



HAL
open science

Estimation non linéaire par ondelettes : régression et survie

Mohamed Bouatou

► **To cite this version:**

Mohamed Bouatou. Estimation non linéaire par ondelettes : régression et survie. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1997. Français. NNT : . tel-00004921

HAL Id: tel-00004921

<https://theses.hal.science/tel-00004921>

Submitted on 20 Feb 2004

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.

THÈSE

présentée par

Mohamed Bouatou

pour obtenir le grade de Docteur
de l'Université de Joseph Fourier de Grenoble
(Arrêtés ministériels du 5 juillet 1984 et
du 30 Mars 1992)

Spécialité : mathématiques appliquées

Estimation non linéaire par ondelettes : Régression et Survie

Soutenue le 28 Mars 1997 devant la commission d'examen :

Bernard Van-Cutsem	président de jury
Christine Thomas-Agnan	rapporteurs
Pierre Jacob	
Gérard Grégoire	examineurs
Pierre Vial	
A. Antoniadis	directeur de thèse

Mes remerciements vont à ma mère, Lalla Touria el Ouazzani, qui m'a appris que le respect d'autrui est une vertu, à mon père, Mohamed-Hosseïn, qui m'a élevé dans la dignité, à ma soeur Amina et mes deux frères Zakaria et Nasser pour leur générosité, à mes oncles Ahmed, Kacem et Abdesslam pour leurs in-tarissables encouragements, à mes enseignants de mathématique qui m'ont fait partager leur passion pour les maths et à messieurs Anestis Antoniadis et Marc Halin pour leur confiance.

Bâtir une théorie avant d'avoir des données est une erreur monumentale : insensiblement on se met à torturer les faits pour qu'ils collent avec la théorie alors que ce sont les théories qui doivent coller avec les faits.

Conan Doyle.

Table des matières

I	Régression et ondelettes	7
1	Régression par ondelettes de Haar	9
1.1	Régressogramme et partition dyadique	12
1.2	Estimateur basé sur le système de Haar	13
1.3	Convergence des estimateurs	15
1.3.1	Remarque	20
1.4	Construction de la base optimale	20
1.5	Simulations	23
1.6	Lien avec l'analyse multirésolution	34
2	Régression par paquets d'ondelettes	37
2.1	Régression par ondelettes sur l'intervalle	38
2.1.1	Approximation de la fonction de régression	39
2.1.2	Estimation de la fonction de régression	42
2.2	Régression par paquets d'ondelettes	46
2.2.1	Décomposition en paquets d'ondelettes d'une fonction	48
2.3	Généralisation de l'Algorithme 1	51
2.3.1	Description de la procédure	51
2.3.2	Simulations	53
2.4	Tests et Seuillages	58
2.4.1	Simulations	62
2.5	Estimation par Bootstrap	67
2.5.1	Généralisation	70
3	Ondelettes et problèmes de survie	81
3.1	Écriture de l'estimateur et erreur quadratique	83
3.2	Comportement asymptotique du MISE de $\hat{\lambda}_p$	89
3.3	Estimation non linéaire par seuillage uniforme	93
3.4	Décomposition du MISE de $\hat{\lambda}$	94
3.5	Comportement de la partie linéaire	95

3.6	Comportement de la partie non linéaire	95
3.6.1	Quelques résultats	95
3.6.2	Comportement asymptotique de s_1	96
3.6.3	Comportement asymptotique de $(d_{kl} - \hat{d}_{kl})^2$	97
3.6.4	Comportement asymptotique de s_2	98
3.6.5	Comportement asymptotique de s_3	101
3.7	Simulations numériques	105
3.7.1	Comportement asymptotique du MISE	105
3.7.2	Quelques graphiques	107

Annexes **117**

A Rappels sur les paquets d'ondelettes **117**

A.1	Introduction aux ondelettes	117
A.2	Analyse multirésolution (AMR) et ondelettes	118
A.2.1	Analyse multirésolution	119
A.2.2	Ondelettes	120
A.2.3	Remarque	120
A.3	Filtres miroirs en quadrature (QMF) et ondelettes	122
A.3.1	Exemple	124
A.4	Base de paquets d'ondelettes	125
A.4.1	Aspect séparatif des QMF	125
A.4.2	Paquets d'ondelettes	127
A.5	Analyse multirésolution sur l'intervalle	130
A.5.1	Ondelettes sur l'intervalle	130
A.5.2	Construction des bases de paquets d'ondelettes sur l'intervalle	132
A.6	Aspect numérique	133

Bibliographie **135**

Liste des figures

1.1	Partition dyadique associée à $r = 12$	12
1.2	Arbre binaire associé à la procédure.	23
1.3	La fonction de régression est donnée par le modèle 1 Elle est représentée en trait continu et son estimateur en pointillé.	28
1.4	La fonction de régression est donnée par le modèle 2 Elle est représentée en trait continu et son estimateur en pointillé.	29
1.5	La fonction de régression est donnée par le modèle 3 Elle est représentée en trait continu et son estimateur en pointillé.	30
1.6	La fonction de régression est la fonction $f(x) = -15(x - 1)(x - 0.5)x + 1.5$. Dans ce cas r est égal à 8.	31
1.7	Dans ce cas r est égal à 16.	32
1.8	La fonction de régression est représentée en trait continu et son estimateur en pointillé. Dans ce cas r est égal à 32.	33
2.1	Décomposition en paquets d'ondelettes à l'ordre l	50
2.2	Arbre binaire associé à la procédure.	52
2.3	Estimation obtenue pour une décomposition au niveau 3 avec le même niveau de bruit et le même nombre de points échantillons que la figure 1.3 du premier chapitre.	54
2.4	Estimation obtenue pour une décomposition au niveau 4 de la multirésolution. Le nombre de points échantillons N est égal à 256 et le rapport signal bruit est égal à 2.	55
2.5	Estimation obtenue par les ondelettes de Daubechies d'ordre 3 avec un échantillon de taille 256.	56
2.6	256 points échantillons ont été utilisés avec un rapport signal-bruit égal à 2.	56
2.7	Estimation obtenue par les ondelettes de Daubechies d'ordre 6.	57
2.8	256 points échantillons ont été utilisés avec un rapport signal-bruit égal à 2.	57
2.9	La valeur du risque a été fixée à 0.01.	64
2.10	Pour ces deux figures le risque est égal à 0.1.	64

2.11	La valeur du risque est égale à 0.01.	65
2.12	Tout comme les figures ci-dessus le nombre des observations est égal à 256. Le risque du test est égal à 0.1.	65
2.13	Le nombre d'observations est égal à 256 alors que le rapport signal/bruit est égal à 1.	66
2.14	Pour un rapport signal/bruit=1 la taille de l'échantillon est égale à 512.	66
3.1	Fonctions de taux de hasard Γh_1 et h_2	107
3.2	Estimations du taux de hasard par logspline ($N = 256$). . .	108
3.3	Estimations du taux de hasard par ondelette ($N = 256$ $r = 4$ pour h_1 et $r = 4$ pour h_2).	108
3.4	Estimations du taux de hasard par logspline ($N = 512$). . .	109
3.5	Estimations du taux de hasard par ondelette ($N = 512$ $r = 5$ pour h_1 et $r = 4$ pour h_2).	109
3.6	Estimations du taux de hasard par logspline ($N = 1024$). . .	110
3.7	Estimations du taux de hasard par ondelette ($N = 1024$ $r = 5$ pour h_1 et $r = 4$ pour h_2).	110
3.8	La fonction $f(1 - F)^{-1}$ associée à une loi exponentielle de paramètre 1 censurée à droite par une loi uniforme sur $[0, 1.5]$ (50% des données sont censurées).	111
3.9	Estimations de la fonction ci-dessus par ondelettes de Daubechies d'ordre 3 avec $N = 512$ et d'ordre 6 avec $N = 1024$ respectivement.	111
3.10	Estimation par ondelettes de Daubechies d'ordre 3 avec $N = 1024$ du taux de hasard d'une loi exponentielle de paramètre 1 censurée à droite par une loi uniforme sur $[0, 1.5]$ ($\approx 50\%$ des données sont censurées).	112
3.11	Estimation par ondelettes de Daubechies d'ordre 3 avec $N = 2048$ du taux de hasard d'une loi exponentielle de paramètre 1 censurée à droite par une loi uniforme sur $[0, 1.5]$ ($\approx 52\%$ des données sont censurées).	112
A.1	Décomposition en paquets d'ondelettes à l'ordre j de V_j . . .	129
A.2	Décomposition en ondelettes à l'ordre j de V_j	130

Liste des tables

1.1	Estimation de la moyenne ($\times 10^6$) et du carré de l'erreur standard pour différentes simulations. La variance du bruit Gaussien utilisé dans les simulations est égale à 10^{-4}	25
1.2	Estimation de la moyenne et du carré de l'erreur standard pour différentes simulations. La variance du bruit Gaussien utilisé dans les simulations est égale à 4×10^{-2}	26
1.3	Estimation de la moyenne et du carré de l'erreur standard pour différentes simulations. La variance du bruit Gaussien utilisé dans les simulations est égale à 2.25.	27
2.1	Résultats des simulations pour la fonction de régression f_1	63
2.2	Résultats des simulations pour la fonction de régression f_2	63
3.1	Résultats des simulations pour la fonction de taux de hasard associée à la loi normale.	106
3.2	Résultats des simulations pour la fonction de taux de hasard associée à la loi exponentielle.	106

Introduction générale

L'analyse des résultats d'une expérience aléatoire relève de l'emploi d'un modèle de régression lorsque les observations qui en sont issues peuvent être représentées chacune comme la somme d'un terme systématique dépendant de la valeur prise par une ou plusieurs autres variables et de la réalisation d'une variable aléatoire appelée erreur.

Lorsqu'une expérience est destinée à étudier une telle représentation le résultat est en général une série de valeurs de la variable réponse obtenues pour des valeurs connues de la variable explicative soit $(y_1, x_1) \dots (y_n, x_n)$. En supposant que les quantités observées y_1, \dots, y_n sont la réalisation d'un vecteur aléatoire \mathbf{Y} dont chacune des composantes peut s'écrire sous la forme $Y_i = f(x_i) + \varepsilon_i$ la définition d'un modèle statistique de régression équivaut à la donnée d'un ensemble \mathcal{F} de fonctions auquel appartient la fonction de régression inconnue f et d'un ensemble de lois de probabilité auquel appartient la loi du vecteur des erreurs $\varepsilon_1 \dots \varepsilon_n$.

Bien évidemment la définition d'un modèle statistique de régression que nous avons rappelé ci dessus est trop générale pour que l'on puisse proposer un seul type de méthodes d'analyse des données relevant de tels modèles. Afin de définir le cadre de l'étude que nous avons entrepris dans la première partie de cette thèse nous avons imposé plusieurs restrictions. Nous avons entre autres supposé que

- les variables observées sont continues c'est à dire qu'elles prennent leurs valeurs dans des intervalles de \mathbb{R} ;
- les valeurs des variables explicatives sont fixées ;
- les diverses observations de la variable réponse sont indépendantes.

De plus nous nous sommes restreints au cas d'une seule variable explicative. La loi de probabilité des ε_i sera une loi sur \mathbb{R} centrée et de variance σ^2 finie. Les méthodes que nous avons développées relèvent du cadre de la régression non paramétrique dans lequel il est possible de définir des contraintes très faibles sur la forme de la fonction de régression f . Ainsi l'ensemble \mathcal{F}

de fonctions dans lequel nous recherchons la fonction de régression est de dimension infinie.

Il existe bien sûr de nombreuses méthodes pour l'estimation non paramétrique de la fonction de régression. Les plus célèbres étant les méthodes par noyau, par fonction spline et par séries orthogonales. L'objectif de la première partie de cette thèse est de présenter une méthode particulière par schéma récursif associé à un arbre binaire fondée sur des décompositions en ondelettes conduisant à la définition d'estimateurs ayant des qualités analogues à celles des estimateurs obtenus par les méthodes citées ci-dessus et qui de plus a l'avantage d'être d'utilisation très simple.

La méthodologie d'estimation par arbres de régression (CART) initialement développée par Breiman et al [10] a été largement utilisée au cours de ces dernières années. Reposant sur des idées de partitions récursives de l'espace des variables explicatives elle permet de construire une estimation constante par morceaux de la fonction de régression inconnue. Les morceaux correspondants aux nœuds terminaux d'une partition optimale. L'estimateur obtenu n'est autre qu'un régressogramme calculé sur une partition adaptative de l'espace des observations. Dans notre cas la variable explicative n'est pas aléatoire ainsi la recherche d'une partition adaptée de l'espace des observations est remplacée par celle d'une décomposition adaptée de l'espace de définition de la fonction de régression. Il est alors facile de voir que des estimateurs de la régression fondés sur de tels schémas de partition dyadique récursive peuvent être considérés comme des estimateurs ondelettes d'un type particulier avec pour système d'ondelettes le système de Haar. Le chapitre 1 de cette thèse est consacré à établir cette analogie et par là même de proposer un nouvel estimateur non paramétrique de la régression possédant de bonnes propriétés théoriques et dont le calcul numérique est très rapide.

L'estimation de la fonction de régression à l'aide d'une décomposition de Haar a le désavantage de produire un estimateur de forme étagée même lorsque celle-ci est supposée être lisse. Cela provient essentiellement du fait que le système de Haar est composé de fonctions peu régulières. Dans le cadre du traitement des signaux acoustiques une autre méthode fondée également sur des partitions récursives du plan temps-fréquence à l'aide de paquets d'ondelettes a été récemment développée par Coifman et al [19]. Elle permet entre autres de développer pour l'analyse d'un signal donné une famille de bases de fonctions orthogonales bien plus lisses que les fonctions du système de Haar. L'objet du deuxième chapitre de la thèse est d'établir une relation entre ce type de méthodes et la régression par arbres fondée cette fois-ci sur les paquets d'ondelettes. Les algorithmes de calculs sont

essentiellement les mêmes que ceux du chapitre 1. Cela donc permet un calcul de complexité raisonnable pour la recherche du meilleur estimateur issu de cette méthode. De plus, les estimations obtenues sont cette fois beaucoup plus lisses que celle obtenue par la régression à l'aide du système de Haar.

Les relations établies entre CART et décomposition en paquets d'ondelettes permettent de revoir les méthodes de partitions récursives sous un nouvel aspect. Récemment Donoho et Johnstone [25] ont développé des méthodes adaptatives de choix d'une base pour l'estimation d'une fonction de régression. Nous avons adapté ces idées dans le cadre des partitions récursives obtenant ainsi des résultats nouveaux sur les méthodes de partitions binaires récursives dans le chapitre 2. Dans ce même chapitre nous montrons également que des techniques d'échantillonnage de type Bootstrap peuvent être une solution satisfaisante du problème empirique du choix optimal de l'arbre de régression.

L'ensemble des notions et des résultats de cette première partie repose sur une utilisation constante de la théorie des ondelettes et des paquets d'ondelettes. Pour assurer la continuité de l'exposé nous avons donc préféré reporter en annexe A de cette thèse les pré-requis sur les ondelettes nécessaires à la compréhension de ce travail.

La deuxième partie de la thèse est consacrée à l'étude de l'estimation non paramétrique d'une fonction de hasard par des méthodes d'ondelettes. Parmi les méthodes récentes pour l'analyse non paramétrique des données de survie la méthode développée par Kooperberg et Stone [45] est une des plus prometteuses. Les résultats obtenus sont en général de très bonne qualité mais ceci au prix d'un algorithme de grande complexité qui parfois peut diverger. S'appuyant sur la simplicité et la rapidité avec laquelle sont obtenus les estimateurs par décomposition en ondelettes nous avons donc étendu les méthodes d'estimation par ondelettes au cas des données de survie censurées ou non. Les résultats théoriques démontrant les bonnes propriétés des estimateurs ainsi obtenus font l'objet du dernier chapitre de cette thèse.

Notations et abréviations

x	un scalaire réel.
\mathbf{x}	un vecteur réel.
$[x]_+$	le sup de x et 0 .
$ \mathbf{x} _{2,N}^2$	$= \frac{1}{N} \sum_{i=1}^N x_i^2$.
I_N	La matrice identité d'ordre N .
$[\beta]$	partie entière de β .
$\ f\ _2$	$(\int f^2(x) dx)^{\frac{1}{2}}$.
$MISE(\hat{f})$	$E \int (\hat{f} - f)^2(x) dx$ $\Gamma \hat{f}$ étant une estimation de f .
$u_N = \mathcal{O}(v_N)$	$ \frac{u_N}{v_N} $ reste borné quand N tend vers l'infini.
$u_N = o(v_N)$	$\lim_{N \rightarrow \infty} \frac{u_N}{v_N} = 0$.

Première partie

Régression et ondelettes

Chapitre 1

Régression par ondelettes de Haar

Le modèle de régression que nous étudions dans ce chapitre peut être mathématiquement décrit par les équations et les définitions ci-après. Supposons que le vecteur observé (y_1, y_2, \dots, y_N) soit une réalisation du vecteur aléatoire (Y_1, Y_2, \dots, Y_N) dont chacune des composantes s'écrit sous la forme

$$Y_i = f(x_i) + \varepsilon_i, \quad i = 1, 2, \dots, N, \quad (1.1)$$

où les “instants” d'observation x_i sont connus (le plan d'expérience est déterministe) f est une fonction régulière d'une variable x dans ce premier chapitre cela veut dire Lipschitzienne et où les termes d'erreur ε_i sont des variables aléatoires réelles indépendantes centrées et de même variance σ^2 . Sans perte de généralité nous supposons que les x_i sont dans $[0, 1]$ et nous supposons de plus qu'ils sont équirépartis.

Notre objectif est d'étudier le comportement asymptotique de deux estimateurs particuliers de la fonction de régression inconnue f l'un de type régressogramme et l'autre de type projection. De tels estimateurs ont été considérés et étudiés par plusieurs auteurs dont Čencov [13] Rafajlowicz [59] Cieselski [15] Bosq [9] Collomb [20] Scott [65] pour ne citer que ceux là.

Afin d'établir un lien entre les deux types d'estimation le régressogramme sera défini sur une partition dyadique de l'intervalle $[0, 1]$ le support de f et la base orthogonale qui conduira par la suite à la définition de l'estimateur de projection sera la base de Haar de $L^2([0, 1])$. Ce choix particulier permettra ainsi d'étudier conjointement le comportement asymptotique du MISE (erreur quadratique moyenne intégrée) des deux estimateurs.

Les régressogrammes ont été au même titre que les histogrammes puisqu'ils s'y apparentent largement abordés dans la littérature et ce par plusieurs auteurs. Ils sont basés essentiellement sur la donnée d'une partition du support de la fonction de régression. En effet soit une partition de $[0, 1]$ en K_N intervalles nous estimons sur $[0, 1]$ la fonction de régression par une fonction qui dans chaque intervalle de la partition est constante. Quand chaque constante est égale à la moyenne des observations y_i dont les instants x_i appartiennent à l'intervalle considéré le régressogramme est dit associé à la moyenne arithmétique et quand la longueur des éléments de la partition est la même nous parlons de régressogramme uniforme.

Dans le cas d'un plan d'expérience déterministe les estimateurs de type série orthogonale (ou projection) furent introduits par Čencov [13]. L'idée sous jacente à ce type d'estimation est de développer la fonction de régression en une série de fonctions orthogonales ω_ν où $\nu \in \mathbb{N}$

$$f(\cdot) \sim \sum_{\nu=1}^{\infty} \left(\int f(u) \omega_\nu(u) du \right) \omega_\nu(\cdot) = \sum_{\nu=1}^{\infty} \langle f, \omega_\nu \rangle \omega_\nu(\cdot),$$

puis de remplacer les coefficients de Fourier de ce développement par leurs estimateurs

$$\langle \widehat{f}, \widehat{\omega}_\nu \rangle = \frac{1}{N} \sum_{i=1}^N y_i \omega_\nu(x_i).$$

Nous définissons alors un estimateur de f du type

$$\widehat{f}_r(\cdot) = \sum_{\nu=1}^r \langle \widehat{f}, \widehat{\omega}_\nu \rangle \omega_\nu(\cdot), \quad (1.2)$$

où r détermine le nombre de fonctions orthogonales du développement retenu pour l'estimation de f .

Le choix de r revêt une importance considérable pour l'efficacité de l'estimateur. Ainsi un choix de r trop grand implique la présence de fluctuations aléatoires trop importantes (l'estimateur a une grande variance) alors qu'un choix de r trop petit limite les aléas mais introduit des biais d'approximation

trop importants (l'estimateur est fortement biaisé). Ainsi divers auteurs ont abordé le problème du choix approprié de ce paramètre de lissage : nous citerons par exemple Tarter et Kronmal [70] et Rutkowski [62]. Nous ouvrons une parenthèse pour signaler que très souvent le choix de r est fait de sorte à minimiser asymptotiquement l'erreur quadratique moyenne intégrée (MISE).

Dans le cadre des ondelettes de Haar cas que nous allons étudier nous utiliserons la méthodologie introduite par Bickel [7] puis réactualisée par Donoho et al [27] qui consiste non seulement à tronquer le développement mais également à modifier ou éliminer de manière adéquate certains des coefficients empiriques du développement tronqué obtenu par (1.2) et ce afin de construire une estimation optimale et adaptée. Cela revient à choisir une sous famille de fonctions orthogonales contenue dans la base tronquée $S(r) = \{\omega_1, \omega_2, \dots, \omega_r\}$ qui a servi à définir l'estimateur $\hat{f}_r(\cdot)$. Afin de résoudre le choix optimal de S nous nous proposons par la suite d'utiliser les méthodes de régression par arbres binaires variantes adaptées à notre problème de techniques de régression telles que CART [10] ou MARS [34].

Avant de poursuivre nous allons fixer le cadre et les notations nécessaires pour la suite. Nous supposons donc et ce tout au long de ce chapitre que la fonction inconnue de régression f est à support dans $[0, 1]$ et qu'elle est de carré intégrable. Nous noterons par $L^2([0, 1])$ l'espace de Hilbert réel des fonctions réelles définies sur $[0, 1]$ et de carré intégrable.

A toute partition dyadique de l'intervalle $[0, 1]$ nous associerons un estimateur du type régressogramme qu'on notera par \hat{f}^{reg} . D'autre part nous associerons à toute partition de ce type une famille particulière d'ondelettes en identifiant l'ondelette à son support. Cette famille d'ondelettes est celle de Haar. Les ondelettes de Haar permettent d'obtenir une base orthonormée de $L^2([0, 1])$ et conduisent à la définition d'un estimateur de type projection \hat{f}_r .

Nous introduirons dans les deux premiers paragraphes de ce chapitre respectivement les estimateurs \hat{f}^{reg} et $\hat{f}_r(\cdot)$ puis dans le paragraphe 1.3 nous établirons un premier résultat prouvant que les deux estimateurs en question sont asymptotiquement équivalents et nous calculerons la vitesse de convergence vers 0 de leur MISE. Ceci nous permettra d'en déduire la valeur asymptotiquement optimale de r correspondant à la valeur minimum du MISE. Partant de la famille $S(r^*)$ nous donnerons un algorithme qui sera associé à un schéma récursif et à un critère de sélection du type Tarter

et Kronmal permettant un choix optimal au sens défini précédemment de la sous famille optimale de $S(r^*)$.

Le dernier paragraphe regroupera un ensemble de simulations commentées.

1.1 Régressogramme et partition dyadique

L'intérêt des estimateurs de type régressogramme réside en la simplicité de leur écriture mais ceci ne diminue en rien leur qualité. En effet la vitesse de convergence vers 0 de l'erreur de tels estimateurs est comparable à celle obtenue par des méthodes plus sophistiquées et plus complexes et ce lorsque peu de contraintes de régularité sont imposées à la vraie fonction de régression. Ci-après nous définissons le régressogramme usuel associé à la moyenne arithmétique tel qu'il est défini par G. Collomb [20] et obtenu à partir d'une partition dyadique. Cet estimateur sera noté tout au long de ce chapitre par \hat{f}^{reg} .

Pour tout entier naturel non nul r il existe un unique couple d'entiers (μ_r, δ_r) tel que $r = 2^{\mu_r} + \delta_r$ avec $\delta_r < 2^{\mu_r}$. Partant de cette remarque nous associons à tout r une partition de $[0, 1]$ en intervalles dyadiques par

$$\varphi(r) = \cup_{j=1}^r [\tau_{j-1}, \tau_j),$$

avec

$$\tau_j = \begin{cases} \frac{j}{2^{\mu_r+1}} & \text{si } j = 0, 1, \dots, 2\delta_r \\ \frac{j-\delta_r}{2^{\mu_r}} & \text{si } j = 2\delta_r + 1, \dots, r. \end{cases}$$

A titre d'exemple et pour mieux comprendre la suite considérons le cas $r = 12$. Il lui correspond $\mu_r = 3$ et $\delta_r = 4$. La partition associée est illustrée par le graphique suivant

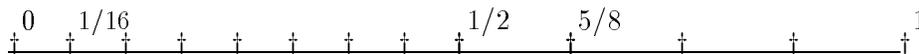


Figure 1.1 : Partition dyadique associée à $r = 12$.

Les points représentés sur ce segment de droite sont les différents τ_j , $j = 0 \dots 12$.

Pour une partition dyadique $\varphi(r)$ de $[0, 1]$ et x dans l'intervalle $[0, 1]$ nous désignerons par $I(x)$ l'élément de $\varphi(r)$ contenant x . La densité des

“instants” d’observation sera définie par

$$\hat{F}_N(x) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{I(x)}(x_i),$$

où $\mathbb{1}_A(\cdot)$ désigne la fonction indicatrice de l’ensemble A .

L’estimateur \hat{f}^{reg} est alors défini en tout point x tel que $\hat{F}_N(x) \neq 0$ par

$$\hat{f}^{reg}(x) = \frac{1}{N \hat{F}_N(x)} \sum_{i=1}^N Y_i \mathbb{1}_{I(x)}(x_i).$$

Afin que \hat{f}^{reg} soit bien défini en tout point x de $[0, 1]$ il faut au moins autant d’intervalles que de points Γ ainsi r doit vérifier $2^{\mu_r+1} < N$ (μ_r est tel que $r = 2^{\mu_r} + \delta_r$) Γ condition que nous imposerons tout au long de ce chapitre.

Pour l’étude qui suivra au paragraphe 1.3 Γ nous préférons l’écriture suivante pour \hat{f}^{reg} Γ

$$\hat{f}^{reg}(x) = \frac{1}{N} \sum_{j=1}^r \frac{1}{\hat{F}_N(x)} \left\{ \sum_{i=1}^N Y_i \mathbb{1}_{[\tau_{j-1}, \tau_j)}(x_i) \right\} \mathbb{1}_{[\tau_{j-1}, \tau_j)}(x). \quad (1.3)$$

Maintenant que nous avons défini l’estimateur de type régressogramme Γ nous ferons de même Γ dans le paragraphe suivant Γ avec l’estimateur de type projection Γ fondé sur la base de Haar.

1.2 Estimateur basé sur le système de Haar

Avant de donner l’estimateur de type projection fondé sur le système de Haar Γ nous rappelons brièvement la construction de la base orthonormée de Haar dans $L^2([0, 1])$. C’est l’exemple le plus simple de base orthonormée d’ondelette construite à partir de la fonction d’échelle $\varphi = \mathbb{1}_{[0,1]}$. A cette fonction d’échelle est associée la fonction $\psi = \varphi(2\cdot) - \varphi(2\cdot - 1)$ qui n’est autre que l’ondelette mère de Haar qui prend la valeur 1 sur la première moitié $[0, \frac{1}{2})$ et -1 sur la deuxième moitié $[\frac{1}{2}, 1]$; nous construisons par dilatation puis par translation les fonctions

$$\psi_\mu^\delta = 2^{\mu/2} \psi(2^\mu \cdot - \delta),$$

pour tout couple $(\mu, \delta) \in \mathbb{Z} \times \mathbb{Z}$ à partir de la seule fonction ψ .

Il est facile de voir que le système de fonctions $\{\psi_\mu^\delta, \mu, \delta \in \mathbb{Z}\}$ est un système orthonormé et complet dans $L^2(\mathbb{R})$: le système de Haar est une base

orthonormée de $L^2(\mathbb{R})$. Cependant il est construit sur une ondelette mère peu régulière puisque discontinue. Nous utiliserons au deuxième chapitre des bases d'ondelettes plus régulières.

Pour en déduire un système orthonormé complet dans $L^2([0, 1])$ il suffit de prendre les restrictions des ψ_μ^δ à l'intervalle $[0, 1]$; ceci revient à considérer la famille $\mathcal{B} = \{\omega_\nu, \nu = 1, 2, \dots, \infty\}$ dont chaque élément est défini par :

$$\forall \nu \in \mathbb{N}^*,$$

$$\begin{aligned} \omega_1 &= \mathbb{1}_{[0,1]}, \\ \omega_2 &= \mathbb{1}_{[0, \frac{1}{2}]}(x) - \mathbb{1}_{[\frac{1}{2}, 1]} = \psi, \\ \text{et } \forall \nu \geq 2, \omega_{\nu+1} &= \psi_\nu = 2^{\mu/2} \omega_2(2^\mu \cdot -\delta), \end{aligned}$$

où pour $\nu = 2^\mu + \delta$ nous posons :

$$\psi_\nu = \psi_\mu^\delta.$$

Le système $\{\omega_\nu, \nu \in \mathbb{N}^*\}$ constitue une base orthonormée de $L^2([0, 1])$. D'ailleurs sa construction repose de façon centrale sur la notion d'analyse de multirésolution qui préconise que tout signal peut être construit par raffinements successifs c'est à dire par l'ajout de détails à une première approximation. Ainsi soient $V_0^{[0,1]}$ le sous espace engendré par $\omega_1(x)$ l'espace d'approximation et $W_j^{[0,1]}$ celui engendré par $\{\omega_\nu, \nu = 2^j + 1, \dots, 2^{j+1}\}$ l'espace de détails. Le choix du système de base \mathcal{B} revient à la décomposition de $L^2([0, 1])$ en somme directe de $V_0^{[0,1]}$ et $W_j^{[0,1]}$ $j = 0, 1, \dots, \infty$ associés à la multirésolution en ondelettes de Haar sur l'intervalle $[0, 1]$ [voir annexe A].

Les notations étant fixées nous désignerons par $S(r)$ à la fois le système et l'espace engendré par $\{\omega_\nu, \nu = 1 \dots r\}$ et par $f_r(\cdot)$ la projection de $f(\cdot)$ sur cet espace. Nous posons ci-après pour $\nu = 1, 2, \dots, r$

$$a_\nu = \langle f, \omega_\nu \rangle = \int_0^1 f(x) \omega_\nu(x) dx,$$

les coefficients de la décomposition de f sur $S(r)$.

En estimant chaque coefficient par

$$\hat{a}_\nu = \frac{1}{N} \sum_{i=1}^N y_i \omega_\nu(x_i),$$

l'estimation de $f_r(x)$ basée sur l'échantillon $y_1, y_2 \dots y_N$, est alors donnée par

$$\begin{aligned} \hat{f}_r(x) &= \sum_{\nu=1}^r \hat{a}_\nu \omega_\nu(x) \\ &= \frac{1}{N} \sum_{i=1}^N y_i \sum_{\nu=1}^r \omega_\nu(x_i) \omega_\nu(x). \end{aligned}$$

Notons que \hat{f}_r n'est autre qu'un estimateur de type noyau généralisé. En effet en posant $K_r(x', x) = \sum_{\nu=1}^r \omega_\nu(x') \omega_\nu(x)$ nous avons l'expression suivante pour \hat{f}_r

$$\hat{f}_r(x) = \frac{1}{N} \sum_{i=1}^N y_i K_r(x_i, x).$$

Cette écriture permet à présent d'établir le lien entre \hat{f}_r et l'estimateur \hat{f}^{reg} défini au paragraphe précédent.

Comme pour r donné nous avons

$$K_r(x', x) = \begin{cases} \frac{1}{\tau_j - \tau_{j-1}} & \text{si } x' \text{ et } x \in [\tau_{j-1}, \tau_j) \\ 0 & \text{sinon} \end{cases}$$

les τ_j étant les instants associés à la partition $\varphi(r)$ définie dans le paragraphe 1.1 une écriture analogue à celle obtenue pour \hat{f}^{reg} (voir (1.3)) peut être alors établie pour \hat{f}_r . En effet

$\forall x \in [0, 1]$

$$\hat{f}_r(x) = \frac{1}{N} \sum_{j=1}^r \frac{1}{\tau_j - \tau_{j-1}} \left\{ \sum_{i=1}^N y_i \mathbb{1}_{[\tau_{j-1}, \tau_j)}(x_i) \right\} \mathbb{1}_{[\tau_{j-1}, \tau_j)}(x). \quad (1.4)$$

Ainsi si x appartient à $[\tau_{j-1}, \tau_j)$ nous avons

$$\hat{f}_r(x) = \frac{1}{N(\tau_j - \tau_{j-1})} \sum_{i=1}^N y_i \mathbb{1}_{[\tau_{j-1}, \tau_j)}(x_i),$$

alors que

$$\hat{f}^{reg}(x) = \frac{1}{N\hat{F}_N(x)} \sum_{i=1}^N y_i \mathbb{1}_{[\tau_{j-1}, \tau_j)}(x_i).$$

De plus ces deux expressions seront identiques sur $[\tau_{j-1}, \tau_j)$ chaque fois que $\hat{F}_N(x) = \tau_j - \tau_{j-1}$.

1.3 Convergence des estimateurs

Dans les deux paragraphes précédents nous avons défini \hat{f}^{reg} et \hat{f}_r et établi le lien qui existe entre ces deux estimateurs. Nous nous proposons maintenant d'étudier leurs qualités. Nous montrerons ci-après que ces deux estimateurs partagent asymptotiquement les mêmes propriétés puis nous établirons des conditions suffisantes sur r afin que \hat{f}^{reg} et \hat{f}_r coïncident sur le domaine de définition de f ($[0, 1]$ pour le cas étudié).

Nous nous concentrons d'abord sur l'erreur quadratique moyenne intégrée (MISE) des deux estimateurs. Le lemme 1.1 nous permettra par la suite de réduire le calcul du MISE de l'un ou l'autre des deux estimateurs au calcul du MISE d'un seul d'entre eux.

Dans ce qui suit nous notons par I_j l'ensemble des indices i tels que les x_i correspondants sont dans $[\tau_{j-1}, \tau_j]$ (i.e $I_j = \{i; x_i \in [\tau_{j-1}, \tau_j]\}$) Les τ_j étant associés à la partition $\wp(r)$ et le cardinal de I_j sera noté $\#(I_j)$. Il est aisé de voir que

$$\forall x \in [\tau_{j-1}, \tau_j] \quad N\hat{F}_N(x) = \#(I_j).$$

D'autre part pour toutes suites $(u_n)_{n \in \mathbb{N}}$ et $(v_n)_{n \in \mathbb{N}}$ telles que $|\frac{u_n}{v_n}|$ reste borné quand n tend vers l'infini nous noterons $u_n = \mathcal{O}(v_n)$.

Lemme 1. 1. *Si f appartient à $L^2([0, 1])$, alors*

$$E \int_0^1 [\hat{f}^{reg}(x) - \hat{f}_r(x)]^2 dx = \mathcal{O}\left(\frac{r^2}{N^2}\right).$$

Démonstration

Posons

$$\begin{aligned} Z^{(j)} &= \left(\frac{1}{N(\tau_j - \tau_{j-1})} - \frac{1}{\#(I_j)}\right) \sum_{i=1}^N Y_i \mathbb{1}_{[\tau_{j-1}, \tau_j]}(x_i) \\ &= \left(\frac{1}{N(\tau_j - \tau_{j-1})} - \frac{1}{\#(I_j)}\right) \sum_{\{i \in I_j\}} Y_i. \end{aligned}$$

Par (1.3) et (1.4) nous obtenons

$$\hat{f}_r(x) - \hat{f}^{reg}(x) = \sum_{j=1}^r Z^{(j)} \mathbb{1}_{[\tau_{j-1}, \tau_j]}(x).$$

L'orthogonalité des $\mathbb{1}_{[\tau_{j-1}, \tau_j]}$ entraîne que

$$\int_0^1 \left(\sum_{j=1}^r Z^{(j)} \mathbb{1}_{[\tau_{j-1}, \tau_j]}(x)\right)^2 dx = \sum_{j=1}^r (\tau_j - \tau_{j-1}) [Z^{(j)}]^2.$$

Ainsi

$$\begin{aligned} E \int_0^1 (\hat{f}^{reg} - \hat{f}_r)^2(x) dx &= \sum_{j=1}^r (\tau_j - \tau_{j-1}) E[Z^{(j)}]^2 \\ &= \sum_{j=1}^r (\tau_j - \tau_{j-1}) (var[Z^{(j)}] + E^2[Z^{(j)}]). \end{aligned}$$

Pour tout $j = 1 \dots r$ nous avons

$$var[Z^{(j)}] = \sigma^2 \sum_{\{i \in I_j\}} \left[\frac{1}{\#(I_j)} - \frac{1}{N(\tau_j - \tau_{j-1})}\right]^2.$$

Comme pour $N \rightarrow \infty$

$$\begin{aligned} \left| \frac{1}{\#(I_j)} - \frac{1}{N(\tau_j - \tau_{j-1})} \right| &= \left| \frac{N(\tau_j - \tau_{j-1}) - \#(I_j)}{\#(I_j)N(\tau_j - \tau_{j-1})} \right| \\ &\leq \frac{\mathcal{O}(1)}{\#(I_j)N(\tau_j - \tau_{j-1})}, \end{aligned}$$

nous obtenons

$$\text{var}[Z^{(j)}] = \mathcal{O}\left(\frac{r^3}{N^3}\right),$$

et par suite

$$\sum_{j=1}^r (\tau_j - \tau_{j-1}) \text{var}[Z^{(j)}] = \mathcal{O}\left(\frac{r^3}{N^3}\right). \quad (1.5)$$

D'autre part $\forall j = 1, 2, \dots, r$ nous avons

$$\begin{aligned} E^2[Z^{(j)}] &= \left[\frac{1}{\#(I_j)} - \frac{1}{N(\tau_j - \tau_{j-1})} \right]^2 \left(\sum_{\{i \in I_j\}} f(x_i) \right)^2 \\ &= \left(\frac{1}{\#(I_j)} \sum_{\{i \in I_j\}} f^2(x_i) \right) \times \mathcal{O}\left(\frac{1}{N^2(\tau_j - \tau_{j-1})^2}\right). \end{aligned}$$

La dernière égalité est due au fait que

$$\left(\frac{1}{\#(I_j)} \sum_{i \in I_j} f(x_i) \right)^2 \leq \frac{1}{\#(I_j)} \sum_{i \in I_j} f^2(x_i).$$

Comme f appartient à $L^2([0, 1])$

$$\frac{1}{\#(I_j)} \sum_{\{i \in I_j\}} f^2(x_i) < \infty.$$

Nous en déduisons que

$$\sum_{j=1}^r (\tau_j - \tau_{j-1}) E^2[Z^{(j)}] = \mathcal{O}\left(\frac{r^2}{N^2}\right). \quad (1.6)$$

Les expressions (1.5) et (1.6) entraînent de ce fait l'assertion du lemme 1.1 ■

Le lemme 1.1 établi nous désignerons dans tout ce qui suit par ${}_r\hat{f}$ l'un ou l'autre des estimateurs \hat{f}^{reg} et \hat{f}_r de f .

Ci-après nous nous intéresserons au comportement du MISE de ${}_r\hat{f}$. A cette fin nous décomposons le MISE en deux parties l'une associée à la contribution du carré du biais et l'autre à celle de la variance. Ainsi nous avons

$$\begin{aligned} MISE({}_r\hat{f}) &= E \int_0^1 ({}_r\hat{f}(x) - f(x))^2 dx \\ &= E \int_0^1 ({}_r\hat{f}(x) - E[{}_r\hat{f}(x)])^2 dx + \int_0^1 (E[{}_r\hat{f}(x)] - f(x))^2 dx. \end{aligned}$$

Nous allons étudier séparément chacun des deux termes de l'égalité précédente. Pour cela nous établissons les deux lemmes suivants :

Lemme 1. 2. *Sous l'hypothèse du lemme 1.1 et avec les notations de ce chapitre, nous avons*

$$E \int_0^1 ({}_r\hat{f}(x) - E[{}_r\hat{f}(x)])^2 dx = \frac{r\sigma^2}{N} (1 + \mathcal{O}(\frac{1}{N})).$$

Démonstration

Le fait que les ω_ν forment un système orthonormé entraîne que

$$\begin{aligned} \int_0^1 ({}_r\hat{f}(x) - E[{}_r\hat{f}(x)])^2 dx &= \int_0^1 [\sum_{\nu=1}^r (\hat{a}_\nu - E[\hat{a}_\nu]) \omega_\nu(x)]^2 dx \\ &= \sum_{\nu=1}^r (\hat{a}_\nu - E[\hat{a}_\nu])^2. \end{aligned}$$

En prenant l'espérance des deux membres de l'égalité le calcul se ramène à la détermination de la variance de chacun des \hat{a}_ν .

Ainsi :

$$\begin{aligned} \text{var}[\hat{a}_\nu] &= \text{var}[\frac{1}{N} \sum_{i=1}^N Y_i \omega_\nu(x_i)] \\ &= \frac{\sigma^2}{N^2} \sum_{i=1}^N \omega_\nu^2(x_i) \\ &= \frac{2^\mu \sigma^2}{N^2} \#\{i; x_i \in [\frac{\delta-1}{2^\mu}, \frac{\delta}{2^\mu}]\} \\ &= \frac{\sigma^2}{N} (1 + \mathcal{O}(\frac{1}{N})). \end{aligned}$$

Le lemme 1.2 suit par sommation sur ν . ■

Pour pouvoir contrôler le biais de l'estimateur Γ dû essentiellement à l'erreur d'approximation de f par son développement tronqué nous allons exiger un comportement régulier de f . Plus précisément nous supposons que f appartient à la classe des fonctions Lipschitziennes c'est à dire qu'il existe une constante positive C telle que

$$\forall x, x' \in [0, 1] \quad |f(x) - f(x')| \leq C|x - x'|.$$

Lemme 1. 3. *Si f est Lipschitzienne et C sa constante de Lipschitz, alors*

$$\int_0^1 (E[{}_r\hat{f}(x)] - f(x))^2 dx = \sum_{\nu>r} a_\nu^2 + \mathcal{O}(\frac{r^2}{N^2}) \leq \frac{C^2}{3r^2} + \mathcal{O}(\frac{r^2}{N^2}).$$

Démonstration

Du fait de l'orthogonalité des fonctions de base nous avons

$$\begin{aligned} \int_0^1 (E[{}_r\hat{f}(x)] - f(x))^2 dx &= \int_0^1 (E[{}_r\hat{f}(x)] - f_r(x))^2 dx \\ &\quad + \int_0^1 (f_r(x) - f(x))^2 dx \\ &= \int_0^1 (E[{}_r\hat{f}(x)] - f_r(x))^2 dx + \sum_{\nu=r+1}^\infty a_\nu^2 \\ &= B_1 + B_2. \end{aligned}$$

Le terme du biais B_1 s'écrit

$$B_1 = \int_0^1 \left[\frac{1}{N} \sum_{i=1}^N f(x_i) K_r(x_i, x) - \int_0^1 f(s) K_r(s, x) ds \right]^2 dx.$$

Comme

$$\left| \frac{1}{N} \sum_{i=1}^N f(x_i) K_r(x_i, x) - \int_0^1 f(s) K_r(s, x) ds \right| \leq C \frac{r}{N},$$

alors

$$B_1 = \mathcal{O}\left(\frac{r^2}{N^2}\right).$$

Pour le terme d'approximation B_2 nous utilisons ci-après une majoration en valeur absolue de chacun des coefficients a_ν . Par définition des ψ_μ^δ nous avons

$$\begin{aligned} \forall \nu \geq 1, \quad a_\nu &= \int_0^1 f(s) \omega_\nu(s) ds \\ &= \sqrt{2^\mu} \left[\int_{\frac{\delta-1}{2^\mu}}^{\frac{2\delta-1}{2^\mu}} (f(s) - f(s + \frac{1}{2^{\mu+1}})) ds \right]. \end{aligned}$$

Le caractère Lipschitzien de f permet d'écrire

$$\begin{aligned} |a_\nu| &\leq \sqrt{2^\mu} \int_{\frac{\delta-1}{2^\mu}}^{\frac{2\delta-1}{2^\mu}} |f(s) - f(s + \frac{1}{2^{\mu+1}})| ds \\ &\leq \sqrt{2^\mu} \int_{\frac{\delta-1}{2^\mu}}^{\frac{2\delta-1}{2^\mu}} \frac{C}{2^{\mu+1}} ds \\ &\leq \frac{C\sqrt{2^\mu}}{2^{2\mu+2}}. \end{aligned}$$

Ainsi

$$a_\nu^2 \leq \frac{C^2}{16} 2^{-3\mu},$$

ce qui donne en majorant chaque terme de la somme

$$\begin{aligned} \sum_{\nu=r+1}^{\infty} a_\nu^2 &= \sum_{\delta=\delta_r+1}^{2^{\mu_r}-1} a_\nu^2 + \sum_{\mu=\mu_r+1}^{\infty} \sum_{\delta=0}^{2^\mu-1} a_\nu^2 \\ &\leq \sum_{\delta=\delta_r+1}^{2^{\mu_r}-1} \frac{C^2}{16} 2^{-3\mu_r} + \sum_{\mu=\mu_r+1}^{\infty} \sum_{\delta=1}^{2^\mu-1} \frac{C^2}{16} 2^{-3\mu} \\ &\leq \frac{C^2}{12r^2} + \frac{C^2}{16} (2^{\mu_r} - \delta_r) 2^{-3\mu_r} \\ &\leq \frac{C^2}{12r^2} + \frac{C^2}{16} 2^{-2\mu_r} \\ &\leq \frac{C^2}{12r^2} + \frac{C^2}{4} 2^{-2(\mu_r+1)} \\ &\leq \frac{C^2}{12r^2} + \frac{C^2}{4r^2} \\ &= \frac{C^2}{3r^2}. \quad \blacksquare \end{aligned}$$

Grâce aux lemmes 1.1-3 nous pouvons maintenant énoncer le théorème suivant :

Théorème 1. 1. *Si f est une fonction Lipschitzienne sur l'intervalle $[0, 1]$ et C sa constante de Lipschitz, alors, avec les notations de ce chapitre, nous avons, quand N tend vers l'infini,*

$$MISE(r\hat{f}) \leq \frac{r\sigma^2}{N} \left(1 + \mathcal{O}\left(\frac{1}{N}\right)\right) + \frac{C^2}{3r^2} + \mathcal{O}\left(\frac{r^2}{N^2}\right).$$

Le MISE est minimum pour $r = \left(\frac{C^2}{3\sigma^2}\right)^{1/3} N^{1/3}$ et sa vitesse optimale de convergence vers 0 est de $\mathcal{O}(N^{-2/3})$.

Il est à noter que les deux valeurs optimales du théorème 1.1 sont obtenues en équilibrant le terme du carré du biais et le terme de la variance.

1.3.1 Remarque

Dans cette remarque nous nous proposons d'évaluer le nombre des x_i dans chaque intervalle $[\tau_{j-1}, \tau_j]$ élément de $\varphi(r)$ pour un choix de r dépendant de N et où chaque x_i est égal à $\frac{2i-1}{2N}$. Il est clair que

$$\#\{i; x_i \in [\tau_{j-1}, \tau_j]\},$$

n'est autre que le nombre de i pour lesquels

$$\begin{cases} \frac{j-1}{2^{\mu_r+1}} \leq \frac{2i-1}{2N} < \frac{j}{2^{\mu_r+1}} & \text{lorsque } j = 1 \dots 2\delta_r + 1 \quad (1) \\ \frac{j-1-\delta_r}{2^{\mu_r}} \leq \frac{2i-1}{2N} < \frac{j-\delta_r}{2^{\mu_r}} & \text{lorsque } j = 2\delta_r + 2 \dots r \quad (2). \end{cases}$$

Ainsi si r est tel qu'il existe l appartenant à \mathbb{N} et que $\frac{N}{2^{\mu_r+1}} = l$, (1) est équivalent à

$$lj - l + \frac{1}{2} \leq i < lj + \frac{1}{2},$$

d'où

$$\#\{i; x_i \in (\tau_{j-1}, \tau_j)\} = l.$$

Nous remarquons que dans le cas où $\frac{N}{2^{\mu_r+1}}$ appartient à \mathbb{N} nous avons

$$\#\{i; x_i \in [\tau_{j-1}, \tau_j]\} = N(\tau_j - \tau_{j-1}),$$

et donc

$$\hat{F}_N(x) = \tau_j - \tau_{j-1}.$$

Dans ce cas l'estimateur de projection et le régressogramme définis respectivement par (1.3) et (1.4) sont identiques sur l'intervalle $[0, 1]$.

1.4 Construction de la base optimale

Au vu des paragraphes 1.1-3 il est possible de construire un estimateur de type projection basé sur le système de Haar chaque fois que nous disposons d'une partition en intervalles dyadiques du support de la fonction de

régression. Notre objectif ci-après est de définir une partition dyadique φ de $[0, 1]$ le support de la fonction de régression f et de lui associer l'estimateur correspondant de type projection orthogonale sur la base de Haar de $L^2([0, 1])$. Le choix de la partition φ sera fait de sorte que l'estimateur associé soit localement adapté à la régularité de f .

Nous entreprenons dans ce qui suit de construire à l'aide d'un schéma récursif la partition φ de $[0, 1]$. A cette fin nous allons définir de manière récursive une suite de partitions emboîtées conduisant à la définition de la partition la mieux adaptée à l'estimation de f . Le lecteur désireux de plus de détails sur les estimateurs de la fonction de régression par des schémas de partitions récursives pourra se reporter aux travaux de Gordon et Olshen [37] [38] [39].

Le fait de se limiter à des partitions dyadiques est particulièrement approprié au cadre que nous avons développé précédemment. Remarquons simplement ici que les estimateurs fondés sur des schémas de partitions récursives apparaissent localement comme des moyennes non pondérées des observations associées aux instants présents dans chaque élément de la partition ; ainsi si tous les éléments de la partition étaient de même longueur l'estimateur lisserait de manière identique les données et donc ne s'adapterait pas à la régularité locale de la fonction f .

Dans les paragraphes 1.1-3 nous avons étudié des estimateurs de la régression non adaptatifs possédants de bonnes propriétés asymptotiques. Ainsi par le théorème 1.1 nous avons établi que pour obtenir une valeur optimale du MISE de cet estimateur la longueur des éléments de φ devra donc être de l'ordre $r^* = \mathcal{O}(N^{-\frac{1}{3}})$.

La procédure que nous proposons consiste à fixer une valeur r pour r^* et à construire $\varphi = \varphi^*(r)$ une partition de $[0, 1]$ au plus aussi fine que $\varphi(r)$ partition définie au paragraphe 1.1. A cette fin nous construirons une suite de partitions emboîtées $(\varphi^{(r,i)}, i = 1, 2, \dots, \mu_r + 1)$ telle que $\varphi^{(r,1)} = [0, 1]$ et $\varphi^{(r,\mu_r+1)} = \varphi^*(r)$ (r étant égal à $2^{\mu_r} + \delta_r$). Chaque partition $\varphi^{(r,i)}$ est déterminée à partir d'un ensemble d'indices $\mathcal{I}^{(r,i)}$ dont chaque élément $\nu = 2^\mu + \delta$ correspond à l'intervalle $I_\nu = [\frac{\delta-1}{2^\mu}, \frac{\delta}{2^\mu}]$.

Les différents $\mathcal{I}^{(r,i)}$ forment une suite d'ensembles emboîtés $\mathcal{I}^{(r,1)} \subseteq \mathcal{I}^{(r,2)} \subseteq \dots \mathcal{I}^{(r,\mu_r+1)} = \mathcal{I}^*(r)$ où $\mathcal{I}^{(r,1)} = \{1\}$; le passage de $\mathcal{I}^{(r,i)}$ à $\mathcal{I}^{(r,i+1)}$ et de surcroît de $\varphi^{(r,i)}$ à $\varphi^{(r,i+1)}$ s'effectue en considérant tous les éléments de $\mathcal{I}^{(r,i)}$ et en décidant via un critère de décision bien défini pour chaque I_ν si oui ou non il doit être partagé en deux nouveaux intervalles $I_{\nu_g} = [\frac{\delta-1}{2^\mu}, \frac{2\delta-1}{2^{\mu+1}})$ et $I_{\nu_d} = [\frac{2\delta-1}{2^{\mu+1}}, \frac{\delta}{2^{\mu+1}})$.

Une fois que le choix de $\mathcal{I}^*(r)$ fait l'estimateur de projection orthogonale adapté est alors déterminé par la décomposition suivant les éléments ω_ν dont

les indices ν appartiennent à $\mathcal{I}^*(r)$.

Afin de justifier le choix du critère de décision retenu nous remarquerons que la valeur de a_ν donne une idée précise sur le comportement de la restriction de la fonction de régression à l'intervalle I_ν . En effet puisque

$$a_\nu = 2^{\frac{\mu}{2}} \int_{\frac{\delta}{2^\mu}}^{\frac{2\delta-1}{2^{\mu+1}}} (f(s) - f(s + \frac{1}{2^{\mu+1}})) ds,$$

le théorème des accroissements finis entraîne l'existence de $c(s)$ dans $]\frac{\delta}{2^\mu}, \frac{2\delta-1}{2^{\mu+1}}[$ tel que

$$|a_\nu| \approx |2^{-(\frac{\mu}{2}+1)} \int_{\frac{\delta}{2^\mu}}^{\frac{2\delta-1}{2^{\mu+1}}} f'(c(s)) ds|,$$

f' étant la dérivée de f .

Ainsi nous constatons que là où la fonction f est régulière les coefficients $|a_\nu|$ seront petits alors que des variations importantes de f' entraînent de grandes valeurs de a_ν . De ce fait seront retenus uniquement les coefficients \hat{a}_ν tels que \hat{a}_ν^2 est supérieur à un certain seuil L (ceux pour lesquels $\hat{a}_\nu^2 \geq L$). Néanmoins la règle de décision devra conserver l'aspect récursif de la partition.

Le travail de Tarter et Kronmal [70] suggère de prendre $L = 2\frac{\hat{\sigma}^2}{N}$ où $\hat{\sigma}^2$ est un estimateur de la variance du bruit du modèle (1.1). Dans le cas étudié et afin de préserver l'analogie avec l'analyse multirésolution nous avons modifié ce critère. Ainsi au lieu de prendre \hat{a}_ν^2 nous prenons une estimation de $E(\hat{a}_\nu^2)$ que nous comparons au seuil L ; l'estimation de $E(\hat{a}_\nu^2)$ étant la moyenne arithmétique des $\hat{a}_{\nu'}^2$, dont le support contient I_ν (support de ω_ν).

Afin de faciliter l'écriture de l'algorithme correspondant à la procédure décrite ci-dessus nous prendrons $r = 2^{\mu_r}$. Nous définissons par $ant(\cdot)$ la fonction qui prend pour argument un élément ν et restitue son antécédent dans l'arbre binaire figure 1.2 et par var une estimation consistante de la variance.

ALGORITHME 1

Retenir \hat{a}_1

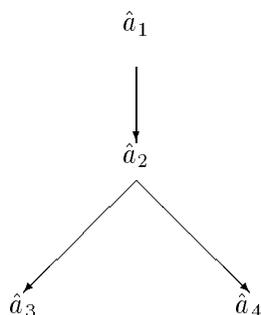
- * Pour $j_{puis} = 0, 1, \dots, \mu_r$
- * Pour $j = 1, 2, \dots, 2^{j_{puis}}$
- * Faire
- * $ind_0 = 2^{j_{puis}} + j$
- * $ind = ind_0$
- * Faire jusqu'à $\hat{a}_{ind}^2 > \frac{2 \times var}{N}$

```

*           ind = ant(ind)
*           Fin de Faire
*           Si moyarithmétique ( $\hat{a}_{ind_0}^2, \hat{a}_{ant(ind_0)}^2, \dots, \hat{a}_{ind}^2$ ) >  $\frac{2 \times var}{N}$ 
*           Alors retenir tous les éléments  $\hat{a}_{ind_0}, \hat{a}_{ant(ind_0)}, \dots, \hat{a}_{ind}$ 
*           Autrement  $j = j + 1$ 
*           Fin de Si
*           Fin de Pour ;
*            $j_{puis} = j_{puis} + 1$ 
*           Fin de Pour.

```

Les r coefficients $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_r$ sont associés à un arbre binaire de la manière suivante :



...

Figure 1.2 : Arbre binaire associé à la procédure.

Au vu de la procédure décrite dans l'algorithme 1 et au vu du critère de sélection choisi si un premier coefficient \hat{a}_{ν_1} et un deuxième \hat{a}_{ν_2} appartenant à la branche issue de \hat{a}_{ν_1} sont retenus tous les coefficients de l'arbre binaire ci-dessus qui joignent \hat{a}_{ν_1} et \hat{a}_{ν_2} seront retenus. Ainsi si nous retenons \hat{a}_5 et \hat{a}_2 nous retiendrons forcément \hat{a}_3 .

1.5 Simulations

Dans le paragraphe 1.3 nous avons montré que la variance des \hat{a}_ν est proportionnelle à σ^2 en général inconnue. De plus nous venons de voir que le choix d'une partition adaptative repose sur celui d'un sous ensemble de

coefficients \hat{a}_ν jugés suffisamment importants relativement à leur variance. Il est donc essentiel de disposer d'une estimation efficace de σ^2 qui ne doit pas dépendre de la méthode choisie pour estimer f . Il existe dans la littérature plusieurs estimateurs de la variance reposant uniquement sur des hypothèses de régularité de f . Nous en avons retenu trois.

Le premier travail a donc consisté en l'étude de la qualité de trois estimateurs de la variance du bruit pour un modèle de régression non paramétrique. Il s'agit de l'estimateur de J. Rice [61] de l'estimateur de T. Gasser [36] et de l'estimateur de P. Hall [40]. Les expressions de ces trois estimateurs sont rappelées ci-dessous ; puis pour chacun des trois modèles ci-après nous donnons un tableau regroupant le calcul de la moyenne et du carré de l'erreur standard des différentes estimations de la variance du bruit. Le nombre de simulations pour chaque modèle est de 200 alors que le nombre de points tests N s'étend de 50 à 250.

Nous avons retenu trois fonctions de régression définies sur l'intervalle $[0, 1]$

- modèle 1 :

$$f(x) = x \cdot (x - 0.3) \cdot (x - 0.5) \cdot (x - 1.) + 0.2.$$

- modèle 2 :

$$f(x) = \sin(6.28314 \cdot x) \cdot \sin(6.28314 \cdot (x - 0.3)) \cdot \sin(6.28314 \cdot (x - 0.5)).$$

- modèle 3 :

$$f(x) = 5 \cdot \sin(6.28 \cdot x).$$

A partir des N variables aléatoires définies par (1.1) nous définissons l'estimateur de Rice

$$\hat{\sigma}^2 = \frac{1}{2(N-1)} \sum_{i=1}^{N-1} (Y_i - Y_{i-1})^2,$$

l'estimateur de Gasser

$$\hat{\sigma}^2 = \frac{2}{3(N-2)} \sum_{i=2}^{N-1} \left(\frac{1}{2} Y_{i-1} + \frac{1}{2} Y_{i+1} - Y_i \right)^2,$$

l'estimateur de Hall

$$\hat{\sigma}^2 = \frac{1}{N-2} \sum_{i=1}^{N-2} (d_0 Y_i + d_1 Y_{i+1} + d_2 Y_{i+2}),$$

avec $d_0 = 0.8090$, $d_1 = -0.5$ et $d_2 = -0.3090$.

Les trois estimateurs définis ci-dessus font partie de la famille des estimateurs de la variance pour des modèles homoscedastiques de régression non paramétrique. Basés sur des différences entre les observations, ils ne requièrent pas une estimation préalable de la fonction de régression. L'estimateur de Rice est l'unique estimateur de ce type qui fait intervenir les différences entre uniquement deux observations : c'est un estimateur du premier ordre. L'estimateur de Gasser, tout comme celui de Rice, est basé sur des différences symétriques (symétrie des poids qui pondèrent les observations), cependant il utilise les différences entre trois observations successives, ce qui en fait un estimateur du second ordre. Contrairement à l'estimateur de Gasser et bien qu'ils soient tous les deux du second ordre, l'estimateur de Hall n'est pas basé sur des différences symétriques, il est par contre optimal au sens de l'erreur moyenne quadratique. Pour le modèle 1, la variance du bruit Gaussien utilisée dans les simulations est de 10^{-4} (ce qui correspond à un rapport signal bruit presque égal à 4). Les résultats sont les suivants :

méthode	Gasser	Rice	Hall
50 points			
moyenne	100.059	99.6978	99.4070
erreur-standard au carré	7.61626E-06	5.43302E-06	4.51146E-06
100 points			
moyenne	100.694	100.541	100.251
erreur-standard au carré	4.37786E-06	3.29094E-06	2.72090E-06
150 points			
moyenne	100.345	100.378	100.5
erreur-standard au carré	2.58307E-06	1.99188E-06	1.69957E-06
250 points			
moyenne	100.954	100.647	100.153
erreur-standard au carré	1.78872E-06	1.41127E-06	1.20047E-06

Table 1.1 : Estimation de la moyenne ($\times 10^6$) et du carré de l'erreur standard pour différentes simulations. La variance du bruit Gaussien utilisé dans les simulations est égale à 10^{-4} .

Les mêmes simulations avec le modèle 2 ont donné le tableau 1.2 pour un bruit Gaussien de variance 0.04Γ soit un rapport signal bruit égal à 2.

méthode	Gasser	Rice	Hall
50 points			
moyenne	4.01332E-02	4.25125E-02	4.63664E-02
erreur-standard au carré	1.21947E-04	8.78727E-05	7.73191E-05
100 points			
moyenne	4.02849E-02	4.08652E-02	4.17408E-02
erreur-standard au carré	7.00480E-05	5.25437E-05	4.34925E-05
150 points			
moyenne	4.01392E-02	4.04444E-02	4.09338E-02
erreur-standard au carré	4.13286E-05	3.18956E-05	2.73109E-05
250 points			
moyenne	4.03818E-02	4.03643E-02	4.03261E-02
erreur-standard au carré	2.86183E-05	2.25779E-05	1.92031E-05

Table 1.2 : Estimation de la moyenne et du carré de l'erreur standard pour différentes simulations. La variance du bruit Gaussien utilisé dans les simulations est égale à 4×10^{-2}

Pour le troisième et dernier modèle la variance du bruit Gaussien a été de 2.25 ce qui correspond un rapport signal bruit égal à 5. Les résultats des simulations sont regroupés dans le tableau ci-dessous :

méthode	Gasser	Rice	Hall
50 points			
moyenne	2.25180	2.33724	2.46124
erreur-standard au carré	0.385560	0.278296	0.231968
100 points			
moyenne	2.26564	2.28641	2.31405
erreur-standard au carré	0.221628	0.166458	0.137619
150 points			
moyenne	2.25776	2.26893	2.28757
erreur-standard au carré	0.130764	1.00801E-01	8.62584E-02
250 points			
moyenne	2.27147	2.26851	2.24476
erreur-standard au carré	9.05476E-02	7.14464E-02	4.87280E-02

Table 1.3 : Estimation de la moyenne et du carré de l'erreur standard pour différentes simulations. La variance du bruit Gaussien utilisé dans les simulations est égale à 2.25.

Au vu des résultats obtenus ci-dessus il est clair que l'utilisation de l'estimateur de Gasser pour des échantillons de faible taille est préférable aux deux autres alors que pour des échantillons de taille moyenne la qualité de l'estimateur de Hall est supérieure aux autres.

L'objet des simulations qui suivent est d'analyser Γ par des graphiques Γ le comportement de l'estimateur de la fonction de régression Γ développé au cours de ce chapitre.

Les trois figures ci-dessous 1.3-5 représentent l'estimation obtenue pour des échantillons de 256 points Γ et pour chacun de ces cas Γ la valeur de r a été fixée à 16. Étant donné la taille des échantillons Γ nous avons utilisé l'estimateur de Hall Γ pour estimer la variance du bruit Γ dans l'algorithme 1.

Dans toutes les figures données ci-après Γ nous représentons à la fois la fonction de régression Γ les observations et l'estimation obtenue par l'application de l'algorithme 1.

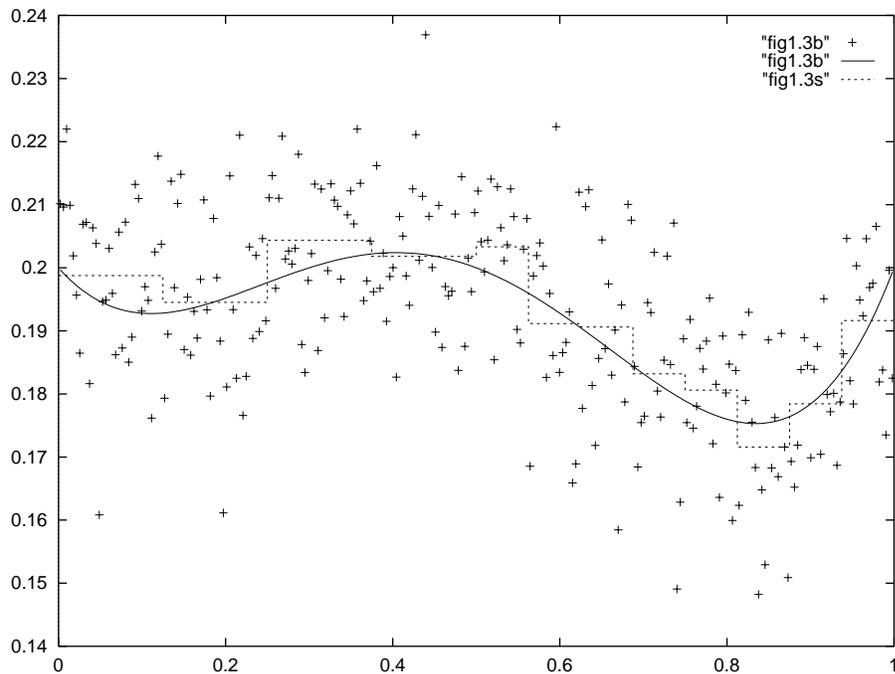


Figure 1.3 : La fonction de régression est donnée par le modèle 1 Γ elle est représentée en trait continu et son estimateur en pointillé.

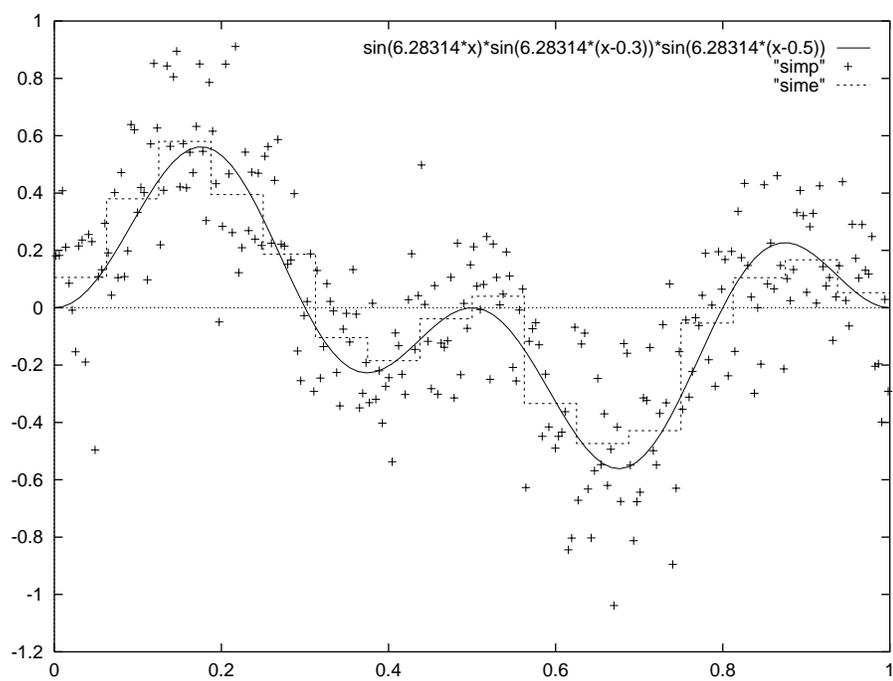


Figure 1.4 : La fonction de régression est donnée par le modèle 2 Elle est représentée en trait continu et son estimateur en pointillé.

Tout comme pour les deux exemples précédents le rapport signal/bruit est égal à 2.

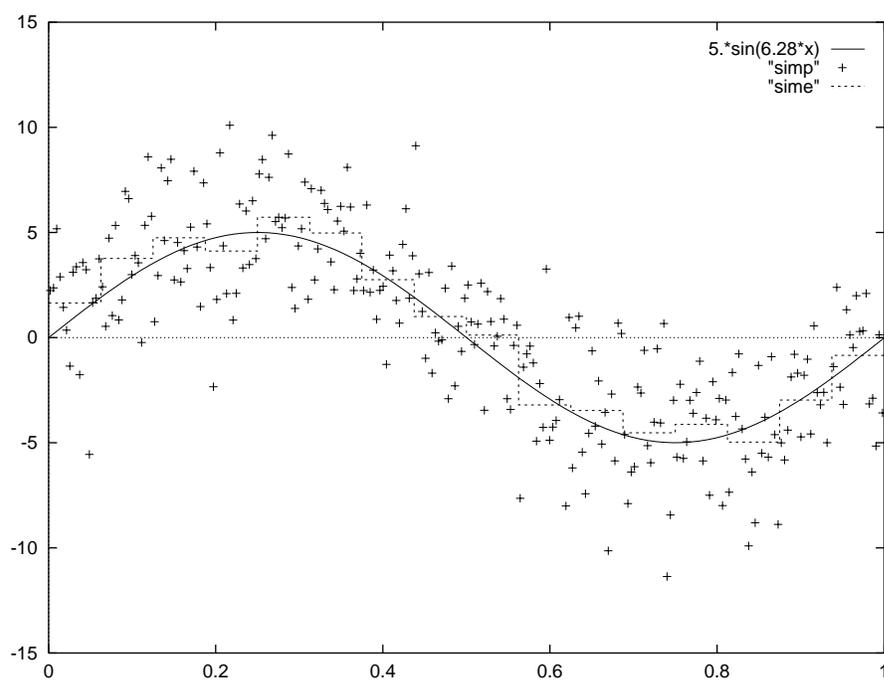


Figure 1.5 : La fonction de régression est donnée par le modèle $5 \sin(6.28x)$ Elle est représentée en trait continu et son estimateur en pointillé.

Comparaison avec d'autres résultats

Dans cet exemple la fonction de régression est une fonction polynômiale (exemple extrait de [49]). Sur le graphe sont représentés les 100 points échantillons $y_i = f(x_i) + \varepsilon_i$ où ε_i suit une loi normale centrée avec un écart-type égal à 0.1. La fonction de régression est représentée par un trait continu et l'estimateur obtenu par l'algorithme 1 est en pointillé.

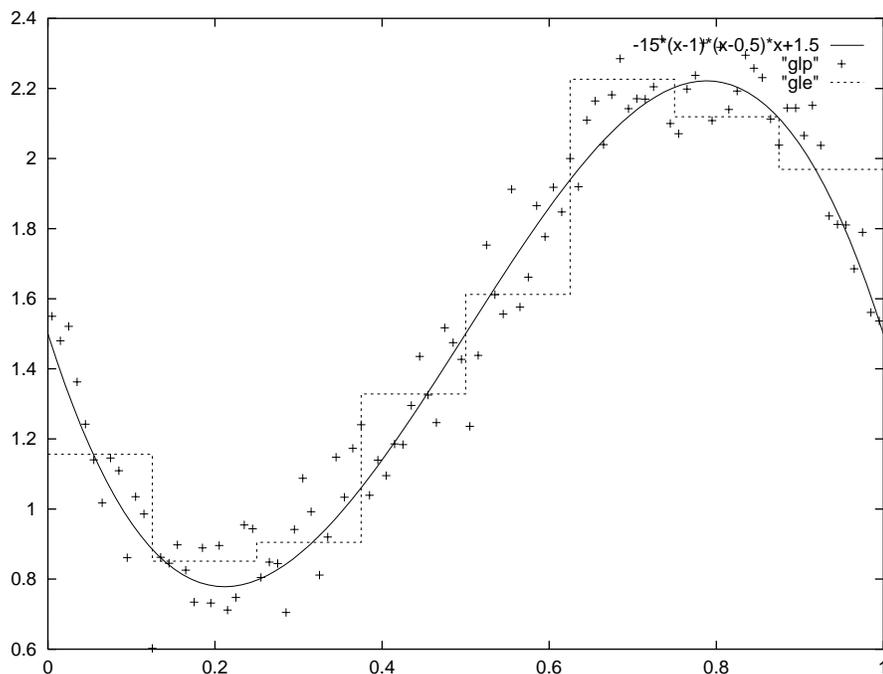


Figure 1.6 : La fonction de régression est la fonction $f(x) = -15(x - 1)(x - 0.5)x + 1.5$. Dans ce cas r est égal à 8.

Dans cet exemple la fonction de régression est une fonction ayant deux points de discontinuité aux points 0.3 et 0.5. Sur le graphe sont représentés les 256 points échantillons $y_i = f(x_i) + \varepsilon_i$ où ε_i suit une loi normale centrée avec un écart-type de 0.096 la fonction de régression est représentée par un trait continu et l'estimateur obtenu par l'algorithme 1 est en pointillé.

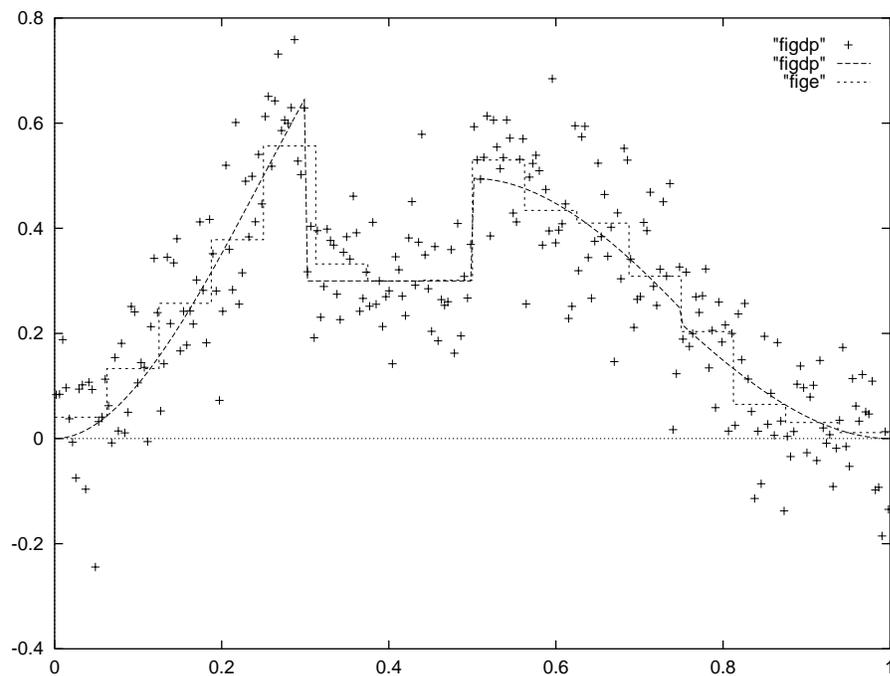


Figure 1.7 : Dans ce cas r est égal à 16.

La fonction de régression utilisée est définie par

$$f(x) = \begin{cases} 4x^2(3 - 4x) & \text{si } x \leq 0.3 \\ 0.3 & \text{si } 0.3 < x \leq 0.5 \\ \frac{4}{3}x(4x^2 - 10x + 7) - 1.5 & \text{si } 0.5 < x \leq 0.75 \\ \frac{16}{3}x(x - 1)^2 & \text{si } 0.75 < x. \end{cases}$$

Le nuage de points de la figure ci-dessous représente les 1000 points issus du modèle $y = \sin(4/t) + 1.5 + \varepsilon$ avec ε de loi normale centrée d'écart-type égal à 0.2.

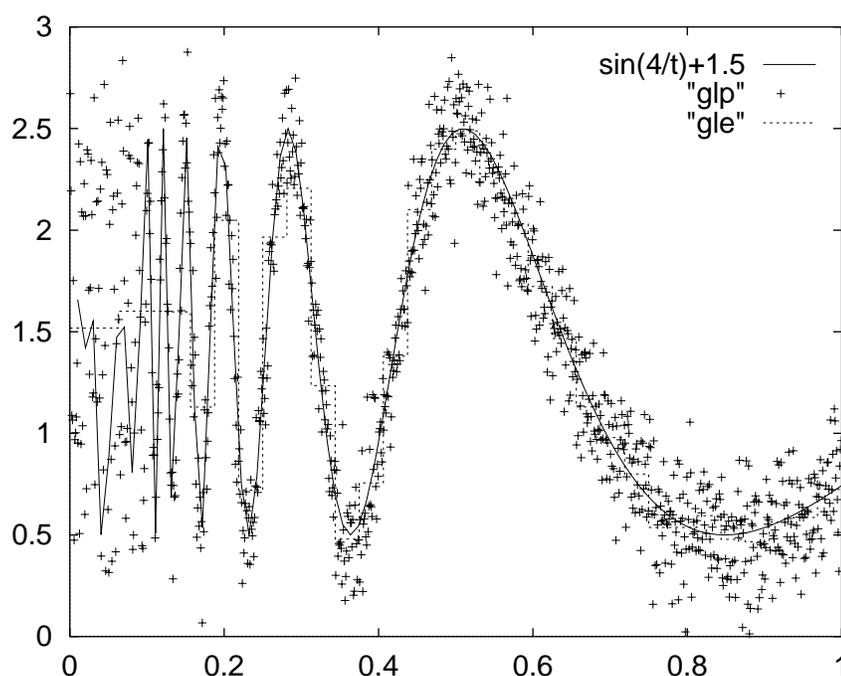


Figure 1.8 : La fonction de régression est représentée en trait continu et son estimateur en pointillé. Dans ce cas r est égal à 32.

Les estimateurs obtenus pour les figures 1.6 et 1.8 sont comparables aux résultats de Mammen et Van de Geer[49] avec l'avantage que la méthode proposée est plus rapide et moins complexe.

Les graphiques ci-dessus donnent une idée sur le caractère adaptatif de la procédure décrite dans l'algorithme 1. En effet la largeur des intervalles constituant la partition du support de la fonction de régression Γ varie selon la régularité de la fonction : plus la fonction est régulière plus la largeur de l'intervalle est grande. L'estimateur obtenu s'apparente à l'estimateur du type noyau à fenêtre variable.

Au vu des résultats présentés dans ce paragraphe et des diverses simulations effectuées pour tester l'influence de l'estimation de la variance sur l'estimateur de la fonction de régression obtenu par notre méthode nous concluons que :

1. La méthode est localement adaptative et produit une estimation approchée de la vraie fonction en dépit de l'intensité du bruit.
2. Quel que soit l'estimateur de la variance nous obtenons des estimations similaires et ce pour les 3 estimateurs de variance étudiés.
3. Malgré le manque de lissitude de la fonction estimation (fonction en escaliers) cette dernière donne une idée précise sur l'allure de la fonction de régression quand les intervalles de la partition du support contiennent suffisamment de données.

1.6 Lien avec l'analyse multirésolution

Nous avons signalé au paragraphe 1.2 que le système de Haar est associé à une structure de multirésolution que nous nous proposons de rappeler ci-après.

Du fait que les $W_j^{([0,1])}$ définis dans le paragraphe 1.2 de ce chapitre soient orthogonaux nous pouvons construire une suite de sous espaces emboîtés de $L^2([0, 1])$ par

$$V_j^{([0,1])} = V_{j-1}^{([0,1])} \oplus W_j^{([0,1])} \quad \text{pour } j \in \mathbb{N},$$

les $V_j^{([0,1])}$ ainsi obtenus forment une analyse multirésolution sur l'intervalle $[0, 1]$ telle qu'elle est définie dans l'annexe A.

D'autre part soit \hat{P}_{V_j} un estimateur du projecteur sur $V_j^{([0,1])}$ et définissons $\varphi_{j,k} = 2^{\frac{j}{2}} \omega_1(2^j \cdot -k)$ dont le support noté $I_{j,k}$ est égal à l'intervalle $[\frac{k}{2^j}, \frac{k+1}{2^j}]$; on peut aisément montrer que chaque $V_j^{([0,1])}$ est engendré par les φ_{jk} , $k = 0, 1, \dots, 2^j - 1$.

La procédure décrite dans ce chapitre consiste à fixer un j_{opt} qui détermine la longueur optimale des intervalles $I_{j,k}$ puis à définir un estimateur

$$\hat{f}_{j_{opt}}(\varphi) = \sum_{(j,k) \in \varphi} \hat{P}_{V_j}(f) \mathbb{1}_{I_{j,k}},$$

où

$$\varphi = \{(j, k) \in \mathbb{N} \times \mathbb{Z} : 1 \leq j \leq j_{opt} \text{ et } 0 \leq k \leq 2^j - 1\}$$

est tel que $\cup_{(j,k) \in \varphi} I_{j,k} = [0, 1]$.

L'application de l'algorithme 1 conduit à la détermination d'un choix optimal φ^* ce qui correspond à un choix optimal $\hat{f}_{j_{opt}}(\varphi^*)$.

Une première possibilité pour généraliser cet algorithme à des ondelettes admettant une meilleure régularité est d'utiliser une décomposition qui puisse préserver l'écriture de l'estimateur. On pourrait se contenter de remplacer les estimateurs empiriques des $\langle f, \omega_\nu \rangle$ obtenus pour Haar par ceux obtenus pour une ondelette quelconque mais malheureusement les supports des ondelettes ω_ν et $\omega_{\nu+1}$ ne sont pas disjoints ce qui entraîne une corrélation entre les coefficients de la décomposition en ondelettes.

Une deuxième possibilité est d'utiliser une décomposition en paquets d'ondelettes ; c'est cette dernière qui sera retenue et fera l'objet du deuxième chapitre.

Chapitre 2

Régression par paquets d'ondelettes

Dans le chapitre précédent le problème d'estimation de la fonction de régression a été abordé à l'aide d'un schéma de partitions récursives binaires. Grâce aux décompositions en ondelettes de Haar nous avons pu établir un algorithme rapide pour l'évaluation numérique de l'estimateur et déterminer ses propriétés statistiques. Néanmoins cet estimateur présente les défauts d'être constant par morceaux et discontinu sur les bords des intervalles de la partition qui le définit avec pour conséquence une limitation de ses qualités d'approximation lorsque la fonction de régression est régulière.

Nous nous proposons de décrire une méthode adaptative d'estimation assez apparentée à l'approche par partition récursive. Elle sera associée à des décompositions en ondelettes régulières.

L'approche que nous avons adoptée dans ce chapitre repose sur les décompositions de la fonction de régression dans une "librairie" de bases associées à des paquets d'ondelettes. Elle se réalise en trois étapes :

1. Le choix de la meilleure base dans l'ensemble des bases défini par la décomposition en paquets d'ondelettes.
2. La sélection des coefficients dans cette base les plus importants pour

l'approximation de la fonction de régression.

3. L'utilisation des coefficients retenus pour la reconstruction de l'estimateur.

Le choix d'un ensemble de bases issues d'une décomposition en paquets d'ondelettes nous permettra non seulement d'adopter un schéma récursif de calcul établissant le lien avec l'approche adoptée au chapitre précédent mais donnera aussi lieu à quelques approches adaptatives.

Étant donné que nous utiliserons des bases de paquets d'ondelettes régulières nous consacrerons le premier paragraphe de ce chapitre à l'étude des qualités asymptotiques de l'estimateur de type ondelettes lorsque l'estimation de la régression repose sur des analyses multirésolutions régulières.

Pour réaliser une méthode adaptative d'estimation nous nous proposons ensuite de sélectionner une base orthonormée complète dans une famille de bases orthonormées produite par une analyse par paquets d'ondelettes de $L^2([0, 1])$; l'intérêt d'une librairie de bases de paquets d'ondelettes réside dans le fait qu'elle conduit à des schémas récursifs associés à des arbres binaires.

Pour choisir la meilleure base nous utiliserons trois approches distinctes. La première s'inspire de techniques utilisées par Tarter et Kronmal dans le cadre de l'estimation d'une densité ; nous concluons cette approche par l'énoncé d'un algorithme qui généralise l'algorithme 1 du premier chapitre. La deuxième approche associera des techniques de test et seuillage ce qui permettra à la fois d'éliminer les coefficients de la décomposition jugés peu importants et de modifier ceux qui sont retenus afin de diminuer l'erreur d'estimation. Finalement nous concluons ce chapitre en adaptant le principe du bootstrap à notre problème d'estimation.

Tout au long de ce chapitre nous ferons usage de la décomposition du vecteur des observations finie en paquets d'ondelettes. Les notions essentielles pour ce type de décomposition sont rappelées dans l'annexe A.

2.1 Régression par ondelettes sur l'intervalle

Nous adopterons le même modèle qu'au chapitre 1 ; le vecteur des observations sera noté $\mathbf{y} = (y_i)_{i=1}^N$. Chaque composante y_i de ce vecteur est une réalisation d'une variable aléatoire

$$Y_i = f(x_i) + \varepsilon_i, \quad i = 1, \dots, N$$

où ε_i est une variable aléatoire de moyenne nulle et de variance σ^2 et $x_i = \frac{i}{N}$. Nous supposons tout au long de ce chapitre que les Y_i sont des variables indépendantes et que le support de f est l'intervalle $[0, 1]$.

En posant $F_i = f(x_i)$ nous pouvons écrire l'équation vectorielle :

$$\mathbf{y} = \mathbf{F} + \underline{\varepsilon}, \quad (2.0)$$

avec $\mathbf{F} = (F_i)_{i=1}^N$ et $\underline{\varepsilon} = (\varepsilon_i)_{i=1}^N$.

Notre objectif ci-après est de proposer un estimateur \hat{f} de la fonction de régression sur l'intervalle $[0, 1]$ lorsque cette dernière est régulière par morceaux ; en fait nous estimerons les valeurs que prend f aux points x_i .

2.1.1 Approximation de la fonction de régression

Dans ce qui suit nous considérons une fonction d'échelle du type Daubechies à support compact minimal notée φ . Nous rappelons qu'en définissant par V_j le sous espace fermé engendré par les translations à pas entiers de la fonction $\varphi_{j0} = 2^{\frac{j}{2}}\varphi(2^j \cdot)$ la suite V_j est une analyse multirésolution de $L^2(\mathbf{R})$ (voir **Annexe A**). Soit alors ψ l'ondelette mère de Daubechies associée à φ .

Les ondelettes de Daubechies sont généralement indexées par un entier L et sont notées par ${}_L\psi$. L'indice L permet de déduire à la fois la régularité de la fonction d'échelle et celle de l'ondelette mère et la longueur de leur support. En effet en notant par ${}_L\varphi$ la fonction d'échelle nous avons :

1. Le support de ${}_L\varphi$ est inclus dans $[0, 2L - 1]$.
2. ${}_L\varphi$ est p -régulière avec $p \approx 0.2L$ (i.e $\int_{\mathbf{R}} |x|^p |{}_L\hat{\varphi}(x)| < \infty$).
3. Le support de ${}_L\psi$ est inclus dans $[-L, L - 1]$.
4. La fonction ${}_L\psi$ a la même régularité que ${}_L\varphi$.
5. ${}_L\psi$ admet des moments nuls jusqu'à l'ordre L ; i.e :

$$\int x^k {}_L\psi(x) dx = 0, \quad k = 0, 1, \dots, L, \quad \int x^{L+1} {}_L\psi(x) dx \neq 0.$$

Pour r entier strictement supérieur à 2 prenons $L = r - 1$; cette condition revient à exclure le cas des ondelettes de Haar ($L = 1$) de l'étude ci-après. Afin d'alléger les notations $\Gamma_{r-1}\varphi$ et ${}_{r-1}\psi$ seront notées Γ dans tout ce qui suit $\Gamma\varphi$ et ψ respectivement. De plus Γ nous supposons que ces dernières ont le même support Γ l'intervalle $[-L, L - 1]$ en l'occurrence ; ceci est possible par simple translation du support de φ .

Pour tout $(j, k) \in \mathbb{Z} \times \mathbb{Z}$ posons

$$\varphi_{jk} = 2^{\frac{j}{2}}\varphi(2^j \cdot -k),$$

$$\text{et } \psi_{jk} = 2^{\frac{j}{2}}\psi(2^j \cdot -k).$$

V_j est alors le sous espace fermé engendré par $\{\varphi_{jk}, k \in \mathbb{Z}\}$. Soit W_j le sous espace fermé engendré par $\{\psi_{jk}, k \in \mathbb{Z}\}$; par construction W_j est le supplémentaire orthogonal de V_j dans V_{j+1} (i.e : $V_{j+1} = V_j \oplus W_j$). Au vu de la structure multirésolution définie par les V_j

$$\mathcal{B}_J^{\mathbb{R}} = \{\varphi_{jk}, \psi_{jk}, j = J, J + 1, \dots, k \in \mathbb{Z}\}$$

est une base orthonormée de $L^2(\mathbb{R})$.

Comme la fonction de régression du modèle 1.1 est à support compact Γ $[0, 1]$ en l'occurrence l'utilisation d'une base orthonormée de $L^2(\mathbb{R})$ conduit à une dégradation de l'estimation aux extrémités de l'intervalle $[0, 1]$. Afin de pallier à ce problème nous utiliserons une base orthonormée d'ondelettes de $L^2([0, 1])$. Malheureusement et contrairement au cas des ondelettes de Haar la restriction des éléments de cette base à $[0, 1]$ ne conduit pas à une base orthonormée d'ondelettes de $L^2([0, 1])$.

Néanmoins par le travail de Cohen et al [17] il est possible de construire une base orthonormée d'ondelettes de $L^2([0, 1])$ de type Daubechies Γ via la construction d'une analyse multirésolution sur l'intervalle $[0, 1]$ (cette construction est largement détaillée dans l'annexe A). Cohen et al construisent pour tout j un système orthonormé $\{\tilde{\varphi}_{jk}, k = 1, 2 \dots 2^j\}$ (certains de ces éléments sont obtenus par translation et dilatation de φ les autres sont construits à la main) et lui associent le sous-espace fermé $V_j^{[0,1]}$ engendré par les éléments de ce système. Par construction les $V_j^{[0,1]}$ définissent une analyse multirésolution sur l'intervalle $[0, 1]$. Ils en déduisent alors un autre système orthonormé $\{\tilde{\psi}_{jk}, k = 1, \dots 2^j\}$ tel que le sous espace fermé $W_j^{[0,1]}$ engendré par ce dernier soit le supplémentaire orthogonal de $V_j^{[0,1]}$ dans

$V_{j+1}^{[0,1]}$; ainsi $W_j^{[0,1]}$ vérifie la relation suivante :

$$V_{j+1}^{[0,1]} = V_j^{[0,1]} \oplus W_j^{[0,1]}.$$

Ceci permet de définir pour tout $J_0 \in \mathbb{N}$

$$\mathcal{B}_{J_0}^{[0,1]} = \{\tilde{\varphi}_{J_0 k}, \tilde{\psi}_{jk}, j = J_0, J_0 + 1, \dots, k = 1, \dots, 2^j\},$$

une base orthonormée de $L^2([0, 1])$ possédant les mêmes propriétés de régularité que la base $\mathcal{B}_{J_0}^{\mathbb{R}}$.

Pour des considérations purement techniques il faut que $2^{J_0} \geq 2L$; ce qui sera le cas dans ce qui suit ci-après. Notons alors et tout au long de ce chapitre par J_0 le plus petit entier j tel que $2^j \geq 2L$.

Quelque soit $j \geq J_0$ la projection de f de $L^2([0, 1])$ sur l'espace d'approximation $V_j^{[0,1]}$ s'écrit

$$f_j = \sum_{k=1}^{2^j} \langle f, \tilde{\varphi}_{jk} \rangle \tilde{\varphi}_{jk}. \quad (2.1)$$

Le lemme 2.1 ci-dessous permet de contrôler l'erreur d'approximation de f par f_j en norme $L^2([0, 1])$.

Lemme 2.1. *Si f une fonction continûment dérivable à l'ordre $(r-1)$ dont la dérivée à l'ordre r est continue par morceaux, alors*

$$\|f - f_j\|_{L^2([0,1])}^2 \leq C_r 2^{-2jr},$$

avec $C_r^{\frac{1}{2}} = (r!(1 - 2^{-2r})^{\frac{1}{2}})^{-1} (\int |y^r \psi(y)| dy) \sup |f^{(r)}|$.

Preuve du lemme 2.1 :

Pour alléger l'écriture de cette démonstration posons pour tout $(l, k) \in \mathbb{N} \times \mathbb{N}$

$$d_{lk} = \langle f, \tilde{\psi}_{lk} \rangle.$$

Nous avons alors

$$\begin{aligned} d_{lk} &= \int \tilde{\psi}_{lk}(x) f(x) dx \\ &= \int 2^{\frac{l}{2}} \tilde{\psi}_k(2^l x) f(x) dx \\ &= 2^{-\frac{l}{2}} \int \tilde{\psi}_k(y) f\left(\frac{y}{2^l}\right) dy, \end{aligned}$$

ainsi

$$2^{\frac{l}{2}} d_{lk} = \int \tilde{\psi}_k(y) f\left(\frac{y}{2^l}\right) dy.$$

Par un développement limité de Taylor à l'ordre r de f au voisinage de 0 et du fait que ψ a des moments nuls jusqu'à l'ordre $(r-1)$ nous avons

$$\begin{aligned} 2^{\frac{l}{2}} d_{lk} &= \int \tilde{\psi}_k(y) \left\{ \sum_{j=0}^{r-1} (j!)^{-1} \left(\frac{y}{2^l}\right)^j f^{(j)}(0) \right. \\ &\quad \left. + ((r-1)!)^{-1} \left(\frac{y}{2^l}\right)^r \left[\int_0^1 (1-t)^{r-1} f^{(r)}\left(\frac{ty}{2^l}\right) dt \right] \right\} dy \\ &= ((r-1)!)^{-1} \int \tilde{\psi}_k(y) \left(\frac{y}{2^l}\right)^r \left[\int_0^1 (1-t)^{r-1} f^{(r)}\left(\frac{ty}{2^l}\right) dt \right] dy \\ &\leq \frac{2^{-lr}}{(r!)} \left(\int_0^1 |y^r \tilde{\psi}_k(y)| dy \right) \sup_x |f^{(r)}(x)|. \end{aligned}$$

Du fait que $\mathcal{B}_{j_0}^{[0,1]}$ soit une base orthonormée de $L^2([0,1])$ alors

$$f - f_j = \sum_{l=j}^{\infty} \sum_{k=1}^{2^l} d_{lk} \tilde{\psi}_{lk}.$$

Ainsi

$$\begin{aligned} \|f - f_j\|_{L^2([0,1])}^2 &= \sum_{l=j}^{\infty} \sum_{k=1}^{2^l} d_{lk}^2 \\ &\leq \sum_{l=j}^{\infty} \sum_{k=1}^{2^l} 2^{-l(2r+1)} (r!)^{-2} \left(\int_0^1 |y^r \tilde{\psi}_k(y)| dy \right)^2 \left(\sup_x |f^{(r)}(x)| \right)^2, \end{aligned}$$

d'où le résultat. ■

Partant du lemme 2.1 nous nous proposons dans le paragraphe suivant d'estimer f par une estimation de sa projection sur $V_j^{[0,1]}$.

2.1.2 Estimation de la fonction de régression

Pour tout réel positif α nous dirons qu'une fonction g de $L^1(\mathbb{R})$ est α -régulière si et seulement si

$$\int_{\mathbb{R}} |x|^\alpha |\hat{g}(x)| dx < \infty,$$

\hat{g} étant la transformée de Fourier de g et nous désignerons par $\mathcal{C}^\alpha(\mathbb{R})$ l'espace des fonctions α -régulières.

Notons (voir [16]) que si nous désignons par \mathcal{H}^α l'espace des fonctions höldériennes d'exposant α nous avons alors l'inclusion suivante

$$\mathcal{C}^\alpha \subset \mathcal{H}^\alpha.$$

L'inclusion dans l'autre sens n'est pas vérifiée. Cependant si nous nous limitons aux fonctions à support compact nous avons :

$$\forall \varepsilon > 0, \mathcal{H}^{\alpha - \frac{1}{2} - \varepsilon} \subset \mathcal{C}^\alpha.$$

Notons \mathbf{s}^0 le vecteur de \mathbb{R}^{2^j} dont chaque composante s_k^0 $k = 1, 2, \dots, 2^j$ est définie par :

$$s_k^0 = \frac{1}{N} \sum_{i=1}^N F_i \tilde{\varphi}_{jk}(x_i), \quad (2.2)$$

les F_i et x_i étant définis dans le paragraphe 2.1. Nous avons le résultat suivant :

Lemme 2.2. *Sous les conditions du lemme 2.1, et si $r > 2$, on a $\forall k = 1, \dots, 2^j$,*

$$|s_k^0 - \langle f, \tilde{\varphi}_{jk} \rangle| \leq 2^{j(p - \frac{1}{2})} N^{-p} (\text{const.} \|f\|_\infty + o(1)).$$

La constante ci-dessus dépend uniquement de la fonction d'échelle φ et de sa régularité, et p est égal à $0.2(r - 1)$.

Preuve du lemme 2.2 :

Nous nous contenterons de démontrer le résultat pour des fonctions $\tilde{\varphi}_{jk}$ définies par simple dilatation et translation de la fonction d'échelle φ i.e :

$$\tilde{\varphi}_{jk} = \varphi_{jk} = 2^{\frac{j}{2}} \varphi(2^j \cdot - k).$$

Le support de φ étant l'intervalle $[-L, L - 1]$ celui de φ_{jk} est l'intervalle $[\frac{-L+k}{2^j}, \frac{L-1+k}{2^j}]$.

D'autre part comme $L = r - 1$ et que φ est de classe $\mathcal{C}^p \subset \mathcal{H}^p$ avec $p \approx 0.2 * L \Gamma$ nous avons :

$$\begin{aligned} \langle f, \varphi_{jk} \rangle - \frac{1}{N} \sum_{i=1}^N \varphi_{jk}(\frac{i}{N}) f(\frac{i}{N}) &= \int_0^1 f(x) \varphi_{jk}(x) dx - \frac{1}{N} \sum_{i=1}^N \varphi_{jk}(\frac{i}{N}) f(\frac{i}{N}) \\ &= \sum_{i=1 + [\frac{-L+k}{2^j}] }^{[\frac{L-1+k}{2^j}]} \int_{\frac{i-1}{N}}^{\frac{i}{N}} (f(x) \varphi_{jk}(x) \\ &\quad - f(\frac{i}{N}) \varphi_{jk}(\frac{i}{N})) dx \\ &= \sum \int_{\frac{i-1}{N}}^{\frac{i}{N}} [\varphi_{jk}(\frac{i}{N}) (f(x) - f(\frac{i}{N})) \\ &\quad + (\varphi_{jk}(x) - \varphi_{jk}(\frac{i}{N})) f(x)] dx. \end{aligned}$$

$[x]$ et $\lceil x \rceil$ ci-dessus désignent respectivement le plus grand entier (le plus petit entier) inférieur ou égal (supérieur ou égal) à x .

Ainsi

$$\begin{aligned}
| \langle f, \varphi_{jk} \rangle - \frac{1}{N} \sum_{i=1}^N \varphi_{jk}(\frac{i}{N}) f(\frac{i}{N}) | &\leq \sum_{i=1}^{\frac{i}{N}} \left[2^{\frac{j}{2}} |\varphi(2^j \frac{i}{N} - k)| |f(x) - f(\frac{i}{N})| \right. \\
&\quad \left. + 2^{j(p+\frac{1}{2})} \sup |f(x)| \text{const.}_p |x - \frac{i}{N}|^p \right] dx \\
&= \sum_{i=1}^{\frac{i}{N}} \left[2^{\frac{j}{2}} |\varphi(2^j \frac{i}{N} - k)| \text{const.}_r |x - \frac{i}{N}|^{r-1} \right. \\
&\quad \left. + 2^{j(p+\frac{1}{2})} |f(x)| \text{const.}_p |x - \frac{i}{N}|^p \right] dx. \\
&= \sum \left[2^{\frac{j}{2}} \sup |\varphi(x)| \frac{\text{const.}_r}{r} N^{-r} \right. \\
&\quad \left. + 2^{j(p+\frac{1}{2})} \sup |f(x)| \frac{\text{const.}_p}{p+1} N^{-(p+1)} \right]. \\
&= 2^{j(p-\frac{1}{2})} N^{-p} \left(\left(\frac{2L-1}{p+1} \right) \text{const.}_p \sup |f| + o(1) \right).
\end{aligned}$$

Pour les autres éléments Γ la démarche ci-dessus ne varie pas sauf que pour chaque cas il convient de préciser le support de $\tilde{\varphi}_{jk}$ (cf. Annexe A). ■

Le lemme 2.2 établit soit \tilde{f}_j l'approximation de f_j définie par :

$$\tilde{f}_j = \sum_{k=1}^{2^j} s_k^0 \tilde{\varphi}_{jk}. \quad (2.3)$$

Il est maintenant naturel d'estimer f_j par

$$\hat{f}_j = \sum_{k=1}^{2^j} \hat{s}_k^0 \tilde{\varphi}_{jk}, \quad (2.4)$$

où \hat{s}_k^0 est obtenu en remplaçant les valeurs F_i par y_i dans (2.2) ; ce qui conduit à :

$$\hat{s}_k^0 = \frac{1}{N} \sum_{i=1}^N y_i \tilde{\varphi}_{jk}(x_i). \quad (2.5)$$

Le calcul du MISE de l'estimateur défini par (2.4) revient à évaluer l'erreur d'approximation de f par $f_j \Gamma$ sa projection sur le $j^{\text{ème}}$ niveau de la multirésolution Γ puis à déterminer le biais et la variance de $\hat{f}_j \Gamma$ l'estimateur de f_j ; ainsi Γ sous les hypothèses requises pour f et φ dans les lemmes 2.1-2 Γ nous établissons le théorème suivant :

Théorème 2. 1. *Sous les conditions du lemme 2.1 et pour l'ondelette définie au paragraphe 2.1, on a*

$$MISE(\hat{f}_j) \leq \frac{2^j}{N} \sigma^2 (1 + o(1)) + \text{const.} \left(\frac{2^j}{N}\right)^{2p} (1 + o(1)) + C_r 2^{-2jr},$$

const. étant une constante qui dépend que de φ et de f .

Ainsi,

1. si $p \geq \frac{1}{2}$ (i.e $r \geq 4$), alors le MISE est optimal pour J_{opt} tel que

$$J_{opt} = \log_2(C_1 N^{\frac{1}{2r+1}}),$$

et sa valeur optimale est au plus égale à $C_r C_1 N^{\frac{-2r}{2r+1}}$, C_1 étant égale à $\left(\frac{2rC_r}{\sigma^2}\right)^{\frac{1}{2r+1}}$ et où C_r est la constante du lemme 2.1.

2. si $p < \frac{1}{2}$, le MISE est optimal pour J_{opt} tel que

$$2^{J_{opt}} = \mathcal{O}(N^{\frac{p}{r+p}}),$$

et sa valeur optimale est égale à $\mathcal{O}(N^{\frac{-2pr}{r+p}})$.

Preuve du théorème 2.1 :

Il est facile de voir que par le lemme 2.2 Il existe une constante *const* telle que :

$$\int_0^1 (\tilde{f}_j - f_j)^2(x) dx = \text{const.} \left(\frac{2^j}{N}\right)^{2p} (1 + o(1)).$$

D'autre part :

$$E[\hat{f}_j] = \tilde{f}_j$$

$$\begin{aligned} \text{et } \text{var}[\hat{s}_k^0] &= \text{var}\left[\frac{1}{N} \sum_{i=1}^N Y_i \tilde{\varphi}_{j,k}\left(\frac{i}{N}\right)\right] \\ &= \frac{\sigma^2}{N^2} \sum_{i=1}^N \tilde{\varphi}_{j,k}^2\left(\frac{i}{N}\right) \\ &= \frac{\sigma^2}{N} \left(\int \tilde{\varphi}_{j,k}^2(x) dx + o(1)\right) \\ &= \frac{\sigma^2}{N} (1 + o(1)). \end{aligned}$$

Par l'égalité de Parseval nous avons :

$$\begin{aligned} E \int_0^1 (\hat{f}_j - f_j)^2(x) dx &= E\left[\sum_{k=1}^{2^j} (\hat{s}_k^0 - \langle f, \varphi_{jk} \rangle)^2\right] \\ &= \sum_{k=1}^{2^j} E\left[(\hat{s}_k^0 - \langle f, \varphi_{jk} \rangle)^2\right] \\ &= \sum_{k=1}^{2^j} \left\{ (E[\hat{s}_k^0] - \langle f, \varphi_{jk} \rangle)^2 + \text{var}[\hat{s}_k^0] \right\} \\ &= \int_0^1 (\tilde{f}_j - f_j)^2(x) dx + \sum_{k=1}^{2^j} \text{var}[\hat{s}_k^0]. \end{aligned}$$

Ainsi

$$E \int_0^1 (\hat{f}_j - f_j)^2(x) dx = \text{const.} \left(\frac{2^j}{N}\right)^{2p} (1 + o(1)) + \frac{2^j}{N} \sigma^2 (1 + o(1)). \quad (2.6)$$

La décomposition suivante

$$MISE(\hat{f}_j) = E \int_0^1 (\hat{f}_j - f_j)^2(x) dx + \int_0^1 (f_j - f)^2(x) dx,$$

le lemme 2.1 et l'identité (2.6) permettent alors de conclure la démonstration du Théorème 2.1 ■

Nous avons établi dans ce paragraphe une procédure d'estimation de la fonction de régression à l'aide d'une décomposition sur une base d'ondelettes à support compact présentant une erreur quadratique moyenne intégrée optimale au sens de Stone [67]. Il est alors naturel d'évoquer le choix de la base optimale permettant une estimation adaptée aux données observées de la fonction de régression.

2.2 Régression par paquets d'ondelettes

Avant de procéder à l'écriture de l'estimateur de type projection basé sur des paquets d'ondelettes nous présentons sommairement la définition d'une base de paquets d'ondelettes sur l'intervalle.

La donnée de la base orthonormée $\mathcal{B}_{J_0}^{[0,1]}$ du paragraphe 2.1.1 est équivalente à la décomposition de $L^2([0, 1])$ en somme orthogonale des sous espaces $V_{J_0}^{[0,1]}$ et $W_l^{[0,1]}$, $l = J_0, J_0 + 1, \dots, \infty$, i.e :

$$L^2([0, 1]) = V_{J_0}^{[0,1]} \oplus W_{J_0+1}^{[0,1]} \oplus \dots$$

Il est par ailleurs évident que pour chaque $j \geq J_0$ en appliquant le schéma ci-après

$$V_l^{[0,1]} = V_{l-1}^{[0,1]} \oplus W_{l-1}^{[0,1]}$$

$$V_{l-1}^{[0,1]} = V_{l-2}^{[0,1]} \oplus W_{l-2}^{[0,1]},$$

pour $l = j, \dots, J_0 + 2$ et où la somme est orthogonale nous obtenons une unique décomposition de $V_j^{[0,1]}$ de la forme :

$$V_j^{[0,1]} = V_{J_0}^{[0,1]} \oplus W_{J_0+1}^{[0,1]} \oplus \dots \oplus W_{j-1}^{[0,1]}.$$

Une telle décomposition est désignée par décomposition en ondelettes de $V_j^{[0,1]}$. Il s'ensuit que

$$\mathcal{B}_{J_0,j}^{[0,1]} = \{\tilde{\varphi}_{J_0k}, \tilde{\psi}_{lk}, l = J_0, J_0 + 1, \dots, j - 1, k = 1, \dots, 2^l\},$$

est une base orthonormée de $V_j^{[0,1]}$.

Nous pourrions également décomposer les $W_l^{[0,1]}$ dans le schéma de la décomposition en ondelettes. Nous obtiendrions ainsi :

$$V_j^{[0,1]} = V_{j-1}^{[0,1]} \oplus W_{j-1}^{[0,1]},$$

puis

$$\begin{cases} V_{j-1}^{[0,1]} &= V_{j-2}^{[0,1]} \oplus W_{j-2}^{[0,1]} \\ W_{j-1}^{[0,1]} &= W_{j-2,2}^{[0,1]} \oplus W_{j-2,3}^{[0,1]} \end{cases}$$

En posant $W_{j-2,1}^{[0,1]} = W_{j-2}^{[0,1]}$ et $W_{j-2,0}^{[0,1]} = V_{j-2}^{[0,1]} \Gamma$ nous avons alors

$$V_j = W_{j-2,0}^{[0,1]} \oplus W_{j-2,1}^{[0,1]} \oplus W_{j-2,2}^{[0,1]} \oplus W_{j-2,3}^{[0,1]},$$

la somme étant orthogonale.

Soit alors le schéma suivant

$$W_{lk}^{[0,1]} = W_{l,2k}^{[0,1]} \oplus W_{l,2k+1}^{[0,1]},$$

$l = j, j - 1, \dots, J_0 + 1$, et $k = 0, 1, \dots, 2^l - 1 \Gamma$ où $W_{l,0} = V_l$ et $W_{l,1} = W_l \Gamma$ et la somme est orthogonale. Son application conduit à la décomposition suivante

$$V_j^{[0,1]} = V_{J_0}^{[0,1]} \oplus W_{J_0,1}^{[0,1]} \oplus \dots \oplus W_{J_0,2^{j-J_0}-1}^{[0,1]}.$$

Une telle décomposition est appelée décomposition en paquets d'ondelettes de $V_j^{[0,1]}$ au niveau J_0 ou encore décomposition en paquets d'ondelettes de $V_j^{[0,1]}$ à l'ordre $j - J_0$; elle sera notée par **DDPO**.

Concrètement Γ la **DDPO** équivaut à la construction (voir annexe A) d'une suite d'éléments de base de paquets d'ondelettes

$$\{\omega_{lk}, k = 1, 2, \dots, 2^{J_0}, l \geq J_0\},$$

telle que $\omega_{J_0 k} = \tilde{\varphi}_{J_0 k}$ et $\omega_{J_0+1, k} = \tilde{\psi}_{J_0 k}$. Par construction même $\Gamma V_j^{[0,1]}$ est engendré par $\{\omega_{J_0 k}, k = 1, 2, \dots, 2^{J_0}\}$ et chaque $W_{J_0 l}^{[0,1]} \Gamma l = 1, \dots, 2^{j-J_0} - 1 \Gamma$ est engendré par $\{\omega_{l k}, k = 1, 2, \dots, 2^{J_0}\}$. Nous en déduisons que

$$\mathcal{B}_j = \{\omega_{l k}, J_0 \leq l \leq 2^{j-J_0} - 1, k = 1, 2, \dots, 2^{J_0}\}$$

est une base orthonormée de $V_j^{[0,1]}$.

2.2.1 Décomposition en paquets d'ondelettes d'une fonction

Une des propriétés caractéristiques d'une analyse multirésolution réelle associée à des ondelettes (voir paragraphe 2.1.1) est que pour tout $j_1 \Gamma j_2$ dans $\mathbb{Z} \Gamma$ si $V_{j_1} \Gamma j_1^{\text{me}}$ niveau de la multirésolution Γ est engendré par $\{\varphi_{j_1 k}, k \in \mathbb{Z}\}$ alors V_{j_2} est engendré par $\{2^{\frac{j_2-j_1}{2}} \varphi_{j_1 k}(2^{j_2-j_1} \cdot), k \in \mathbb{Z}\}$. Cette propriété est préservée dans la construction de Cohen et al [17] des ondelettes sur l'intervalle.

Ainsi Γ pour le choix du système

$$\{2^{\frac{j-J_0}{2}} \tilde{\varphi}_{J_0 k}(2^{j-J_0} \cdot), k = 1, 2, \dots, 2^j\},$$

comme base de $V_j^{[0,1]} \Gamma$ soit \tilde{f}_j défini par

$$\tilde{f}_j = \sum_{k=1}^{2^j} s_k^0 2^{\frac{j-J_0}{2}} \tilde{\varphi}_{J_0 k}(2^{j-J_0} \cdot),$$

les s_k^0 étant définis par (2.2) où nous remplaçons $\tilde{\varphi}_{j k}$ par $2^{\frac{j-J_0}{2}} \tilde{\varphi}_{J_0 k}(2^{j-J_0} \cdot)$. La fonction $\tilde{f}_j \Gamma$ ainsi définie Γ correspond à une approximation de la projection de f sur $V_j^{[0,1]}$.

Posons dans tout ce qui suit $g = \tilde{f}_j$. La **DDPO** de g consiste à écrire g sous la forme :

$$g(x) = g_{11}(x) + g_{12}(x).$$

Les fonctions g_{11} et $g_{12} \Gamma$ respectivement dans $V_{j-1}^{[0,1]}$ et $W_{j-1}^{[0,1]} \Gamma$ sont définies par :

$$g_{11}(x) = \sum_{k=1}^{2^{j-1}} s_k^{1,1} 2^{\frac{j-J_0-1}{2}} \omega_{J_0 k}(2^{\frac{j-J_0-1}{2}} x),$$

$$\text{et } g_{12}(x) = \sum_{k=1}^{2^{j-1}} s_k^{1,2} 2^{\frac{j-J_0-1}{2}} \omega_{J_0+1, k}(2^{\frac{j-J_0-1}{2}} x).$$

En réitérant la procédure un nombre de fois égal à l nous obtenons une nouvelle écriture de g :

$$g(x) = \sum_{m=1}^{2^l} g_{l,m}(x), \quad (2.7)$$

avec

$$g_{l,m}(x) = \sum_{k=1}^{2^{j-l}} s_k^{l,m} 2^{\frac{j-J_0-l}{2}} \omega_{J_0+m-1,k} \left(2^{\frac{j-J_0-l}{2}} x \right). \quad (2.8)$$

Chaque $s_k^{l,m}$ défini ci-dessus s'écrit

$$s_k^{l,m} = \frac{1}{N} \sum_{i=1}^N F_i 2^{\frac{j-J_0-l}{2}} \omega_{J_0+m-1,k} \left(2^{\frac{j-J_0-l}{2}} x_i \right).$$

Cette décomposition équivaut à la construction d'une matrice orthogonale $W_j^{(l)}$ d'ordre 2^j et dont l'application au vecteur \mathbf{s}^0 définit la **DDPO** de g (i.e : \tilde{f}_j) à l'ordre l (ou encore décomposition de g au niveau $j-l$) :

$$\mathbf{s}^{(l)} := W_j^{(l)} \mathbf{s}^0, \quad (2.9)$$

de version empirique

$$\hat{\mathbf{s}}^{(l)} := W_j^{(l)} \hat{\mathbf{s}}^0.$$

Il est maintenant aisé de voir qu'une estimation de f est donnée par :

$$g^{\text{est}} = \sum_{m=1}^{2^l} g_{l,m}^{\text{est}}, \quad (2.10)$$

où

$$g_{l,m}^{\text{est}} = \sum_{k=1}^{2^{j-J_0-l}} \hat{s}_k^{l,m} 2^{\frac{j-J_0-l}{2}} \omega_{J_0+m-1,k} \left(2^{\frac{j-J_0-l}{2}} \cdot \right). \quad (2.11)$$

Remarque 2. 1. Le vecteur $\mathbf{s}^{(l)}$, défini par (2.9) et son analogue $\hat{\mathbf{s}}^{(l)}$, ont une décomposition en blocs qui correspond à la **DDPO** de g .

$$\mathbf{s}^{(l)} = [\mathbf{s}^{l,1}, \mathbf{s}^{l,2}, \dots, \mathbf{s}^{l,2^l}]'$$

$$\hat{\mathbf{s}}^{(l)} = [\hat{\mathbf{s}}^{l,1}, \hat{\mathbf{s}}^{l,2}, \dots, \hat{\mathbf{s}}^{l,2^l}]'$$

$s_k^{l,m}$ et $\hat{s}_k^{l,m}$ sont les $k^{\text{ème}}$ composantes respectives des blocs $\mathbf{s}^{l,m}$ et $\hat{\mathbf{s}}^{l,m}$, de dimension 2^{j-l} .

Les paramètres l et m utilisés dans (2.11) déterminent respectivement l'ordre de la décomposition et le numéro du paquet.

En effet la décomposition à l'ordre l de \mathbf{s}^0 conduit au schéma suivant :

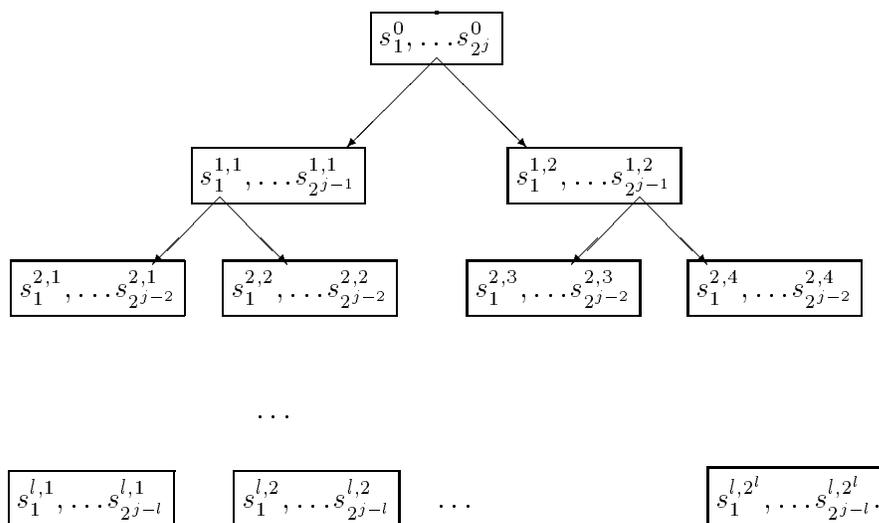


Figure 2.1 : Décomposition en paquets d'ondelettes à l'ordre l .

Grâce à la décomposition en paquets d'ondelettes l'estimateur de la projection de f sur $V_j^{[0,1]}$ (resp. V_j) admet plusieurs écritures équivalentes chacune associée à une base particulière de $V_j^{[0,1]}$ (resp. V_j). Nous nous proposons d'en choisir la plus adaptée et cela afin d'obtenir une estimation adaptative de la fonction de régression. L'ensemble des procédures que nous allons définir dans les paragraphes 2.3-5 reposent sur le schéma suivant :

1. Partant du vecteur des observations \mathbf{y} et par une décomposition discrète en paquets d'ondelettes (**DDPO**) nous déterminons le vecteur suivant :

$$\mathbf{v} = W\mathbf{y}, \quad (2.12)$$

où W est une matrice orthogonale.

2. Nous définirons un opérateur non linéaire \mathbb{T} voire procédure afin d'annuler ou de corriger certaines composantes de \mathbf{v} ; il en résulte $\mathbf{v}' = \mathbb{T}(\mathbf{v})$.
3. L'estimation $\hat{\mathbf{F}}$ de \mathbf{F} définie dans le paragraphe 2.1 est alors obtenue par l'application de la matrice transposée W^T à \mathbf{v}' ; d'où une estimation des valeurs échantillonnées de f aux points x_i .

Notre technique d'estimation se résume donc à définir :

$$\hat{\mathbf{F}} = (W^T \circ \top \circ W)\mathbf{y}.$$

2.3 Généralisation de l'Algorithme 1

2.3.1 Description de la procédure

Compte tenu de la remarque émise au paragraphe 1.6 du chapitre 1 à propos du lien établi entre l'estimateur de type projection basé sur le système de Haar et l'analyse multirésolution nous allons adapter le schéma de l'algorithme 1 à la **DDPO**. Pour cela nous identifierons dans un premier temps les paramètres de contrôle de l'algorithme 1 ; ainsi la longueur optimale des intervalles dyadiques 2^{-J} par exemple correspond au niveau de résolution désiré d'ordre J . Aux coefficients de la transformée en ondelettes discrète et finie seront substitués leurs analogues obtenus par la **DDPO** ; quand au seuil utilisé dans le critère de sélection de l'algorithme 1 il sera tout simplement remplacé par une estimation de la variance des paquets.

Pour cette première procédure la définition du vecteur \mathbf{v} défini par (2.12) nécessite deux étapes :

1. A partir du vecteur des observations \mathbf{y} nous construisons par (2.4) et (2.5) une estimation \hat{f}_J de la projection sur $V_J^{[0,1]}$ de f la fonction de régression du modèle défini en début de ce chapitre. Le choix de la valeur de J est donné par le Théorème 2.1 et conduit à une valeur asymptotiquement optimale du MISE.
2. Cette estimation est ensuite décomposée en paquets d'ondelettes à l'ordre l selon le schéma décrit dans le paragraphe précédent. Le vecteur obtenu suite à la décomposition en paquets d'ondelettes est noté \mathbf{v} .

Nous obtenons ainsi 2^l paquets contenant chacun 2^{J-l} coefficients ; les différents paquets sont alors associés aux nœuds d'un arbre binaire de la façon suivante :

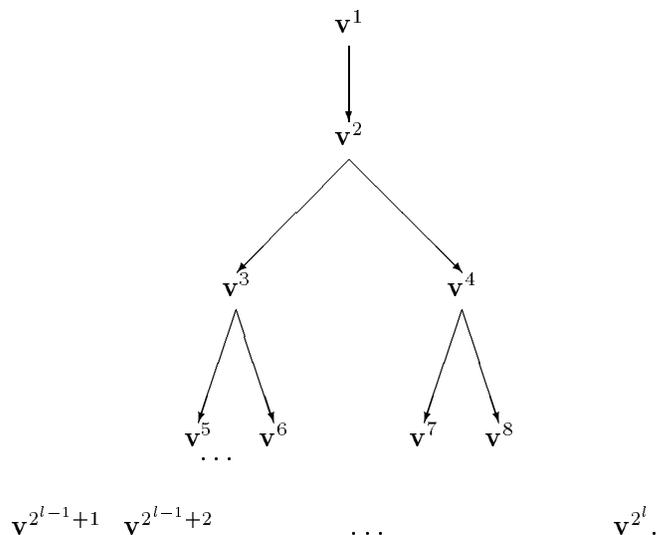


Figure 2.2 : Arbre binaire associé à la procédure.

Par l'utilisation de l'algorithme ci-après nous procédons alors à une sélection des paquets par une technique analogue à celle utilisée dans l'algorithme 1 du chapitre 1.

ALGORITHME 2

Retenir v^1
 Pour $j_{puis} = 1, 2 \dots l - 1$
 * Pour $j = 1, 2 \dots 2^{j_{puis}-1}$
 * Faire
 * $ind_0 = 2^{j_{puis}} + j$
 * $ind = ind_0$
 * Faire jusqu'à $|v^{ind}|^2 > \frac{2^{J-l+1} var}{N}$
 * $ind = ant(ind)$
 * Fin de Faire
 * Si moyarithmétique ($|v^{ind_0}|^2, |v^{ant(ind_0)}|^2, \dots, |v^{ind}|^2$) $> \frac{2^{J-l+1} var}{N}$
 * Alors retenir tous les éléments $v^{ind_0}, v^{ant(ind_0)}, \dots, v^{ind}$
 * Autrement $j = j + 1$
 * Fin de Si
 * Fin de Pour ;
 * $j_{puis} = j_{puis} + 1$
 Fin de Pour.

Dans l'algorithme 2 le paramètre var correspond à une estimation de la variance du bruit. Par ailleurs la fonction ant est définie de \mathbb{N} dans \mathbb{N} par :

$$ant(j) = \left[\frac{j+1}{2} \right] \text{ où } [x] \text{ désigne la partie entière de } x.$$

L'application de cet algorithme permet alors de construire un vecteur \mathbf{v}' qui par reconstitution conduit au calcul de l'estimateur de f via (2.10) et (2.11).

2.3.2 Simulations

Les simulations suivantes ont pour but d'illustrer la procédure d'estimation associée à l'algorithme 2. Les ondelettes utilisées dans nos simulations sont les ondelettes de Daubechies d'ordre $L = 1, 3, 6$.

L'estimation de la variance du bruit est obtenue par la méthode de Hall ; ce choix est largement commenté au chapitre 1. Pour le choix de l'ordre de la **DDPO** nous avons pris la plus grande valeur possible.

Afin de comparer les résultats obtenus par l'utilisation de l'algorithme 1 pour le système de Haar et la procédure décrite ci-dessus associée aux paquets d'ondelettes de Haar nous avons utilisé une des fonctions de régression utilisées au chapitre 1.

L'ensemble des simulations a été effectué avec une seule et même fonction de régression

$$f(x) = x(x - 0.3)(x - 0.5)(x - 1.) + 0.2.$$

Les données sont issues du modèle $Y_i = f(x_i) + \varepsilon_i$ avec ε_i i.i.d de loi Gaussienne centrée et de variance σ^2 . La fonction f a été discrétisée en 256 points équidistants sur l'intervalle $[0, 1]$. Pour l'ensemble des simulations le rapport signal/bruit a été choisi égal à 2.

Dans un premier temps Γ nous comparerons les résultats obtenus via l'utilisation de l'algorithme 1 du chapitre 1 et la procédure décrite ci-dessus associée aux paquets d'ondelettes de Haar. Cette comparaison est illustrée par les figures 2.3 et 2.4.

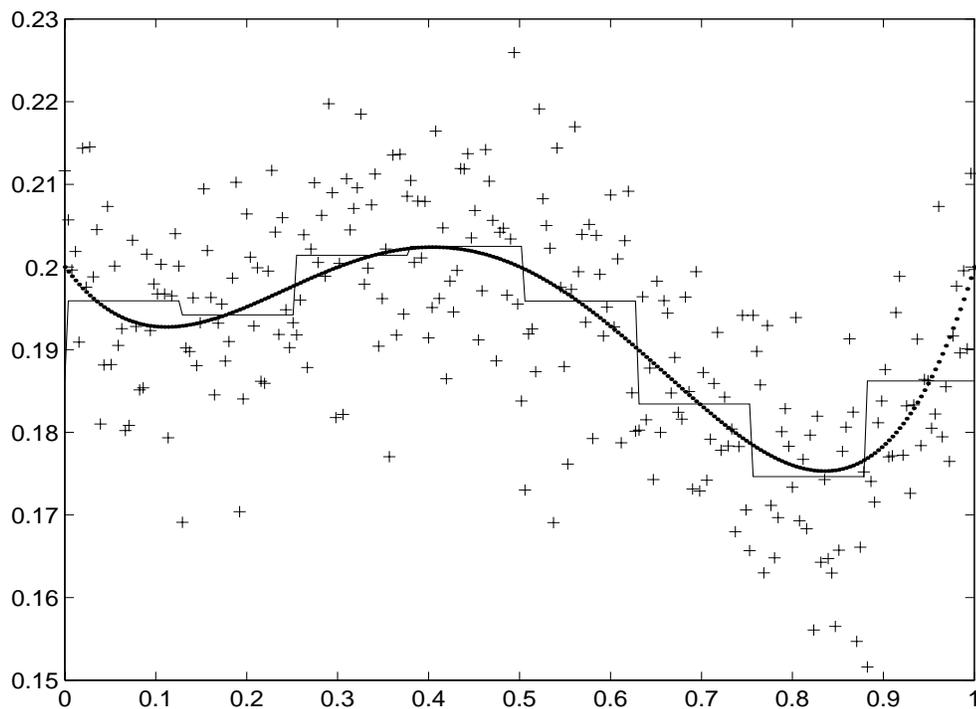


Figure 2.3 : Estimation obtenue pour une décomposition au niveau 3 avec le même niveau de bruit et le même nombre de points échantillons que la figure 1.3 du premier chapitre.

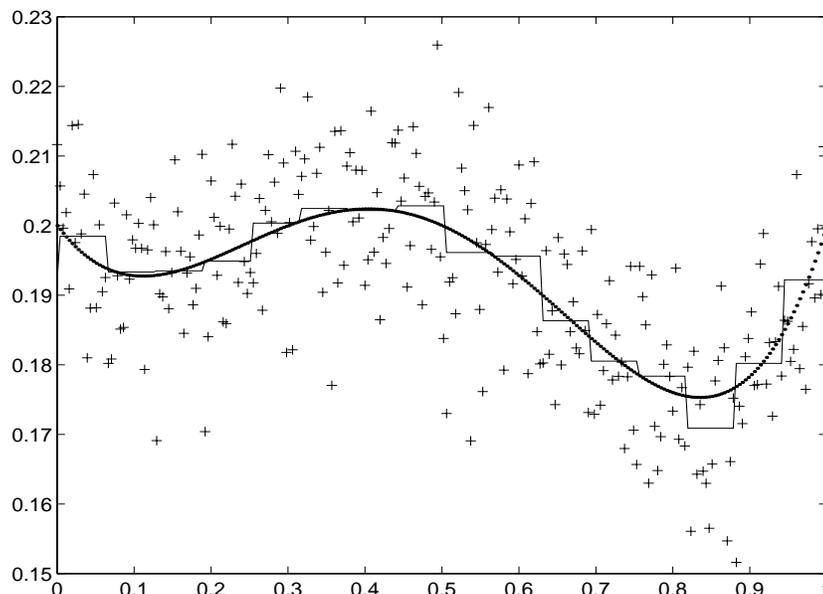


Figure 2.4 : Estimation obtenue pour une décomposition au niveau 4 de la multirésolution. Le nombre de points échantillons N est égal à 256 et le rapport signal bruit est égal à 2.

Contrairement à l'estimateur adaptatif basé sur le système de Haar (voir chapitre 1) l'estimation basée sur les paquets d'ondelettes produit des intervalles de même longueur ; cependant il s'adapte parfaitement à la régularité de la fonction de régression. Ainsi sur l'intervalle $[0.25, 0.5]$ l'estimateur de Haar (chapitre 1) produit deux intervalles de même longueur 0.125 (la valeur optimale des intervalles est égale à 0.0625) et deux constantes dont les valeurs sont proches (voir fig 1.3) ; une estimation identique est obtenue pour le deuxième estimateur avec un niveau égal à 3 (voir fig 2.3) alors que pour un niveau égal à 4 (voir fig 2.4) l'estimateur basé sur les paquets d'ondelettes produit quatre constantes dont deux sont quasiment identiques ce qui revient à une estimation sur l'intervalle $[0.3125, 0.4375]$ de longueur 0.125 (la valeur optimale des intervalles ayant été choisie égale à 0.0625).

Ci-après nous utilisons des ondelettes de Daubechies régulières. Pour les figures 2.5-6 et 2.7-8 le niveau de résolution a été fixé à 3 alors que l'ordre de l'ondelette est successivement égal à 3 puis à 6 (i.e $L = 3$ et $L = 6$). Dans chacun de ces deux cas (figures 2.5 et 2.7) nous avons d'abord testé l'efficacité de la procédure sur des données non bruitées.

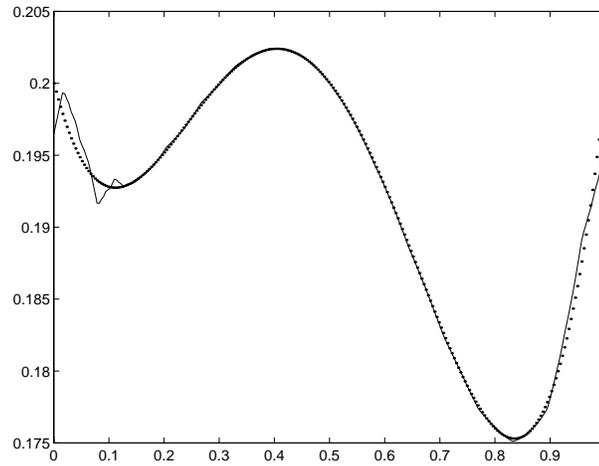


Figure 2.5 : Estimation obtenue par les ondelettes de Daubechies d'ordre 3 avec un échantillon de taille 256.

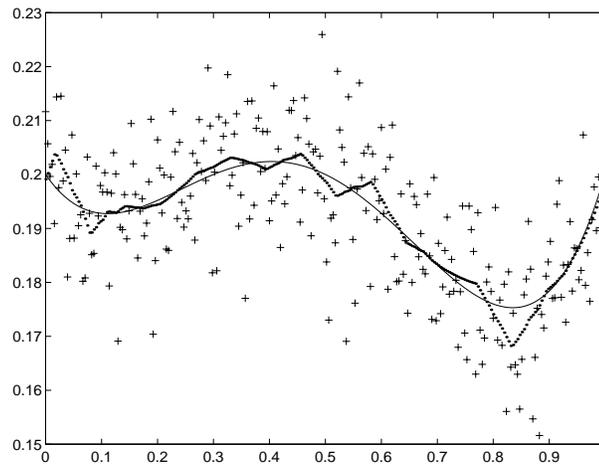


Figure 2.6 : 256 points échantillons ont été utilisés avec un rapport signal-bruit égal à 2.

La fonction de régression est représentée en trait continu et son estimation en pointillé.

Pour les figures 2.5 et 2.6 (ainsi que 2.7 et 2.8 ci-après) les paquets retenus via l'application de l'algorithme 2 pour le signal avec bruit sont les mêmes que ceux retenus pour le signal sans bruit.

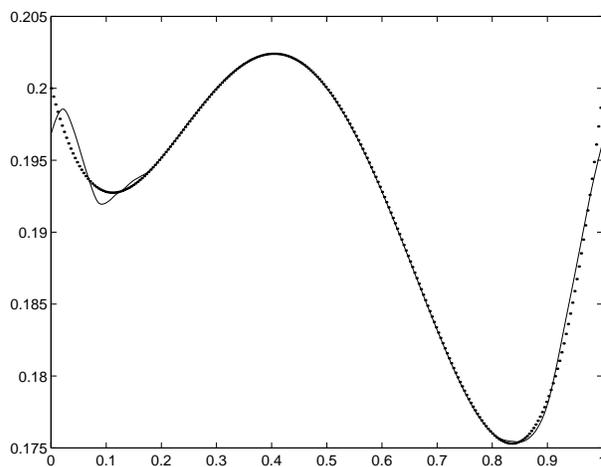


Figure 2.7 : Estimation obtenue par les ondelettes de Daubechies d'ordre 6.

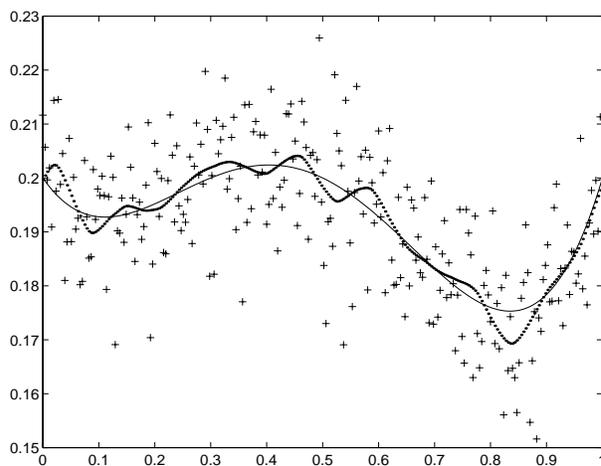


Figure 2.8 : 256 points échantillons ont été utilisés avec un rapport signal-bruit égal à 2.

Compte tenu de l'ensemble des simulations effectuées nous remarquons

1. La lissitude de l'estimateur augmente avec la régularité de l'ondelette utilisée conduisant à une amélioration de la qualité de l'estimateur proposé.
2. Contrairement à beaucoup d'estimateurs de régression la qualité de l'estimateur obtenu par l'application de l'algorithme 2 ne se dégrade pas trop aux extrémités du support de la fonction de régression.

2.4 Tests et Seuillages

Dans ce paragraphe nous supposons que le bruit du modèle (2.0) est Gaussien et que N le nombre des observations est une puissance de 2 ; ainsi il existe n entier naturel tel que $N = 2^n$.

Plutôt que d'estimer f par un estimateur de sa projection sur $V_j^{[0,1]}$ nous préférons décomposer le vecteur des coefficients empiriques par paquets d'ondelettes à partir de la résolution la plus fine. Pour sélectionner la meilleure base nous utiliserons des tests et des techniques de seuillage. Nous estimerons d'abord $P_{V_n}^{[0,1]}(f)$. Pour cela nous utiliserons des ondelettes à support compact de type coiflet (cf. Annexe A). Les coiflets n'ont jusque là pas été associés à une AME sur l'intervalle et de ce fait une correction des effets de bord s'impose et nous pourrons utiliser l'une des méthodes décrites dans [17]. Nous pouvons toutefois considérer que nous estimons la fonction de régression sur un intervalle strictement inclus dans $[0, 1]$. Nous utiliserons alors une estimation de la projection sur V_n au lieu de $V_n^{[0,1]}$ en se basant sur la donnée du vecteur \mathbf{F} . Posons :

$$\mathbf{s}^0 = 2^{-\frac{n}{2}} \mathbf{F}, \quad (2.13)$$

$$\text{et } \hat{\mathbf{s}}^0 = 2^{-\frac{n}{2}} \mathbf{y}.$$

Le choix du vecteur \mathbf{s}^0 défini ci-dessus est justifié par le résultat suivant :

Lemme 2. 3 (A. Antoniadis) *Sous les hypothèses suivantes :*

- H_1 : f appartient à l'espace de Sobolev $H^\beta([0, 1])$ $\beta > \frac{1}{4}$.
- H_2 : l'AME est r -régulière, i.e : la fonction d'échelle est continûment dérivable à l'ordre r et ses dérivées sont à décroissance rapide .

Alors pour une coiflet d'ordre L telle que $L > \beta + 1$, et pour tout $L \leq k \leq 2^n - L$, nous avons :

$$2^{\frac{n}{2}} | \langle f, \varphi_{n,k}(\cdot) \rangle - s_k^0 | \leq \mathcal{O}(2^{-n(\beta + \frac{1}{2})}).$$

Partant du lemme 2.3 nous déterminerons la **DDPO** à l'ordre l du vecteur \mathbf{s}^0 . Soit $W_n^{(l)}$ la matrice orthogonale associée à cette décomposition. Dans ce qui suit nous fixerons l à $J = n - J_{opt}$ où J_{opt} est la valeur optimale du

Théorème 2.1 et nous poserons $W = W_n^{(J)}$.

En appliquant W de part et d'autre de la relation (2.0) nous obtenons :

$$\begin{aligned}\mathbf{v} = W\hat{\mathbf{s}}^0 &= 2^{-\frac{n}{2}}W\mathbf{y} \\ &= 2^{-\frac{n}{2}}W(\mathbf{F} + \underline{\boldsymbol{\varepsilon}}) \\ &= 2^{-\frac{n}{2}}(\Theta + \mathbf{z}),\end{aligned}$$

où $\Theta = W\mathbf{F}$ et $\mathbf{z} = W\underline{\boldsymbol{\varepsilon}}$.

Comme W est orthogonale \mathbf{z} a la même structure de variance-covariance que $\underline{\boldsymbol{\varepsilon}}$ et de plus \mathbf{z} est un vecteur Gaussien.

Nous nous proposons ci-après de construire une estimation adaptée de Θ par l'application d'un opérateur non linéaire T . Ce dernier aura pour principale tâche de détecter les blocs de \mathbf{v} qui ne contiennent essentiellement que du bruit et de les éliminer (i.e : annuler leurs composantes) puis de "corriger" ceux qui contiennent le signal pour aboutir ainsi à une bonne estimation de Θ . Nous allons donc :

1. Opérer un test sur chacun des paquets en utilisant le travail de J. Fan [31].
2. Adapter les paquets en utilisant la technique de seuillage définie par Donoho et Johnstone [27].

Le résultat de ces opérations conduit à une version débruitée \mathbf{v}' de \mathbf{v} à laquelle nous associerons par reconstitution l'estimateur \hat{f} de f .

Pour faciliter l'écriture nous supprimerons l'indexation par J et nous adopterons la notation :

$$\mathbf{v} = [\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(2^J)}]'$$

$$\Theta = [\Theta^{(1)}, \Theta^{(2)}, \dots, \Theta^{(2^J)}]'$$

et

$$\mathbf{z} = [\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \dots, \mathbf{z}^{(2^J)}]'$$

Chaque paquet admet une écriture vectorielle de la forme

$$\mathbf{v}^{(m)} = 2^{-\frac{n}{2}}(\Theta^{(m)} + \mathbf{z}^{(m)}).$$

La première étape revient donc à tester si le vecteur moyenne β d'une loi normale de dimension p est nul par la simple donnée d'un vecteur échantillon $\xi \sim N(\beta, \sigma^2 I_p)$ contre l'alternative que l'une au moins de ses composantes est non nulle ; I_p est la matrice identité d'ordre p . Dans notre cas $p = 2^{n-J}$. Pour simplifier nous supposons que l'écart-type σ est donné et sans perte de généralité qu'il est égal à 1.

Nous définissons la statistique suivante :

$$\hat{T}_H^* = \sum_{i=1}^p \xi_i^2 \mathbb{1}_{(|\xi_i| > \delta)}.$$

Cette statistique permet pour un choix approprié de δ d'exclure tout ξ_i correspondant à $\beta_i = 0$. Ainsi le problème ci-dessus se réduit à tester si les composantes les plus importantes de β sont nulles sans aucune connaissance a priori sur leur nombre et leur localisation. Ceci correspond tout à fait à la nature de notre problème. Pour plus de détails le lecteur est renvoyé au travail de J. Fan [31].

Les valeurs asymptotiques de l'espérance et de la variance de cette statistique sont respectivement :

$$\mu_p = \sqrt{\frac{2}{\pi}} a_p \delta (1 + \delta^{-2}) \text{ et } \sigma_p^2 = \sqrt{\frac{2}{\pi}} a_p \delta^3 (1 + 3\delta^{-2}).$$

Soient c et d deux constantes positives. Pour le choix

$$\delta = \sqrt{2 \log\left(\frac{p}{a_p}\right)} \text{ avec } a_p = c(\log p)^d,$$

nous avons le résultat suivant :

Théorème 2. 2 (J. Fan) *Sous l'hypothèse nulle et à condition que :*

$$a_p (\log p)^{-\frac{1}{2}} \rightarrow \infty \text{ quand } p \rightarrow \infty,$$

\hat{T}_H^* converge en loi vers une loi normale centrée réduite.

La région critique au risque α associée à ce test est alors donnée par

$$\{\hat{T}_H = \sigma_p^{-1}(\hat{T}_H^* - \mu_p) \geq \Phi^{-1}(1 - \alpha)\},$$

$\Phi^{-1}(1 - \alpha)$ étant le $(1 - \alpha)$ quantile de la loi normale centrée réduite.

La première étape permet donc la détection des paquets à composantes nulles ou jugés comme tels suite au test décrit ci-dessus. La seconde étape a

pour but de corriger l'estimation des $2^{-\frac{n}{2}}\Theta^{(m)}$ (estimation donnée par $\mathbf{v}^{(m)}$). Elle est fondée sur une procédure de seuillage des composantes de chacun des paquets ayant été jugés significatifs pour l'estimation de la fonction de régression.

La méthode de seuillage que nous avons adoptée est issue des travaux de D. Donoho et Ian M. Johnstone [27]. Elle consiste en l'application de la fonction $\eta(t) := \text{sgn}(t)(|t| - \lambda)_+$ à chacune des composantes de $\mathbf{v}^{(m)}$ pour une valeur appropriée de λ .

Partant du fait que pour p variables aléatoires indépendantes \mathbf{z}_i de loi Gaussienne centrée réduite nous avons :

$$Pr\{\max_i |\mathbf{z}_i| > (2 \log p)^{\frac{1}{2}}\} \rightarrow 0 \text{ quand } p \rightarrow \infty,$$

nous prendrons pour le cas étudié $\lambda = (2 \log(p))^{\frac{1}{2}}$. Un tel choix conduit à la suppression de la majeure partie de l'erreur stochastique dans \mathbf{v} .

Pour plus de détails sur cette technique de seuillage le choix de la valeur du seuil et les propriétés asymptotiques de tels estimateurs nous renvoyons le lecteur au travail de Donoho et Johnstone [27].

Pour résumer l'application de la procédure décrite dans ce paragraphe conduit au vecteur \mathbf{v}' définie par

$$\frac{1}{\sigma}(\mathbf{v}')^m = \left(\eta\left(\frac{v_k^m}{\sigma}\right)\right)_{k=1}^{2^n - J},$$

et ce pour les $\mathbf{v}^{(m)}$ non nuls au vu du test effectué et pour le reste par $(\mathbf{v}')^m = 0$.

2.4.1 Simulations

Pour illustrer le comportement asymptotique de l'estimateur défini dans ce paragraphe nous regroupons ci-après quelques simulations.

Pour l'ensemble des simulations effectuées nous avons utilisé une coiffet d'ordre 5.

Les données utilisées sont issues du modèle $Y_i = f(x_i) + \varepsilon_i$ où les ε_i sont i.i.d de loi Gaussienne. Nous avons utilisé deux fonctions de régression f_1 et f_2 afin de tester la qualité de notre estimateur. Ces deux fonctions sont définies par

$$f_1(x) = \begin{cases} 4x^2(3 - 4x) & \text{si } x \leq 0.3 \\ 0.3 & \text{si } 0.3 < x \leq 0.5 \\ \frac{4}{3}x(4x^2 - 10x + 7) - 1.5 & \text{si } 0.5 < x \leq 0.75 \\ \frac{16}{3}x(x - 1)^2 & \text{si } 0.75 < x. \end{cases}$$

et

$$f_2(x) = 3 \cos(2\pi x) + 3\pi^2(x^2 - x + \frac{1}{6}).$$

La première est une fonction polynômiale par morceaux présentant deux discontinuités aux points $x = 0.3$ et $x = 0.75$. La deuxième est infiniment différentiable telle que $f_1(0) = f_1(1)$ mais $f_1'(0) \neq f_1'(1)$.

Chacune des deux fonctions de régression a été discrétisée en 256 et 512 points sur l'intervalle $[0, 1]$ choisis de manière à ce qu'il soient équidistants. Les valeurs de σ^2 variance du bruit ont été déterminées de sorte à avoir le même rapport signal/bruit pour chacun des deux cas. Ce rapport sera égal à 3 et 1.

Notre procédure dépend du paramètre α le risque du test. Pour les simulations qui suivent ci-après nous avons pris deux valeurs 0.1 et 0.01. D'autre part pour l'ensemble des simulations effectuées l'ordre de la **DDPO** est fixé à 3.

Pour étudier la qualité de l'estimateur que nous proposons et l'incidence que peut avoir le choix de la valeur du risque sur l'estimation nous utiliserons le critère d'erreur quadratique moyenne (MSE) définie par :

$$MSE = \frac{1}{N} \sum_{i=1}^N (f(x_i) - \hat{f}(x_i))^2,$$

où $\hat{f}(x_i)$ est l'estimation de f au point x_i .

Pour chacune des deux fonctions nous avons simulé 100 échantillons de taille 256 et 512 et nous avons estimé la fonction de régression. Le critère RMSE est alors défini comme la racine carrée positive de la moyenne sur l'ensemble des 100 échantillons des valeurs du MSE.

L'ensemble des valeurs du RMSE obtenues pour la fonction de régression f_1 (resp. f_2) est regroupé dans le tableau 2.1 (resp. 2.2). Pour les simulations des tableaux 2.1 et 2.2 le rapport signal/bruit (s/b) est égal à 3 et 1 et l'ordre de la décomposition est égal à 3. Les tableaux de droite correspondent à des échantillons de taille 512 et ceux de gauche à 256.

	$\alpha = 0.1$	$\alpha = 0.01$
$s/b = 3$	0.0013	0.0013
$s/b = 2$	0.0017	0.0017
$s/b = 1$	0.0539	0.0539

	$\alpha = 0.1$	$\alpha = 0.01$
$s/b = 3$	0.0009	0.0009
$s/b = 2$	0.0012	0.0012
$s/b = 1$	0.0447	0.0458

Table 2.1 : Résultats des simulations pour la fonction de régression f_1 .

	$\alpha = 0.1$	$\alpha = 0.01$
$s/b = 3$	0.6473	0.6473
$s/b = 2$	0.6607	0.6607
$s/b = 1$	0.84	0.84

	$\alpha = 0.1$	$\alpha = 0.01$
$s/b = 3$	0.3254	0.3263
$s/b = 2$	0.3346	0.3346
$s/b = 1$	0.5718	0.6

Table 2.2 : Résultats des simulations pour la fonction de régression f_2 .

Nous complétons cette étude par quelques graphiques. Pour les figures 2.8-9 la figure de gauche (resp. de droite) est obtenue pour un rapport signal-bruit égal à 3 (resp. 2) et le nombre de points est fixé à 256.

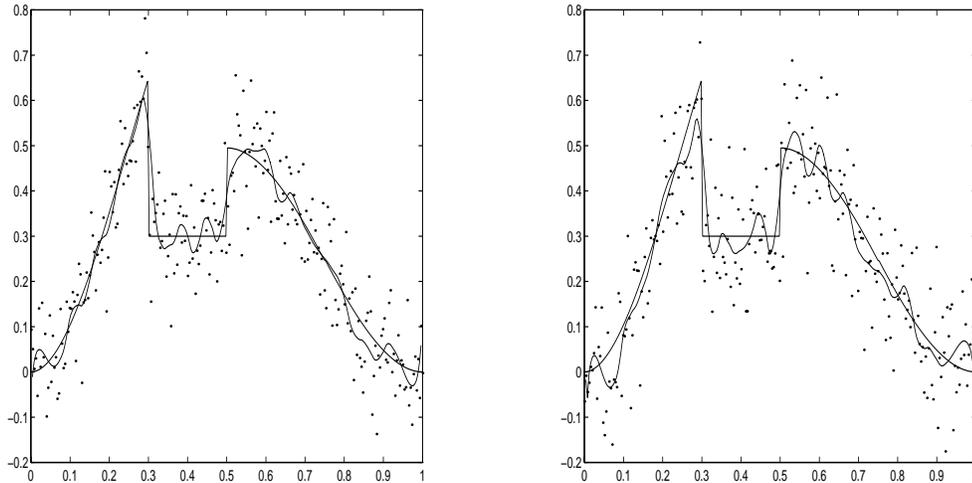


Figure 2.9 : La valeur du risque a été fixée à 0.01.

La fonction de régression pour ce cas est définie par :

$$f_1(x) = \begin{cases} 4x^2(3 - 4x) & \text{si } x \leq 0.3 \\ 0.3 & \text{si } 0.3 < x \leq 0.5 \\ \frac{4}{3}x(4x^2 - 10x + 7) - 1.5 & \text{si } 0.5 < x \leq 0.75 \\ \frac{16}{3}x(x - 1)^2 & \text{si } 0.75 < x \end{cases}$$

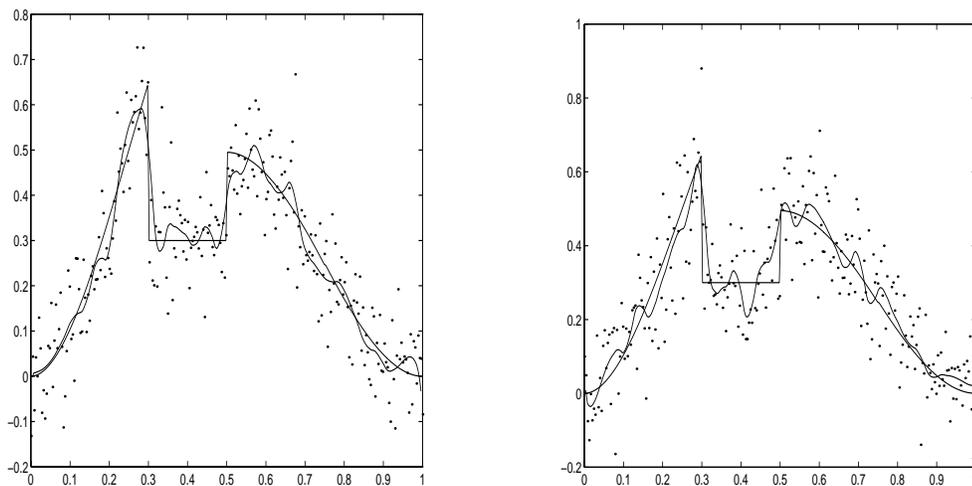


Figure 2.10 : Pour ces deux figures le risque est égal à 0.1.

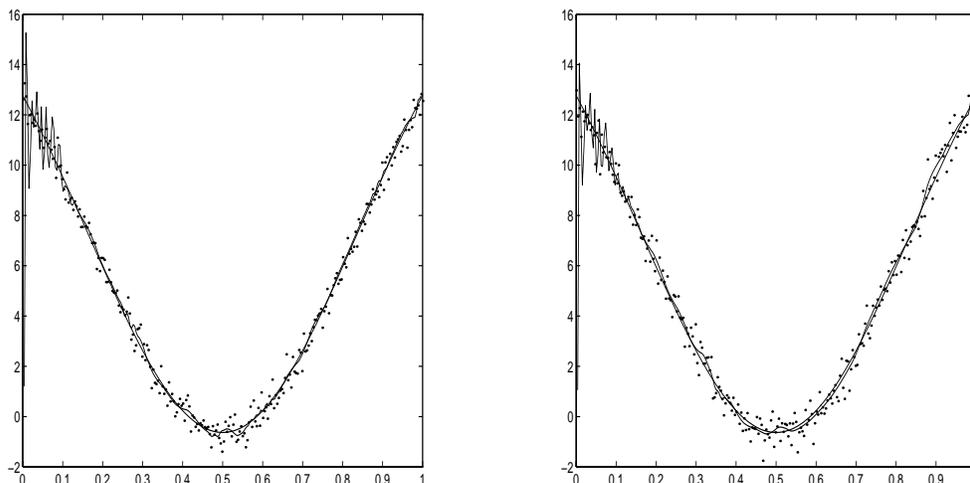


Figure 2.11 : La valeur du risque est égale à 0.01.

Nous avons pris Γ comme deuxième exemple pour nos simulations Γ la fonction suivante :

$$f_2(x) = 3 \cos(2\pi x) + 3\pi^2(x^2 - x + \frac{1}{6}).$$

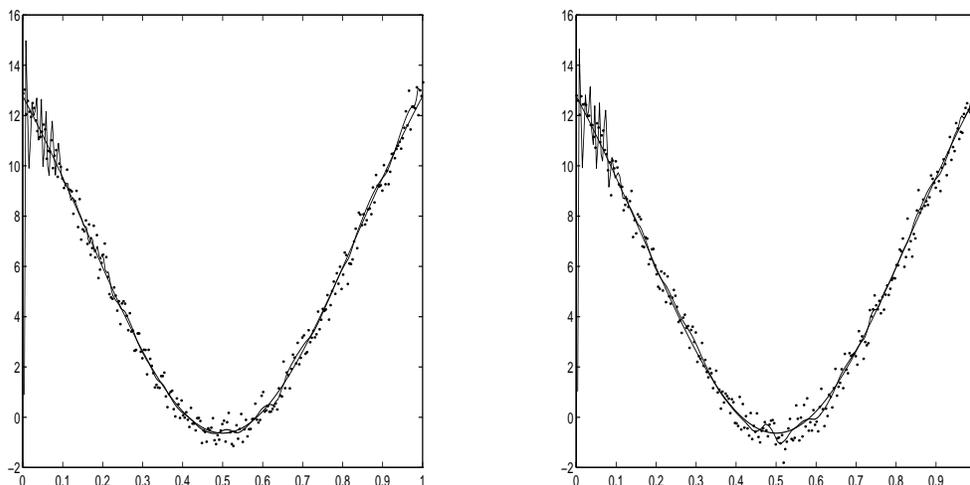


Figure 2.12 : Tout comme les figures ci-dessus Γ le nombre des observations est égal à 256. Le risque du test est égal à 0.1.

Afin de mesurer le comportement de l'estimateur proposé en fonction de la taille de l'échantillon dont nous disposons, nous avons fixé le risque à 0.1 et nous avons fait varier le nombre d'observations de 256 à 512.

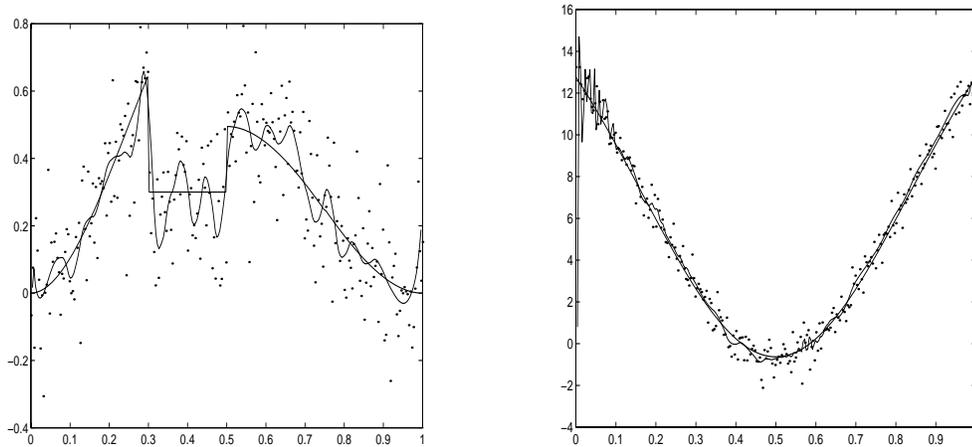


Figure 2.13 : Le nombre d'observations est égal à 256 alors que le rapport signal/bruit est égal à 1.

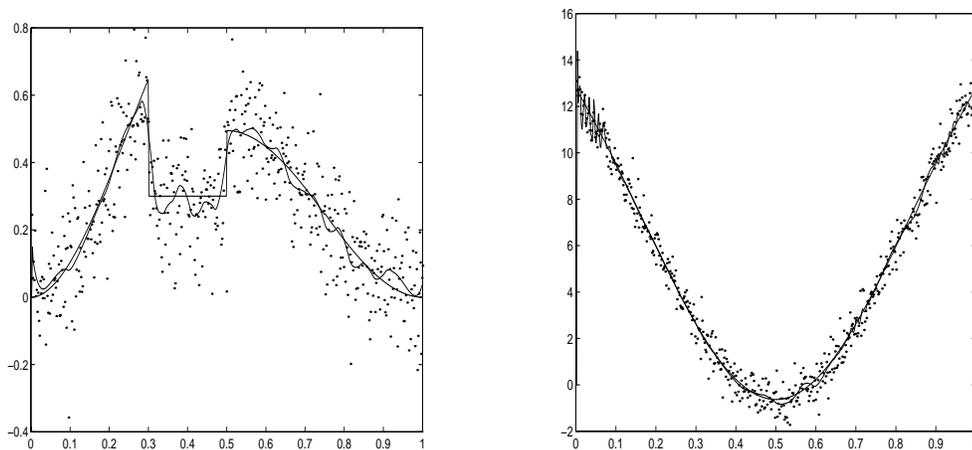


Figure 2.14 : Pour un rapport signal/bruit=1 la taille de l'échantillon est égale à 512.

Le choix de la valeur du paramètre α a une légère influence sur la qualité de l'estimateur que nous proposons. En effet, au vu des différentes figures et des tableaux 2.1 et 2.2 pour un rapport signal/bruit égal à 3 la valeur $\alpha = 0.1$ donne d'aussi bons résultats que le cas $\alpha = 0.01$ (fig 2.9-10 et 2.11-12). Pour un rapport signal/bruit égal à 2 les figures obtenues pour la valeur $\alpha = 0.01$ sont légèrement meilleurs que ceux obtenus pour $\alpha = 0.1$. Ceci s'expliquerait par le fait qu'en diminuant le rapport signal/bruit nous augmentons la contribution du bruit dans l'estimation et de ce fait le nombre de paquets ne contenant essentiellement que du bruit. L'inclusion de ces derniers conduit forcément à une dégradation de la qualité de l'estimateur. Cette dernière hypothèse est infirmée par les valeurs des RMSE calculées dans les tableaux 2.1 et 2.2. La qualité de l'estimateur proposé augmente avec la taille de l'échantillon. Ainsi même pour un rapport signal/bruit égal à 1 la méthode proposée produit une estimation de bonne qualité mais uniquement si la taille de l'échantillon est importante ($N = 512$).

2.5 Estimation par Bootstrap

Pour cette troisième et dernière procédure nous nous plaçons sous les conditions du paragraphe 2.4.

Soit W la matrice orthogonale correspondant à la décomposition **DDPO** au niveau j par exemple de la multirésolution associée à $L^2(\mathbb{R})$ ce qui correspond à une décomposition à l'ordre $n - j$. L'application de la matrice W à l'identité (2.0) conduit à la relation suivante

$$\begin{aligned} W\mathbf{y} &= W(\mathbf{F} + \boldsymbol{\varepsilon}) \\ \text{i.e : } \mathbf{v} &= \boldsymbol{\Theta} + \mathbf{z} \end{aligned}$$

où \mathbf{z} est un vecteur Gaussien $\mathcal{N}(0, \sigma^2 I_N)$ I_N étant la matrice identité d'ordre N .

L'objectif de cette procédure est de construire une estimation de \mathbf{F} par l'estimation de $\boldsymbol{\Theta}$ car W est orthogonal en annulant certaines composantes de \mathbf{v} . Un estimateur obtenu par cette technique sera désigné par estimateur de type V-S (Variable Selection).

Il s'agit ci-après d'utiliser des techniques de Bootstrap pour estimer l'erreur quadratique réelle définie par

$$L_N(\hat{\boldsymbol{\Theta}}_N, \boldsymbol{\Theta}) = |\hat{\boldsymbol{\Theta}}_N - \boldsymbol{\Theta}|_{2,N}$$

$\hat{\Theta}_N$ étant un estimateur de $\Theta\Gamma$ puis de sélectionner l'estimateur de type V-S qui minimise l'erreur quadratique estimée.

Dans un travail récent Γ R. Beran [6] montre que les techniques classiques de Bootstrap (construction d'un échantillon à partir de $\mathcal{N}(\mathbf{v}, \hat{\sigma}^2 I_N)$ où $\hat{\sigma}^2$ est un estimateur de la variance σ^2 du modèle (2.0)) Γ conduisent à la construction d'estimateurs de type V-S dont le risque asymptotique est fortement biaisé. Pour remédier à ce problème Γ il propose de construire un échantillon à partir de la distribution $\mathcal{N}(\tilde{\mathbf{v}}, \hat{\sigma}^2 I_N)$ où $\tilde{\mathbf{v}}$ est obtenu en appliquant un seuillage "souple" aux composantes de \mathbf{v} . Ceci permet de corriger le biais et produit un bon estimateur de type V-S noté $\hat{\Theta}_{N,B}$. Nous nous proposons d'adapter son approche à notre cas.

Fixons la valeur de $j\Gamma$ niveau de la multirésolution Γ à J_{opt} la valeur optimale du Théorème 2.1. Soit $J = n - J_{opt}$; nous définissons alors pour tout $m = 0, 1, 2, \dots, 2^J$

$$a_m^J = \frac{m}{2^J}.$$

Soit $K(p, J)\Gamma p \leq 2^J\Gamma$ l'ensemble contenant $[0, 1]$ et les ensembles A à p éléments tels que

$$A = \bigcup_{k=2}^p]B_k^J, B_{k+1}^J] \cup [a_0^J, a_1^J]$$

où

$$B_k^J \in \{a_m^J \mid m = 1, 2, 3 \dots 2^J - 1\}.$$

D'après la remarque 2.1 Γ le vecteur \mathbf{v} admet l'écriture suivante

$$\mathbf{v} = (\mathbf{v}^1, \mathbf{v}^2, \dots, \mathbf{v}^{2^J});$$

elle correspond à la décomposition en paquets d'ondelettes du vecteur \mathbf{v} à l'ordre J . Avec cette écriture Γ nous pouvons identifier chaque paquet \mathbf{v}^m à l'intervalle $]a_m^J, a_{m+1}^J]$. L'estimateur de type V-S que nous proposons est alors défini par

$$\hat{\Theta}_N(A) = (\eta_1(A)\mathbf{v}^1, \eta_2(A)\mathbf{v}^2, \dots, \eta_{2^J}(A)\mathbf{v}^{2^J}), \quad (2.14)$$

où chaque η_i est défini par :

$$\eta_i(A) = \begin{cases} 1 & \text{si } a_i^J \in A \\ 0 & \text{sinon} \end{cases}$$

Pour tout A dans $K(p, J)\Gamma$ nous définissons

- $\nu_N(A) = N^{-1} \sum_{m: a_m^J \in A} |\Theta^m|^2$,
- $\hat{\tau}_N(A) = N^{-1} \sum_{m: a_m^J \in A} |\mathbf{v}^m|^2$,
- $\mu_N(A) = N^{-1} \sum_{m: a_m^J \in A} 1$,

et nous désignerons par $\mu(A)$ la mesure de Lebesgue.

Nous construisons alors Γ pour tout A dans $K(p, J)\Gamma$ le vecteur

$$\tilde{\mathbf{v}}(A) = (\tilde{\mathbf{v}}^1, \tilde{\mathbf{v}}^2, \dots, \tilde{\mathbf{v}}^{2^J})$$

tel que

$$\tilde{\mathbf{v}}^m = \begin{cases} \mathbf{v}^m & \text{si } a_m^J \in A \\ \hat{s}_N^{\frac{1}{2}}(A^c) \mathbf{v}^m & \text{sinon} \end{cases}$$

où

$$\hat{s}_N(A^c) = [1 - \frac{\hat{\sigma}_N^2 \mu_N(A^c)}{\hat{\tau}_N(A^c)}]_+ \text{ et } [x]_+ = \sup(x, 0).$$

Un tel seuillage conduit à une diminution Γ voire suppression Γ du bruit dans certains paquets de \mathbf{v} .

Soit \mathbf{v}^* un vecteur aléatoire dont la distribution conditionnelle sachant \mathbf{v} est une loi normale $\mathcal{N}(\tilde{\mathbf{v}}(A), \hat{\sigma}_N^2 I_N)\Gamma \hat{\Theta}_{N,B}(A)$ l'estimateur de type V-S défini à partir de \mathbf{v}^* et soit

$$\hat{L}_N(\hat{\Theta}_{N,B}(A), \tilde{\mathbf{v}}(A)) = |\hat{\Theta}_{N,B}(A) - \tilde{\mathbf{v}}(A)|_{2,N},$$

l'erreur quadratique estimée par bootstrap de $\hat{\Theta}_{N,B}(A)$. Nous avons alors :

$$E_{\text{cond}}[\hat{L}_N(\hat{\Theta}_{N,B}(A), \tilde{\mathbf{v}}(A))] \text{ converge en probabilité vers } \rho(A),$$

$\rho(A)$ étant la valeur asymptotique du risque réel de $\hat{\Theta}_N(A)\Gamma$ i.e $\rho(A) = \lim_N E[L_N(\hat{\Theta}_N(A), \Theta)]$.

Ce problème étant résolu Γ nous allons pouvoir définir l'estimateur Bootstrap. A cet effet Γ posons :

$$\hat{R}(A, \hat{\sigma}_N^2) = E_{\text{cond}}[\hat{L}_N(\hat{\Theta}_{N,B}(A), \tilde{\mathbf{v}}(A))].$$

Nous avons (voir Beran) :

$$\hat{R}(A, \hat{\sigma}_N^2) = \hat{\sigma}_N^2 \mu_N(A) + \hat{s}_N(A^c) \hat{\tau}(A^c);$$

ceci permet en minimisant l'estimateur du risque défini ci-dessus sur l'ensemble des éléments de $K(p, J)\Gamma$ de définir l'estimateur Bootstrap $\mathbf{v}' = \hat{\Theta}_{N,B}(A^*)$ où :

$$A^* = \operatorname{argmin}_{A \in K(p,J)} \hat{R}(A, \hat{\sigma}_N^2).$$

Pour un estimateur consistant de la variance du bruit Γ comme ceux introduits au paragraphe 1.5 du chapitre 1 et en désignant par *plim* la limite en probabilité Γ nous avons alors le résultat suivant :

Théorème 2. 3 (Beran) *Soit ν une mesure bornée sur $K(p, J)$ tel que:*

$$\lim_{N \rightarrow \infty} \left(\sup_{K(p,J)} |\nu_N - \nu| \right) = 0,$$

alors :

$$\operatorname{plim}_{N \rightarrow \infty} \left(\sup_{K(p,J)} |\hat{R}(\cdot, \hat{\sigma}_N^2) - \rho(\cdot)| \right) = 0,$$

ρ étant défini pour tout $A \in K(p, J)$ par:

$$\rho(A) = \sigma^2 \mu(A) + \nu(A^c).$$

De même :

$$\operatorname{plim}_{N \rightarrow \infty} L_N(\hat{\Theta}(A), \Theta(A)) = \min_{A \in K(p,J)} \rho(A).$$

2.5.1 Généralisation

Il est clair que la méthode décrite ci-dessus peut être facilement étendue au cas plus général où nous disposons des décompositions en paquets d'ondelettes effectuées à l'ordre $j = 0, 1, 2, \dots, J$, obtenues par l'application des matrices $W_n^{(j)}$ correspondantes (voir fig 2.1).

La dépendance en A des paramètres et mesures définies dans le paragraphe ci-dessus est modifiée en une dépendance en φ ; ce dernier étant défini par :

$$\varphi = \bigcup_{k=1}^{p-1} [B_k, B_{k+1}[\bigcup [a_1^j, a_2^j[$$

avec

$$B_k \in \{a_m^j = \frac{m}{2^j}, \forall j = 1, 2, \dots, J, m = 2, 3 \dots 2^j\}.$$

Les éléments φ sont restreints à vérifier une deuxième condition :

$$\forall l \neq l' [B_l, B_{l+1}[\cap [B_{l'}, B_{l'+1}[= \emptyset.$$

Le sous-ensemble des partitions φ de $[0, 1]$ vérifiant les deux conditions ci-dessus est noté par $K(p, J)$; le résultat établi dans le paragraphe précédent reste alors valable.

Commentaires:

Un problème qui reste sans solution dans le développement de cette procédure est le choix de p qui peut s'avérer crucial. Il est difficile d'imaginer une technique permettant d'estimer ce paramètre qui est directement lié aux nombres de coefficients nuls dans la décomposition en paquets d'ondelettes à chaque niveau de résolution de l'AME et ce malgré le fait que toutes les méthodes d'estimation utilisées en régression par ondelettes s'appuient de manière heuristique sur le fait qu'un nombre très limité de ces coefficients contribuent réellement à l'estimation de la fonction de régression. Ce qui est en total accord avec les simulations pratiques.

D'autre part il subsiste une difficulté de taille dans le cas généralisé quand à la conception d'une procédure algorithmique très simple et ce à cause du cardinal de $K(p, J)$.

Deuxième partie

Analyse de survie

INTRODUCTION

Dans différents domaines de recherche on est amené à traiter des données de survie éventuellement incomplètes ou censurées. Ceci est le cas par exemple en recherche médicale où des patients sujets d'une expérience y prennent part à des moments distincts et peuvent à cause d'une censure quitter l'expérience avant terme. La censure dans ce cas peut être la mort du patient à cause de facteurs aléatoires extérieurs à l'étude. D'autres domaines de recherche sont concernés par de telles données ; on citera par exemple les sciences économiques, physiques et sociales, l'ingénierie ou encore la fiabilité des systèmes.

De ce fait il est primordial dans l'étude de toute fonction caractérisant des données de ce type de prendre en considération l'information fournie par la censure quand celle-ci se produit.

Parmi toutes les fonctions qu'on peut définir on s'est intéressé en particulier à la fonction de taux de hasard, la fonction de distribution et la fonction de densité. C'est principalement cette dernière qui a suscité le plus grand intérêt.

Dans le cas censuré les méthodes d'estimation utilisées sont identiques à celles développées pour des données non censurées sauf que les problèmes qui en découlent sont plus difficiles du point de vue mathématique. De plus du fait même de la définition de la fonction de hasard l'estimation de la fonction de hasard est analogue à celle de la fonction de densité. Ainsi toutes (ou presque) les techniques utilisées pour estimer cette dernière ont été adaptées à l'estimation de la fonction de hasard.

1. Notations et définitions :

Soient T_1, T_2, \dots, T_N N variables représentant les durées de vie de N sujets et C_1, C_2, \dots, C_N les instants de censures associés (variables

aléatoires ou constantes). Les T_i sont des variables aléatoires positives indépendantes et identiquement distribuées de fonction de distribution commune F_T . On suppose que F_T est absolument continue de densité f_T . La fonction de hasard de T est alors définie en tout instant t tel que $F_T(t) < 1$ par $h(t) = \frac{f_T(t)}{1-F_T(t)}$.

On parle de censure à droite quand les variables observées sont définies par $\{(X_i, \delta_i), i = 1, 2, \dots, N\}$ où :

$$X_i = \min(T_i, C_i) \quad \delta_i = \begin{cases} 1 & \text{si } T_i \leq C_i \\ 0 & \text{sinon.} \end{cases}$$

Les variables δ_i indicateurs de NON-censure permettent de ce fait de distinguer les données censurées parmi les X_i .

La nature des C_i définit celle de la censure ; ainsi on distingue trois types de censure :

- i) Les C_i sont des constantes : on parle de données censurées dans le temps. Si de plus ces constantes sont égales la censure est de type I.
- ii) La censure est dite de type II si pour tout i $C_i = T_{(k)}$ où $T_{(k)}$ est la $k^{\text{ème}}$ statistique d'ordre associée à T_i .
- iii) Dans le cas où les variables C_i sont indépendantes des T_i et issues d'une même distribution F_C (X_i, δ_i) est un échantillon de censure aléatoire à droite.

C'est sous les conditions de ce troisième type de censure que la plupart des estimateurs de la fonction de hasard ont été étudiés.

Avant de conclure ce paragraphe on définit la fonction de survie par $S(t) = 1 - F_T(t)$ et on retiendra que son estimateur de maximum de vraisemblance (ou estimateur de Kaplan Meier) noté \hat{S}_N a fait l'objet d'une large étude pour des données censurées. Il est à noter par ailleurs qu'une bonne partie des estimateurs de la fonction de densité et parfois de hasard pour des données censurées sont basés sur ce dernier.

2. Estimation de la fonction de hasard

On donne ci-après quelques estimateurs de la fonction de hasard dans le cas censuré. Le cas non censuré s'en déduit en prenant δ_i égal à 1 pour tout i . Les estimateurs proposés ont été étudiés pour des données de survie aléatoirement censurées à droite.

Estimateurs du type noyau : Du fait que l'on disposait d'estimateurs consistants pour les fonctions de densité et de survie Földes et al [33] ont proposé un estimateur de la fonction de hasard de la forme

$$\hat{h}_0(x) = \frac{\hat{f}_T(x)}{\hat{S}_N(x) + \frac{1}{N}}, \quad x \geq 0,$$

le terme $\frac{1}{N}$ dans le dénominateur évitant la division par zéro. La fonction \hat{f}_T est un estimateur du type noyau ou histogramme de la fonction de densité f_T .

L'estimateur du type noyau initialement introduit par Watson et Leadbetter [74] dans le cas non censuré fut adapté par Tanner et Wong [69] au cas censuré. Ainsi ils ont défini

$$\hat{h}_1(x) = \frac{1}{N+1} \sum_{i=1}^N (1 - \tilde{F}_N(X_i))^{-1} \delta_i K_l(x - X_i),$$

avec

$$\tilde{F}_N(x) = \frac{1}{N+1} \sum_{i=1}^N \mathbb{1}_{[X_i \leq x]}$$

et K est un noyau symétrique d'intégrale 1 telle que $K_l(x) = l^{-1}K(\frac{x}{l})$; la dépendance de l en N est omise. L'estimateur ainsi défini est fortement consistant sous certaines conditions sur le noyau K . De plus par la technique de projection de Hajek ils ont montré que $\hat{h}_1(x)$ est de loi asymptotiquement normale.

Sous le modèle de Koziol où il est supposé que $1 - F_C = (1 - F_T)^\beta$, $\beta > 0$ McNichols et Padgett [51] ont proposé et étudié la consistance de l'estimateur suivant

$$\hat{h}_2(x) = l^{-1} \int \frac{K((x-t)/l)}{\hat{S}_N(t)} d\hat{F}_N(t)$$

où

$$\hat{F}_N(t) = 1 - \hat{S}_N(t).$$

Un autre estimateur du type noyau fût étudié par Tanner à travers la définition des statistiques d'ordre $(X_{(1)}, X_{(2)}, \dots, X_{(N)})$ associés à (X_1, X_2, \dots, X_N) et de leurs indicateurs de censure $(\delta_{(1)}, \delta_{(2)}, \dots, \delta_{(N)})$.

Soit d_k la distance entre x et la $k^{\text{ème}}$ plus proche observation non censurée (i.e : $\delta_i = 1$) parmi les X_i . Il définit

$$\hat{h}_3(x) = (2d_k)^{-1} \sum_{i=1}^N \frac{\delta_{(i)}}{N+1-i} K\left(\frac{x - X_{(i)}}{2d_k}\right).$$

Un tel estimateur fait intervenir les données observées dans le choix de la largeur de la fenêtre du noyau et de ce fait le degré de lissage au voisinage de x . Tanner [68] montre que cet estimateur est fortement consistant.

Enfin Blum et Susarla [8] définissent l'estimateur suivant

$$\hat{h}_4(x) = \frac{f_T^*(x)}{(1+N)(1 - \hat{F}_N(x)) - 1},$$

où

$$f_T^*(x) = (Nl)^{-1} \sum_{i=1}^N K\left(\frac{x - X_i}{l}\right) \mathbb{1}_{[\delta_i=1]}.$$

Méthodes de vraisemblance et approche Bayésienne :

On citera la méthode de vraisemblance utilisée par Marshall et Proschan [50] pour le cas de données non censurées puis reprise par Padgett et Wei [57] pour une famille particulière de données censurées où la fonction h est croissante. Mykytyn et Santner [54] ont étendu le problème au cas où le taux de hasard est décroissant et ont montré des résultats de consistance pour leur estimateur. Par ailleurs il y a eu quelques travaux qui ont porté sur l'estimation de la fonction de hasard par des techniques bayésiennes.

Les problèmes de survie étant un cas particulier de processus ponctuel sous le modèle multiplicatif de Aalen-Ramlau-Hansen [60]

a utilisé l'approche par processus ponctuel pour étudier l'estimateur du type noyau de la fonction de hasard et montrer sa consistance en norme $L^2(\mathbb{R})$. De leur part Antoniadis et Grégoire [2] ont utilisé cette approche pour construire un estimateur de la fonction de hasard par la méthode de vraisemblance pénalisée en utilisant la racine carré de la fonction de hasard et ont montré qu'un tel estimateur est aussi consistant.

Régression par fonctions de splines :

Récemment Kooperberg Stone et Truong [45] ont proposé un estimateur positif de la fonction log-hasard par régression sur un système de fonctions splines. A cette fin pour tout entier d strictement positif soit G un espace linéaire de dimension d dont les éléments sont des fonctions définies sur $[0, \infty[$; de plus ces fonctions sont bornées sur $[0, \infty[$. Le fait d'estimer la fonction log-hasard est justifié par la remarque suivante : un modèle basé sur la fonction log-hasard admet une fonction de vraisemblance concave ce qui assure l'existence du maximum de vraisemblance.

Soit $\mathcal{B} = \{B_1, B_2, \dots, B_d\}$ une base de G la régression de $\beta(t) = \log(\alpha(t))$ sur cette base s'écrit

$$\beta(t) = \sum_{j=1}^d \beta_j B_j(t).$$

L'estimation du vecteur $\beta = (\beta_1, \beta_2, \dots, \beta_d)$ est alors obtenue par maximum de vraisemblance.

Désignons par $l(\beta)$ la fonction log-vraisemblance associée aux données (X_i, δ_i) et au modèle ci-dessus On a

$$l(\beta) = \sum_{i=1}^N \left(\delta_i \alpha(X_i, \beta) - \int_0^{X_i} \exp(\alpha(u, \beta)) du \right).$$

L'estimateur de maximum de vraisemblance étant défini par $\hat{\beta} = \arg\max l(\beta)$ ce qui induit $\hat{\beta}(t) = \beta(t, \hat{\beta})$ comme estimateur de maximum de vraisemblance pour $\beta(t)$.

Finalement pour résoudre le problème du choix de G les auteurs ont proposé une procédure similaire à celle définie par C.A.R.T [10] ; elle consiste en la construction d'une suite emboîtée dans

un sens que les auteurs ont défini d'espaces G appartenant à une famille \mathcal{G} .

On choisit alors le modèle correspondant à l'espace G^* l'élément de la suite Γ de manière à minimiser le critère d'information d'Akaike. Il est à noter que le choix des auteurs s'est porté sur un système de base de fonctions splines cubiques.

Dans le troisième chapitre nous abordons le problème de l'estimation de la fonction de hasard lorsque celle-ci est à support compact et régulière par morceaux. Cette dernière sera notée λ et sera étudiée pour des données de survie avec ou sans censure.

Pour ce faire partant d'une analyse de multirésolution et de la base d'ondelettes associée nous construisons un système de base orthonormée faisant intervenir un paramètre de dilatation p et nous définissons l'estimateur linéaire de projection orthogonale qui lui est naturellement associé. Pour un tel estimateur comme il est de coutume nous nous intéressons au comportement asymptotique de l'erreur quadratique moyenne intégrée (MISE).

Nous montrons qu'un estimateur non linéaire $\hat{\lambda}$ obtenu par simple modification d'une partie des coefficients de la décomposition de λ sur le système de base choisi admet un comportement en MISE identique à celui obtenu pour l'estimateur par noyau pour des fonctions de régression et de densité à savoir

$$MISE(\hat{\lambda}) \sim C_1 N^{-1} p + C_2 p^{-2r},$$

où p^{-1} par analogie désigne la fenêtre du noyau.

Chapitre 3

Ondelettes et problèmes de survie

Nous supposons que nous disposons de N variables aléatoires positives $T_i, i = 1, \dots, N$ indépendantes et identiquement distribuées de fonction de répartition F_T . Cette dernière est supposée absolument continue de fonction de densité f_T . Éventuellement les T_i seront aléatoirement censurées à droite : ceci revient à se donner N variables aléatoires indépendantes C_1, C_2, \dots, C_N indépendantes des T_i et de fonction de répartition commune F_C .

Les observations consistent en la réalisation de N couples de variables aléatoires $(X_i, \delta_i), i = 1, 2, \dots, N$ définis par

$$X_i = \min(T_i, C_i) \quad \text{et} \quad \delta_i = \mathbb{1}_{[T_i \leq C_i]}.$$

En tout point x tel que $F_T(x) < 1$ la fonction ou taux de hasard est définie par $h(x) = \frac{f_T(x)}{1-F_T(x)}$.

Les variables X_i sont indépendantes et identiquement distribuées de fonction de répartition F telle que $1-F = (1-F_T)(1-F_C)$. Nous définissons alors

$$\hat{F}_N(x) = \frac{1}{N+1} \sum_{i=1}^N \mathbb{1}_{[X_i \leq x]},$$

comme estimateur empirique de F .

Avant de poursuivre nous notons η la fonction définie par

$$\eta(x) = \frac{(1 - F_C(x))f_T(x)}{Q(x)}$$

où $Q(x)$ est une fonction croissante vérifiant $0 < Q(x) \leq 1$. L'intérêt de la fonction η est qu'elle englobe selon sa formulation les fonctions de densité et de hasard dans le cas de données censurées à droite mais aussi celui des données non censurées. Ceci dépend exclusivement du choix de la fonction $Q(x)$; ainsi nous pouvons distinguer trois cas :

1. Si $Q(x) = 1 - F(x)$ et F_C a sa masse à l'infini (i.e : $F_C(x) = 0 \forall x \in \mathbb{R}$) ce qui entraîne que $F(x) = F_T(x)$ alors $\eta(x)$ n'est autre que la fonction de hasard d'un modèle de survie sans censure.
2. Si $Q(x) = 1 - F(x)$ ou $Q(x) = 1 - F_C(x)$ nous avons respectivement la fonction de hasard et la fonction de densité dans le cas d'une censure aléatoire à droite.
3. Si F_C a sa masse à l'infini et $Q(x) = 1$ alors η est une fonction de densité pour des données non censurées.

Dans les paragraphes qui suivent nous nous intéressons à l'estimation par projection sur un système de base orthonormée d'ondelettes de la fonction η lorsque celle-ci correspond à la fonction de hasard soit $Q(x) = 1 - F(x)$. Pour des raisons d'intégrabilité de la fonction η nous définissons pour ε_0 strictement positif donné T_{ε_0} par

$$T_{\varepsilon_0} = \sup\{x \mid Q(x) > \varepsilon_0\}$$

et nous nous intéresserons plutôt à l'estimation de $\lambda(x) = \eta(x) \mathbb{1}_{[0, T_{\varepsilon_0}]}(x)$ que nous supposons être de carré intégrable. Pour des raisons de commodité ε_0 étant fixé nous noterons par la suite T la valeur de T_{ε_0} .

3.1 Écriture de l'estimateur et erreur quadratique

Soit une fonction d'échelle φ Höldérienne à support compact et la multirésolution associée de $L^2(\mathbb{R})\Gamma$ et soit ψ l'ondelette mère correspondante ; elle même est à support compact. De plus nous supposons que la fonction ondelette admet des moments nuls jusqu'à un certain ordre L donné c'est à dire nous supposons que

$$\int x^k \psi(x) dx = 0 \quad \forall k = 0, 1, \dots, L \text{ et } \int x^{L+1} \psi(x) dx \neq 0.$$

Il est à noter que les ondelettes de Daubechies vérifient les propriétés requises ci-dessus (voir paragraphe 2.1.1).

En définissant

$$\varphi_l = \varphi(\cdot - l) \quad \text{et} \quad \psi_{kl} = \psi(2^k \cdot - l),$$

le système $\{\varphi_l, \psi_{kl}, l \in \mathbb{Z} \text{ et } k = 0, 1, \dots\}$ est une base orthonormée de $L^2(\mathbb{R})$. Pour tout réel strictement positif p nous définissons les fonctions

$$\tilde{\varphi}_l = p^{\frac{1}{2}} \varphi(p \cdot - l) \quad \text{et} \quad \tilde{\psi}_{kl} = p_k^{\frac{1}{2}} \psi(p_k \cdot - l),$$

où $p_k = p2^k$.

Notons $V_k^p \Gamma$ l'espace engendré par les $\tilde{\varphi}_{kl} = p_k^{\frac{1}{2}} \varphi(p_k \cdot - l) \Gamma l \in \mathbb{Z} \Gamma$ la suite $(V_k^p, k \in \mathbb{Z})$ est une multirésolution de $L^2(\mathbb{R})$ (différente de celle construite par φ). En définissant par W_k^p le supplémentaire orthogonal de V_k^p dans V_{k+1}^p (i.e : $V_{k+1}^p = V_k^p \oplus W_k^p$) ΓW_k^p est alors engendré par $\{\tilde{\psi}_{kl}, l \in \mathbb{Z}\}$.

Il est à noter que la multirésolution définie par la suite $(V_k^p, k \in \mathbb{Z})$ est associée à la fonction d'échelle $\tilde{\varphi} = p^{\frac{1}{2}} \varphi(p \cdot)$ et est définie sur le réseau $\Gamma = \mathbb{Z}/p$. En effet $\Gamma \tilde{\varphi}_l = \tilde{\varphi}(\cdot - \frac{l}{p})$ et donc contrairement à la multirésolution définie par $\varphi \Gamma V_0^p$ est engendré par les translations de $\tilde{\varphi}$ sur \mathbb{Z}/p et non pas \mathbb{Z} .

Nous construisons de la sorte une nouvelle base orthonormée pour $L^2(\mathbb{R})\Gamma$ engendrée par

$$\{\tilde{\varphi}_l, \tilde{\psi}_{kl}, k = 0, 1, \dots \text{ et } l \in \mathbb{Z}\}.$$

Ainsi toute fonction $\lambda \Gamma$ de carré intégrable Γ admet une décomposition orthogonale sur cette base

$$\lambda = \sum_l c_l \tilde{\varphi}_l + \sum_{k=0}^{\infty} \sum_l d_{kl} \tilde{\psi}_{kl},$$

avec

$$c_l = \int \tilde{\varphi}_l(y) \lambda(y) dy \quad \text{et} \quad d_{kl} = \int \tilde{\psi}_{kl}(y) \lambda(y) dy.$$

Les coefficients c_l et d_{kl} représentent respectivement les coefficients de la décomposition de λ en “approximation” et en “détails”.

L’utilisation d’un tel système de base orthonormé pour $L^2(\mathbf{R})$ a été suggérée par Patil et Hall [41] pour l’estimation de la fonction de densité dans le cadre de données non censurées.

Un estimateur linéaire de λ déterminé respectivement à partir des X_i dans le cas non censuré ou des (X_i, δ_i) dans le cas censuré est basé sur les estimateurs empiriques \hat{c}_l et \hat{d}_{kl} de c_l et d_{kl} et s’écrit comme suit :

$$\tilde{\lambda}_q = \sum_l \hat{c}_l \tilde{\varphi}_l + \sum_{k=0}^{q-1} \sum_l \hat{d}_{kl} \tilde{\psi}_{kl}.$$

Il correspond à l’estimation de la projection de λ sur V_q^p .

Dans un premier temps nous étudierons le comportement asymptotique des estimateurs des coefficients de la décomposition en ondelettes de λ dans le premier et deuxième cas évoqués ci-dessus. Au préalable nous donnons l’écriture des \hat{c}_l et \hat{d}_{kl} pour chacun des deux cas ; à cet effet nous

$$\omega = \mathbb{1}_{[0,T]}.$$

Ainsi :

Premier cas

$$\hat{c}_l = \frac{1}{N+1} \sum_{i=1}^N \tilde{\varphi}_l(X_i) (1 - \hat{F}_N(X_i))^{-1} \omega(X_i) \quad (3.1)$$

et

$$\hat{d}_{kl} = \frac{1}{N+1} \sum_{i=1}^N \tilde{\psi}_{kl}(X_i) (1 - \hat{F}_N(X_i))^{-1} \omega(X_i). \quad (3.2)$$

Deuxième cas

$$\hat{c}_l = \frac{1}{N+1} \sum_{i=1}^N \tilde{\varphi}_l(X_i) (1 - \hat{F}_N(X_i))^{-1} \omega(X_i) \delta_i \quad (3.1)'$$

et

$$\hat{d}_{kl} = \frac{1}{N+1} \sum_{i=1}^N \tilde{\psi}_{kl}(X_i) (1 - \hat{F}_N(X_i))^{-1} \omega(X_i) \delta_i. \quad (3.2)'$$

Remarque 3. 1. :

1. Notons qu'en posant $\hat{\lambda}_p = \sum_l \hat{c}_l \tilde{\varphi}_l$, $\hat{\lambda}_p$ admet une écriture analogue à celle obtenue pour l'estimateur linéaire par la méthode du noyau généralisé [69]. En effet, posons $K(x, y) = \sum_l \varphi_l(x) \varphi_l(y)$,

$$\begin{aligned} \hat{\lambda}_p(x) &= \frac{p}{N+1} \sum_{i=1}^N \sum_l \varphi(pX_i - l) \varphi(px - l) \omega(X_i) (1 - \hat{F}_N(X_i))^{-1} \\ &= \frac{p}{N+1} \sum_{i=1}^N K(px, pX_i) \omega(X_i) (1 - \hat{F}_N(X_i))^{-1} \end{aligned} \quad (3.3)$$

et donc par analogie, le paramètre p^{-1} correspond à la largeur de fenêtre du noyau généralement noté par h . Cette analogie ne s'arrête pas là, puisque p^{-1} apparaîtra dans l'évaluation du MISE de $\hat{\lambda}_p$ et son interprétation sera équivalente au rôle de h , pour un estimateur par noyau.

2. L'estimateur $\hat{\lambda}_p$ n'est autre que l'estimateur $\tilde{\lambda}_q$ défini ci-dessus avec $q = 0$.
3. La structure de multirésolution associée à $\tilde{\varphi}$, permet l'utilisation de l'algorithme pyramidal, ce qui facilite considérablement l'estimation des coefficients \hat{c}_l et \hat{d}_{kl} .

Dans le lemme 3.2Γ énoncé ci-aprèsΓ nous étudions le comportement asymptotique des coefficients \hat{c}_l pour chacun des deux cas qui nous intéressent. A cette finΓ nous avons besoin du résultat suivant : soient $X_{(i)}$ les statistiques d'ordre associées aux X_i et soient $f_{(i)}$ leur fonctions de densité respectives ; posons pour tout réel a $\Gamma I_{a,N} = \sum_{i=1}^N (N - i + 1)^{-a} f_{(i)}$. Nous avons :

Lemme 3. 1 (Watson et Leadbetter) *Quand N tend vers l'infini, on a*

$$NI_{2,N} \rightarrow \frac{f}{(1 - F)^2}.$$

Dans tout ce qui suit nous posons $\alpha = \lambda(1 - F)^{-1}$.

Lemme 3. 2. Pour \hat{c}_l défini par (3.1) ou (3.1)', on a

$$E(\hat{c}_l) = c_l - \int \tilde{\varphi}_l(y) \lambda(y) F^N(y) dy \quad (3.4)$$

et

$$\begin{aligned} \text{var}(\hat{c}_l) = & \int I_{2,N}(y) \tilde{\varphi}_l^2(y) dy + 2 \int \int_{0 \leq y \leq z} \left\{ \frac{1 - F^N(y)}{1 - F(y)} F^N(z) \right. \\ & \left. - \frac{F^N(z) - F^N(y)}{F(z) - F(y)} \right\} f(y) h(z) \tilde{\varphi}_l(y) \tilde{\varphi}_l(z) dy dz. \end{aligned} \quad (3.5)$$

On en déduit que, lorsque N tend vers l'infini, on a

$$E(\hat{c}_l) = c_l + \mathcal{O}(N^{-1} p^{-\frac{1}{2}})$$

et

$$N \text{var}(\hat{c}_l) \rightarrow \int \alpha \left(\frac{y+l}{p} \right) \varphi^2(y) dy.$$

Les techniques utilisées pour la démonstration du lemme 3.2 pour des données censurées permettront par la suite de se limiter au seul cas non censuré (premier cas).

Preuve du lemme 3.2

Premier Cas

Chacun des coefficients c_l respectivement d_{kl} admet une écriture en fonction des $X_{(i)}$

$$\hat{c}_l = \sum_{i=1}^N \tilde{\varphi}_l(X_{(i)}) (N + 1 - i)^{-1} \omega(X_{(i)}) \quad (3.6)$$

et

$$\hat{d}_{kl} = \sum_{i=1}^N \tilde{\psi}_{kl}(X_{(i)}) (N + 1 - i)^{-1} \omega(X_{(i)}). \quad (3.7)$$

Comme

$$f_{(i)} = NC_{N-1}^{i-1}[F]^{i-1}[1-F]^{N-i}f,$$

alors

$$\begin{aligned} I_{1,N} &= \sum_{i=1}^N (N+1-i)^{-1} NC_{N-1}^{i-1} F^{i-1} (1-F)^{N-i} f \\ &= \sum_{k=0}^{N-1} C_N^k F^k (1-F)^{N-k} \frac{f}{1-F} \\ &= \frac{f}{1-F} (1-F^N). \end{aligned}$$

Or en l'absence de censure $F_T = F$ et donc $\omega I_{1,N} = \lambda(1-F^N)$. Ainsi

$$\begin{aligned} E(\hat{c}_l) &= \frac{1}{N+1} \sum_{i=1}^N E[\tilde{\varphi}_l(X_{(i)}) \omega(X_{(i)}) (1-\hat{F}_N(X_{(i)}))^{-1}] \\ &= \sum_{i=1}^N (N+1-i)^{-1} \int \tilde{\varphi}_l(y) \omega(y) f_{(i)}(y) dy \\ &= \int \tilde{\varphi}_l(y) \omega(y) I_{1,N}(y) dy \\ &= c_l - \int \tilde{\varphi}_l(y) \lambda(y) F^N(y) dy. \end{aligned}$$

Du fait que sur l'intervalle $[0, T]$ on ait $F < 1$ F^N décroît vers 0 quand N tend vers l'infini et par le théorème de la convergence dominée nous obtenons que

$$E(\hat{c}_l) \rightarrow c_l \text{ quand } N \rightarrow \infty.$$

De plus comme $(1-F)^{-1} \leq \varepsilon_0^{-1}$ et que $\int \tilde{\varphi}_l(y) dy = p^{-\frac{1}{2}}$

$$\begin{aligned} \int \tilde{\varphi}_l(y) \lambda(y) F^N(y) dy &= \int \tilde{\varphi}_l(y) (1-F(y))^{-1} f(y) F^N(y) dy \\ &= \mathcal{O}(N^{-1} p^{-\frac{1}{2}}). \end{aligned}$$

La fonction d'échelle φ est à support compact et telle que $\int \varphi(x) dx = 1$ le calcul de la variance est donné par un résultat de Watson et Leadbetter [74] dont nous rappelons l'énoncé :

Résultat 1. Soit $b(x)$ une fonction bornée telle que $\int b(x) dx < \infty$ et $b(x) = o(x^{-1})$ quand $|x| \rightarrow \infty$. Pour toute fonction

$$r(x) = \sum_{j=1}^N (N+1-j)^{-1} l^{-1} b\left(\frac{x-X_{(j)}}{l}\right),$$

on a :

$$\begin{aligned} \text{var}(r(x)) &= \int I_{2,N}(y) l^{-2} b^2\left(\frac{x-y}{l}\right) dy + 2 \int \int_{0 \leq y \leq z} \left\{ \frac{1-F^N(y)}{1-F(y)} F^N(z) \right. \\ &\quad \left. - \frac{F^N(z)-F^N(y)}{F(z)-F(y)} \right\} f(y) h(z) l^{-2} b\left(\frac{x-y}{l}\right) b\left(\frac{x-z}{l}\right) dy dz. \end{aligned}$$

D'autre part par le lemme 3.1 $\Gamma N I_{2,N}(y)$ tend vers $(1 - F(y))^{-1}$. Il s'ensuit par le théorème de la convergence dominée que le terme dominant de $N \text{var}(\hat{c}_l)$ en l'occurrence $N \int I_{2,N}(y) \tilde{\varphi}_l^2(y) dy$ converge vers $\int \lambda(y) (1 - F(y))^{-1} \tilde{\varphi}_l^2(y) dy$.

Pour montrer que le terme dominant dans (3.5) est le bon il suffit de remarquer que chacun des termes

$$I_1 = \int \int_{0 \leq y \leq z} \left\{ \frac{1 - F^N(y)}{1 - F(y)} F^N(z) \right\} f(y) h(z) \tilde{\varphi}_l(y) \tilde{\varphi}_l(z) dy dz$$

et

$$I_2 = \int \int_{0 \leq y \leq z} \left\{ \frac{F^N(z) - F^N(y)}{F(z) - F(y)} \right\} f(y) h(z) \tilde{\varphi}_l(y) \tilde{\varphi}_l(z) dy dz$$

admet un comportement asymptotique en $o(N^{-1} p^{-\frac{1}{2}})$.

Deuxième cas

Pour montrer le résultat du lemme 3.2 dans le cas censuré nous remarquerons d'abord que

$$E[\delta_i / (X_i = y)] = \frac{(1 - F_C(y)) f_T(y)}{f(y)}.$$

Par l'équivalent de (3.6) dans le cas censuré nous avons

$$\begin{aligned} E(\hat{c}_l) &= \frac{1}{N+1} \sum_{i=1}^N E[\tilde{\varphi}_l(X_{(i)}) \omega(X_{(i)}) (1 - \hat{F}_N(X_{(i)}))^{-1} \delta_{(i)}] \\ &= \sum_{i=1}^N E[E(\tilde{\varphi}_l(X_{(i)}) \omega(X_{(i)}) (1 - \hat{F}_N(X_{(i)}))^{-1} \delta_{(i)} / (X_{(i)} = y))] \\ &= \int \tilde{\varphi}_l(y) \omega(y) \frac{(1 - F_C(y)) f_T(y)}{f(y)} I_{1,N}(y) dy \\ &= \int \tilde{\varphi}_l(y) \omega(y) \frac{(1 - F_C(y)) f_T(y)}{f(y)} \frac{f(y)}{1 - F(y)} (1 - F^N(y)) dy. \end{aligned}$$

Comme $(1 - F) = (1 - F_C)(1 - F_T)$ nous obtenons

$$E(\hat{c}_l) = \int \tilde{\varphi}_l(y) \omega(y) \eta(y) (1 - F^N(y)) dy,$$

et nous retombons sur le résultat obtenu pour le cas non censuré.

Dans le cas censuré le résultat 1 est généralisé par Tanner et Wong [69] ; nous en déduisons de ce fait le même résultat pour la variance que dans le cas non censuré. ■

Dans la proposition 3.1 nous montrons que $\hat{\lambda}_p$ défini par (3.3) est un estimateur consistant en moyenne quadratique de $\lambda_p = \sum_l c_l \tilde{\varphi}_l(x)$. L'analogie de ce résultat a été obtenu pour l'estimation par noyau linéaire dans le cas non censuré par Watson et Leadbetter et par Tanner et Wong dans le cas censuré.

Proposition 3. 1.

$$E(\hat{\lambda}_p) = \lambda_p + \mathcal{O}(N^{-1}p^{\frac{1}{2}})$$

$$\text{et } \text{var}(\hat{\lambda}_p) \rightarrow 0.$$

Preuve de la proposition 3.1 :

Par (3.3) et le fait que $\sum_l \varphi_l = 1$ nous montrons que $\hat{\lambda}_p$ est un estimateur sans biais. De plus le calcul de la variance de $\hat{\lambda}_p$ est identique à celui de \hat{c}_l et le terme dominant est le même. ■

3.2 Comportement asymptotique du MISE de $\hat{\lambda}_p$

Rappelons la définition du MISE de $\hat{\lambda}_p$

$$MISE(\hat{\lambda}_p) = E \int (\hat{\lambda}_p - \lambda)^2(x) dx.$$

Par la définition de $\hat{\lambda}_p$ et vu l'orthonormalité du système de base nous avons la décomposition suivante

$$MISE(\hat{\lambda}_p) = \sum_l E(\hat{c}_l - c_l)^2 + \sum_{k=0}^{\infty} \sum_l d_{kl}^2.$$

Le premier terme de la décomposition du MISE correspond à l'erreur quadratique d'estimation alors que le second présente l'erreur d'approximation.

Dans ce qui suit nous supposons que λ est de classe $\mathcal{C}^{(r-1)}$ et que sa dérivée à l'ordre r est continue par morceaux et nous noterons $\kappa = (r!)^{-1} \int \psi(y) y^r dy$ et $C_1 = \int \alpha(y) dy$

Théorème 3. 1. *Supposons que l'on dispose d'une ondelette mère dont les moments jusqu'à l'ordre $(r - 1)$ sont nuls et d'une fonction d'échelle toutes*

les deux à support compact, si $p \rightarrow \infty$ et que $pN^{-1} \rightarrow 0$ quand $N \rightarrow \infty$, alors :

$$MISE(\hat{\lambda}_p) = C_1 N^{-1} p + p^{-2r} \kappa^2 (1 - 2^{-2r})^{-1} \int |\lambda^{(r)}|^2(x) dx + \mathcal{O}(p^{-2r}) + o(pN^{-1}).$$

Remarque 3. 2. :

1. Le calcul du MISE ci-dessus justifie la remarque 3.1 du paragraphe 3.1.
2. La formule du MISE pour l'estimateur linéaire correspond au résultat classique $MISE \sim C_1 N^{-1} p + C_2 p^{-2r}$. Cependant à cause du terme $\mathcal{O}(p^{-2r})$, la constante C_2 ne peut être donnée explicitement. Néanmoins, pour $p = \mathcal{O}(N^{\frac{1}{2r+1}})$, le MISE de $\hat{\lambda}_p$ admet un comportement asymptotique en $\mathcal{O}(N^{-\frac{2r}{2r+1}})$.
3. Les valeurs des intégrales utilisées dans l'expression du MISE sont finies du fait des conditions imposées à λ .

Preuve du Théorème 3.1

Considérons le terme correspondant à l'erreur d'estimation et notons que

$$\sum_{\Gamma} E(\hat{c}_l - c_l)^2 = \sum_{\Gamma} [\text{var}(\hat{c}_l) + (E(\hat{c}_l) - c_l)^2].$$

D'après le lemme 3.2 chaque terme de la somme est dominé par $\text{var}(\hat{c}_l)$; le terme dominant de la somme est alors $\sum_{\Gamma} \text{var}(\hat{c}_l)$.

Or

$$N \text{var}(\hat{c}_l) \rightarrow \int \alpha\left(\frac{y+l}{p}\right) \varphi^2(y) dy$$

et comme $\int \varphi^2(x) dx = 1$ il s'ensuit :

$$\sum_{\Gamma} E(\hat{c}_l - c_l)^2 = C_1 N^{-1} p + o(N^{-1} p).$$

Posons

$$s_0 = \sum_{k=0}^{\infty} \sum_{\Gamma} d_{kl}^2.$$

Avant de procéder à une majoration de chacun des termes de s_0 nous noterons que pour k fixé seul un nombre fini des termes d_{kl} sont non nuls. En effet λ et ψ sont à support compact et donc pour k fixé ce nombre est

inférieur à $2^k LA$ où $2L$ et $2A$ sont les longueurs respectives du support de ψ et de f .

Pour tout $(k, l) \in \Gamma$ nous avons :

$$\begin{aligned} d_{kl} &= \int \tilde{\psi}_{kl}(x) \lambda(x) dx \\ &= \int p_k^{\frac{1}{2}} \psi(p_k x - l) \lambda(x) dx \\ &= p_k^{-\frac{1}{2}} \int \psi(y) \lambda\left(\frac{y+l}{p_k}\right) dy, \end{aligned}$$

soit

$$p_k^{\frac{1}{2}} d_{kl} = \int \psi(y) \lambda\left(\frac{y+l}{p_k}\right) dy.$$

Par un développement limité de Taylor à l'ordre r et du fait que ψ a des moments nuls jusqu'à l'ordre $(r-1) \in \Gamma$ nous obtenons

$$\begin{aligned} p_k^{\frac{1}{2}} d_{kl} &= \int \psi(y) \sum_{j=0}^{r-1} (j!)^{-1} \left(\frac{y}{p_k}\right)^j \lambda^{(j)}\left(\frac{l}{p_k}\right) \\ &\quad + ((r-1)!)^{-1} \left(\frac{y}{p_k}\right)^r \left[\int_0^1 (1-t)^{r-1} \lambda^{(r)}\left(\frac{l+ty}{p_k}\right) dt \right] dy \\ &= ((r-1)!)^{-1} \int \psi(y) \left(\frac{y}{p_k}\right)^r \left[\int_0^1 (1-t)^{r-1} (\lambda^{(r)}\left(\frac{l+ty}{p_k}\right) - \lambda^{(r)}\left(\frac{l}{p_k}\right)) dt \right. \\ &\quad \left. + \int_0^1 (1-t)^{r-1} \lambda^{(r)}\left(\frac{l}{p_k}\right) dt \right] dy \\ &= (r!)^{-1} p_k^{-r} \int \psi(y) y^r dy [\lambda^{(r)}\left(\frac{l}{p_k}\right) + \Delta_{kl}] \end{aligned}$$

où Δ_{kl} vérifie que $\sup_l |\Delta_{kl}| \rightarrow 0$ quand $N \rightarrow \infty$.

Comme $\kappa = (r!)^{-1} \int \tilde{\psi}(y) y^r dy \in \Gamma$ nous avons

$$d_{kl} = \kappa p_k^{-(r+\frac{1}{2})} (\lambda^{(r)}\left(\frac{l}{p_k}\right) + \Delta_{kl}). \quad (3.8)$$

Nous en déduisons que

$$s_0 = \sum_{k=0}^{\infty} \sum_l \kappa^2 p_k^{-(2r+1)} (\lambda^{(r)}\left(\frac{l}{p_k}\right) + \Delta_{kl})^2,$$

et donc pour N suffisamment grand Γ le terme Δ_{kl}^2 est négligeable devant $(\lambda^{(r)}\left(\frac{l}{p_k}\right))^2$; ainsi :

$$\begin{aligned} s_0 &= \kappa^2 \sum_{k=0}^{\infty} p_k^{-(2r+1)} \sum_l [\lambda^{(r)}\left(\frac{l}{p_k}\right)]^2 + \mathcal{O}\left(\sum_k p_k^{-2r}\right) \\ &= p^{-2r} (1 - 2^{-2r})^{-1} \kappa^2 \int (\lambda^{(r)})^2(x) dx + \mathcal{O}(p^{-2r}), \end{aligned}$$

puisque $p_k^{-1} \sum_l [\lambda^{(r)}\left(\frac{l}{p_k}\right)]^2 < \infty$. ■

Remarquons au vu du théorème 2.1 qu'il nous est pas possible d'évaluer la valeur optimale de p qui minimise l'erreur quadratique moyenne intégrée de l'estimateur linéaire $\hat{\lambda}_p$. Nous nous proposons maintenant d'estimer λ par une procédure non linéaire.

Soit $j = j(N)$ un niveau de la multirésolution définie par φ et posons $c_{jl} = \langle \lambda, \varphi_{jl} \rangle$ et $\lambda_j = \sum_l c_{jl} 2^{\frac{j}{2}} \varphi(2^j \cdot -l)$.

Notons $\hat{\lambda}_j$ l'estimateur empirique de λ_j obtenu en remplaçant c_{jl} par \hat{c}_{jl} défini par

$$\hat{c}_{jl} = \sum_{i=1}^N \varphi_{jl}(X_{(i)}) (N+1-i)^{-1} \omega(X_{(i)}).$$

Si λ est une fonction r -régulière le biais de l'estimateur $\hat{\lambda}_j$ est de l'ordre de $2^{-j(N)r}$ alors que le terme de la variance (composante stochastique du MISE) est de l'ordre $\frac{2^{j(N)}}{N}$. Nous choisissons donc un niveau de résolution $j_1(N)$ suffisamment petit pour que la composante stochastique du MISE ait le bon taux de convergence (soit $j_1(N) \approx \log_2(N^{\frac{1}{2r+1}})$) puis nous rajoutons certains détails de manière à ce que le terme du biais atteigne le bon taux de convergence. Cependant plus le nombre de termes nécessaires à la définition de l'estimateur augmente plus l'erreur stochastique augmente mais plus le biais diminue.

Pour nos propositions nous proposons de modifier l'estimateur linéaire $\hat{\lambda}_p$ de sorte à inclure certains coefficients associés aux "détails". Ces derniers seront modifiés par la technique de seuillage déjà décrite au chapitre 2 avec pour conséquence la diminution globale de l'erreur stochastique de l'estimateur. Pour équilibrer les deux composantes du MISE l'estimateur en question dépend de deux paramètres q et t respectivement paramètre de troncature correspondant à l'espace de projection V_q^p auquel nous nous plaçons et paramètre de seuillage.

Remarque 3.3. :

1. Les simulations effectuées montrent que l'estimateur par ondelettes est sensible au niveau de résolution primaire $j_1(N)$.
2. Dans l'écriture de $\hat{\lambda}_p$, p tel qu'il est défini correspond à $2^{j_1(N)}$ avec l'avantage que p n'est pas restreint à être un entier : ceci est le principal intérêt dans l'utilisation du système $\tilde{\varphi}$. Nous pouvons nous poser

alors la question suivante : “Comment procéder pour obtenir un taux optimal?” La réponse est donnée par l’analogie avec l’estimation par noyau où la largeur de la fenêtre est choisie de manière à optimiser asymptotiquement la valeur du MISE en équilibrant ses deux composantes (biais et composante stochastique). C’est cet argument qui sera utilisé pour le choix du paramètre p .

3.3 Estimation non linéaire par seuillage uniforme

La technique d’estimation de seuillage utilisée ci-après a été largement étudiée par D. Donoho et I.M Jhonstone [25]. Ils en ont montré toutes les vertus pour l’estimation des fonctions de régression et de densité notamment pour le critère d’erreur min-max : l’estimateur non linéaire qu’ils obtiennent reste optimal pour une large classe de fonctions (Espaces de Besov).

Pour des fonctions lisses λ seul un nombre fini des coefficients d_{kl} de la transformée en ondelettes de λ sont non négligeables et de manière heuristique nous dirons que l’information fournie par les coefficients \hat{d}_{kl} lorsqu’ils ont une grande (resp. faible) amplitude en valeur absolue provient essentiellement de la fonction à estimer (resp. uniquement de l’erreur stochastique). Il est donc tout à fait approprié de supprimer dans l’écriture de $\hat{\lambda}$ défini ci-après les coefficients \hat{d}_{kl} tels que $|\hat{d}_{kl}| < t$ où t reste à définir : nous parlerons alors de seuillage “dur”. Si de plus nous procédons à une modification des coefficients \hat{d}_{kl} pour lesquels $|\hat{d}_{kl}| > t$ nous parlerons alors de seuillage “doux”. Cette technique de seuillage est d’autant plus correcte que nous ne modifierons pas les valeurs des \hat{c}_l qui constituent ce que nous appelons une approximation principale alors que les \hat{d}_{kl} incarnent les détails.

Notre choix s’est porté sur le seuillage “dur”. Ainsi pour t donné nous définissons

$$\hat{\lambda} = \sum_l \hat{c}_l \tilde{\varphi}_l + \sum_{k=0}^{q-1} \sum_l \hat{d}_{kl} \mathbb{1}_{(|\hat{d}_{kl}| > t)} \tilde{\psi}_{kl}, \quad (3.9)$$

soit en posant

$$\hat{\lambda}_p = \sum_l \hat{c}_l \tilde{\varphi}_l,$$

$$\text{et } \hat{\lambda}_d = \sum_{k=0}^{q-1} \sum_l \hat{d}_{kl} \mathbb{1}_{(|\hat{d}_{kl}| > t)} \tilde{\psi}_{kl}(x),$$

nous avons

$$\hat{\lambda}(x) = \hat{\lambda}_p + \hat{\lambda}_d, \quad (3.9)'$$

ce qui constitue une décomposition de $\hat{\lambda}$ en une composante linéaire $\hat{\lambda}_p$ et une composante non linéaire $\hat{\lambda}_d$.

Par le Théorème 3.2 énoncé ci-après nous étendons les travaux de D. Picard et al [28] et de N. Patil et P. Hall [41] sur l'estimation de la fonction de densité par projection non linéaire sur un système orthonormé d'ondelettes à l'estimation de la fonction de hasard pour des données avec ou sans censure.

Soient κ et C_1 les constantes définies au paragraphe 3.2. Nous avons :

Théorème 3. 2. *Sous les conditions du Théorème 3.1 et si de plus q et $p^{2r+1}t^2$ tendent vers l'infini alors que $p_q t^2$ tend vers 0 quand N tend vers l'infini, pour un choix de t tel que $t > C(\frac{\log(N)}{N})^{\frac{1}{2}}$ où C est une constante définie explicitement dans la preuve du théorème, on a*

$$MISE(\hat{\lambda}) = C_1 N^{-1} p + p^{-2r} \kappa^2 (1 - 2^{-2r})^{-1} \int (\lambda^{(r)})^2 + o(N^{-1} p) + o(p^{-2r}).$$

La formule du MISE pour l'estimateur étudié est analogue à celle obtenue pour les estimateurs par noyau et ondelettes pour la fonction de densité. Les termes $C_1 N^{-1} p$ et $p^{-2r} \kappa^2 (1 - 2^{-2r})^{-1} \int (\lambda^{(r)})^2$ représentent les contributions respectives de l'erreur stochastique et du carré du biais.

La suite est consacrée à la preuve du théorème qui se décomposera en plusieurs parties.

3.4 Décomposition du MISE de $\hat{\lambda}$

Nous rappelons que $MISE(\hat{\lambda}) = E \int (\hat{\lambda} - \lambda)^2(x) dx$. Ce qui au vu de l'écriture (3.9) et de l'orthonormalité du système de base conduit à la décomposition suivante

$$\begin{aligned} MISE(\hat{\lambda}) = & \sum_l E(\hat{c}_l - c_l)^2 + \sum_{k=0}^{q-1} \sum_l E\left((\hat{d}_{kl} - d_{kl})^2 \mathbb{1}_{(|\hat{d}_{kl}| > t)}\right) \\ & + \sum_{k=0}^{q-1} \sum_l E\left(d_{kl}^2 \mathbb{1}_{(|\hat{d}_{kl}| < t)}\right) \\ & + \sum_{k=q}^{\infty} \sum_l d_{kl}^2. \end{aligned}$$

Pour chacun des quatre termes ci-dessus nous donnerons son comportement asymptotique. Ce qui permettra par la suite d'en déduire celui du MISE de $\hat{\lambda}$. Nous partagerons le travail en deux : d'une part la partie linéaire qui se réduit à un seul terme puis la partie non linéaire constituée de trois termes.

3.5 Comportement de la partie linéaire

La contribution de la partie linéaire $\hat{\lambda}_p$ dans le calcul du MISE est donnée par la somme suivante

$$\sum_l E(\hat{c}_l - c_l)^2 = \sum_l [\text{var}(\hat{c}_l) + (E(\hat{c}_l) - c_l)^2].$$

Or d'après les résultats du paragraphe 3.2 On a

$$\sum_l E(\hat{c}_l - c_l)^2 = C_1 N^{-1} p + o(N^{-1} p). \quad (3.10)$$

3.6 Comportement de la partie non linéaire

Traisons séparément chacun des trois termes de la partie non linéaire. Posons :

$$s_1 = \sum_{k=q}^{\infty} \sum_l d_{kl}^2.$$

Le terme s_1 correspond à l'erreur d'approximation de λ par sa projection sur le sous-espace V_q^p .

Une fois qu'on estime les coefficients d_{kl} par les \hat{d}_{kl} définis par (3.2) ou son équivalent (3.2) On procède à une sélection en ne retenant que ceux pour lesquels $|\hat{d}_{kl}|$ est supérieur à un seuil t ce qui revient à poser :

$$\begin{aligned} d_{kl}^* &= 0 \text{ si } |\hat{d}_{kl}| < t \\ d_{kl}^* &= \hat{d}_{kl} \text{ sinon.} \end{aligned}$$

De ce fait si on s'intéresse à l'erreur due à l'estimation des d_{kl} par \hat{d}_{kl} on a deux termes :

$$s_2 = \sum_{k=0}^{q-1} \sum_l d_{kl}^2 \mathbb{1}_{[|\hat{d}_{kl}| < t]} \quad (3.11)$$

et

$$s_3 = \sum_{k=0}^{q-1} \sum_l (d_{kl} - \hat{d}_{kl})^2 \mathbb{1}_{[|\hat{d}_{kl}| > t]}. \quad (3.12)$$

3.6.1 Quelques résultats

Nous rappelons deux résultats de la théorie des probabilités et dont nous nous servirons dans les paragraphes ci-après.

Le premier est le suivant :

Lemme 3. 3 (Bretagnolle-Huber) Soit T_1, T_2, \dots, T_N des variables aléatoires indépendantes et identiquement distribuées telles que $E(T_i) = 0, E(T_i^2) \leq \sigma^2$ et $|T_i| < A$. Pour tout m , il existe une constante c_m telle que

$$\begin{aligned} \text{si } m \geq 2, \quad E|N^{-1} \sum_{i=1}^N T_i|^m &\leq c_m \left(\frac{\sigma^2 A^{m-2}}{N^{m-1}} + \frac{\sigma^m}{N^{\frac{m}{2}}} \right) \\ \text{et si } m = 2, \quad E|N^{-1} \sum_{i=1}^N T_i|^m &\leq \frac{\sigma^m}{N^{\frac{m}{2}}}. \end{aligned}$$

Le deuxième est un résultat de la théorie des grandes déviations :

Lemme 3. 4 (Bernstein) Soit N variables \tilde{T}_i indépendantes et telles que $E(\tilde{T}_i) = 0, E(\tilde{T}_i^2) = \sigma_i^2$ et $|\tilde{T}_i| < M$. Si $\sum_{i=1}^N \sigma_i^2 \leq V_N$, alors pour tout $\eta > 0$, on a

$$P\left\{ \left| \sum_{i=1}^N \tilde{T}_i \right| \geq \eta \right\} \leq 2 \exp\left[-\frac{1}{2} \eta^2 / (V_N + \frac{1}{3} M \eta)\right].$$

3.6.2 Comportement asymptotique de s_1

Par (3.8) On a

$$s_1 = \sum_{k=q}^{\infty} \sum_l \kappa^2 p_k^{-(2r+1)} \left(\lambda^{(r)}\left(\frac{l}{p_k}\right) + \Delta_{kl} \right)^2$$

et comme pour N suffisamment grand on peut majorer le terme Δ_{kl}^2 par $(\lambda^{(r)}(\frac{l}{p_k}))^2$ cela entraîne que

$$\begin{aligned} s_1 &\leq 2\kappa^2 \sum_{k=q}^{\infty} p_k^{-(2r+1)} \sum_l \left[\lambda^{(r)}\left(\frac{l}{p_k}\right) \right]^2 \\ &= \mathcal{O}(p_q^{-2r}) \\ &= o(p^{-2r}), \end{aligned}$$

car d'une part q tend vers l'infini avec N et d'autre part

$$p_k^{-1} \sum_l \left[\lambda^{(r)}\left(\frac{l}{p_k}\right) \right]^2 < \infty.$$

■

3.6.3 Comportement asymptotique de $(d_{kl} - \hat{d}_{kl})^2$

Ci-après nous étudions le comportement asymptotique de $E(d_{kl} - \hat{d}_{kl})^2$. Cette étude passe par l'introduction du terme \tilde{d}_{kl} défini par

$$\tilde{d}_{kl} = \frac{1}{N} \sum_{i=1}^N (1 - F(X_i))^{-1} \omega(X_i) \tilde{\psi}_{kl}(X_i). \quad (3.13)$$

Nous montrerons ci-dessous en étudiant successivement le comportement de $E(\hat{d}_{kl} - \tilde{d}_{kl})^2$ et $E(\tilde{d}_{kl} - d_{kl})^2$ que le terme dominant de $E(d_{kl} - \hat{d}_{kl})^2$ est donné par le second terme.

Posons $\alpha_{kl} = \sup_{x \in \text{supp} \psi} \alpha\left(\frac{x+l}{p_k}\right)$ et $f_{kl} = \sup_{x \in \text{supp} \psi} f\left(\frac{x+l}{p_k}\right)$ et $\hat{Q}_N(x) = \max(1 - \hat{F}_N(x), \varepsilon_0)$; $\hat{F}_N(x)$ et ε_0 ayant été définis en début de chapitre. On a d'abord

Lemme 3. 5.

$$E(\hat{d}_{kl} - \tilde{d}_{kl})^2 = o(N^{-1} f_{kl})$$

$$\text{et } E(d_{kl} - \tilde{d}_{kl})^2 \leq N^{-1} \alpha_{kl}.$$

Preuve :

Par définition de \hat{d}_{kl} et \tilde{d}_{kl} (resp. 3.2 et 3.13)

$$\begin{aligned} \hat{d}_{kl} - \tilde{d}_{kl} &= \frac{1}{N+1} \sum_{i=1}^N [(\hat{Q}_N)^{-1}(X_i) - Q^{-1}(X_i)] \tilde{\psi}_{kl}(X_i) \omega(X_i) \\ &\quad - \frac{1}{N(N+1)} \sum_{i=1}^N Q^{-1}(X_i) \tilde{\psi}_{kl}(X_i) \omega(X_i). \end{aligned}$$

Notons d'une part que $\sup(|\hat{Q}_N(x) - Q(x)|) = \mathcal{O}_p(N^{-\frac{1}{2}})$ et d'autre part que $\sup(|\hat{Q}_N(x) - Q(x)|)$ est absolument borné ce qui par l'inégalité de Markov entraîne que

$$E(\sup |\hat{Q}_N(x) - Q(x)|)^2 \text{ tend vers } 0 \text{ quand } N \text{ tend vers l'infini.}$$

Comme $Q(x)$ et $\hat{Q}_N(x)$ sont strictement supérieurs à ε_0 sur $[0, T]$

obtient

$$\begin{aligned}
E(\tilde{d}_{kl} - \hat{d}_{kl})^2 &= \frac{1}{(N+1)^2} E\left[\sum_{i=1}^N ((\hat{Q}_N)^{-1}(X_i) - Q^{-1}(X_i)) \tilde{\psi}_{kl}(X_i) \omega(X_i) \right]^2 \\
&\quad + \mathcal{O}(N^{-3}) \\
&\leq \frac{\text{const.}}{(N+1)^2} E(\sup |\hat{Q}_N(x) - Q(x)|)^2 E\left(\sum_{i=1}^N |\tilde{\psi}_{kl}(X_i) \omega(X_i)| \right)^2 \\
&\leq \frac{\text{const.}}{(N+1)^2} E(\sup |\hat{Q}_N(x) - Q(x)|)^2 N \sup_{x \in \text{supp} \psi} f\left(\frac{x+l}{p_k}\right) \\
&= o(N^{-1} f_{kl}).
\end{aligned}$$

Pour démontrer la deuxième partie du Lemme 3.5 posons Γ dans le cas non censuré Γ

$$T_i = \tilde{\psi}_{kl}(X_i)(1 - F(X_i))^{-1} \omega(X_i) - d_{kl}$$

et

$$T_i = \tilde{\psi}_{kl}(X_i)(1 - F(X_i))^{-1} \omega(X_i) \delta_i - d_{kl}$$

pour le cas censuré. On a ainsi

$$\tilde{d}_{kl} - d_{kl} = \frac{1}{N} \sum_{i=1}^N T_i.$$

Le résultat est alors immédiat par le lemme de Bretagnolle-Huber en prenant $\sigma^2 = \alpha_{kl}$ et $A = 2 \|\psi\|_{\infty} p_k^{\frac{1}{2}} \varepsilon_0^{-1}$. \blacksquare

3.6.4 Comportement asymptotique de s_2

Soit $t_0 = 3t/2$. Nous désignons ci-après par s'_2 et s''_2 les termes

$$s'_2 = \sum_{k=0}^{q-1} \sum_l d_{kl}^2 \mathbb{1}_{[|\tilde{d}_{kl}| < t_0]},$$

et

$$s''_2 = \sum_{k=0}^{q-1} \sum_l d_{kl}^2 \mathbb{1}_{[|\tilde{d}_{kl} - \hat{d}_{kl}| > \frac{1}{2}]}$$

Il est facile de voir que

$$s_2 \leq s'_2 + s''_2.$$

Cette majoration du terme s_2 permettra de traiter séparément chacun des deux termes.

Comportement de s'_2

L'étude du comportement de s'_2 nécessite un calcul algébrique préalable. Notons dans un premier temps que par (3.8) on peut exhiber une constante telle que

$$|d_{kl}| < \text{const.} p_k^{-(r+\frac{1}{2})} \quad (3.14)$$

et donc sous l'hypothèse que $p^{(2r+1)t^2}$ tend vers l'infini avec N , $|d_{kl}|$ est négligeable devant t_0 pour N suffisamment grand. Il en découle que pour tout ε réel strictement positif les sommes s_{21} et s_{22} définies respectivement par

$$s_{21} = \sum_{k=0}^{q-1} \sum_l d_{kl}^2 \mathbb{1}_{[|d_{kl}| < (1+\varepsilon)t_0]}$$

et

$$s_{22} = \sum_{k=0}^{q-1} \sum_l d_{kl}^2 \mathbb{1}_{[|d_{kl}| < (1-\varepsilon)t_0]}$$

sont pour N suffisamment grand telles que $s_{21} = s_{22} = \sum_{k=0}^{q-1} \sum_l d_{kl}^2$.

Comme

$$|\tilde{d}_{kl}| \geq |d_{kl}| - |\tilde{d}_{kl} - d_{kl}|,$$

on a

$$\left. \begin{array}{l} |d_{kl}| > (1+\varepsilon)t_0 \\ |\tilde{d}_{kl} - d_{kl}| < \varepsilon t_0 \end{array} \right\} \Rightarrow |\tilde{d}_{kl}| > t_0.$$

D'autre part

$$|d_{kl}| \geq |\tilde{d}_{kl}| - |\tilde{d}_{kl} - d_{kl}|$$

entraîne

$$\left. \begin{array}{l} |\tilde{d}_{kl}| > t_0 \\ |\tilde{d}_{kl} - d_{kl}| < \varepsilon t_0 \end{array} \right\} \Rightarrow |d_{kl}| > (1-\varepsilon)t_0.$$

Ce qui équivaut en posant

$$\Delta = \sum_{k=0}^{q-1} \sum_l d_{kl}^2 \mathbb{1}_{[|\tilde{d}_{kl} - d_{kl}| > \varepsilon t_0]},$$

à

$$s_{22} - \Delta < s'_2 < s_{21} + \Delta.$$

Nous allons montrer successivement que :

$$s_{21} = s_{22} \sim \kappa^2 (1 - 2^{-2r})^{-1} p^{-2r} \int (\lambda^{(r)})^2(x) dx$$

et que

$$E(\Delta) = o(p^{-2r}).$$

Par (3.8) et du fait que $\sup_{0 \leq k \leq q-1, l} |\Delta_{kl}|$ tend vers 0 on a

$$\begin{aligned}
s_{21} &= \sum_{k=0}^{q-1} \sum_l \kappa^2 p_k^{-(2r+1)} (\lambda^{(r)}(\frac{l}{p_k}) + \Delta_{kl})^2 \\
&= \kappa^2 \sum_{k=0}^{q-1} p_k^{-(2r+1)} \sum_l (\lambda^{(r)}(\frac{l}{p_k}))^2 + o(\sum_{k=0}^{q-1} p_k^{-2r}) \\
&= \kappa^2 \sum_{k=0}^{q-1} p_k^{-2r} \int (\lambda^{(r)}(y))^2 dy + o(\sum_{k=0}^{q-1} p_k^{-2r}) \\
&= \kappa^2 p^{-2r} (1 - 2^{-2r})^{-1} \int (\lambda^{(r)}(y))^2 dy + o(\sum_{k=0}^{q-1} p_k^{-2r}).
\end{aligned}$$

Donc

$$s_{21} = \kappa^2 p^{-2r} (1 - 2^{-2r})^{-1} \int (\lambda^{(r)}(y))^2 dy + o(p^{-2r}). \quad (3.15)$$

Pour le terme $\Delta\Gamma$ on utilisera le lemme de Bernstein. Posons

$$\tilde{Z}_i = \frac{1}{N} \tilde{\psi}_{kl}(X_i) (1 - F(X_i))^{-1} \omega(X_i), \quad (3.16)$$

et de manière analogue Γ pour le cas censuré Γ

$$\tilde{Z}_i = \frac{1}{N} \tilde{\psi}_{kl}(X_i) (1 - F(X_i))^{-1} \omega(X_i) \delta_i \quad (3.16)'$$

et

$$\tilde{T}_i = \tilde{Z}_i - \frac{1}{N} d_{kl} = \frac{T_i}{N}. \quad (3.17)$$

On a alors Γ

$$\tilde{d}_{kl} - d_{kl} = \sum_{i=1}^N \tilde{T}_i. \quad (3.18)$$

En remarquant que $E(\tilde{T}_i) = 0$, $|\tilde{T}_i| < 2N^{-1} p_k^{\frac{1}{2}} \|\psi\|_{\infty} \varepsilon_0^{-1}$ et que $\sum_{i=1}^N \text{var}(\tilde{T}_i) < N^{-1} \sup_{x \in [0, T]} \alpha(x) \Gamma$ le théorème de Bernstein entraîne Γ pour

N suffisamment grand Γ

$$\begin{aligned}
E(\Delta) &= \sum_{k=0}^{q-1} \sum_l d_{kl}^2 P(|\tilde{d}_{kl} - d_{kl}| > \varepsilon t_0) \\
&\leq 2 \sum_{k=0}^{q-1} \sum_l d_{kl}^2 \exp\{-\frac{1}{3} \varepsilon^2 t_0^2 N (\sup_{x \in [0, T]} \alpha(x))^{-1}\} \\
&= o(s_{21}).
\end{aligned}$$

L'inégalité ci-dessus est vérifiée car $p_q t^2$ tend vers 0 quand N tend vers l'infini.

En combinant ce résultat avec (3.15) on obtient

$$E(s'_2 - \kappa^2 p^{-2r} (1 - 2^{-2r})^{-1} \int (\lambda^{(r)})^2(y) dy) = o(p^{-2r}).$$

Comportement de s''_2

Par définition même de s''_2 on a

$$E(s''_2) = \sum_{k=0}^{q-1} \sum_l d_{kl}^2 P[|\tilde{d}_{kl} - \hat{d}_{kl}| > \frac{t}{2}].$$

Par application du lemme de Bernstein et d'après les résultats du paragraphe 3.6.3 le comportement asymptotique de ce terme est dominé par celui de $E(\Delta)$. On en déduit donc

$$E(s''_2) = o(p^{-2r}),$$

et par conséquent

$$E(s_2 - \kappa^2 p^{-2r} (1 - 2^{-2r})^{-1} \int (\lambda^{(r)})^2(y) dy) = o(p^{-2r}). \quad (3.19)$$

3.6.5 Comportement asymptotique de s_3

Des considérations analogues à celles utilisées pour le terme s_2 montrent que

$$s_3 = o(N^{-\frac{2r}{2r+1}}). \quad (3.20)$$

En effet posons

$$s'_3 = \sum_{k=0}^{q-1} \sum_l E[(d_{kl} - \tilde{d}_{kl})^2 \mathbb{1}_{[|\hat{d}_{kl}| > t]}], \quad (3.21)$$

et

$$s''_3 = \sum_{k=0}^{q-1} \sum_l E[(\hat{d}_{kl} - \tilde{d}_{kl})^2 \mathbb{1}_{[|\hat{d}_{kl}| > t]}]. \quad (3.22)$$

Remarquons que $s_3 \leq 2(s'_3 + s''_3)$; partant de cette inégalité et du lemme 3.5 il suffit de montrer que

$$s'_3 = o(N^{-\frac{2r}{2r+1}}), \quad (3.23)$$

car en procédant de la même manière on peut montrer que ce résultat est vrai aussi pour s''_3 .

Soit τ et ξ deux réels positifs tels que $\tau + \xi = 1$. Il est facile de voir que s'_3 est majoré par l'addition des trois termes définis ci-dessous :

$$s'_{31} = \sum_{k=0}^{q-1} \sum_l E\{(\tilde{d}_{kl} - d_{kl})^2 \mathbb{1}_{[|d_{kl}| > \tau t]}\},$$

$$s'_{32} = \sum_{k=0}^{q-1} \sum_l E\{(\tilde{d}_{kl} - d_{kl})^2 \mathbb{1}_{[|\tilde{d}_{kl} - d_{kl}| > \frac{1}{2}\xi t]}\}$$

et

$$s'_{33} = \sum_{k=0}^{q-1} \sum_l E\{(\tilde{d}_{kl} - d_{kl})^2 \mathbb{1}_{[|\hat{d}_{kl} - \tilde{d}_{kl}| > \frac{1}{2}\xi t]}\}.$$

Ci-après on étudie le comportement asymptotique de chacun de ces trois termes.

Le fait que α et ψ soient à support compact entraîne que

$$p_k^{-1} \sum_l \alpha_{kl} < \infty \text{ uniformément en } k.$$

Or d'après le lemme 3.5

$$E(\tilde{d}_{kl} - d_{kl})^2 \leq N^{-1} \alpha_{kl}$$

et par (3.14)

$$|d_{kl}| < \text{const.} p_k^{-(r+\frac{1}{2})}.$$

Posons

$$k_0 = \arg \max_{k < q} \{p_k < \text{const.} (\tau t)^{-\frac{2}{2r+1}}\}.$$

Nous avons alors

$$\begin{aligned} s'_{31} &\leq N^{-1} \sum_{k=0}^{q-1} \sum_l \alpha_{kl} \mathbb{1}_{\left[p_k < \text{const.} (\tau t)^{-\frac{2}{2r+1}}\right]} \\ &= N^{-1} \sum_{k=0}^{q-1} (\sum_l \alpha_{kl} p_k^{-1}) p_k \mathbb{1}_{\left[p_k < \text{const.} (\tau t)^{-\frac{2}{2r+1}}\right]} \\ &\leq \text{const}' . N^{-1} \sum_{k=0}^{q-1} p_k \mathbb{1}_{\left[p_k < \text{const.} (\tau t)^{-\frac{2}{2r+1}}\right]} \\ &\leq \text{const}' . N^{-1} \sum_{k=0}^{k_0} p_k \mathbb{1}_{\left[p_{k_0} < \text{const.} (\tau t)^{-\frac{2}{2r+1}}\right]} \\ &= \mathcal{O}(N^{-1} t^{-\frac{2}{2r+1}}) \\ &= o(N^{-\frac{2r}{2r+1}}), \end{aligned}$$

la dernière égalité étant due au fait que $N^{\frac{1}{2}}t$ tend vers l'infini avec N .

Soit a un réel positif strictement supérieur à 1 et a^* son conjugué (i.e : $\frac{1}{a} + \frac{1}{a^*} = 1$). L'inégalité de Bretagnolle-Huber donne pour tout (k, l) :

$$E(|\tilde{d}_{kl} - d_{kl}|^{2a^*}) \leq \text{const.}(a^*, \tilde{\psi}, \alpha) ([N^{-a^*} + N^{1-2a^*} p_k^{a^*-1}]). \quad (3.24)$$

En choisissant cette fois $M = 2N^{-1} p_q^{\frac{1}{2}} \|\psi\|_{\infty} \varepsilon_0^{-1}$ et $V_N = N^{-1} \alpha_{kl}$ dans le lemme de Bernstein on a pour $0 < \varepsilon < 1/\Gamma N$ suffisamment large et d donné

$$\begin{aligned} \sum_{\Gamma} (P[|\tilde{d}_{kl} - d_{kl}| > \frac{1}{2}\xi t])^{\frac{1}{a}} &\leq \sum_{\Gamma} 2 \exp\{-\frac{1}{8}(1-\varepsilon)\xi^2 t^2 N \alpha_{kl}^{-1} a^{-1}\} \\ &= \mathcal{O}(\sum_{\Gamma} \alpha_{kl} (Nt^2)^{-1} \exp\{-\frac{1}{8}(1-3\varepsilon)\xi^2 t^2 N \alpha_{kl}^{-1} a^{-1}\}) \\ &= \mathcal{O}(\sum_{\Gamma} \alpha_{kl} (Nt^2)^{-1} \exp\{-\frac{1}{8}(1-3\varepsilon)\xi^2 t^2 N (\sup \alpha(x))^{-1} a^{-1}\}) \\ &= \mathcal{O}(\sum_l \alpha_{kl} (Nt^2)^{-1} N^{-d}) \\ &= \mathcal{O}(p_k t^{-2} N^{-(d+1)}), \end{aligned}$$

et ce pour un choix de $\varepsilon \Gamma a$ et ξ tels que $\exp\{-\frac{1}{8}(1-3\varepsilon)\xi^2 N (\sup \alpha(x))^{-1} t^2 a^{-1}\} \leq N^{-d}$.

L'inégalité de Hölder dans un premier temps et le fait que pour N tendant vers l'infini $p_q t^2$ tend vers 0 et $N^{\frac{1}{2}}t$ tend vers l'infini (ce qui entraîne que $N^{-1} p_q$ tend vers 0) on obtient

$$\begin{aligned} s'_{32} &\leq \sum_{k=0}^{q-1} \sum_{\Gamma} [E(|\tilde{d}_{kl} - d_{kl}|^{2a^*})]^{\frac{1}{a^*}} (P[|\tilde{d}_{kl} - d_{kl}| > \frac{\xi t}{2}])^{\frac{1}{a}} \\ &= \mathcal{O}[\sum_{k=0}^{q-1} p_k (N^{-1} + N^{\frac{1}{a^*}-2} p_q) N^{-(d+1)} t^{-2}] \\ &= o(N^{\frac{1}{a^*}-d}). \end{aligned}$$

En prenant d égal à $\frac{2r}{2r+1} + \frac{1}{a^*}$ on obtient

$$s'_{32} = o(N^{-\frac{2r}{2r+1}}). \quad (3.25)$$

Le choix évoqué ci-dessus au vu de l'hypothèse que $t > C(\frac{\log(N)}{N})^{\frac{1}{2}} \Gamma$ est alors possible en imposant $0 < \varepsilon < \frac{1}{3} \Gamma \xi < 1/\Gamma a \approx 1$ et $C \geq (8(\sup_{x \in [0, T]} \alpha(x))^{\frac{2r}{2r+1}})^{\frac{1}{2}}$.

Finalement comme $\sup(|\hat{Q}_N(x) - Q(x)|) = \mathcal{O}_p(N^{-\frac{1}{2}})$ et $Q(x) > \varepsilon_0$ sur $[0, T]$ on a :

$$|\tilde{d}_{kl} - \hat{d}_{kl}| \leq \left| \frac{\text{const.}}{N+1} \sum_{i=1}^N \tilde{\psi}_{kl}(X_i) \omega(X_i) \right|,$$

et donc

$$P[|\tilde{d}_{kl} - \hat{d}_{kl}| \geq \frac{\xi t}{2}] \leq P\left[\frac{\text{const.}}{N+1} \sum_{i=1}^N |\tilde{\psi}_{kl}(X_i)\omega(X_i)| \geq \frac{\xi t}{2}\right].$$

La même démarche que celle utilisée pour $s'_{32}\Gamma$ prouve que

$$s'_{33} = o(N^{-\frac{2r}{2r+1}}) \quad (3.26)$$

Du comportement asymptotique de $s'_{31}\Gamma$ de (3.24) et (3.25) nous déduisons le résultat (3.20).

En combinant (3.10) et les résultats des paragraphes 3.6.1 et 3.6.3-5 nous complétons la preuve du Théorème 3.2. ■

3.7 Simulations numériques

Dans ce paragraphe nous avons regroupé quelques résultats dans le but de comparer l'estimateur de la fonction de taux de hasard que nous proposons avec celui de Kooperberg et al [45]. Ce dernier est retenu pour la qualité de ses résultats pratiques.

Notre choix s'est porté sur deux fonctions de densité f_1 loi Gaussienne de moyenne et variance respectivement égales à 20 et à 1 et f_2 une loi exponentielle de paramètre 1. Les fonctions de taux de hasard correspondantes seront notées respectivement par h_1 et h_2 . Pour le premier choix h_1 est infiniment différentiable et pour le deuxième h_2 est une constante.

La fonction d'échelle utilisée tout au long de cette étude est du type Daubechies à support minimal. Pour assurer un minimum de régularité pour notre estimateur nous avons pris des ondelettes d'ordre 3 et 5.

Trois paramètres sont nécessaires pour la détermination de l'estimateur non linéaire de la fonction de hasard défini au paragraphe 3 de ce chapitre. Il s'agit de q paramètre de troncature t paramètre de seuillage et p paramètre de dilatation.

Le paramètre q est égal à $[log(N)] - 1$ ($[x]$ désignant la partie entière de x) où N est la taille de l'échantillon dont on dispose. Le paramètre p est déterminé par Plug-in : comme p est choisi de sorte à optimiser la valeur théorique du MISE sa valeur dépend de la fonction de hasard inconnue ; une première estimation par ondelettes de la fonction de hasard est déterminée puis dans l'écriture de p la fonction de hasard est remplacée par son estimation. Finalement t est égal à $(8(\sup_{x \in [0, T]} \alpha(x)) \frac{2r}{2r+1})^{\frac{1}{2}} (\frac{log(N)}{N})^{\frac{1}{2}}$.

3.7.1 Comportement asymptotique du MISE

Pour chacune des densités f_1 et f_2 nous avons simulé des échantillons de taille 256 et 1024. La première partie consiste en l'étude du comportement asymptotique de l'estimateur non linéaire proposé et la comparaison de ses performances avec celles de l'estimateur logspline. Nous utilisons le critère d'erreur quadratique moyenne noté MSE et défini par :

$$MSE(h_i) = \frac{1}{N} \sum_{j=1}^N (h_i(t_j) - \hat{h}_i(t_j))^2,$$

où $\hat{h}_i(t_j)$ est l'estimateur de h_i à l'instant t_j et N le nombre d'observations dont on dispose.

Pour chacun des trois cas Γ nous avons calculé l'estimateur ondelette et logspline pour 50 échantillons de même taille Γ tout en fixant à 4 la valeur de r (i.e l'ordre de l'ondelette de Daubechies est égal à 3) Γ pour l'estimation de h_2 et à 6 pour celle de h_1 . Nous définissons le critère RMSE comme la racine carré de la moyenne sur l'ensemble des 50 échantillons Γ des valeurs du MSE.

L'ensemble des valeurs du RMSE Γ obtenues pour la fonction de hasard Γ h_1 (resp. h_2) Γ sont regroupées dans le tableau 3.1 (resp. 3.2).

Taille de l'échantillon	256	512	1024
Ondelette	0.043	0.008	0.0065
Logspline	0.1016	0.0908	0.0742

Table 3.1 : Résultats des simulations pour la fonction de taux de hasard associée à la loi normale.

La fonction de hasard est estimée sur l'intervalle $[18, 22]$.

Taille de l'échantillon	256	512	1024
Ondelette	0.0584	0.0505	0.0311
Logspline	0.0282	0.0425	0.0110

Table 3.2 : Résultats des simulations pour la fonction de taux de hasard associée à la loi exponentielle.

Dans ce cas Γ nous nous sommes restreints à l'estimation de la fonction de hasard sur $[0, 5]$.

Au vu des simulations présentées ci-dessus Γ nous remarquons

1. Pour la fonction de hasard régulière h_1 Γ l'estimateur ondelette proposé a un comportement asymptotique légèrement supérieur à celui de logspline. Ce qui n'est pas du tout le cas pour l'estimation de la fonction de hasard h_2 .
2. La qualité des deux estimateurs augmente avec la taille de l'échantillon.

3.7.2 Quelques graphiques

Dans cette deuxième partie nous comparons graphiquement la qualité des estimateurs ondelette et logspline [45] et ce pour chacun des trois échantillons. Cette comparaison est illustrée par les figures 3.3-7.

Les courbes que nous nous proposons d'estimer sont données par les 2 figures ci-dessous. Elles correspondent aux fonctions de taux de hasard des densités f_1 et f_2 .

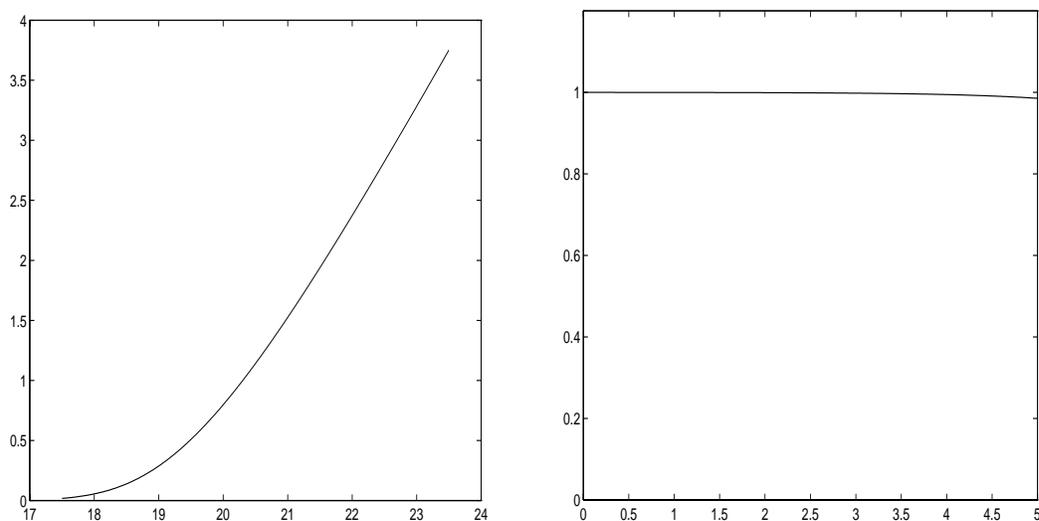


Figure 3.1 : Fonctions de taux de hasard h_1 et h_2 .

Ci-après nous donnons les figures obtenues par nos simulations. Pour chacune des fonctions de hasard nous déterminons son estimateur ondelette et logspline et ce pour des échantillons de taille 256, 512 et 1024. Ces figures permettront de compléter l'étude du comportement asymptotique de l'estimateur de hasard proposé et sa comparaison avec l'estimateur logspline.

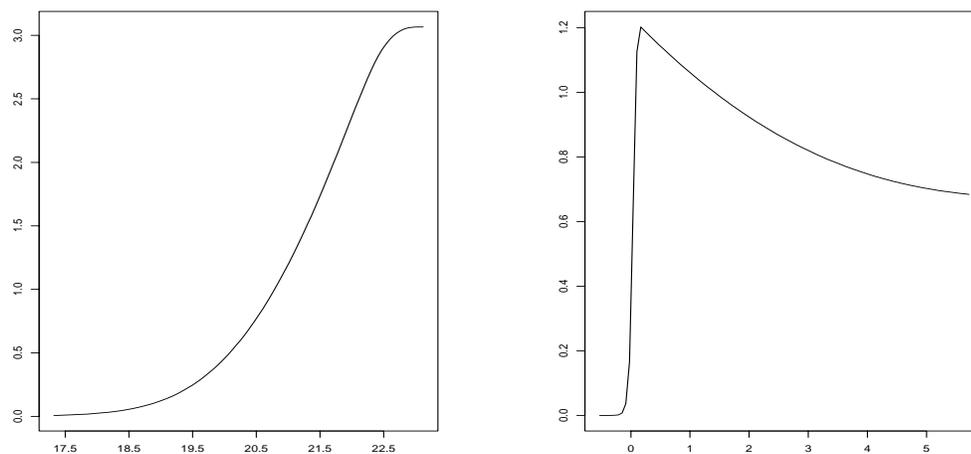


Figure 3.2 : Estimations du taux de hasard par logspline ($N = 256$).

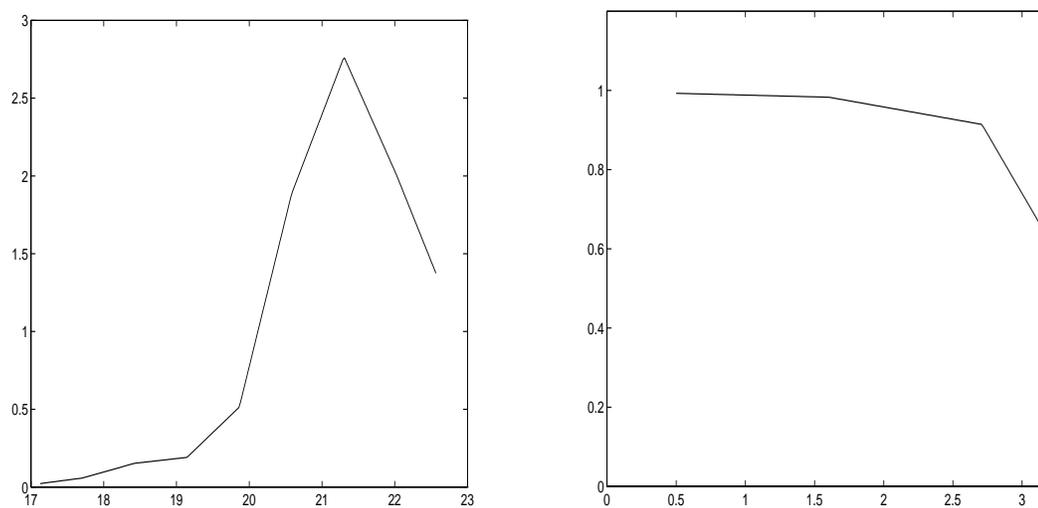


Figure 3.3 : Estimations du taux de hasard par ondelette ($N = 256\Gamma r = 4$ pour h_1 et $r = 4$ pour h_2).

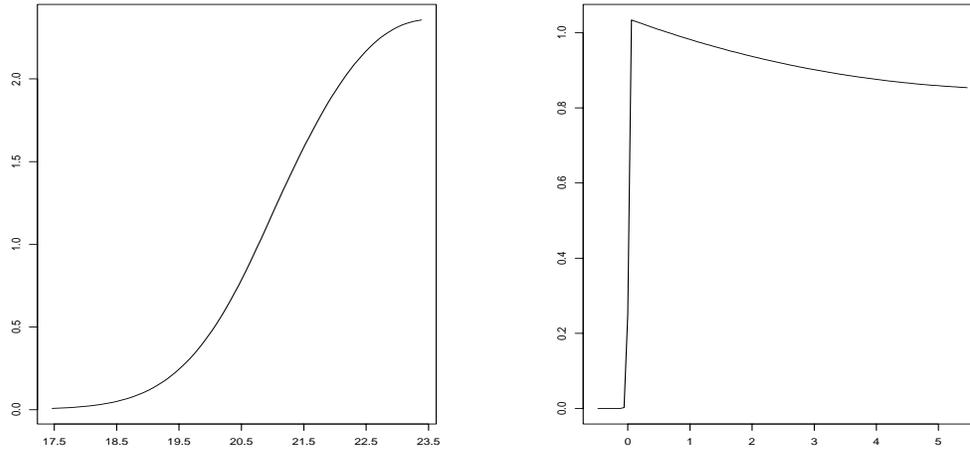


Figure 3.4 : Estimations du taux de hasard par logspline ($N = 512$).

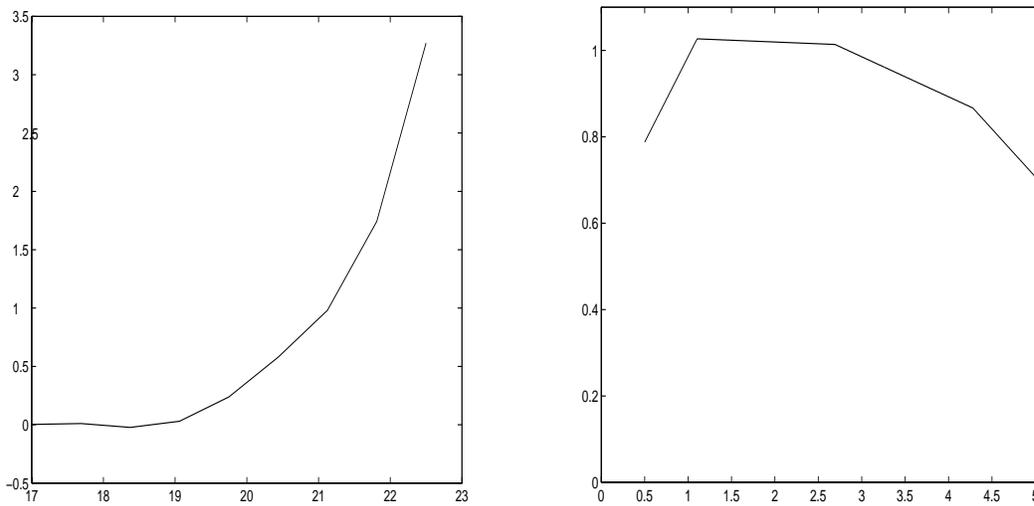


Figure 3.5 : Estimations du taux de hasard par ondelette ($N = 512\Gamma r = 5$ pour h_1 et $r = 4$ pour h_2).

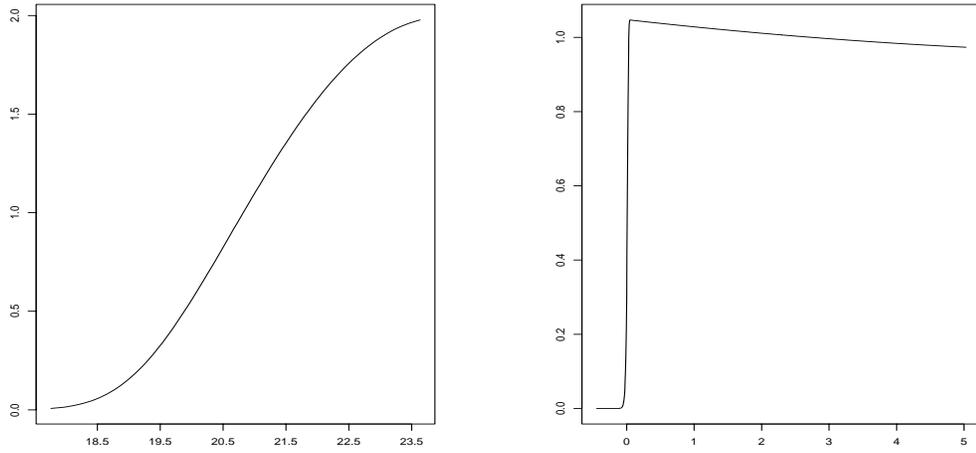


Figure 3.6 : Estimations du taux de hasard par logspline ($N = 1024$).

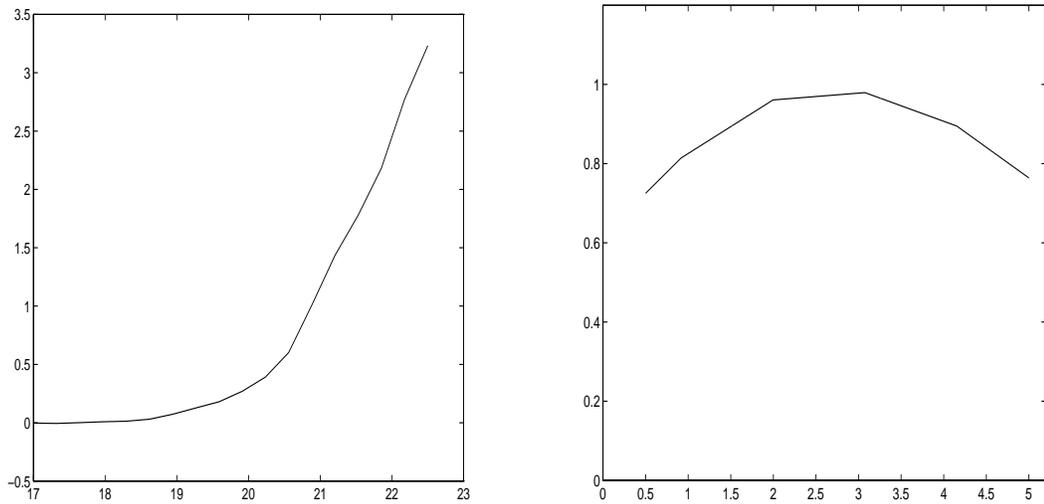


Figure 3.7 : Estimations du taux de hasard par ondelette ($N = 1024\Gamma$ $r = 5$ pour h_1 et $r = 4$ pour h_2).

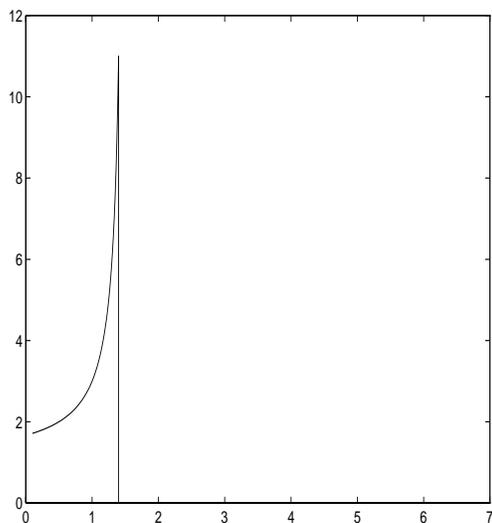


Figure 3.8 : La fonction $f(1 - F)^{-1}$ associée à une loi exponentielle de paramètre 1 censurée à droite par une loi uniforme sur $[0, 1.5]$ (50% des données sont censurées).

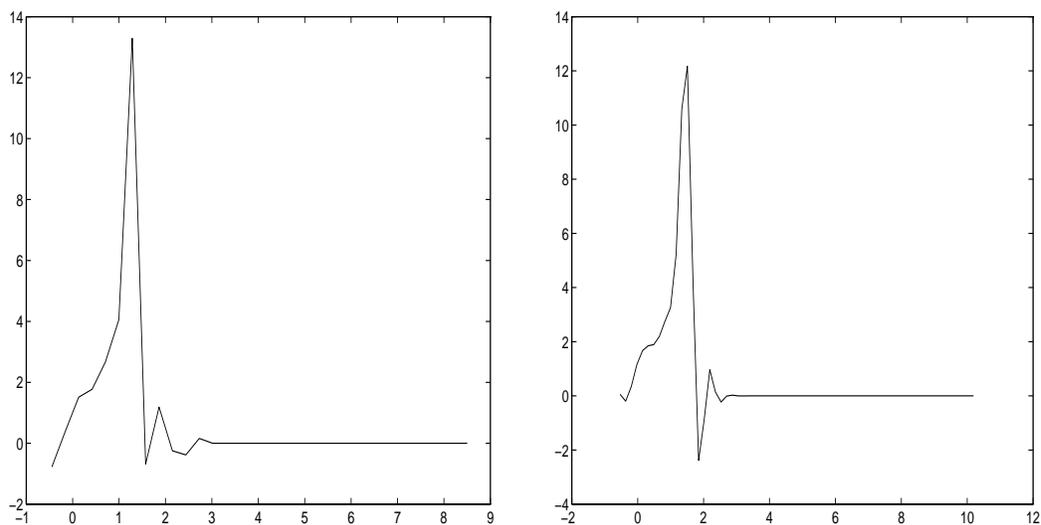


Figure 3.9 : Estimations de la fonction ci-dessus par ondelettes de Daubechies d'ordre 3 avec $N = 512$ et d'ordre 6 avec $N = 1024$ respectivement.

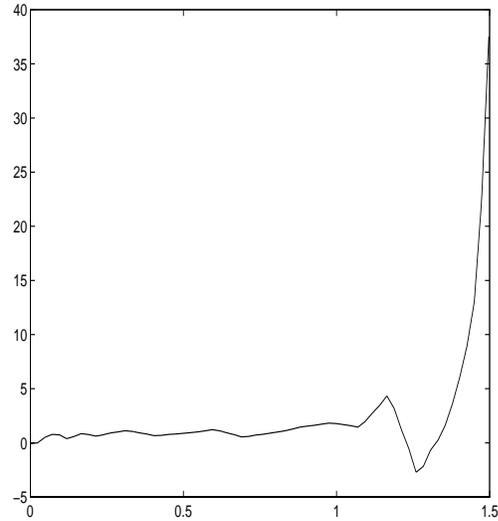


Figure 3.10 : Estimation par ondelettes de Daubechies d'ordre 3 avec $N = 1024$ du taux de hasard d'une loi exponentielle de paramètre 1 censurée à droite par une loi uniforme sur $[0, 1.5]$ ($\approx 50\%$ des données sont censurées).

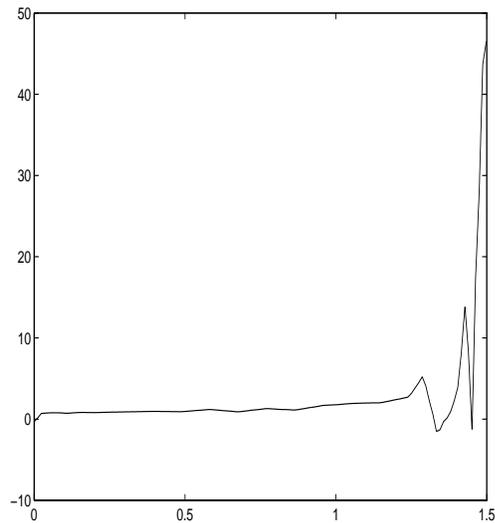


Figure 3.11 : Estimation par ondelettes de Daubechies d'ordre 3 avec $N = 2048$ du taux de hasard d'une loi exponentielle de paramètre 1 censurée à droite par une loi uniforme sur $[0, 1.5]$ ($\approx 52\%$ des données sont censurées).

Les échantillons simulés et qui ont servi à la détermination des graphiques (3.10) et (3.11) n'ont pas été tronqués ce qui produit un léger effet de bord dû essentiellement au fait que la fonction de survie est proche de 0 au point 1.5 .

Conclusion

L'estimateur non-linéaire de type ondelettes de la fonction de hasard a les 3 propriétés fondamentales requises dans tout estimateur et que peu d'estimateurs réunissent. Ces propriétés sont :

1. L'estimateur défini admet un comportement asymptotique au sens de Stone. Mieux encore le résultat obtenu dans le théorème généralise ceux obtenus par les méthodes de type noyau à des fonctions régulières par morceaux.
2. En pratique et comme le prouvent les simulations numériques l'estimateur défini est d'aussi bonne voire de meilleure qualité que l'estimateur logspline ce dernier étant une bonne référence en la matière. De plus l'estimateur étudié est toujours défini et sur tout l'intervalle des observations ce qui n'est pas toujours le cas de la méthode logspline qui peut diverger.
3. L'utilisation de l'algorithme pyramidal dans notre procédure d'estimation conduit à une évaluation de l'estimateur de type ondelettes largement plus rapide que celle utilisée pour les autres estimateurs.

Annexes

Annexe A

Rappels sur les paquets d'ondelettes

Les bases de paquets d'ondelettes ont été définies par R. Coifman et Y. Meyer [18] pour traiter et analyser des signaux. Ils ont ainsi construit une nouvelle famille de bases orthonormées de $L^2(\mathbb{R})$ conduisant à une nouvelle décomposition de $L^2(\mathbb{R})$ en somme directe de sous-espaces où chaque sous-espace est engendré par un paquet d'éléments de bases.

Dans cette annexe on rappellera le lien existant entre l'analyse multirésolution réelle, les filtres miroirs en quadrature et la construction d'une base orthonormée d'ondelettes de $L^2(\mathbb{R})$ puis on explicitera la construction des bases de paquets d'ondelettes.

On conclura alors par la construction de bases d'ondelettes et de paquets d'ondelettes de $L^2([0, 1])$ associés à une analyse multirésolution sur l'intervalle $[0, 1]$.

A.1 Introduction aux ondelettes

Bien que le système de Haar soit le plus vieil exemple de bases orthonormées d'ondelettes (1909) ça n'est que bien plus tard que l'analyse par ondelettes est implicitement introduite par A.P. Calderon ; elle sera redécouverte par le géophysicien J. Morlet dans un travail sur l'étude des signaux sismiques enregistrés durant des prospections pétrolières et explicitée par A. Grossmann.

La transformée en ondelettes est une représentation temps-échelle (ou position-échelle) dont le principe est de décrire l'évolution temporelle d'un

signal relativement à une échelle d'observation ; elle s'apparente à la transformée de Fourier mais les exponentielles complexes $\exp(ix\xi)$ indexées par les fréquences $\xi \in \mathbb{R}$ y sont remplacées par les ondelettes ψ_I indexées par la collection de tous les intervalles $I \subset \mathbb{R}$. Ces ondelettes sont toutes des copies d'une même fonction ψ régulière et ayant une forte décroissance à l'infini.

La transformée en ondelettes Γ comme la transformée de Fourier Γ permet par un choix judicieux de ψ d'écrire toute fonction f de $L^2(\mathbb{R})$ comme une somme de termes de la forme $c_I \psi_I$ où les coefficients c_I se calculent comme des coefficients de Fourier :

$$c_I = \langle f, \psi_I \rangle := \int_{\mathbb{R}} f(x) \psi_I(x) dx.$$

Il existe deux types de transformées en ondelettes :

1. La transformée discrète.
2. La transformée continue.

Étant donné que dans ce travail on utilise uniquement la transformée discrète on se limitera ci-après à ce cas.

Meyer et Lemarié [47] ont montré qu'il est possible de choisir l'ondelette ψ régulière et de carré intégrable Γ de sorte que la collection des $\psi_{jk} = 2^{\frac{j}{2}} \psi(2^j \cdot - k)$ Γ et k appartenant à \mathbb{Z} obtenues par translation et dilatation de ψ forme une base orthonormée de $L^2(\mathbb{R})$. Ainsi Γ toute fonction f de $L^2(\mathbb{R})$ s'écrit

$$f = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \langle f, \psi_{jk} \rangle \psi_{jk}.$$

Les coefficients c_{jk} Γ définis par

$$c_{jk} = \langle f, \psi_{jk} \rangle = \int f(x) \psi_{jk}(x) dx,$$

sont appelés coefficients d'ondelettes de f et la collection des c_{jk} est la transformée en ondelettes discrète de f .

La collection des intervalles I est dans ce cas réduite aux seuls intervalles dyadiques.

A.2 Analyse multirésolution (AMR) et ondelettes

La première construction de bases orthonormées d'ondelettes est due à Y. Meyer en 1985 Γ mais c'est surtout après que le lien entre la construction

d'une telle base et l'analyse multirésolution ait été établi que les choses ont évolué. On rappelle ci-après la définition d'une **AMR** réelle et comment on peut en déduire la construction d'une base orthonormée d'ondelettes de $L^2(\mathbb{R})$.

A.2.1 Analyse multirésolution

Définition A. 1. Une analyse multirésolution (ou multi-échelle) de $L^2(\mathbb{R})$ est une suite de sous-espaces fermés V_j de $L^2(\mathbb{R})$, $j \in \mathbb{Z}$, vérifiant les conditions suivantes :

1. $V_j \subset V_{j+1}$, $\forall j \in \mathbb{Z}$.

2. L'intersection des V_j est réduite à 0 et leur réunion est dense dans $L^2(\mathbb{R})$, i.e :

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\} \text{ et } \overline{\bigcup_{j \in \mathbb{Z}} V_j} = L^2(\mathbb{R}). \quad (\text{A.1})$$

3. Pour toute fonction f dans $L^2(\mathbb{R})$,

$$f \in V_j \iff f(2 \cdot) \in V_{j+1}. \quad (\text{A.2})$$

4. Il existe une fonction φ dans V_0 , dite fonction d'échelle ou fonction père, telle que $\{\varphi_{0k}, k \in \mathbb{Z}\}$ est une base inconditionnelle de V_0 , avec $\varphi_{0k} = \varphi(\cdot - k)$.

Rappelons qu'une famille (e_j) est une base inconditionnelle d'un espace de Hilbert \mathcal{H} si et seulement et si les deux propriétés suivantes sont satisfaites :

- il existe deux constantes C_1 et C_2 strictement positives telles que pour toute suite finie de scalaires réels (α_j) :

$$C_1 \left(\sum |\alpha_j|^2 \right)^{\frac{1}{2}} \leq \left\| \sum \alpha_j e_j \right\|_2 \leq C_2 \left(\sum |\alpha_j|^2 \right)^{\frac{1}{2}}.$$

- l'ensemble des combinaisons linéaires finies $\sum \alpha_j e_j$ est dense dans \mathcal{H} .

Remarque A. 1. :

1. Au vu de (A.2), pour tout j , on a :

$$\{\varphi_{jk}, k \in \mathbb{Z}\} \text{ est une base inconditionnelle de } V_j, \quad (\text{A.3})$$

avec $\varphi_{jk} = 2^{\frac{j}{2}} \varphi(2^j \cdot - k)$.

2. Chaque espace V_j est invariant par les translations de pas $k2^{-j}$,
i.e $f \in V_j \rightarrow f(\cdot - k2^{-j}) \in V_j, \forall k \in \mathbb{Z}$.

Une analyse ainsi définie permet d'écrire toute fonction f de $L^2(\mathbb{R})$ comme limite d'une suite de fonctions f_j où chaque f_j appartient à V_j . En effet soit P_j le projecteur orthogonal sur V_j et soit f_j la projection orthogonale de f sur V_j (i.e $f_j = P_j(f)$) par les deux premières conditions de la **Définition A.1** On a :

$$\lim_{j \rightarrow \infty} f_j = f.$$

A.2.2 Ondelettes

La construction de la fonction ondelette ψ est étroitement liée à la définition d'une suite de sous-espaces de $L^2(\mathbb{R})$ dits de détails. En effet par la définition même de l'analyse multirésolution il est naturel de construire une suite $W_j, j \in \mathbb{Z}$ telle que chaque W_j soit le supplémentaire orthogonal de V_j dans V_{j+1} c'est à dire

$$V_{j+1} = V_j \oplus W_j. \quad (\text{A.4})$$

Ainsi par (A.1) On a

$$L^2(\mathbb{R}) = \bigoplus_{j \in \mathbb{Z}} W_j \quad (\text{A.5})$$

Soit Q_j le projecteur orthogonal sur W_j par (A.4) et (A.2) on en déduit que

$$Q_j = P_{j+1} - P_j, \quad (\text{A.6})$$

et que W_j est une version de W_0 au niveau j de la multirésolution :

$$f \in W_0 \iff f(2^j \cdot) \in W_j. \quad (\text{A.7})$$

S. Mallat et Y. Meyer voir [22] montrent alors l'existence d'une fonction ψ appelée ondelette mère dont les translations par les entiers engendrent W_0 . Tout comme V_j W_j est alors engendré par les $\psi_{jk} = 2^{\frac{j}{2}} \psi(2^j \cdot - k)$ les versions obtenues par dilatation et translation de ψ . Ainsi par (A.5) la famille $\{\psi_{jk}, j \in \mathbb{Z}, k \in \mathbb{Z}\}$ est une base inconditionnelle de $L^2(\mathbb{R})$.

Au paragraphe A.3 On donnera la construction explicite de ψ .

A.2.3 Remarque

Sans perte de généralité on peut choisir la fonction d'échelle φ telle que $\{\varphi_{0k}, k \in \mathbb{Z}\}$ soit une base orthonormée de V_0 plutôt qu'une base inconditionnelle. En effet il est possible par l'équation ci-après :

$$\hat{\tilde{\varphi}}(\xi) = \hat{\varphi}(\xi) \left[\sum |\hat{\varphi}(\xi + 2k\pi)|^2 \right]^{-\frac{1}{2}} \quad (\text{A.8})$$

de construire $\tilde{\varphi}$ dont les translations par les entiers forment une base orthonormée de V_0 ($\hat{\tilde{\varphi}}(\xi)$ et $\hat{\varphi}(\xi)$ ci-dessus désignent les transformées de Fourier respectives de $\tilde{\varphi}$ et φ).

La fonction d'échelle possède alors les propriétés suivantes :

$$\int_{\mathbf{R}} \tilde{\varphi}(x) dx = \hat{\tilde{\varphi}}(0) = 1. \quad (\text{A.9})$$

$$\sum_{k \in \mathbb{Z}} |\hat{\tilde{\varphi}}(\xi + 2k\pi)|^2 = 1. \quad (\text{A.10})$$

Dès lors on supposera dans tout ce qui suit que $\{\varphi_{0k}, k \in \mathbb{Z}\}$ est une base orthonormée de V_0 . L'AMR est alors dite orthogonale.

Exemple d'AMR :

Un exemple d'analyse multirésolution est donné par la suite des :

$$V_j = \{g \in L^2(\mathbf{R}) : g \text{ constante sur } [2^{-j}k, 2^{-j}(k+1)[, \forall k \in \mathbb{Z}\},$$

pour laquelle les trois premières conditions de la **Définition A.1** sont trivialement vérifiées.

Pour tout j dans \mathbb{Z} P_j est alors défini par :

$$\begin{aligned} P_j(f) &= \sum_{k \in \mathbb{Z}} (2^j \int_{2^{-j}k}^{2^{-j}(k+1)} f(x) dx) \mathbb{1}_{[2^{-j}k, 2^{-j}(k+1)[} \\ &= \sum \langle \varphi_{jk}, f \rangle \varphi_{jk}. \end{aligned}$$

En choisissant comme fonction d'échelle l'indicatrice de $[0, 1[$ on dispose d'une base orthonormée pour V_0 ce qui prouve que les V_j forment bien une **AMR** ; à cette dernière on associe alors la fonction de Haar comme ondelette mère. Ceci est facile à vérifier en utilisant (A.6) et la définition de W_j . En effet

$$\varphi_{j-1,k} = \frac{1}{\sqrt{2}} \{\varphi_{j,2k} + \varphi_{j,2k+1}\}.$$

Il s'ensuit que W_j est engendré par les $\varphi_{j,2k} - \varphi_{j,2k+1}$ et qu'un choix judicieux de ψ est donné par :

$$\psi = \varphi(2 \cdot) - \varphi(2 \cdot - 1) = \mathbb{1}_{[0, \frac{1}{2}[} - \mathbb{1}_{[\frac{1}{2}, 1[}.$$

Exemples d'ondelettes :

Il existe différentes familles d'ondelettes caractérisées par des propriétés sur le support, la symétrie, l'orthogonalité ou la régularité de la fonction

d'échelle et de l'ondelette mère. Dans la famille des ondelettes à support compact Γ on s'intéressera ci-après à celle construite par I. Daubechies [22] Γ $L\varphi$ ΓL étant un entier naturel non nul Γ possédant les propriétés suivantes :

- $L\varphi$ est à support compact $[0, 2L - 1]$.
- $L\varphi$ est μL -régulière i.e $\int_{\mathbb{R}} |\xi|^{\mu L} |L\hat{\varphi}(\xi)| d\xi < \infty$ avec $\mu \approx 0.1936$.
- L'ondelette associée admet des moments nuls à l'ordre $L\Gamma$

$$\forall k = 0, 1, \dots, L, \int_{-\infty}^{\infty} x^k \psi(x) dx = 0.$$

Parmi toutes les ondelettes à support compact Γ vérifiant la troisième propriété ci-dessus Γ les $L\varphi$ sont celles qui ont le plus petit support ; de fait elles sont dites à support minimal. La construction de telles ondelettes est longuement détaillée par I. Daubechies [22] et est partiellement reprise dans l'exemple **A.3.1**.

Un autre exemple d'ondelettes à support compact est celui des coiffets. Elles sont définies de manière à ce que la fonction d'échelle admette Γ tout comme l'ondelette mère Γ des moments nuls à l'ordre $L - 1$ et ce au détriment d'un support plus large que celui de $L\varphi$.

Dans tout ce qui suit Γ on dira qu'une fonction φ génère une **AMR** réelle si Γ en définissant par V_j l'espace engendré par $\{2^{\frac{j}{2}}\varphi(2^j \cdot -k), k \in \mathbb{Z}\}$ Γ la suite $(V_j)_{j \in \mathbb{Z}}$ vérifie les quatre conditions de la **Définition A.1**.

A.3 Filtres miroirs en quadrature (QMF) et ondelettes

En pratique Γ il est plus facile de définir une **AMR** en partant d'une fonction $\varphi \in V_0$ et en exploitant le fait que $V_{-1} \subset V_0$. Ce qui se traduit par l'existence d'une séquence (h_k) telle que

$$\frac{1}{\sqrt{2}}\varphi\left(\frac{x}{2}\right) = \sum_k h_k \varphi(x - k) \quad (\text{A.11})$$

$$\text{où } h_k = \int_{\mathbb{R}} \frac{1}{\sqrt{2}}\varphi\left(\frac{x}{2}\right)\varphi(x - k) dx, \quad (\text{A.12})$$

i.e : $h_k = \langle \varphi_{-1,0}, \varphi_{0k} \rangle$.

Partant de la remarque ci-dessus on aimerait imposer à φ et donc à la séquence h_k des conditions suffisantes pour aboutir à une **AMR**.

A cette fin les liens établis par S. Mallat et Y. Meyer entre l'analyse multirésolution et les algorithmes pyramidaux (voir [22]) montrent que le calcul numérique de la transformée en ondelettes fait appel à la donnée unique d'un filtre miroir en quadrature **QMF** au sens défini par Smith et Barnwell [5] (voir ci-après). Réciproquement s'il satisfait certaines conditions un **QMF** permet de générer une **AMR**.

Soit $(h_k)_{k \in \mathbb{Z}}$ un filtre discret quelconque suite réelle indexée par \mathbb{Z} sa fonction de transfert définie par

$$m_0(\xi) := \frac{1}{\sqrt{2}} \sum_{k \in \mathbb{Z}} h_k e^{-ik\xi}, \quad (\text{A.13})$$

et soit m_1 la fonction de transfert associée au filtre $(g_k)_{k \in \mathbb{Z}}$ définie par $g_k = (-1)^k h_{1-k}$ le filtre conjugué de h_k . On a

$$m_1(\xi) := e^{-i\xi} \overline{m_0(\xi + \pi)}. \quad (\text{A.14})$$

Définition A. 2. On désigne par "filtres **QMF**" les fonctions 2π -périodiques m_0 et m_1 , définis ci-dessus, vérifiant

$$|m_0(\xi)|^2 + |m_1(\xi)|^2 = 1, \quad (\text{A.15})$$

$$i.e : |m_0(\xi)|^2 + |m_0(\xi + \pi)|^2 = 1. \quad (\text{A.16})$$

Cette définition établie on va montrer comment la donnée d'une **AMR** permet de définir un **QMF** et on va donner la construction explicite de l'ondelette mère.

Pour ce considérons une **AMR** et soit φ sa fonction d'échelle vérifiant (A.11). En appliquant la transformée de Fourier à (A.11) on a

$$\hat{\varphi}(2\xi) = m_0(\xi) \hat{\varphi}(\xi). \quad (\text{A.17})$$

Ainsi en introduisant la relation (A.17) dans (A.10) où on remplace $\tilde{\varphi}$ par φ on obtient

$$\sum_{k \in \mathbb{Z}} |\hat{\varphi}(2\xi + 2k\pi)|^2 = \sum_{k \in \mathbb{Z}} (|m_0(\xi)|^2 + |m_0(\xi + \pi)|^2) |\hat{\varphi}(\xi + 2k\pi)|^2.$$

De la sorte on montre que (m_0, m_1) défini par la suite h_k elle même définie par (A.12) est un **QMF**.

D'autre part il est prouvé voir Daubechies [22] par exemple que si ψ est solution de l'équation suivante

$$\hat{\psi}(2\xi) = m_1(\xi)\hat{\varphi}(\xi), \quad (\text{A.18})$$

alors $\{\psi(\cdot - k), k \in \mathbb{Z}\}$ est une base orthonormée de W_0 . La construction de l'ondelette mère est alors obtenue par la relation suivante

$$\frac{1}{\sqrt{2}}\psi\left(\frac{x}{2}\right) = \sum_k g_k \varphi(x - k). \quad (\text{A.19})$$

Réciproquement sous quelles conditions peut-on construire une **AMR** à partir d'un **QMF**?

Avant de répondre à cette question il est intéressant de relever que l'itération de (A.17) produit l'identité suivante

$$\hat{\varphi}(\xi) = \prod_{k=1}^{+\infty} m_0\left(\frac{\xi}{2^k}\right). \quad (\text{A.20})$$

[le produit converge si $m_0(0) = 1$ et m_0 est régulière (les coefficients de Fourier de m_o sont à décroissance rapide) donc lipschitzienne en 0].

La fonction d'échelle est alors déterminée en calculant le membre de droite de (A.20) et en lui appliquant la transformée inverse de Fourier. On en déduit alors l'ondelette mère par (A.18).

Le problème évoqué ci-dessus a été initialement soulevé par S. Mallat sous la forme suivante :

Soit une fonction m_o régulière et 2π -périodique vérifiant $m_0(0) = 1$ et $|m_0(\xi)|^2 + |m_0(\xi + \pi)|^2 = 1$; la fonction φ solution de l'identité (A.20) nous permet-elle de générer une **AMR**?

La réponse est non des contre-exemples ayant été exhibés. A. Cohen[16] cependant établit en exigeant plus de ce filtre miroir en quadrature des conditions nécessaires et suffisantes sur la fonction m_0 permettant de générer une **AMR**.

A.3.1 Exemple

Un exemple particulier d'ondelettes construites à partir d'une suite de coefficients est celui des ondelettes à support compact définies dans le deuxième paragraphe de cette annexe ; le principe est simple : on part d'un entier L et de $2L$ coefficients $h_0, h_1, \dots, h_{2L-1}$ tels que le polynôme trigonométrique

$$m_0(\xi) = h_0 + h_1 e^{i\xi} + \dots + h_{2L-1} e^{i(2L-1)\xi}$$

vérifie en plus de (A.16) la condition $m_0(0) = 1$. Ces deux conditions assurent l'existence d'une suite orthonormée $\varphi(\cdot - k)$, $k \in \mathbb{Z}$ de fonctions de $L^2(\mathbb{R})$ telle que

$$\frac{1}{\sqrt{2}}\varphi\left(\frac{x}{2}\right) = h_0\varphi(x) + h_1\varphi(x-1) + \dots + h_{2L-1}\varphi(x-(2L-1)), \quad (\text{A.21})$$

et de plus :

$$\int_{-\infty}^{\infty} \varphi(x) dx = 1.$$

Par (A.21) on construit φ et on montre que son support est inclus dans l'intervalle $[0, 2L-1]$ et par (A.19) on construit l'ondelette mère. La fonction d'échelle (resp. ondelette mère) obtenue dans cet exemple est notée par $L\varphi$ (resp. $L\psi$) et est désignée par fonction d'échelle de Daubechies (resp. ondelette de Daubechies) à support minimal.

A.4 Base de paquets d'ondelettes

A.4.1 Aspect séparatif des QMF

L'interprétation qu'on peut faire des **QMF** diffère selon l'usage prévu ; dans ce qui suit on abordera l'aspect "séparatif" de tels filtres i.e : un **QMF** produit systématiquement une dichotomie de tout espace Hilbertien \mathcal{H} séparable et de dimension infinie muni d'une base orthonormée $\{e_k, k \in \mathbb{Z}\}$. Une dichotomie simple est donnée par $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ où \mathcal{H}_1 est engendré par (e_{2k}) et \mathcal{H}_2 est engendré par (e_{2k+1}) .

Dans ce qui suit on va s'intéresser au cas général. On va montrer que la donnée d'un **QMF** est équivalente à la décomposition de \mathcal{H} en somme directe de deux sous espaces \mathcal{H}_1 et \mathcal{H}_2 où chaque \mathcal{H}_i est muni d'une base orthonormée induite de la suite (e_k) .

Pour ce supposeons que l'on dispose de deux suites dans $l^2(\mathbb{Z})$ en l'occurrence (u_l) et (v_l) et que l'on construise une suite (f_k) définie par :

$$\begin{cases} f_{2k} & = \sum_l u_{2k-l} e_l \\ f_{2k+1} & = \sum_l v_{2k-l} e_l. \end{cases} \quad (\text{A.22})$$

Posons \mathcal{H}_1 l'adhérence du sous espace engendré par $\{f_{2k}, k \in \mathbb{Z}\}$ et \mathcal{H}_2 celui engendré par $\{f_{2k+1}, k \in \mathbb{Z}\}$.

Quelles conditions faut-il imposer à (u_l) et (v_l) pour que les deux propositions ci-après soient vraies?

- f_k est une base orthonormée de \mathcal{H} .

- \mathcal{H}_1 est le supplémentaire orthogonal de \mathcal{H}_2 (i.e : $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$).

Il se trouve que la deuxième proposition est une conséquence de la première et le résultat suivant apporte une solution (voir Coifman et al [18]) :
soit m_0 et m_1 définies par

$$\begin{cases} m_0(\xi) &= \frac{1}{\sqrt{2}} \sum_l u_l e^{-il\xi} \\ m_1(\xi) &= \frac{1}{\sqrt{2}} \sum_l v_l e^{-il\xi}. \end{cases} \quad (\text{A.23})$$

Proposition A. 1. *Les 3 propositions suivantes sont équivalentes :*

1. $(f_k)_{k \in \mathbb{Z}}$ est un système orthonormé dans \mathcal{H} .
2. $(f_k)_{k \in \mathbb{Z}}$ est une base orthonormée de \mathcal{H} .

3.

$$\forall \xi \in [0, 2\pi) : S(\xi) = \begin{pmatrix} m_0(\xi) & m_1(\xi) \\ m_0(\xi + \pi) & m_1(\xi + \pi) \end{pmatrix}$$

est une matrice unitaire.

L'exemple cité en début de paragraphe correspond à $m_0(\xi) = \frac{1}{\sqrt{2}}$ et $m_1(\xi) = \frac{\exp -i\xi}{\sqrt{2}}$.

La fonction m_0 (resp. m_1) est la fonction de transfert du filtre F_0 (resp. F_1) associée à la séquence (u_l) (resp. (v_l)). Les deux filtres F_0 et F_1 étant définis comme suit :

$$\forall \alpha \in l^2(\mathbb{Z}) : (F_0\alpha)_k = \sum_n u_{n-k} \alpha_n$$

$$\text{et } (F_1\alpha)_k = \sum_n v_{n-k} \alpha_n.$$

Définissons par D l'opérateur de décimation tel que :

$$(D\alpha)_k = \alpha_{2k},$$

et soit $(\beta_k)_{k \in \mathbb{Z}}$ la suite définie par

$$(\beta_{2k}) = D(F_0[\alpha])$$

$$\text{et } (\beta_{2k+1}) = D(F_1[\alpha]).$$

Si les filtres F_0 et F_1 définis ci-dessus sont tels que

$$\sum \alpha_k e_k = \sum \beta_{2k} f_{2k} + \sum \beta_{2k+1} f_{2k+1}, \quad (\text{A.24})$$

(F_0, F_1) est alors appelé par abus de langage “filtres miroirs en quadrature” ou simplement **QMF** ; (A.24) étant équivalente à la condition imposée dans la définition A.2 aux fonctions de transfert.

Il est ainsi possible de remplacer la condition sur la matrice S dans la proposition A.1 par la condition que (F_0, F_1) est un **QMF**.

En ce qui concerne l'analyse multirésolution définie au paragraphe A.2. On dispose de deux bases orthonormées de V_j étant fixé. La première est donnée par $e_k = \varphi_{jk}, k \in \mathbb{Z}$, et la seconde est (f_k) telle que $f_{2k} = \varphi_{j-1,k}$ l'espace engendré par cette suite est V_{j-1} et $f_{2k+1} = \psi_{j-1,k}$ l'espace engendré par cette suite est W_{j-1} . Ces deux bases sont liées par (A.22) via l'utilisation des suites (h_k) et (g_k) définies dans le paragraphe A.3 voir (A.11) et (A.19). Le **QMF** associé est alors noté (H, G) . Ainsi W_{j-1} est le supplémentaire orthogonal de V_{j-1} dans V_j .

A.4.2 Paquets d'ondelettes

Construction de bases de paquets d'ondelettes :

On donne dans ce qui suit une nouvelle famille de bases orthonormées de $L^2(\mathbb{R})$ dont la construction est due à Coifman et Meyer [18].

Considérons le **QMF** (H, G) associé à une **AMR** et posons $\omega_0 = \varphi$ et $\omega_1 = \psi$.

L'application du schéma récursif suivant :

$$\begin{cases} \omega_{2m} &= \sqrt{2} \sum_{-\infty}^{\infty} h_k \omega_m(2 \cdot -k) \\ \omega_{2m+1} &= \sqrt{2} \sum_{-\infty}^{\infty} g_k \omega_m(2 \cdot -k), \end{cases} \quad (\text{A.25})$$

ou son équivalent

$$\begin{cases} \omega_{2m}(\cdot - l) &= H\{\omega_m(2 \cdot -k)\}(l) \\ \omega_{2m+1}(\cdot - l) &= G\{\omega_m(2 \cdot -k)\}(l), \end{cases} \quad (\text{A.25})'$$

permet de définir une suite de fonctions orthonormées $\omega_m, m \in \mathbb{Z}$.

Il est à noter qu'en prenant $m = 0$ dans l'équation (A.25) φ est l'unique solution de (A.11) telle que (A.9) soit vérifiée ce qui justifie le choix de ω_0 et par suite celui de ω_1 (voir A.19). D'autre part on peut réindexer les différents ω en posant dans un premier temps $E = \{0, 1\}^{\mathbb{N}}$ et E_j le sous-ensemble de E tel que :

$$\forall \epsilon \in E_j, 0 = \epsilon_{j+1} = \epsilon_{j+2} = \epsilon_{j+3} = \dots$$

Ainsi si $m = \epsilon_1 + 2\epsilon_2 + \dots + 2^j \epsilon_j \Gamma$ alors $\omega_m = \omega_\epsilon$ avec ϵ dans $E_j \Gamma$ et sa transformée de Fourier est alors donnée par l'équation :

$$\hat{\omega}_\epsilon(\xi) = \prod_{l=1}^j m_{\epsilon_l} \left(\frac{\xi}{2^l} \right) \hat{\varphi} \left(\frac{\xi}{2^j} \right).$$

Ceci est l'analogie de l'équation (A.17) (resp. (A.18)) que vérifie φ (resp. ψ).

On conclut cette construction par l'énoncé du théorème A.1.

Théorème A. 1 (Coifman et Meyer) *Pour tout j dans \mathbb{N} ,*

$$\{\omega_\epsilon(x - k), (\epsilon, k) \in E_j * \mathbb{Z}\} \text{ est un système orthonormé de } V_j.$$

*Il s'ensuit que $\{\omega_m(x - k), (m, k) \in \mathbb{N} * \mathbb{Z}\}$ est une base orthonormée de $L^2(\mathbb{R})$.*

Décomposition en paquets d'ondelettes :

La construction de bases de paquets d'ondelettes dans $L^2(\mathbb{R})$ conduit à une décomposition de ce dernier en somme directe de sous espaces fermés. Cette décomposition sera tout au long de ce cette annexe désignée par décomposition en paquets d'ondelettes. Dans ce qui suit nous allons établir le principe d'une telle décomposition.

Soit $(V_j)_{j \in \mathbb{Z}}$ une **AMR** réelle et soit φ et ψ la fonction d'échelle et l'ondelette mère associées. Rappelons que chaque V_j est engendré par $\{\varphi_{jk} = 2^{\frac{j}{2}} \varphi(2^j \cdot -k), k \in \mathbb{Z}\}$ et que l'application du **QMF** (H, G) associée à l'**AMR** conduit à la décomposition suivante

$$V_j = V_{j-1} \oplus W_{j-1},$$

où V_{j-1} est engendré par $\{\varphi_{j-1,k} = 2^{\frac{j-1}{2}} \varphi(2^{j-1} \cdot -k), k \in \mathbb{Z}\}$ et W_{j-1} est engendré par $\{\psi_{j-1,k} = 2^{\frac{j-1}{2}} \psi(2^{j-1} \cdot -k), k \in \mathbb{Z}\}$.

Alors que la décomposition en ondelettes se contente de décomposer les espaces d'approximations V_j et garde intacts les espaces de détails W_j la décomposition en paquets d'ondelettes opère à la fois sur les espaces de détails et sur les espaces d'approximations tout en utilisant le même filtre (H, G) .

Ainsi

$$V_j = V_{j-1} \oplus W_{j-1},$$

$$\begin{cases} V_{j-1} = V_{j-2} \oplus W_{j-2} \\ W_{j-1} = V_{j-2,2} \oplus V_{j-2,3}. \end{cases}$$

Ce qui est équivalent à Γ

1. V_{j-2} (resp. W_{j-2}) est engendré par les $2^{\frac{j-2}{2}} \omega_0(2^{j-2} \cdot -k)$ (resp. $2^{\frac{j-2}{2}} \omega_1(2^{j-2} \cdot -k)$) $\Gamma k \in \mathbb{Z}$.
2. $V_{j-2,2}$ est engendré par $2^{\frac{j-2}{2}} \omega_2(2^{j-2} \cdot -k)$, $k \in \mathbb{Z}$.
3. $V_{j-2,3}$ est engendré par $2^{\frac{j-2}{2}} \omega_3(2^{j-2} \cdot -k)$, $k \in \mathbb{Z}$.

Posons dans tout ce qui suit $V_{l,0} = V_l$ et $V_{l,1} = W_l \Gamma$ pour tout l dans \mathbb{Z} . Pour mieux illustrer la décomposition en paquets d'ondelettes associée à la construction des bases définies par la formule de récurrence (A.25) on donne le schéma suivant : soit $j \in \mathbb{Z}$

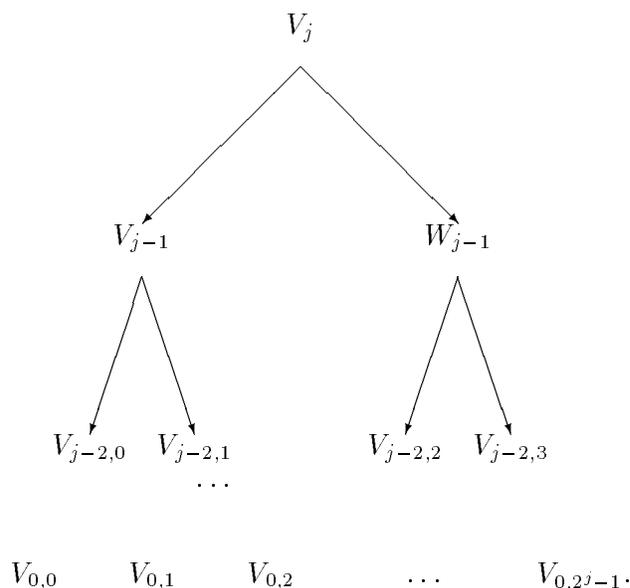


Figure A.1 : Décomposition en paquets d'ondelettes à l'ordre j de V_j .

Notons que chaque V_{0m} est engendré par $\{\omega_m(\cdot - k), k \in \mathbb{Z}\} \Gamma \omega_m$ étant défini par (A.25). Ainsi le **Théorème A.1** est équivalent à l'écriture suiv-

ante

$$V_j = V_{0,0} \oplus V_{0,1} \oplus \dots \oplus V_{0,2^j-1}.$$

Ce schéma décrit de manière claire la différence entre la décomposition en ondelettes est celle en paquets d'ondelettes. La première consiste à décomposer V_n en V_{n-1} et W_{n-1} puis à réitérer cette procédure uniquement pour V_{n-1} puis V_{n-2} jusqu'à l'obtention de V_0 alors que la deuxième applique cette décomposition à V_{n-1} mais aussi à W_{n-1} et redécompose chacun des sous espaces obtenus en utilisant le même **QMF** celui associé à l'**AMR**. Afin de bien comprendre cette différence on rappelle le schéma correspondant à la décomposition en ondelettes

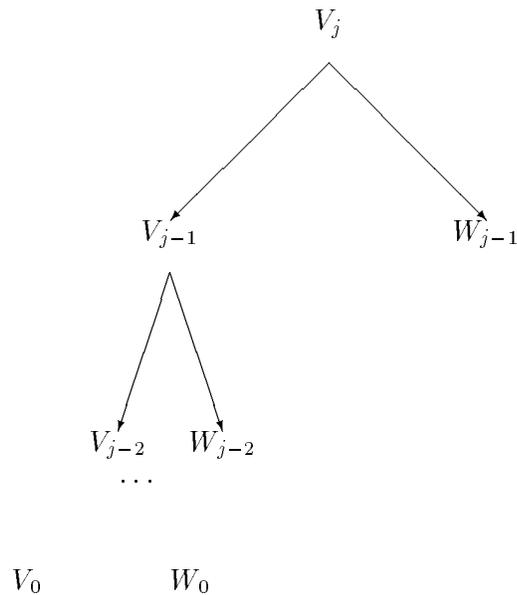


Figure A.2 : Décomposition en ondelettes à l'ordre j de V_j .

A.5 Analyse multirésolution sur l'intervalle

A.5.1 Ondelettes sur l'intervalle

Tout ce qui a été introduit ci-dessus concerne des espaces de fonctions à supports réels non réduits à un intervalle. Cependant dans la plupart des cas pratiques le support de la fonction étudiée coïncide avec un intervalle fermé et borné ($[0, 1]$ par exemple). L'utilisation d'un sous-système de base

déduit d'une **AMR** réelle dégrade considérablement l'estimation de la fonction aux bords de son support ; pour remédier à ce problème Γ a été conçu l'**AMR** sur l'intervalle.

Avant d'atteindre cet objectif Γ différentes solutions ont été proposées dans l'optique d'adapter la multirésolution réelle à l'intervalle $[0, 1]$.

Une première solution consiste à considérer que toute fonction à support sur $[0, 1]$ peut être prolongée à \mathbb{R} tout entier en posant $f(x) = 0$ pour $x \notin [0, 1]$ Γ permettant ainsi l'utilisation d'une **AMR** réelle. Cependant un tel prolongement produit une discontinuité aux points 0 et 1.

Une deuxième solution se résume à définir des versions périodisées d'ondelettes à partir des φ_{jk} et ψ_{jk} Γ et d'analyser la fonction définie sur $[0, 1]$ par de telles ondelettes. Ceci équivaut à prolonger f en une fonction périodique de période 1 à support réel Γ et à utiliser une **AMR** réelle.

Chaque fois que f n'est pas périodique Γ cette méthode hérite du problème de discontinuité déjà évoqué pour la première solution et ce malgré le fait que les versions ondelettes périodisées déduisent une **AMR** sur l'intervalle $[0, 1]$ et constituent une base orthonormée de $L^2([0, 1])$.

En 1991 Γ Y. Meyer [53] construisit une famille de bases orthonormées d'ondelettes dans $L^2([0, 1])$ Γ possédant des propriétés de régularité identiques à celles des ondelettes de Daubechies à support compact. Ces bases d'ondelettes sont associées à une **AMR** sur l'intervalle $[0, 1]$.

Sa construction conduit à la définition d'un espace d'approximation $V_j^{[0,1]}$ Γ engendré par $2^j + 2L - 2$ éléments de base Γ et de son supplémentaire orthogonal dans $V_{j+1}^{[0,1]}$ Γ $W_j^{[0,1]}$ Γ engendré Γ quant à lui Γ par exactement 2^j éléments de base ; du fait que ces deux sous-espaces n'ont pas la même dimension mais aussi du fait que le nombre de fonctions d'échelles à chaque résolution n'est pas forcément une puissance de 2 Γ l'utilisation de la décomposition en paquets d'ondelettes sur l'intervalle n'est pas possible. D'autre part Γ sa construction nécessite la résolution d'un système mal conditionné.

Pour pallier à ces deux problèmes A. Cohen Γ I. Daubechies et P. Vial [17] ont récemment introduit une nouvelle approche pour la construction de bases orthonormées en ondelettes associées à une multirésolution sur l'intervalle.

Le point de départ est une fonction φ Γ du type Daubechies de support $[-L, L-1]$ (on obtient une telle fonction par simple translation de ${}_L\varphi$ définie dans A.3.1) ; ils définissent alors L fonctions de bord φ_k^g , $k = 0, 1 \dots L-1$ pour l'extrémité gauche 0 et autant de fonctions φ_k^d , $k = -1 \dots -L$ pour l'autre extrémité 1 ; chacune des fonctions φ_k^g (resp. φ_k^d) a pour support l'intervalle $[0, L+k]$ (resp. $[k-L+1, 0]$).

La construction de ces fonctions de bord est faite de manière à préserver

toutes les propriétés des ondelettes de Daubechies qui leur sont associés dont notamment le fait qu'elles engendrent les polynômes d'ordre $L - 1$. Ainsi en définissant :

$$\varphi_{jk}^d = 2^{\frac{j}{2}} \varphi_k^d(2^j \cdot), \quad k = -1, \dots, -L,$$

$$\text{et } \varphi_{jk}^g = 2^{\frac{j}{2}} \varphi_k^g(2^j \cdot), \quad k = 0, 1 \dots L - 1.$$

Il est montré dans [17] que pour tout j tel que $2^j \geq 2L\Gamma$ les fonctions d'échelles dites fonctions intérieures φ_{jk} , $k = L, \dots, 2^j - L + 1\Gamma$ déterminent avec $\varphi_{j,2^j-k}^1 = \varphi_{j,-k}^d(\cdot - 1)$, $k = 1 \dots L\Gamma$ et $\varphi_{jk}^0 = \varphi_{jk}^g(\cdot)$, $k = 0, \dots, L - 1\Gamma$ un système orthonormé à 2^j éléments et l'espace engendré par ces derniers est noté $V_j^{[0,1]}$.

D'autre part la structure de multirésolution ($V_j \subset V_{j+1}$) est conservée et ceci est dû au fait que par (A.11) et (A.19) :

$$\forall k = L, \dots, 2^j - L + 1\Gamma$$

$$\begin{cases} \frac{1}{\sqrt{2}} \varphi(2^j \cdot - k) &= \sum_{l=2k-L+1}^{L+2k} h_{l-2k+L-1} \varphi(2^{j+1} \cdot - l) \\ \frac{1}{\sqrt{2}} \psi(2^j \cdot - k) &= \sum_{l=2k-L+1}^{L+2k} g_{l-2k+L-1} \varphi(2^{j+1} \cdot - l) \end{cases} \quad (\text{A.26})$$

et qu'il est en de même pour φ_{jk}^0 et ψ_{jk}^0 $k = 0, \dots, L - 1\Gamma$ et $\varphi_{j,2^j-k}^1$ et ψ_{jk}^1 $k = 2^j - L, \dots, 2^j - 1$. En effet des coefficients $H_{k,l}^i$ $G_{k,l}^i$ $h_{k,m}^i$ et $g_{k,m}^i$ ($i = 0, 1$) sont calculés de sorte que :

$$\begin{cases} \varphi_{jk}^i = \sum_{l=0}^{L-1} H_{k,l}^i \varphi_{j+1,l}^i + \sum_{m=L}^{L+2k} h_{k,m}^i \varphi_{j+1,m}^i \\ \psi_{jk}^i = \sum_{l=0}^{L-1} G_{k,l}^i \varphi_{j+1,l}^i + \sum_{m=L}^{L+2k} g_{k,m}^i \varphi_{j+1,m}^i. \end{cases}$$

L'ensemble conduit à une multirésolution sur l'intervalle ($L^2(\mathbb{R})$) est remplacé par $L^2([0, 1])$ où chacun des sous-espaces $V_j^{[0,1]}$ et $W_j^{[0,1]}$ défini par $V_{j+1}^{[0,1]} = V_j^{[0,1]} \oplus W_j^{[0,1]}$ est engendré par un système orthonormé contenant 2^j éléments. On en déduit alors une base orthonormée d'ondelettes héritant des propriétés des ondelettes à support compact du type Daubechies.

A.5.2 Construction des bases de paquets d'ondelettes sur l'intervalle

La construction d'une base d'ondelettes sur l'intervalle correspond à la donnée d'un filtre **QMF** (H', G') qui agit comme (H, G) Γ filtre associé à

l'AMR réelle sur les fonctions intérieures mais nécessite des modifications pour les fonctions de bord.

Ainsi en se plaçant à un niveau $j > J_0$ tel que $2^{J_0-1} < 2L$ et $2^{J_0} \geq 2L$ la construction de bases orthonormées de paquets d'ondelettes sur l'intervalle consiste tout simplement à reprendre le schéma du paragraphe (A.4.2) et à utiliser le filtre (H', G') et les fonctions $\varphi_{j,k}^i$ et $\psi_{j,k}^i$ ($i = 0, 1$) dans (A.25)'.

En désignant par $\tilde{\varphi}_{jk}$ les éléments de la base d'ondelettes sur l'intervalle définies par Cohen et al [17] on définit alors une base de paquets d'ondelettes que l'on notera par ω_{jk} . Ainsi $\omega_{J_0 k} = \tilde{\varphi}_{J_0 k}$ et $\omega_{J_0+1, k} = \tilde{\psi}_{J_0+1, k}$.

On conclut ce paragraphe par l'énoncé du théorème suivant

Théorème A. 2. *Pour tout entier naturel $j \geq J_0$, $\{\omega_{lk}(x), l = J_0, J_0 + 1, \dots, J_0 + 2^{j-J_0} - 1, k = 1, 2, \dots, 2^{J_0}\}$ est un système orthonormé de $V_j^{[0,1]}$. Par la structure même de l'AMR, on en déduit que $\{\omega_{jk}(x), j \geq J_0, k = 1, 2, \dots, 2^{J_0}\}$ est une base orthonormée de $L^2([0, 1])$.*

La preuve de ce théorème est immédiate et analogue à celle du théorème A.1.

A.6 Aspect numérique

On désigne par transformée en ondelettes discrète et finie la décomposition en ondelettes discrète d'un signal discret de longueur finie.

L'idée d'une telle transformée est de décomposer tout vecteur \mathbf{v} en composantes principales $\mathbf{a}^j = M_H^j \mathbf{a}^{j-1}$ et en détails $\mathbf{d}^j = M_G^j \mathbf{a}^{j-1}$ en partant de $\mathbf{a}^0 = \mathbf{v}$. Les matrices M_H^j et M_G^j sont définies respectivement à partir du filtre passe-bas H associé à la fonction d'échelle de l'AMR et le filtre passe-haut G associé à l'ondelette mère et réduisent de moitié la longueur de \mathbf{a}^j ; les filtres en question sont explicitement définies dans le paragraphe A.4. Ainsi pour un vecteur \mathbf{v} de longueur 2^J on détermine sa TODF par :

$$\hat{\mathbf{v}} = [\mathbf{d}^1, \dots, \mathbf{d}^J, \mathbf{a}^J],$$

et par le vecteur des longueurs :

$$\mathbf{b} = [\text{longueur}(\mathbf{a}^0), \text{longueur}(\mathbf{d}^1), \dots, \text{longueur}(\mathbf{d}^J)].$$

Il est alors possible de reconstruire \mathbf{v} à partir de $\hat{\mathbf{v}}$ et \mathbf{b} en réitérant $\mathbf{a}^{j-1} = (M_H^j)^T \mathbf{a}^j + (M_G^j)^T \mathbf{d}^j$ pour $j = J, \dots, 1$.

La TODF est alors réduite à la détermination des matrices associées aux

filtres (h_k) et (g_k) associés à l'AMR.

Dans notre cas on s'intéresse à la construction de matrices orthogonales qui facilitent l'écriture et l'interprétation de la transformée en ondelettes discrète. Ceci sera possible chaque fois qu'une version analogue à la **FDT** transformée de Fourier discrète ou **FFT** transformée de Fourier rapide est utilisée pour construire la **TODF** sous réserve que l'on porte une attention très particulière à ce qui se passe aux extrémités.

Une façon de construire de telles matrices pour un vecteur \mathbf{v} de longueur 12 par exemple lorsqu'on dispose d'un filtre $l_{(\cdot)}$ de longueur 6 et dont chaque composante $l_i = i$ est d'utiliser une version périodique ; ainsi la matrice associée au filtre s'écrit :

$$\begin{pmatrix} 3 & 4 & 5 & 6 & & & & & & & & \\ 1 & 2 & 3 & 4 & 5 & 6 & & & & & & \\ & & 1 & 2 & 3 & 4 & 5 & 6 & & & & \\ & & & & 1 & 2 & 3 & 4 & 5 & 6 & & \\ & & & & & & 1 & 2 & 3 & 4 & 5 & 6 \\ & & & & & & & & 1 & 2 & 3 & 4 \end{pmatrix}$$

D'autres versions sont présentées et étudiées par Carl Taswell [71]. L'avantage de celle présentée ci-dessus est qu'elle conduit à une transformation orthogonale et permet une reconstruction exacte de \mathbf{v} à partir de $\hat{\mathbf{v}}$ et de \mathbf{b} .

Bibliographie

- [1] Antoniadis (A.)ΓBerruyer (J.) et Carmona (R.). – *Régression non linéaire et applications*. – EconomicaΓesaΓParis.Γ1993.
- [2] Antoniadis (A.) et Grégoire (G.). – Penalized likelihood estimation for rates with censored survival data. *Scand. J. Stat.*Γ vol. 17Γ 1990Γ pp. 43–63.
- [3] Antoniadis (A.)ΓGrégoire (G.) et McKeague (I.W.). – Wavelet methods for curve estimation. *JASA*Γ vol. 89Γ n° 428Γ 1994Γ pp. 1340–1353.
- [4] Auscher (P.)ΓWeiss (G.) et Wickerhauser (M.V.). – Local sine and cosine of Coifman and Meyer and the construction of smooth wavelets. pp. 237–256. – Academic PressΓInc.Γ1988.
- [5] Barnwell (T.P.) et Smith (M.J.T.). – Exact reconstruction technique for tree structured subband coders. *IEEE trans. ASSP 30*Γ vol. 5Γ 1982Γ pp. 751–765.
- [6] Beran (R.). – *Bootstrap Variable Selection and Confidence Sets*. – Rapport techniqueΓ1994. University of CaliforniaΓBerkley.
- [7] Bickel (P. J.). – Minimax estimation of a normal mean subject to doing well at a point. *In: In Recent Advances in Statistics*Γ pp. 511–528. – Academic PressΓNew YorkΓ1983.
- [8] Blum (J. R.) et Susarla (V.). – Maximal derivation theory of density and failure rate function estimates under random censorship. pp. 213–222. – P. R. KrishniahΓ1980.
- [9] Bosq (D.) et Lecoutre. (J-P.). – *Théorie de l'estimation fonctionnelle*. – EconomicaΓesaΓParisΓ1987.
- [10] Breiman (L.)ΓFreidman (J. H.)ΓOlshen (R. A.) et Stone. (C. J.). – *Classification and Regression Trees*. – Wadsworth InternationalΓBelmontΓCAΓ1984.

- [11] Buckley (M. J.) et Eagleson (J. K.). – The estimation of residuals variance in nonparametric regression. *Biometrika* Γ vol. 75 Γ 1988 Γ pp. 189–199.
- [12] Buja (H.) et Tibshirani (R.). – Linear smoothers and additive models. *Ann. Stat.* Γ vol. 17 Γ 1989 Γ pp. 595–645.
- [13] Čencov. – Frequency polygons: theory and applications. *JASA* Γ vol. 80 Γ 1962 Γ pp. 595–645.
- [14] Cieselski (Z.) et Domsta (J.). – Construction of an orthogonal basis in C^m and $W^p(I^d)$. *Studia Mathematica* Γ vol. 6 Γ 1972 Γ pp. 212–224.
- [15] Cieselski (Z.) et Domsta (J.). – Estimates for the spline orthonormal functions and for their derivatives. *Studia Mathematica* Γ vol. 6 Γ 1972 Γ pp. 315–320.
- [16] Cohen (A.). – Construction de bases α -höldériennes. *Revista matemática Iberoamericana* Γ vol. 6 Γ 1990 Γ pp. 91–108.
- [17] Cohen (A.) Γ Daubechies (I.) et Vial (P.). – Wavelets on the interval and fast wavelet transforms. *App. and comp. harmonic analysis* Γ vol. 1 Γ 1993 Γ pp. 54–81.
- [18] Coifman (R. R.) et Meyer (Y.). – Orthonormal wave packets basis. – ftp anonyme Γ ceres.math.yale.edu.
- [19] Coifman (R. R.) Γ Meyer (Y.) et Wickhauser (V.). – Wavelet analysis and signal processing. – ftp anonyme Γ ceres.math.yale.edu.
- [20] Collomb (Gerard). – *Estimation non paramétrique en régression par la méthode de noyau*. – Thèse de PhD Γ Université Paul Sabatier de Toulouse Γ 1976.
- [21] Craven (P.) et Wahba (G.). – Smoothing noisy data with spline functions. *Numer. Math.* Γ vol. 31 Γ 1979 Γ pp. 377–403.
- [22] Daubechies (I.). – Orthonormal bases of compactly supported wavelets. *Commun. Pure Appl. Math* Γ vol. 41 Γ 1988 Γ pp. 909–996.
- [23] Davis (P.J.) et Rabinovitz. (P.). – *Methods of numerical integration*. – 1975.
- [24] De Boor. – *A practical guide to splines*. – Springer Γ New York Γ 1978.

- [25] Donoho (D.) et Johnstone (I.). – *Minimax estimation via Wavelet shrinkage*. – Rapport technique n° 402Γjuillet 1992.
- [26] Donoho (D.) et Johnstone (I.). – *Ideal denoising in an orthonormal basis chosen from a library of bases*. – Rapport technique n° 461ΓSeptember 1994.
- [27] Donoho (D. L.) et Johnstone (I. M.). – *Ideal Spatial Adaptation by Wavelet Shrinkage*. – Rapport technique n° 400ΓJuillet 1992.
- [28] Donoho (D. L.)ΓJohnstone (I. M.)ΓKerkycharian (G.) et Picard (D.). – *Density estimation by wavelet thresholding*. – Rapport technique n° 426ΓJuin 1993. <http://playfair.stanford.edu/subjects/wavelets.html>.
- [29] Doob. – *Stochastic Processes*. – WileyΓNew YorkΓ1953.
- [30] Engel (J.). – Density estimation with haar series. *Stat. and Prob. Letters*Γvol. 9Γ1990Γpp. 111–117.
- [31] Fan (J.). – *Test of Significance based on Wavelet Thresholding and Neyman's Truncation*. – Rapport techniqueΓUniversité Catholique de LouvainΓmay 1994.
- [32] Finbar (O'Sullivan). – Nonparametric estimation of relative risk using splines and cross validation. *Siam J. Sci. Stat. Comput*Γvol. 9Γ1988Γpp. 531–542.
- [33] Földes (A.)ΓRejto (L.) et Winter (B. B.). – Strong consistency properties of non parametric estimators for randomly censored data. *Periodica Mathematica Hungarica*Γvol. 12Γ1981Γpp. 15–29.
- [34] Friedman (J. H.). – Multiple adaptive regression splines. *Ann. Stat.*Γvol. 19Γ1991Γpp. 1–141.
- [35] Friedman (J. H.) et Stuezle (W.). – Projection pursuit regression. *JASA*Γvol. 76Γ1981Γpp. 817–823.
- [36] Gasser (T.)ΓSroka (L.) et Jennen-Steinmitz (C.). – Residual variance and residual pattern in nonlinear regression. *Ann. Stat.*Γvol. 73Γ1986Γpp. 625–633.
- [37] Gordon (L.) et Olshen (R. A.). – Asymptotically efficient solution to the classification problem. *Ann. Stat.*Γvol. 6Γ1978Γpp. 515–533.

- [38] Gordon (L.) et Olshen (R. A.). – Consistent nonparametric regression from recursive partitioning schemes. *J. Multivariate Anal.* Γvol. 10Γ1980Γpp. 611–627.
- [39] Gordon (L.) et Olshen (R. A.). – Almost surely consistent nonparametric regression from recursive partitioning schemes. *J. Multivariate Anal.* Γvol. 15Γ1984Γpp. 147–163.
- [40] Hall (P.). – Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* Γvol. 77Γ1990Γpp. 521–528.
- [41] Hall (P.) et Patil (P.). – Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Ann. Stat.* Γvol. 23Γ1995Γpp. 905–928.
- [42] Huang (C.). – Estimation of a projection-pursuit type regression model. *Ann. Stat.* Γvol. 19Γ1991Γpp. 142–157.
- [43] Johnstone (I. M.) et Silverman (W.). – Wavelet threshold estimators for data with correlated noise. – 1994. <http://playfair.stanford.edu/subjects/wavelets.html>.
- [44] Kooperberg (C.) et Stone (C. J.). – Logspline density estimation for censored data. *J. of Computational and Graphical Statistics* Γvol. 1Γ1992Γpp. 301–328.
- [45] Kooperberg (C.) ΓStone (C. J.) et Truong (Y. K.). – Hazard regression. *JASA* Γvol. 90Γ1995Γpp. 78–94.
- [46] Krishnaiah (P. R.) et Kanal (L.N.). – Classification pattern recognition ... In: *Handbook of statistics*. – 1987.
- [47] Lemarié (P.G.) et Meyer (Y.). – Ondelettes et bases hilbertiennes. *Revista matematica iberoamericana* Γvol. 2Γ1986Γpp. 1–17.
- [48] Loh et al. – Piecewise-polynomial regression trees. *Statistica Sinica* Γvol. 4Γ1994Γpp. 143–167.
- [49] Mammen (E.) et van de Geer (S.). – Locally adaptive regression splines. – july 1993.
- [50] Marshall (A. W.) et Proschan (F.). – Maximum likelihood estimation for distributions with monotone failure rate. *Ann. Math. Stat.* Γvol. 36Γ1965Γpp. 69–77.

- [51] McNichols (D.T.) et Padgett (W.J.). – *Hazard rate estimation under the Koziol-Green model of random censorship*. – Rapport technique n° 79ΓUniversity of south CarolinaΓmai 1983.
- [52] Meyer (Y.). – Construction de bases ortonormées d'ondelettes. *Revista matematica iberoamericana*Γvol. 4Γ1988Γpp. 31–39.
- [53] Meyer (Y.). – Ondelettes sur l'intervalle. *Revista matematica iberoamericana*Γvol. 7Γ1991Γpp. 115–133.
- [54] Mykytyn (S.) et Santner (T. A.). – Maximum likelihood estimation of the survival function based on censored data under hazard rate assumptions. *Comm. Stat. Theory Meth. A*Γvol. 10Γ1981Γpp. 1369–1387.
- [55] Nason (G. P.) et Silverman (B. W.). – The stationary wavelet transform. – 1995. <http://www.mathsoft.com.wavelets.html>.
- [56] Nussbaum (M.). – Spline smoothing in regression models and asymptotically efficiency in L_2 . *Ann. Stat.*Γvol. 13Γ1985Γpp. 984–997.
- [57] Padgett (W. J.) et Wei (L. J.). – Maximum likelihood estimation of the distribution function with increasing failure rate based on censored observations. *Biometrika*Γvol. 67Γ1980Γpp. 470–474.
- [58] Quak (E.) et Weyrich (N.). – Decomposition and reconstruction algorithms for spline wavelets on a bounded interval. *Applied and computational harmonic analysis*Γvol. 1Γ1994Γpp. 217–231.
- [59] Rafajlowicz (E.). – Nonparametric orthogonal series estimators of regression :A class attaining the optimal convergence in L_2 . *Stat. and Prob. Letters*Γvol. 5Γ1985Γpp. 219–224.
- [60] Ramlau-Hansen (H.). – Smoothing counting process intensities by means of kernel functions. *Ann. Stat.*Γvol. 11Γ1983Γpp. 453–466.
- [61] Rice (J.). – Bandwith choice for nonparametric regression. *Ann. Stat.*Γvol. 12Γ1984Γpp. 1215–1230.
- [62] Rutkowsky. – On system identification by non parametric function fitting. *I.E.E.E Trans. Aut. Control*Γvol. 27Γ1982Γpp. 225–227.
- [63] Schwartz. – Estimation of probability density by an orthogonal series. *Ann. Math. Statistics*Γvol. 38Γ1967Γpp. 595–645.
- [64] Schwartz (G.). – Estimating the dimension of a model. *Ann. Stat.*Γvol. 6Γ1978Γpp. 461–464.

- [65] Scott. – Frequency polygons: theory and its applications. *JASA* Γvol. 80 Γ1985 Γpp. 515–533.
- [66] Speckman (P.). – Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Stat.* Γvol. 13 Γ1985 Γpp. 970–983.
- [67] Stone (C. J.). – Nonparametric regression and its applications (with discussion). *Ann. Stat.* Γvol. 5 Γ1977 Γpp. 595–645.
- [68] Tanner (M. A.). – A note on the variable kernel estimator of the hazard function from randomly censored data. *Ann. Stat.* Γvol. 11 Γ1983 Γpp. 994–998.
- [69] Tanner (M. A.) et Wong (W.H.). – The estimation of the hazard function from randomly censored data by the kernel method. *Ann. Stat.* Γvol. 11 Γ1983 Γpp. 989–993.
- [70] Tarter (M.) et Kronmal (R.). – On multivariate density estimates based on orthogonal expansions. *Ann. Stat.* Γvol. 4 Γ1970 Γpp. 718–722.
- [71] Taswell (C.) et McGill (K. C.). – Wavelet transform. *ACM transactions on Mathematical Software* Γvol. 20 Γn° 3 ΓSeptember 1994 Γpp. 398–412.
- [72] Vidakovic (B.). – Non linear wavelet shrinkage with bayes rules and bayes factors. – <http://www.mathsoft.com.wavelets.html>.
- [73] Wahba. – A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Ann. Stat.* Γvol. 13 Γ1985 Γpp. 1378–1402.
- [74] Watson (G. S.) et Leadbetter (M. R.). – Hazard analysis ii. *Sankhya Ser. A* Γvol. 26 Γ1964 Γpp. 110–116.

Résumé

Dans ce travail nous proposons de nouvelles approches d'estimation fonctionnelle pour des problèmes de régression et d'analyse de données de survie et ce par l'utilisation de techniques adaptatives et non linéaires fondées sur des décompositions en ondelettes. Les estimateurs qui en découlent combinent les techniques d'estimation par projection orthogonale et celles de seuillage ou arbre de régression.

Tout au long de ce travail l'accent est mis sur l'importance que revêt le choix de la base optimale parmi une famille de bases d'ondelettes ou de paquets d'ondelettes exploitant au mieux la structure d'analyse multiéchelle qui leur est associée.

Abstract

This dissertation is concerned with the problem of regression and hazard function estimation under the regression models as well as the survival analysis ones. As a contribution to solve this kind of problems we introduce new adaptive non linear methods obtained via a decomposition on wavelet and wave-packet bases defined on a real multiscale. Our estimators are based on the tree regression or thresholding and the linear projection approaches.

Mots clés : Ondelettes, Paquets d'ondelettes, Régression par arbre, Fonctions de Haar, Données de survie.

Keywords : Wavelets, Wavepackets, Tree regression, Haar functions, Survival analysis.