analysis* **14** (1984) 1–16.
- [20] P. HALL. Integrated square error properties of kernel estimators of regression functions. *The Annals of Statistics* **12** (1984) 241–260.
- [21] P. HALL. On projection pursuit regression. *The Annals of Statistics* **17** (1989) 573–588.
- [22] P. HALL. *The bootstrap and edgeworth expansion*. Springer series in Statistics. Springer, New York Berlin Heidelberg 1992.
- [23] P. HALL, C. HEYDE. *Martingale limit theory and its application*. Probability and mathematical statistics. Academic press, New York 1980.
- [24] W. HÄRDLE, P. HALL, H. ICHIMURA. Optimal smoothing in single-index models. *The annals of Statistics* **21** (1993) 157–178.
- [25] W. HÄRDLE, J. HOROWITZ. Direct semiparametric estimation of single-index models with discret covariates. *J. Amer. Statist. Assoc.* **91** (1996) 1632–1640.
- [26] W. HÄRDLE, A. KOROSTELEV. Search of significant variables in nonparametric additive regression. *Biometrika* **83** (1996) 541–549.
- [27] W. HÄRDLE, H. LIANG, J. GAO. *Partially linear models*. Physica Verlag and Springer Verlag, Germany 2000.
- [28] W. HÄRDLE, E. MAMMEN. Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **4** (1993) 1926–1947.
- [29] W. HÄRDLE, S. SPERLICH, V. SPOKOINY. Structural tests for additive regression. *J. Amer. Stat. Acc.* **96**, no.456 (2001) 1333–1347.
- [30] W. HÄRDLE, V. SPOKOINY, S. SPERLICH. Semiparametric single index versus fixed link function modelling. *The Annals of Statistics* **25** (1997) 212–243.
- [31] W. HÄRDLE, T. STOKER. Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association* **84** (1989) 986–995.

- [32] J. HART. *Nonparametric smoothing and lack-of-fit tests*. Springer Verlag, New York. 1997.
- [33] T. HASTIE, R. TIBSHIRANI. Generalized linear models. *Statistical Science* **1** (1986) 297–318.
- [34] N. HECKMAN. Spline smoothing in partly linear models. *J. Royal Statist. Soc., Series B* **48** (1986) 244–248.
- [35] J. HOROWITZ, V. SPOKOINY. An adaptative rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69** (2001) 599–631.
- [36] J. HOROWITZ, V. SPOKOINY. An adaptive, rate-optimal test of a parametric mean regression model against a nonparametric alternative. *Econometrica* **69** (2001) 599–631.
- [37] M. HRISTACHE, A. JUDITSKY, J. POLZEHL, V. SPOKOINY. Structure adaptive approach for dimension reduction. *Ann. Statist.* **29, no. 6** (2001) 1537–1566.
- [38] M. HRISTACHE, A. JUDITSKY, V. SPOKOINY. Direct estimation of the index coefficient in a single-index model. *Ann. Statist.* **29, no.3** (2001) 595–623.
- [39] I. IBRAGIMOV, R. HAS’MINSKII. *Statistical estimation : asymptotic theory*. Springer, New York Berlin 1980.
- [40] H. ICHIMURA. Semiparametric least squares and weighted sls estimation of single-index models. *Journal of Econometrics* **58** (1993) 71–120.
- [41] Y. INGSTER. Asymptotically minimax hypothesis testing for nonparametric alternatives i. *Mathematical Methods of Statistics* **2** (1993) 85–114.
- [42] Y. INGSTER. Asymptotically minimax hypothesis testing for nonparametric alternatives ii. *Mathematical Methods of Statistics* **2** (1993) 171–189.
- [43] Y. INGSTER. Asymptotically minimax hypothesis testing for nonparametric alternatives iii. *Mathematical Methods of Statistics* **2** (1993) 249–268.
- [44] T. LEDWINA. Data-driven version of neyman’s smooth test of fit. *J. Amer. Stat. Ass.* **89, no. 427** (1994) 1000–1005.
- [45] H. LIERO. Strong uniform consistency of nonparametric regression function estimates. *Probability theory and related fields* **82** (1989) 587–614.
- [46] E. MAMMEN, S. VAN DE GEER. Penalized estimation in partial linear models. *Ann. Statist.* **25** (1997) 1014–1035.
- [47] J. NELDER, R. WEDDURBURN. Generalized linear models. *J. Royal Statist. Soc.* **135** (1972) 370–384.
- [48] D. NOLAN, D. POLLARD. U-processes : Rates of convergence. *The Annals of Statistics* **15** (1987) 780–799.
- [49] J. OPSOMER, D. RUPPERT. A root-n consistent estimators for semi-parametric additive models. *Journal of Computational and Graphical Statistics* **8** (1999) 715–732.

- [50] A. PAKES, D. POLLARD. Simulation and the asymptotics of optimization estimators. *Econometrica* **57** (1989) 1027–1057.
- [51] D. POLLARD. *Convergence of Stochastic Processes*. Statistics. Springer-Verlag, New York 1984.
- [52] J. POWELL, J. STOCK, T. STOKER. Semiparametric estimation of index coefficient. *Econometrica* **57** (1989) 1403–1430.
- [53] R. SHERMAN. Maximal inequalities for degenerate u-processes with applications to optimization estimators. *The Annals of Statistics* **22** (1994) 439–459.
- [54] T. SHIVELY, R. KOHN, S. WOOD. Variable selection and function estimation in additive nonparametric regression using a data-based prior. *Journal of the American Statistical Association* **94** (1999) 777–794.
- [55] V. SPOKOINY. Data driven testing the fit of linear models. *Math. Methods of Statistics* **10, no. 4** (2001) 465–497.
- [56] V. SPOKOINY. Variance estimation for high-dimensional regression models. *J. of Multivariate Analysis* (2002) 1–23.
- [57] C. STONE. An asymptotically optimal window selection rule for kernel density estimates. *Annals os statistics* **12** (1984) 1285–1297.
- [58] D. STONE. Additive regression and other nonparametric models. *The annals of Statistics* **13** (1985) 535–552.
- [59] W. STUTE. Nonparametric model checks for regression. *Ann. Statist.* **25, no. 2** (1997) 613–641.
- [60] A. VAN DER VAART, J. WELLNER. *Weak convergence and empirical processes, with applications to statistics*. Statistics. Springer, New York 1996.
- [61] M. WAND, M. JONES. *Kernel Smoothing*. Chapman and Hall, London 1995.