

Université de Montréal

Estimation non paramétrique bayésienne de  
courbes de croissance

par

Cindy Ubartas

Département de mathématiques et de statistique  
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures  
en vue de l'obtention du grade de  
Maître ès sciences (M.Sc.)  
en statistique

août 2007



QA

3

U54

2007

V.018

## AVIS

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

## NOTICE

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

**Estimation non paramétrique bayésienne de  
courbes de croissance**

présenté par

**Cindy Ubartas**

a été évalué par un jury composé des personnes suivantes :

*Alejandro Murua*

---

(président-rapporteur)

*Jean-François Angers*

---

(directeur de recherche)

*Jonathan Taylor*

---

(membre du jury)

Mémoire accepté le:

*6 août 2007*

---

## SOMMAIRE

---

Dans ce mémoire, nous nous intéressons à l'estimation de fonctions monotones non décroissantes dans un contexte de régression bayésienne non paramétrique et, tout particulièrement, à celle d'une courbe moyenne représentant les courbes de croissance d'une population d'enfants donnée. Pour ce faire, nous nous inspirons d'un article (voir Merleau *et al.*, 2007) dans lequel les auteurs proposent l'estimation d'hydrogrammes cumulatifs à partir de splines monotones.

La première étape consiste donc à obtenir une courbe moyenne à partir de l'ensemble des courbes de la population étudiée en s'assurant que la courbe résultante soit représentative de la majorité de celles-ci. Pour ce faire, le calcul direct de la hauteur moyenne de tous les enfants en un temps donné se trouve à être la technique la plus simple afin de construire cette courbe. Cependant, nous devons tenir compte du fait que la croissance humaine est une conséquence de phénomènes complexes qui ne surviennent pas au même moment chez tous les enfants. Par exemple, la plupart des enfants subissent une poussée de croissance au début de la puberté, mais l'âge exact auquel nous l'observons varie d'un enfant à l'autre. Ainsi, la présence de ces différences physiologiques nous oblige à appliquer une transformation du temps afin de nous assurer que l'âge auquel de tels événements se produisent coïncide pour toutes les courbes avant de calculer la hauteur moyenne. Cette procédure se nomme synchronisation événementielle. Son application nous permet d'éviter que les tendances observées dans la population ne soient plus perceptibles sur la courbe de référence à cause du fait que certains enfants grandissent rapidement à un âge plus précoce que d'autres.

Par la suite, nous présentons deux modèles dans le but d'estimer la courbe résultante. Dans le premier cas, nous suggérons un modèle *a priori* pour les observations associées à la courbe moyenne. Dans le second cas, nous proposons plutôt un modèle *a priori* pour les observations de chacune des courbes individuelles. De plus, chacun des deux modèles est d'abord développé à l'aide de splines monotones et, ensuite, avec une modification d'une base d'ondelettes. La raison pour laquelle nous n'utilisons pas directement une base d'ondelettes, mais bien une transformation de celle-ci est qu'elle nous permet d'obtenir des fonctions qui, comme les courbes de croissance, sont monotones. Finalement, le comportement des estimateurs développés est étudié à l'aide d'un exemple utilisant des données réelles et d'une courte simulation.

## SUMMARY

---

In this master's thesis, we are interested by the estimation of nondecreasing monotone functions in a nonparametric Bayesian regression context. Particularly, we want to estimate an average curve representing growth curves for a given population of children. To do so, we generalize a method described by Merleau *et al.* (2007) in which the authors propose the estimation of cumulative hydrograms by using monotone splines.

The first step consists of obtaining an average curve coming from all the curves of the studied population while being assured that the resulting curve is representative to the majority of them. To do so, the direct calculation of the average height of all the children in a given time is the easiest technique to build this curve. However, we must keep in mind that the human growth is a consequence of complex phenomena that do not occur at the same time for each child. By example, most of the children have a growth spurt at the beginning of puberty, but the exact age at which we can observe it depends from one child to another. So, the presence of these different physiologies obliges us to apply a time transformation to be sure that the age to which the events happen are at the same times for all the curves before calculating the average height. This procedure is called the landmark registration. Its application helps us avoid that the observed tendencies in the population are not perceptible on the reference curve because of the fact that certain kids grow up much faster at an earlier age than the others.

Afterwards, we describe two models that we will use to estimate the resulting curve. For the first one, we suggest a prior model for the observations linked with the average curve. For the second model, we rather propose a prior model for the observations from each of the individual curves. Furthermore, each of the two

models is first developed with monotone splines and then with a modification of a wavelet basis. The reason for which we do not use directly a wavelet basis, but a transformation of this one is that it permits us to obtain functions that, like the growth curves, are monotones. Finally, the behaviour of the estimators developed is studied using an example involving real data and a small simulation.

## MOTS CLÉS

---

Fonctions monotones non décroissantes, courbes de croissance, synchronisation, splines intégrés, ondelettes non négatives, ondelettes intégrées.

## KEYWORDS

---

Monotone nondecreasing functions, growth curves, registration, integrated splines, nonnegative wavelets, integrated wavelets.

# TABLE DES MATIÈRES

---

<b>Sommaire</b> .....	iii
<b>Summary</b> .....	v
<b>Mots clés</b> .....	vii
<b>Keywords</b> .....	viii
<b>Liste des figures</b> .....	xii
<b>Liste des tableaux</b> .....	xiv
<b>Remerciements</b> .....	1
<b>Introduction</b> .....	2
<b>Chapitre 1. Analyse fonctionnelle</b> .....	5
1.1. Les bases polynomiales .....	6
1.2. Les bases de splines .....	8
1.2.1. M-splines et splines intégrés .....	9
1.2.2. Choix de l'ordre $k$ du spline .....	14
1.2.3. Choix des noeuds $t_j$ .....	14
1.3. Les bases d'ondelettes .....	17
1.3.1. Bases d'ondelettes : propriétés générales .....	17
1.3.2. Les ondelettes de Daubechies .....	21
1.3.3. Décomposition d'une fonction à l'aide d'ondelettes .....	22
1.3.4. Les ondelettes biorthogonales .....	23
1.3.5. Les ondelettes non négatives .....	26

<b>Chapitre 2. Modélisation d'une courbe de croissance de référence pour une population</b> .....	29
2.1. Synchronisation événementielle .....	29
2.2. Modèle hiérarchique de Bayes et décompositions conditionnelles ...	33
2.3. Modélisation par base de splines .....	36
2.3.1. Modèle <i>a priori</i> .....	36
2.3.1.1. Modèle 1 : sur la moyenne des observations .....	36
2.3.1.2. Modèle 2 : sur les observations directement .....	38
2.3.2. Estimateur de la courbe de référence .....	39
2.3.2.1. Modèle 1 : sur la moyenne des observations .....	39
2.3.2.2. Modèle 2 : sur les observations directement .....	40
2.4. Modélisation à l'aide de fonctions d'échelle positives .....	45
2.4.1. Développement analytique de la fonction $\rho_{r,l}^*(x)$ pour les ondelettes de Haar .....	47
2.4.2. Modèle <i>a priori</i> .....	50
2.4.2.1. Modèle 1 : sur la moyenne des observations .....	51
2.4.2.2. Modèle 2 : sur les observations directement .....	54
2.4.3. Estimateur de la courbe de référence .....	55
2.4.3.1. Modèle 1 : sur la moyenne des observations .....	55
2.4.3.2. Modèle 2 : sur les observations directement .....	56
2.5. Estimation avec contraintes .....	57
2.5.1. Coefficients positifs .....	57
2.5.2. Dérivées positives .....	59
<b>Chapitre 3. Applications</b> .....	61
3.1. Exemple avec données réelles .....	61
3.1.1. Synchronisation événementielle .....	62

3.1.2. Modélisation.....	66
3.2. Application des modèles développés à des fonctions contenant un certain degré de variation.....	73
<b>Conclusion.....</b>	<b>81</b>
<b>Annexe A. Densités.....</b>	<b>A-i</b>
A.1. Distributions normales.....	A-i
A.1.1. Distribution normale $\mathcal{N}(\mu, \sigma^2)$ .....	A-i
A.1.2. Distribution normale multidimensionnelle $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .....	A-i
A.2. Distribution inverse gamma $\mathcal{IG}(\alpha, \beta)$ .....	A-ii
A.3. Distribution t de Student multidimensionnelle $T_p(\nu, \boldsymbol{\theta}, \boldsymbol{\Sigma})$ .....	A-ii
<b>Annexe B. Programmation.....</b>	<b>B-i</b>
<b>Bibliographie.....</b>	<b>B-i</b>

## LISTE DES FIGURES

---

1.1	Relation entre $\xi$ et $t$ . . . . .	11
1.2	Exemples de fonctions (a) M-splines (b) I-splines . . . . .	14
1.3	Exemples de la fonction d'échelle (bleu pâle) et de la mère des ondelettes (bleu foncé) (a) Ordre 1 (b) Ordre 2 . . . . .	22
2.1	Exemples de variation avec la fonction sinus (a) Phase (b) Amplitude	30
2.2	Exemples de la fonction $\rho_r(2^j x - l)$ (a) Pour $j = l = 0$ (b) Pour $j = 1$ et $l = 0$ (c) Pour $j = 0$ et $l = 1$ . . . . .	49
2.3	Exemples de la fonction $\rho_{r,l}^*(x)$ (a) Pour $j = l = 0$ (b) Pour $j = 1$ et $l = 0$ (c) Pour $j = 0$ et $l = 1$ . . . . .	50
3.1	Courbes de croissance de dix garçons choisis aléatoirement . . . . .	62
3.2	Effet de la synchronisation : Données originales (trait plein gris), Données transformées (carrés rouges) (a) Garçon 1 (b) Garçon 2 . . . . .	63
3.3	Temps transformés en fonction de ceux observés (a) Garçon 1 (b) Garçon 2 . . . . .	64
3.4	Fonctions synchronisées pour les garçons (a) Vitesses (b) Hauteurs . . . . .	64
3.5	Fonctions synchronisées pour les filles (a) Vitesses (b) Hauteurs . . . . .	65
3.6	Courbes moyennes obtenues après la standardisation (a) Garçons (b) Filles . . . . .	66
3.7	Exemples du logarithme de la marginale en fonction de $n_0$ pour les garçons (a) Modèle 1 - splines (b) Modèle 2 - ondelettes intégrées . . . . .	69

3.8	Estimation (couleur) de la courbe moyenne (points) avec les splines pour les garçons (a) Modèle 1 (b) Modèle 2 .....	72
3.9	Estimation (couleur) de la courbe moyenne (points) avec les splines pour les filles (a) Modèle 1 (b) Modèle 2 .....	72
3.10	Estimation (couleur) de la courbe moyenne (points) avec les ondelettes intégrées pour les garçons (a) Modèle 1 (b) Modèle 2 .....	73
3.11	Estimation (couleur) de la courbe moyenne (points) avec les ondelettes intégrées pour les filles (a) Modèle 1 (b) Modèle 2 .....	73
3.12	Estimation (couleur) de la courbe moyenne (points) pour les filles avec les ondelettes intégrées en utilisant l'estimateur des moindres carrés pondérés .	73
3.13	Estimation de la fonction A : fonction réelle (trait plein noir), estimation moyenne (trait plein en couleur) et bruit (points noirs) (a) Modèle 1 - splines (b) Modèle 2 - splines (c) Modèle 1 - ondelettes intégrées (d) Modèle 2 - ondelettes intégrées .....	78
3.14	Estimation de la fonction B : fonction réelle (trait plein noir), estimation moyenne (trait plein en couleur) et bruit (points noirs) (a) Modèle 1 - splines (b) Modèle 2 - splines (c) Modèle 1 - ondelettes intégrées (d) Modèle 2 - ondelettes intégrées.....	79
3.15	Estimation de la fonction C : fonction réelle (trait plein noir), estimation moyenne (trait plein en couleur) et bruit (points noirs) (a) Modèle 1 - splines (b) Modèle 2 - splines (c) Modèle 1 - ondelettes intégrées (d) Modèle 2 - ondelettes intégrées.....	80

## LISTE DES TABLEAUX

---

3.1	Résultats en utilisant les splines. ....	70
3.2	Résultats en utilisant les ondelettes intégrées. ....	70
3.3	Description des fonctions utilisées. ....	74
3.4	Résultats associés aux fonctions A, B et C. ....	75

## REMERCIEMENTS

---

Tout d'abord, merci à mon directeur de recherche pour la patience et la compréhension dont il a toujours fait preuve. J'ai trouvé très agréable de travailler avec quelqu'un qui réussit à me faire sourire même lorsque je suis déprimée. Merci également à tous les étudiants qui m'ont aidée au cours de mes études et grâce à qui certains cours paraissaient beaucoup moins longs... Finalement, merci à mon copain Julien qui est capable de m'endurer quotidiennement et qui a même lu mon mémoire au complet même si sa compréhension s'est arrêtée à la première page du premier chapitre !

# INTRODUCTION

---

L'estimation de fonctions monotones non décroissantes est un problème couramment rencontré en pratique, que ce soit, par exemple, lorsqu'il est question de fonctions de répartition ou de courbes de croissance. Le principal défi à relever lors de la modélisation de telles fonctions consiste en l'obtention d'estimations qui sont elles-mêmes croissantes, ce qui ne peut souvent pas être garanti par les estimateurs de régression classiques.

Pour y parvenir, plusieurs méthodes ont déjà été élaborées. L'une des plus simples est sûrement la régression isotonique (voir Barlow *et al.*, 1972) qui n'est jamais devenue très populaire, puisqu'elle produit des estimateurs en escaliers, ce qui n'est pas trop désirable lorsque nous sommes en présence de fonctions lisses. D'autres techniques plus complexes existent également. Mukerjee (1988) propose, par exemple, une version améliorée de la régression isotonique en lissant l'estimateur obtenu par cette méthode à l'aide d'un noyau approprié.

Dans ce mémoire, nous adoptons une approche bayésienne non paramétrique que nous illustrons surtout par le biais des courbes de croissance humaine. Dans les dernières décennies, plusieurs statisticiens se sont intéressés à l'analyse de ces courbes qui s'avère être très pratique pour les biologistes et les pédiatres. En effet, il est essentiel de posséder une connaissance approfondie de la croissance humaine afin d'être en mesure de détecter, le plus tôt possible, la moindre anomalie dans la croissance d'un enfant. Pour ce faire, les spécialistes ont besoin non seulement de données de qualité, mais également de techniques adéquates pour les analyser.

Parmi les travaux déjà publiés à cet effet, la modélisation employant des polynômes (voir Rao, 1965) a d'abord été privilégiée, mais ce sont les techniques de régression non linéaire (voir Preece et Baines, 1978) qui ont par la suite pris le

dessus. Dernièrement, certains modèles non paramétriques utilisant des bases de fonctions relativement plus complexes telles que les bases de splines ont fait leur apparition (voir Ramsay et Silverman, 2005). C'est vers ce type d'estimation, qui offre généralement d'excellentes approximations, que nous nous enlignons.

Pour avoir une idée de l'évolution d'un individu par rapport à la normale, nous pensons qu'il est important de pouvoir estimer correctement une courbe de référence qui représente la tendance moyenne de la croissance des enfants provenant d'une certaine population. La façon la plus simple de créer cette dernière est de faire la moyenne des tailles des enfants à chacun des temps, mais cela n'est pas toujours approprié. En effet, au même titre que tout phénomène naturel (l'évolution de la température ou du volume d'eau dans les rivières, par exemple), les événements qui caractérisent ces courbes ne surviennent pas exactement au même âge d'un enfant à l'autre. Le fait de prendre la moyenne directement peut donc résulter en une mauvaise représentation des tendances observées sur chacune des courbes. Dans le pire des scénarios, les poussées de croissance que nous observons chez la majorité des enfants à l'adolescence ne seraient plus visibles sur la courbe moyenne. Pour éviter qu'un tel phénomène ne se produise, il est préférable de commencer par synchroniser les courbes par rapport aux événements marquants qui nous intéressent avant d'en faire la moyenne. Ceci se justifie en quelque sorte en disant qu'un temps physiologique plutôt que chronologique est utilisé et qu'il varie d'un enfant à l'autre. Cette procédure se nomme synchronisation événementielle et elle constitue la première étape de nos analyses.

Par la suite, la courbe moyenne résultante est modélisée à partir de deux bases de fonctions qui sont dotées des mêmes propriétés que celle-ci, à savoir la positivité et la monotonie. Les bases sélectionnées sont les I-splines ainsi qu'une modification de la fonction d'échelle de la base d'ondelettes de Daubechies. De plus, deux modèles bayésiens sont présentés. Dans le premier cas, nous suggérons un modèle *a priori* pour les observations associées à la courbe moyenne. Dans le second cas, nous proposons plutôt un modèle *a priori* pour les observations de chacune des courbes individuelles.

L'organisation de ce mémoire se résume donc ainsi. Le premier chapitre est consacré à l'estimation bayésienne non paramétrique d'une fonction. Nous présentons quelques bases qui peuvent être utilisées pour la représenter. Premièrement, la base polynomiale est abordée rapidement. Par la suite, les bases de splines et d'ondelettes sont étudiées de manière plus détaillée, puisqu'il s'agit de celles retenues pour réaliser nos estimations. Au second chapitre, nous introduisons la synchronisation événementielle avant de développer les deux modèles considérés, et ce, pour les splines d'abord et pour la transformation établie à partir de la base d'ondelettes de Daubechies ensuite. Finalement, au troisième et dernier chapitre, nous nous penchons sur le comportement de nos estimateurs à l'aide d'un exemple utilisant des courbes de croissance ainsi que d'une courte simulation.

# Chapitre 1

---

## ANALYSE FONCTIONNELLE

Dans ce mémoire, nous nous intéressons à l'estimation de courbes de croissance dans un contexte d'estimation bayésienne non paramétrique, ce qui signifie que nous disposons de couples de données

$$(x_{ij}, y_{ij}), \quad i = 1, \dots, N, j = 1, \dots, n,$$

où  $n$  est le nombre d'observations associé à chacune des  $N$  fonctions. Nous considérons alors le modèle

$$y_{ij} = g_i(x_{ij}) + \varepsilon_{ij}, \quad i = 1, \dots, N, j = 1, \dots, n, \quad (1.0.1)$$

où  $g_i$  sont des fonctions inconnues à estimer, les  $x_{ij}$  appartiennent à un ensemble borné  $\mathcal{X} \subset \mathbb{R}$  et les  $\varepsilon_{ij}$  représentent les termes d'erreurs. Ces erreurs, centrées à zéro, sont supposées indépendantes et normalement distribuées. De plus, leur variance  $\sigma_i^2$  peut être différente d'une fonction à l'autre. Autrement dit, nous avons  $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_i^2)$ .

Par la suite, nous estimons ces fonctions à l'aide d'une base de fonctions quelconque, c'est-à-dire que nous les exprimons comme

$$g_i(x) = \sum_k \lambda_k b_k(x), \quad (1.0.2)$$

où  $b_k$  appartient à la base pour tout  $k$  et  $\lambda_k \in \mathbb{R}$ . Le but est alors d'estimer les coefficients  $\lambda_k$  et de faire de l'inférence sur ces derniers. Ceci peut se faire en utilisant différentes bases. Parmi les plus simples, nous retrouvons la base polynomiale qui, encore aujourd'hui, demeure l'une des plus naturelles à utiliser. Il en

existe également d'autres parmi lesquelles se trouvent celles des splines et des ondelettes. Dans ce chapitre, nous discutons rapidement de la première, puisqu'elle ne fait pas partie des bases retenues pour nos estimations, avant d'introduire les deux autres de façon beaucoup plus détaillée.

### 1.1. LES BASES POLYNOMIALES

Bien que datant de quelques siècles, ces bases ont joué un rôle central dans la théorie de l'approximation dès le début du développement de celle-ci. En effet, une des principales raisons pour lesquelles elles ont été très populaires dès le départ est que toute fonction  $f$  continue sur un intervalle  $[a, b]$  peut être estimée par un polynôme, défini par  $f(x) = \sum_{i=1}^k a_i x^{i-1}$ , avec une erreur relativement petite. Cette propriété est d'ailleurs connue sous le nom de théorème de Weierstrass (voir Burden et Faires, 2001). Dans la littérature, l'estimation des fonctions  $g_i$ ,  $i = 1, \dots, N$ , est faite à partir de polynômes pouvant être exprimés sous plusieurs formes. À cet effet, une des méthodes les plus simples consiste à obtenir la base en développant ces dernières à l'aide de séries de Taylor autour d'un point choisi. Cette approche a été adoptée par quelques auteurs dont Angers et Delampady (1992) dans le cas où une seule fonction est considérée. L'équation (1.0.1) devient ainsi

$$y_j = g(x_j) + \varepsilon_j, \quad j = 1, \dots, n,$$

où, par hypothèse, la fonction  $g$  est lisse et définie sur  $\mathcal{X}$ .

Dans cet article, la fonction  $g$  est développée autour du point  $x_0$  à l'aide d'un polynôme d'ordre  $m - 1$ , c'est-à-dire que nous retrouvons

$$\begin{aligned} g(x) &= g(x_0) + (x - x_0)g'(x) + \dots + \frac{(x - x_0)^{m-1}}{(m-1)!}g^{(m-1)}(x_0) + R_m(x) \\ &= \Phi(x)^\top \theta + R_m(x), \end{aligned}$$

où

$$\begin{aligned} \Phi(x) &= \left( 1, x - x_0, \frac{(x - x_0)^2}{2!}, \dots, \frac{(x - x_0)^{m-1}}{(m-1)!} \right)^\top, \\ \theta &= (g(x_0), g'(x_0), \dots, g^{(m-1)}(x_0))^\top, \end{aligned}$$

et  $R_m(x)$  est le terme restant. Le modèle employé se résume donc à

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

où  $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$ ,  $\mathbf{X} = (\mathbf{T}, \mathbf{I})$ ,  $\mathbf{T} = (\Phi(x_1), \Phi(x_2), \dots, \Phi(x_n))^\top$ ,  $\boldsymbol{\beta} = (\boldsymbol{\theta}^\top, \mathbf{R}^\top)^\top$  et  $\mathbf{R} = (R_m(x_1), R_m(x_2), \dots, R_m(x_n))^\top$ . Il est étudié à l'aide d'une approche bayésienne hiérarchique dans laquelle les paramètres ont une densité *a priori* diffuse, c'est-à-dire qui possède une très grande variance. Une méthode permettant de choisir l'ordre du développement de Taylor est proposée. De plus, une analyse de sensibilité qui a pour but de déterminer l'influence de la densité *a priori* des hyperparamètres sur le lissage de la fonction  $g$  est présentée.

Les polynômes de Lagrange (voir Burden et Faires, 2001) constituent un deuxième type de polynômes pouvant être utilisés dans un contexte semblable. C'est, par exemple, ce que fait Young (1977). Regardons leur définition.

**Théorème 1.1.1.** *Si nous sommes en présence de  $n$  points distincts  $(x_j, y_j)$ ,  $j = 1, \dots, n$ , alors il existe un unique polynôme  $P(x)$  de degré inférieur ou égal à  $n-1$  passant par ces points, qui est donné par*

$$P(x) = \sum_{j=1}^n y_j L_{n-1,j}(x), \quad (1.1.1)$$

où, pour chaque  $j = 1, \dots, n$ ,

$$L_{n-1,j}(x) = \prod_{l=1, l \neq j}^n \frac{x - x_l}{x_j - x_l}.$$

En observant bien l'équation (1.1.1), nous remarquons que, lors de la construction de tels polynômes, il y a un compromis à faire entre la qualité de l'ajustement et le degré de lissage de la fonction. En effet, plus  $n$  est grand, plus le degré du polynôme résultant est élevé et, par conséquent, plus l'estimation a tendance à osciller entre les points servant à l'interpolation. Ainsi, de par la nature oscillatoire des polynômes de degré élevé, même si une modélisation faisant intervenir un tel polynôme est exacte aux points  $x_1, \dots, x_n$ , elle peut être un mauvais prédicteur de la fonction entre ces points.

Donc, bien qu'elle présente certaines propriétés largement convoitées à des fins de modélisation, comme la facilité avec laquelle elle peut être évaluée, mais surtout dérivée ou intégrée un grand nombre de fois, la base des polynômes a pour principal défaut son manque de flexibilité, c'est-à-dire qu'elle donne de bons ajustements sur de petits intervalles uniquement. Pour contrer ce problème, une approche alternative consiste à diviser l'intervalle en plusieurs sous-intervalles et d'y construire différents polynômes. C'est sur ce principe que se basent les splines qui représentent une option envisageable dans le cas où la représentativité d'une fonction ne peut être assurée par des polynômes et qui sont présentés à la prochaine section.

## 1.2. LES BASES DE SPLINES

Les splines, introduits par Schoenberg (1946), ont commencé à intéresser un grand nombre de chercheurs au début des années 1960. En effet, lorsque ces derniers ont réalisé que les splines constituaient un outil efficace pour modéliser une fonction en utilisant une approche entièrement mathématique, leur popularité a connu un essor considérable. Un développement fulgurant de la théorie liée aux splines s'en est suivi. Parmi les ouvrages publiés à cette époque, mentionnons notamment les travaux de De Boor (1963), de Ahlberg, Nilson et Walsh (1967) et, surtout, de Schoenberg (1964) qui a introduit les splines de lissage classiques. Par la suite, ceux-ci ont été développés en grande partie par Wahba, via ses écrits avec plusieurs collaborateurs. Nous parlons ici des travaux faits par Wahba et Kimeldorf (1971), Wahba et Wold (1975) ainsi que Wahba et Craven (1979).

De nos jours, il existe toute une gamme de fonctions splines couramment employées lors de l'approximation de fonctions. Nous mettons particulièrement l'emphase sur celles que nous utilisons lors des estimations au second chapitre, soient les M-splines et les I-splines. Ces fonctions sont membres de la classe des splines de régression, par opposition aux splines de lissage (voir Eubank, 1988) dont le principe repose sur la recherche de la fonction  $f(x)$  qui minimise la fonction de perte

$$\min_f \left\{ \sum_{i=1}^n [y_i - f(x_i)]^2 + \lambda \int [f''(x)]^2 dx \right\}.$$

Le premier terme dans ce critère de minimisation évalue la qualité de l'ajustement. Le second mesure le degré de lissage de  $f$  qui est contrôlé par le paramètre de lissage  $\lambda$ , qui est toujours supérieur ou égal à 0. Cette distinction étant faite, comme cette technique n'est pas celle que nous privilégions, nous n'élaborons pas plus sur ce sujet. Pour de plus amples informations quant aux mathématiques associées aux splines (de régression ou de lissage), il existe d'excellents ouvrages de référence. Citons, entre autres, Chambers (1977), De Boor (1978) et Schumaker (1981). Nous conseillons également la lecture de Wegman et Wright (1983) qui font une très bonne révision de la littérature portant sur les applications statistiques des splines, en incluant même les travaux de Wahba (1978).

Dans cette section, nous nous concentrons plutôt sur la description des bases utilisées, sans donner tous les détails théoriques qui les entourent. Introduisons donc les fonctions que nous appelons M-splines et I-splines comme le fait Ramsay (1988), à qui nous devons la popularisation de l'usage des splines monotones en statistique.

### 1.2.1. M-splines et splines intégrés

Les splines polynomiaux sont définis comme des polynômes par parties de degré  $k-1$  (ou d'ordre  $k$ ). En d'autres mots, l'intervalle  $[T_0, T_n]$  sur lequel nous désirons faire l'approximation de la courbe est divisé en  $q$  sous-intervalles séparés par des points

$$T_0 = \xi_1 \leq \xi_2 \leq \dots \leq \xi_q = T_n.$$

À l'intérieur de chaque sous-intervalle  $[\xi_j, \xi_{j+1})$ , la fonction spline est définie par un polynôme  $P_j$  de degré  $k-1$ . Certaines conditions de régularité, habituellement caractérisées par l'égalité des dérivées, sont imposées à la jonction de ces polynômes. Nous exigeons donc que

$$(D^{m-1}P_j)(\xi_{j+1}) = (D^{m-1}P_{j+1})(\xi_{j+1}), \quad m = 1, \dots, \nu_j,$$

où la notation  $(D^{m-1}P_j)(\xi)$  désigne la dérivée  $d^{m-1}P/dx^{m-1}$  évaluée à l'argument  $\xi$  lorsque  $m > 1$ . Lorsque  $m = 1$ , nous posons  $(D^0P)(\xi) = P(\xi)$ . Dans le cas le plus souvent rencontré, tous les ordres de continuité  $\nu_j$  sont spécifiés comme étant égaux au degré  $k-1$ , et ce, afin que deux polynômes adjacents aient les mêmes

dérivées jusqu'à l'ordre  $k - 2$ . Dans cette situation, si  $k = 2$ , les splines sont des segments de droite pour lesquels les bornes  $\xi_j$  coïncident. De même, si  $k = 3$ , les splines sont des polynômes quadratiques par parties ayant des dérivées premières qui concordent. De plus, dans le cas particulier où  $v_j = k$  pour tout  $j$ , la fonction spline entière se réduit à un seul polynôme de degré  $k - 1$ . Clarifions cette idée à l'aide d'un exemple. Posons  $k = 2$  et définissons deux polynômes :

$$P_j = a_0 + a_1x,$$

et

$$P_{j+1} = b_0 + b_1x.$$

Si  $v_j = k - 1 = 1$  pour tout  $j$ , alors nous désirons seulement que

$$\begin{aligned} (D^0 P_j)(\xi_{j+1}) &= (D^0 P_{j+1})(\xi_{j+1}) \Rightarrow P_j(\xi_{j+1}) = P_{j+1}(\xi_{j+1}) \\ &\Rightarrow \xi_{j+1} = \frac{a_0 - b_0}{b_1 - a_1}, \end{aligned}$$

ce qui signifie que la valeur des deux polynômes est la même au point  $\xi_{j+1}$ .

D'autre part, si  $v_j = k = 2$  pour tout  $j$ , alors, en plus de la condition citée ci-haut, nous rajoutons celle-ci :

$$(D^1 P_j)(\xi_{j+1}) = (D^1 P_{j+1})(\xi_{j+1}).$$

Donc, les dérivées premières coïncident et, par conséquent, les pentes  $a_1$  et  $b_1$  sont les mêmes. Par la première condition, nous imposons également que  $a_0 = b_0$ . La combinaison des deux exigences précédentes implique nécessairement l'égalité des deux polynômes  $P_j$  et  $P_{j+1}$ . À ce moment, nous ne sommes donc qu'en présence d'un seul et unique polynôme, comme nous l'avons expliqué précédemment.

Notons que les abcisses où se produisent les jonctions de polynômes sont appelées noeuds. Définissons, à présent, le vecteur de ces noeuds par

$$\mathbf{t} = (t_1, \dots, t_{z+k})^\top,$$

où  $z$  est le nombre de paramètres spécifiant la fonction spline qui répond aux hypothèses de continuité désirées. La séquence de noeuds possède alors les propriétés

$$(1) \quad t_1 \leq t_2 \leq \dots \leq t_{z+k},$$

- (2) pour tout  $i$ , il existe une valeur  $j$  telle que  $t_i = \xi_j$ ,
- (3) les caractéristiques de continuité sont déterminées par
- (a)  $t_1 = \dots = t_k = T_0$  et  $T_n = t_{z+1} = \dots = t_{z+k}$ ;
  - (b)  $t_i < t_{i+k}$  pour tout  $i$ ;
  - (c) si  $t_i = \xi_j$  et  $t_{i-1} < \xi_j$ , alors  $t_i = \dots = t_{i+k-v_j-1}$ .

Ceci dit, la séquence de noeuds  $\mathbf{t}$  est dérivée de l'ensemble des points  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_q)$  en plaçant un certain nombre de noeuds aux bornes  $\xi_j$  afin de respecter les hypothèses de continuité à ces mêmes points. Un exemple de relation entre les points  $\xi_j$  et les noeuds est donné dans la figure 1.1.

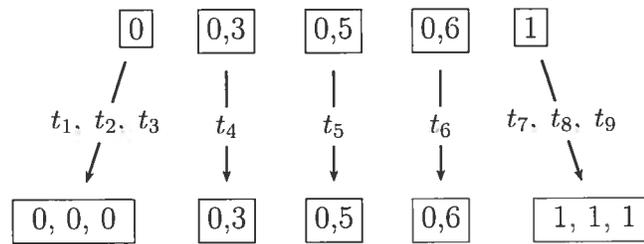


FIG. 1.1. Relation entre  $\boldsymbol{\xi}$  et  $\mathbf{t}$ .

Dans cet exemple, nous avons

$$\begin{aligned}\boldsymbol{\xi} &= (0, 0,3, 0,5, 0,6, 1), \\ \mathbf{t} &= (0, 0, 0, 0,3, 0,5, 0,6, 1, 1, 1).\end{aligned}\tag{1.2.1}$$

Nous remarquons tout de suite, à l'aide de la propriété (a) du point (3), qu'il est ici question de splines quadratiques, donc d'ordre  $k = 3$ . En outre, les noeuds intérieurs sont simples. Nous reviendrons à cet exemple un peu plus tard dans ce chapitre afin d'illustrer l'allure de bases de splines construites à partir de ces noeuds. Pour l'instant, nommons quelques caractéristiques générales des splines qui respectent les trois conditions décrites ci-haut.

Comme un spline de degré  $k - 1$  est un polynôme à tout point  $\xi$ , il est déterminé, en l'absence de contraintes, par  $k$  coefficients sur le sous-intervalle contenant ce point. Cependant, les propriétés de continuité imposent  $v_j$  contraintes linéaires sur les coefficients définissant les polynômes. En fait, le nombre de degrés de liberté total dans un spline d'ordre spécifié est donné par  $z$ . En résumé, il y a un

nombre total de  $z + k$  noeuds à être alloués et le nombre de paramètres associés à un spline étant donnés  $\mathbf{t}$  et  $\xi$  est  $z$ . Dans la situation la plus commune, où  $v_j = k - 1$  pour tout  $j$ , la dimension  $z$  est égale à la somme du nombre de noeuds intérieurs et de l'ordre  $k$ .

Même si l'idée d'un spline représenté comme un polynôme par parties est intuitivement simple, son application exige le développement d'un ensemble de splines  $M_i(\cdot|k, \mathbf{t})$ ,  $i = 1, \dots, z$ , faisant en sorte que tout polynôme par parties ou spline  $f(x)$  d'ordre  $k$  puisse être représenté à l'aide de la combinaison linéaire

$$f(x) = \sum a_i M_i(x|k, \mathbf{t}), \quad \forall x \in \mathcal{X}.$$

La première base développée à cet effet a été construite à partir de fonctions de puissances tronquées définies par

$$\Phi_{jm}(x) = (x - \xi_j)_+^m, \quad m = v_j, \dots, k - 1,$$

où  $u_+ = \max(u, 0)$ .

Par contre, même si la simplicité de ces bases peut sembler attrayante au départ et qu'elles ont effectivement déjà servi lors d'applications (voir Smith, 1979), ces dernières ont le sérieux désavantage de générer des erreurs d'arrondi considérables, sauf pour de très petites valeurs de  $k$  (voir Ramsay, 1988). Ainsi, d'autres bases de splines, dont celle des M-splines qui est particulièrement intéressante en statistique, ont été créées par la suite.

Les M-splines consistent en une famille de fonctions positives ou nulles  $M_i$ ,  $i = 1, \dots, z$ , définies de sorte qu'elles satisfont la normalisation (voir Curry et Schoenberg, 1966)

$$\int_{T_0}^{T_n} M_i(x) dx = 1.$$

Malgré le fait que ces bases peuvent être exprimées en termes de différences divisées de fonctions de puissances tronquées, nous les présentons sous une formule récursive qui les rendent très faciles à manipuler. Nous avons ainsi, pour tout

$$x \in [t_i, t_{i+k}],$$

$$M_i(x|k, \mathbf{t}) = \begin{cases} \frac{1}{t_{i+1} - t_i}, & \text{si } k = 1, \\ \frac{k[(x - t_i)M_i(x|k-1, \mathbf{t}) + (t_{i+k} - x)M_{i+1}(x|k-1, \mathbf{t})]}{(k-1)(t_{i+k} - t_i)}, & \text{sinon.} \end{cases}$$

Remarquons que chaque fonction  $M_i$  est un spline en soit, et par conséquent, un polynôme par parties. D'un point de vue calculatoire, nous conservons donc toutes les propriétés désirables des polynômes énumérées à la section 1.1. Mentionnons que cette base est très fortement liée à celle, bien connue, des B-splines, qui respecte la normalisation alternative

$$\sum_{i=1}^z B_i(x) = 1, \quad \forall x.$$

Cette dernière est d'ailleurs utilisée par He et Shi (1998) dans un contexte fréquentiste afin d'estimer une fonction  $g$  lorsque celle-ci est monotone croissante. Dans cet article, les coefficients  $\lambda_k$  décrits à l'équation (1.0.2) sont déterminés en minimisant la somme des valeurs absolues des erreurs d'estimation aux points observés lorsque la dérivée de chacune de ces estimations est contrainte à être positive. La relation entre ces deux bases est donnée par

$$B_i = \frac{(t_{i+k} - t_i)}{k} M_i.$$

Maintenant, portons notre attention sur les splines monotones qui sont dérivés des précédents. En effet, puisque les M-splines sont des fonctions non négatives, nous pouvons les intégrer pour obtenir des fonctions croissantes. Les splines intégrés, ou I-splines sont donc définis par

$$I_i(x|k, \mathbf{t}) = \int_{T_0}^x \frac{k}{(t_{i+k} - t_i)} B_i(u|k, \mathbf{t}) du = \int_{T_0}^x M_i(u|k, \mathbf{t}) du,$$

ou plus spécifiquement, pour  $j$  allant de  $i$  à  $i+k-1$ , par

$$I_i(x|k, \mathbf{t}) = \begin{cases} 0, & \text{si } x < t_i, \\ \frac{1}{k+1} \sum_{m=i}^j (t_{m+k+1} - t_m) M_m(x|k+1, \mathbf{t}), & \text{si } x \in [t_j, t_{j+1}], \\ 1, & \text{si } x > t_{i+k}. \end{cases}$$

Bien sûr, puisque chaque fonction  $M_i$  est un polynôme par parties de degré  $k - 1$ , toute fonction  $I_i$  construite de cette façon l'est également et elle est de degré  $k$ .

La figure 1.2 illustre ces deux types de fonctions avec les noeuds mentionnés à l'équation (1.2.1). Dans cet exemple,  $z = 6$ . De plus, chaque M-spline est quadratique quand  $k = 3$  et, par conséquent, différent de zéro sur trois intervalles tout au plus.

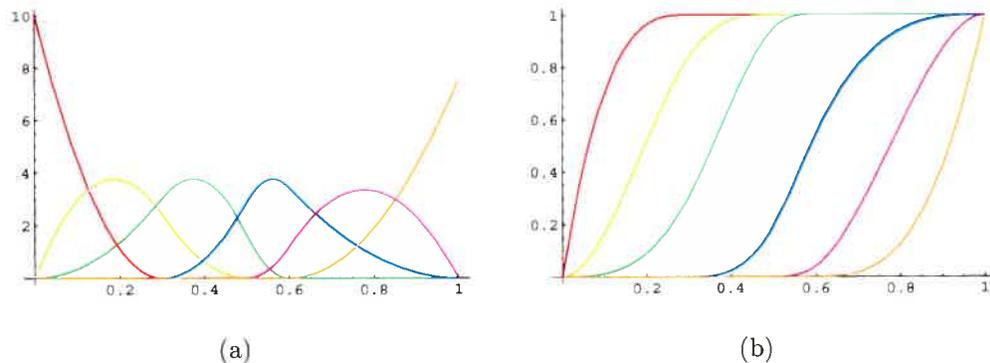


FIG. 1.2. Exemples de fonctions (a) M-splines (b) I-splines

Notons finalement que le nombre total de polynômes par parties avec lesquels nous travaillons, leurs ordres respectifs et le positionnement des noeuds diffèrent selon le type de modélisation désirée. Dans les sections qui suivent, nous discutons brièvement des critères utilisés pour les choisir.

### 1.2.2. Choix de l'ordre $k$ du spline

L'ordre  $k$  que nous utilisons dépend du nombre de dérivées dont nous avons besoin pour faire les calculs à partir de la fonction spline. Par exemple, si nous voulons une dérivée seconde lisse, il est préférable d'utiliser des splines d'ordre supérieur ou égal à 6 (voir Ramsay et Silverman, 2005). Dans notre cas, nous pouvons très bien travailler avec des splines quadratiques, donc d'ordre 3.

Passons désormais au choix des noeuds qui s'avère être un peu plus ardu que celui du degré des splines.

### 1.2.3. Choix des noeuds $t_j$

Plus il y a de noeuds, plus le spline est flexible. Ce principe s'applique localement : si nous avons besoin de beaucoup de flexibilité dans une région particulière, alors

nous plaçons plus de noeuds dans cette région. Cependant, il doit toujours y avoir au moins une donnée dans chaque sous-intervalle déterminé par les noeuds. Les deux techniques les plus simples pour le choix des noeuds sont de prendre des noeuds équidistants en ayant choisi la valeur de  $z$  préalablement ou de placer un noeud entre chaque observation. Par contre, nous voudrions employer le moins de noeuds possible tout en nous assurant que l'emplacement des noeuds nous permette une assez grande flexibilité pour bien détecter les détails de la fonction et ainsi, obtenir une bonne estimation. Mais, les techniques mentionnées précédemment ne répondent pas toujours à ces critères, car la première ne nous garantit pas que les noeuds seront placés à l'endroit où nous avons besoin de la plus grande souplesse, alors que la seconde nous donne un trop grand nombre de noeuds. Dans la littérature, il existe plusieurs autres procédures qui sont moins rudimentaires. Parmi celles fréquemment rencontrées, il y a la méthode MCMC à sauts réversibles imaginée par Green (1995). En partant avec un certain nombre de noeuds fixés d'avance, chaque étape de cet algorithme consiste à l'ajout, la suppression ou le déplacement d'un noeud où chacun de ces mouvements se voit associé à une certaine probabilité de transition. La procédure se poursuit jusqu'à l'atteinte d'une distribution stationnaire et la configuration des noeuds peut alors être choisie en prenant, par exemple, le mode de la chaîne. Notons toutefois que l'application de cette technique, même si elle est parfois bien pratique, n'est pas nécessaire ici, puisque les fonctions que nous devons estimer sont plutôt lisses en général. Dans notre contexte, nous adoptons une approche de sélection de noeuds de type « à rebours » qui commence par placer un noeud entre chaque observation, mais qui en enlève un à la fois jusqu'à l'obtention d'un emplacement des noeuds optimal par rapport aux deux critères indiqués ci-haut, c'est-à-dire la flexibilité et un nombre de noeuds minimal. Pour ce faire, nous devons sélectionner un critère sur lequel nous baser pour enlever un noeud. Idéalement, ce dernier doit tenir compte de la qualité de l'estimation de la fonction en donnant une pénalité pour le nombre de noeuds afin de ne pas trop en conserver.

Nous en avons considéré deux qui répondent à ces exigences : le critère proposé par He et Shi (1998) qui a été conçu spécialement pour la sélection des noeuds

dans les bases splines ainsi que le Bayesian Information Criterion (*BIC*) (voir Schwarz, 1978) qui est bien connu en plus de pouvoir être utilisé à plusieurs sauces. Les voici :

(1) Critère de He et Shi :

$$IC(z) = \log \left( \sum_{i=1}^n |y_i - \hat{g}(x_i)| \right) + 2(z + 2)/n,$$

où  $n$  est le nombre d'observations et  $z$  est le nombre de paramètres estimés.

(2) Critère *BIC* :

$$BIC = n \log \left( \frac{\sum_{i=1}^n (y_i - \hat{g}(x_i))^2}{n} \right) + z \log(n), \quad (1.2.2)$$

où  $n$  et  $z$  sont définis de la même façon que pour le critère précédent.

Pour l'instant, nous nous arrêtons à la définition de ces critères. Leur application sera mise en oeuvre au troisième chapitre, lorsque nous estimerons certaines fonctions en utilisant un jeu de données réelles. Ceci met donc fin à notre discussion sur les splines. La dernière section de ce chapitre est consacrée à l'exploration des bases d'ondelettes. Depuis une quinzaine d'années, la popularité de ces dernières a considérablement augmenté dans le domaine qui nous intéresse, c'est-à-dire l'estimation de fonctions, puisque leur application s'est avérée très bien fonctionner. C'est pourquoi nous désirons les utiliser dans nos analyses. Cependant, nous emploierons plutôt une certaine modification de ces fonctions. En fait, puisque les courbes de croissance sont des fonctions monotones croissantes, l'idéal serait d'avoir accès à une base d'ondelettes qui possède ces propriétés lors de nos estimations, comme c'est le cas pour les I-splines. Malheureusement, il n'en existe pas. Il est néanmoins possible, moyennant quelques transformations, de construire une telle base à partir des ondelettes. Cependant, la base résultante fait partie de la classe des ondelettes biorthogonales qui diffère de celle des ondelettes simples. Ainsi, nous introduisons les bases d'ondelettes brièvement, mais suffisamment pour être en mesure d'expliquer les différences entre ces dernières et les bases d'ondelettes biorthogonales. Nous terminons ensuite cette présentation en discutant de la première transformation que nous effectuons, soit celle qui nous permet d'obtenir des fonctions positives sur tout leur domaine.

### 1.3. LES BASES D'ONDELETTES

Ces bases peuvent être vues comme une version améliorée des bases orthogonales classiques (les séries de Fourier, par exemple), puisqu'elles ont généralement la capacité de saisir le comportement local et global des fonctions plus efficacement que ces dernières. Nous devons la toute première base d'ondelettes à Alfred Haar (1910). Il a toutefois fallu attendre au début des années 1980 pour que Grossman et Morlet (1984) définissent les ondelettes dans un contexte de physique quantique. Cette définition était cependant assez vague et elle a depuis beaucoup évolué, surtout grâce aux travaux de Mallat (1989), qui a fait le lien entre les algorithmes pyramidaux rencontrés dans la théorie du traitement des signaux et les ondelettes, de Meyer (1990, 1992), qui a construit la première base d'ondelettes non triviales, et aussi, de Daubechies (1992), qui a découvert les bases dont tous les éléments ont un support compact. De nos jours, l'application des ondelettes se retrouve dans des domaines très variés. Nous les rencontrons particulièrement dans le traitement des signaux, dans la compression de données et dans l'estimation non paramétrique de données. De plus, dans les années 1990, Donoho et Johnstone (1994, 1995, 1998) ont introduit les ondelettes à une majorité de statisticiens en démontrant leur efficacité à traiter les problèmes de régression et d'estimation de densité dans un contexte non paramétrique. Leur plus grande contribution consiste en la mise au point de techniques de rétrécissement (« shrinkage ») et de seuillage (« thresholding ») qui sont aujourd'hui largement répandues. Pour de plus amples informations quant à la théorie des ondelettes et à leurs applications, nous suggérons Meyer (1993) et Walnut (2002). D'autre part, Härdle *et al.* (1998) ainsi que Vidakovic (1999) font une très bonne synthèse des applications qui existent en statistique. De notre côté, nous faisons une révision rapide des concepts permettant de comprendre la façon de générer une base d'ondelettes.

#### 1.3.1. Bases d'ondelettes : propriétés générales

Il existe plusieurs approches pour construire une base d'ondelettes. Nous présentons celle de Mallat (1989) et Meyer (1990) qui est basée sur la notion d'analyse multirésolution. En fait, les fonctions  $\psi_{j,k}$  qui forment les ondelettes sont dérivées

d'une autre fonction, notée  $\phi$  et appelée fonction d'échelle. Sous cette approche, cette dernière est définie selon le concept d'analyse multirésolution que voici.

**Définition 1.3.1.** Une analyse multirésolution de  $\mathcal{L}_2(\mathbb{R})$  (une fonction  $f$  est dans l'espace  $\mathcal{L}_2(S)$  si  $\int_S f^2(x)dx < \infty$ ) est une suite  $\{V_j\}_{j \in \mathbb{Z}}$  de sous-espaces fermés de  $\mathcal{L}_2(\mathbb{R})$  telle que

$$(i) V_j \subset V_{j+1}, \quad \forall j \in \mathbb{Z},$$

$$(ii) \bigcup_{j \in \mathbb{Z}} V_j \text{ est dense dans } \mathcal{L}_2(\mathbb{R}),$$

$$(iii) \bigcap_{j \in \mathbb{Z}} V_j = \{0\},$$

$$(iv) f(x) \in V_j \iff f(2x) \in V_{j+1}, \quad \forall j \in \mathbb{Z},$$

$$(v) f(x) \in V_j \iff f(x - 2^{-j}k) \in V_j, \quad \forall j, k \in \mathbb{Z},$$

(vi)  $\exists$  une fonction  $h \in V_0$  telle que  $\{h(x - k)\}_{k \in \mathbb{Z}}$  est une base orthonormée de  $V_0$ , c'est-à-dire que  $V_0$  peut s'écrire comme

$$V_0 = \left\{ f \in \mathcal{L}_2(\mathbb{R}) \mid f(x) = \sum_{k \in \mathbb{Z}} c_k h(x - k), \sum_{k \in \mathbb{Z}} c_k^2 < \infty \right\}.$$

Nous pouvons alors énoncer la définition d'une fonction d'échelle.

**Définition 1.3.2.** Une fonction  $\phi \in \mathcal{L}_2(\mathbb{R})$  est dite une fonction d'échelle si elle satisfait les deux propriétés suivantes :

(i) les sous-espaces  $V_j$  de  $\mathcal{L}_2(\mathbb{R})$  définis par

$$V_j = \left\{ f \in \mathcal{L}_2(\mathbb{R}) \mid f(x) = \sum_{k \in \mathbb{Z}} c_{j,k} \phi_{j,k}(x) \right\},$$

où

$$\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k), \quad \forall j, k \in \mathbb{Z},$$

forment une analyse multirésolution de  $\mathcal{L}_2(\mathbb{R})$ ,

(ii) la suite  $\{\phi(x - k)\}_{k \in \mathbb{Z}}$  forme une base orthonormée de  $V_0$ , c'est-à-dire que  $\phi$  satisfait la propriété (vi) de la définition 1.3.1.

Nous disons alors que  $\{V_j\}_{j \in \mathbb{Z}}$  est l'analyse multirésolution engendrée par  $\phi$ .

Étudions maintenant le lien entre la fonction d'échelle  $\phi$  et une autre fonction  $\psi$  appelée mère des ondelettes. À l'aide d'un exemple tiré de Ouellette (2004), illustrons en fait la procédure nous permettant d'obtenir cette fonction  $\psi$  à partir de la fonction d'échelle  $\phi$ .

Posons  $\{V_j\}_{j \in \mathbb{Z}}$ , une analyse multirésolution engendrée par la fonction d'échelle  $\phi$ . Par la propriété (i) de la définition 1.3.1, nous avons

$$\phi \in V_0 \subset V_1,$$

et la fonction  $\phi$  peut être décomposée en séries à partir des éléments de la base de  $V_1$  dénotés  $\{\phi_{1,k}\}_{k \in \mathbb{Z}}$ . En effet, d'après la propriété (i) de la définition 1.3.2, il existe une suite  $\{A_k\}_{k \in \mathbb{Z}}$  de coefficients de carré sommable (c'est-à-dire  $\sum_{k \in \mathbb{Z}} A_k^2 < +\infty$ ) telle que

$$\phi(x) = \sum_{k \in \mathbb{Z}} A_k \phi_{1,k}(x). \quad (1.3.1)$$

De plus,

$$\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k), \quad \forall j, k \in \mathbb{Z}.$$

Nous obtenons alors, pour  $j = 1$ ,

$$\phi_{1,k}(x) = \sqrt{2} \phi(2x - k), \quad \forall k \in \mathbb{Z}.$$

L'équation (1.3.1) devient ainsi

$$\begin{aligned} \phi(x) &= \sum_{k \in \mathbb{Z}} \sqrt{2} A_k \phi(2x - k) \\ &= \sum_{k \in \mathbb{Z}} a_k \phi(2x - k). \end{aligned} \quad (1.3.2)$$

Cette dernière équation est appelée équation de dilatation (voir Müller et Vidakovic, 1999A et 1999B).

À l'aide d'un développement similaire, nous pouvons obtenir une équation semblable à l'équation (1.3.2) pour la mère des ondelettes. Définissons maintenant  $W_0$  comme étant le complément orthogonal de  $V_0$  dans  $V_1$ . Si

$$\psi \in W_0 \subset V_1,$$

alors nous pouvons également développer  $\psi$  en séries à l'aide des éléments de la base de  $V_1$ . Notons par  $\{B_k\}_{k \in \mathbb{Z}}$ , la suite des coefficients de carré sommable telle

que

$$\begin{aligned}
 \psi(x) &= \sum_{k \in \mathbb{Z}} B_k \phi_{1,k}(x) \\
 &= \sum_{k \in \mathbb{Z}} \sqrt{2} B_k \phi(2x - k) \\
 &= \sum_{k \in \mathbb{Z}} b_k \phi(2x - k). \tag{1.3.3}
 \end{aligned}$$

Les coefficients  $a_k$  et  $b_k$ ,  $k \in \mathbb{Z}$  des équations (1.3.2) et (1.3.3) sont nommés filtres et si  $\phi$  est connue, alors la suite de filtres  $\{a_k\}_{k \in \mathbb{Z}}$  l'est également. La mère des ondelettes est construite à partir des filtres de la fonction d'échelle de façon à respecter la contrainte d'orthogonalité imposée par l'analyse multirésolution (propriété (vi) de la définition 1.3.1). Pour ce faire, le choix de la suite  $\{b_k\}_{k \in \mathbb{Z}}$  n'est pas unique. Il est, en effet, possible de construire, à partir d'une seule fonction d'échelle, plusieurs fonctions différentes pouvant agir comme mère des ondelettes. Remarquons cependant que, si nous choisissons

$$b_k = (-1)^k a_{1-k},$$

alors la condition d'orthonormalité est respectée quelle que soit la suite  $\{a_k\}_{k \in \mathbb{Z}}$  (voir Rao et Bopardikar, 1998). Ainsi, l'équation (1.3.3) devient

$$\psi(x) = \sum_{k \in \mathbb{Z}} (-1)^k a_{1-k} \phi(2x - k),$$

qui est une solution générale à la construction de  $\psi$ .

Notons finalement que les fonctions  $\psi_{j,k}$  sont définies par

$$\psi_{j,k} = 2^{j/2} \psi(2^j x - k), \quad \forall j, k \in \mathbb{Z}.$$

Nous avons donc donné un petit aperçu de ce que représente la construction d'une base d'ondelettes. Nous n'ajoutons pas plus de détails puisque, comme nous l'avons déjà indiqué, nous ne travaillons pas directement avec ces bases. Dans la prochaine sous-section, nous précisons les équations (1.3.2) et (1.3.3) dans le cas particulier des ondelettes de Daubechies, qui sont celles que nous employons afin de construire notre base biorthogonale.

### 1.3.2. Les ondelettes de Daubechies

Ce qui fait la particularité de la base d'ondelettes de Daubechies, c'est qu'elle respecte la contrainte suivante : sa suite de filtres contient un nombre fini d'éléments non nuls. Ainsi, tous les éléments orthogonaux de cette base ont un support compact et un certain ordre de régularité. Il existe toute une classe de fonctions d'échelle de différents ordres, une fonction d'ordre plus élevé correspondant à une suite de filtres dont le nombre de coefficients est plus grand.

Dans la base d'ondelettes de Daubechies d'ordre  $q$ , la fonction d'échelle  $\phi(x)$  est définie par

$${}_q\phi(x) = \sum_{k=0}^{2q-1} a_k \times {}_q\phi(2x - k),$$

et son support est  $[0, 2q - 1)$  (voir Alpert, 1992). La mère des ondelettes  $\psi(x)$  correspondante est définie par

$${}_q\psi(x) = \sum_{k=2-2q}^1 (-1)^k a_{1-k} \times {}_q\phi(2x - k),$$

et son support est  $[1 - q, q)$ .

Une ondelette de Daubechies d'ordre  $q$  possède  $2q$  coefficients non nuls dans sa suite de filtres et les valeurs de ces filtres peuvent être calculées à l'aide d'équations récursives (voir Leblanc, 2001). Lorsqu'elles sont connues, l'algorithme pyramidal de Daubechies-Lagarias permet alors d'évaluer les fonctions  ${}_q\phi$  et  ${}_q\psi$  à un certain point  $x$ , avec un degré de précision prédéterminé (voir Vidakovic, 1999). Remarquons finalement que la base d'ondelettes de Daubechies d'ordre un correspond à la base de Haar. Pour ce cas particulier, les équations (1.3.2) et (1.3.3) peuvent s'écrire sous la forme analytique

$$\phi(x) = \begin{cases} 1, & \text{si } 0 \leq x < 1, \\ 0, & \text{sinon,} \end{cases}$$

$$\psi(x) = \begin{cases} 1, & \text{si } 0 \leq x < \frac{1}{2}, \\ -1, & \text{si } \frac{1}{2} \leq x < 1, \\ 0, & \text{sinon.} \end{cases}$$

La figure 1.3 présente la fonction d'échelle et la mère des ondelettes pour cette base ainsi que dans le cas où  $q = 2$ .

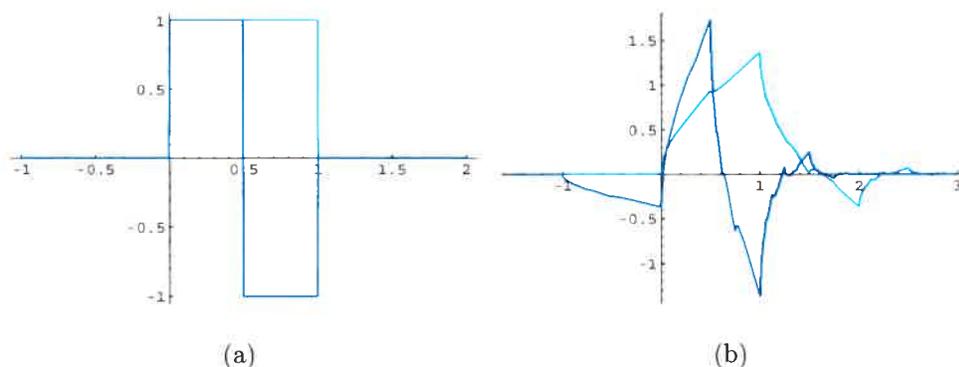


FIG. 1.3. Exemples de la fonction d'échelle (bleu pâle) et de la mère des ondelettes (bleu foncé) (a) Ordre 1 (b) Ordre 2

Nous remarquons que la fonction d'échelle  ${}_1\phi$  semble plus symétrique que  ${}_2\phi$ . En fait, Daubechies (1992) s'est intéressée à la régularité de la famille de fonctions d'échelle qu'elle a créée et elle a démontré que la fonction d'échelle de Haar est la seule à être symétrique et à posséder un support compact simultanément.

Ceci clôt la section dédiée aux ondelettes de Daubechies. Bien que nous n'utilisons pas cette base directement lors de nos estimations, nous montrons, à la section suivante, comment décomposer une fonction à l'aide d'ondelettes. C'est ce qui est fait en général et cela nous aide à comprendre de quelle façon cette décomposition peut être faite lorsque nous sommes en présence d'une base biorthogonale.

### 1.3.3. Décomposition d'une fonction à l'aide d'ondelettes

Nous savons que, si  $\{w_k(x)\}_{k \in \mathbb{Z}}$  est une suite orthonormée de fonctions constituant une base de  $\mathcal{L}_2(\mathbb{R})$ , alors toute fonction  $g \in \mathcal{L}_2(\mathbb{R})$  peut être exprimée sous la forme

$$g(x) = \sum_{k \in \mathbb{Z}} c_k w_k(x),$$

où les coefficients  $c_k$  sont donnés par

$$c_k = \langle g, w_k \rangle = \int_{-\infty}^{\infty} g(x) w_k(x) dx, \quad \forall k \in \mathbb{Z},$$

et où  $x$  fait partie d'un sous-ensemble borné de  $\mathbb{R}$  (voir Müller et Vidakovic, 1999A).

Appliquons ce résultat aux bases d'ondelettes. Nous pouvons démontrer (voir Leblanc, 2001) que, pour tout  $J \in \mathbb{Z}$ ,  $\{\phi_{J,k}(x), \psi_{j,k}(x)\}_{j \geq J, k \in \mathbb{Z}}$  constitue une base orthonormée de  $\mathcal{L}_2(\mathbb{R})$ .

Ainsi, toute fonction  $g \in \mathcal{L}_2(\mathbb{R})$  peut être écrite sous la forme

$$g(x) = \sum_{k \in \mathbb{Z}} \alpha_{J,k} \phi_{J,k}(x) + \sum_{j=J}^{\infty} \sum_{k \in \mathbb{Z}} \beta_{j,k} \psi_{j,k}(x), \quad (1.3.4)$$

où

$$\alpha_{Jk} = \langle g, \phi_{J,k} \rangle = \int_{\mathbb{R}} g(x) \phi_{J,k}(x) dx,$$

et

$$\beta_{j,k} = \langle g, \psi_{j,k} \rangle = \int_{\mathbb{R}} g(x) \psi_{j,k}(x) dx.$$

Notons que, dans l'équation (1.3.4), la fonction est décomposée à partir de la fonction d'échelle et des ondelettes. Par contre, dans le cas des ondelettes orthogonales, les ondelettes  $\{\psi_{j,k}\}_{k \in \mathbb{Z}}$  forment une base orthonormée de  $W_j$  et les fonctions  $\{\phi_{j,k}\}_{k \in \mathbb{Z}}$  forment également une base orthonormée de  $V_j$ . En utilisant le même principe que précédemment, il est donc possible de décomposer une fonction à l'aide de l'une ou l'autre de ces bases. C'est ce que font Chipman *et al.* (1997), qui n'utilisent que les fonctions  $\{\psi_{j,k}\}_{k \in \mathbb{Z}}$ . Dans cette publication, une seule fonction  $g$  est considérée et les observations  $x_i$  sont supposées équidistantes. Un modèle *a priori* correspondant à un mélange de densités normales centrées à 0 est suggéré pour les coefficients. Une étude de simulation compare alors les estimateurs des coefficients d'ondelettes trouvés par les auteurs à ceux proposés par Donoho et Johnstone (1994), et ce, par le biais de quelques fonctions introduites par ces derniers.

L'introduction aux ondelettes étant terminée, regardons en quoi consiste une base d'ondelettes biorthogonales.

### 1.3.4. Les ondelettes biorthogonales

La première base d'ondelettes biorthogonales a été construite par Tchamitchian (1987). Le développement de ces bases est principalement dû au fait qu'il est

impossible de combiner support compact, orthogonalité et symétrie lors de la construction d'une base d'ondelettes de la manière décrite à la section 1.3.1. Cohen, Daubechies et Feauveau (1992) ont donc introduit une classe d'ondelettes plus générale qui allie symétrie et support compact, aux dépens de l'orthogonalité. Dans cette section, nous présentons la théorie liée à ces bases à la manière de Ogden (1997).

Il peut sembler curieux d'utiliser le terme « biorthogonal » pour ce type de base en sachant qu'elle ne respecte pas la propriété d'orthogonalité. En fait, cette expression fait référence à la présence de deux analyses multirésolution différentes dans  $\mathcal{L}_2(\mathbb{R})$ . Nous avons ainsi

$$\dots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset \dots,$$

et

$$\dots \subset \tilde{V}_{-2} \subset \tilde{V}_{-1} \subset \tilde{V}_0 \subset \tilde{V}_1 \subset \dots,$$

la première étant générée par une fonction d'échelle  $\phi$ , alors que la seconde est générée par une fonction d'échelle complémentaire  $\tilde{\phi}$ . De plus, comme dans le cas orthogonal, ces séquences sont respectivement associées aux espaces  $(W_j)_{j \in \mathbb{Z}}$  et  $(\tilde{W}_j)_{j \in \mathbb{Z}}$ , qui sont générés par deux fonctions  $\psi$  et  $\tilde{\psi}$  qui sont aussi complémentaires.

La biorthogonalité est exprimée à travers les relations entre les analyses multirésolution

$$V_j \perp \tilde{W}_j, \quad j \in \mathbb{Z},$$

et

$$\tilde{V}_j \perp W_j, \quad j \in \mathbb{Z}.$$

En termes de fonctions d'échelle et d'ondelettes, cette biorthogonalité prend la forme

$$\phi_{j,k} \perp \psi_{j',k'},$$

$$\phi_{j,k} \perp \tilde{\phi}_{j,k'}, \quad \forall k \neq k',$$

et

$$\psi_{j,k} \perp \tilde{\psi}_{j,k}, \quad \forall j \neq j' \text{ ou } k \neq k'.$$

De plus, comme dans le cas orthogonal, la décomposition de la fonction  $g \in \mathcal{L}_2(\mathbb{R})$  peut s'écrire comme

$$g(x) = \sum_k c_{j,k} \phi_{j,k}(x),$$

mais les coefficients sont maintenant calculés en termes de la fonction d'échelle complémentaire (voir Ogden, 1997) :

$$c_{j,k} = \langle g, \tilde{\phi}_{j,k} \rangle = \int_{-\infty}^{\infty} g(x) \tilde{\phi}_{j,k}(x) dx.$$

Dans le même ordre d'idées, nous pouvons décomposer cette fonction en n'employant que les ondelettes, ce qui donne

$$g(x) = \sum_j \sum_k d_{j,k} \psi_{j,k},$$

où les coefficients sont donnés par

$$d_{j,k} = \langle g, \tilde{\psi}_{j,k} \rangle = \int_{-\infty}^{\infty} g(x) \tilde{\psi}_{j,k}(x) dx.$$

Mentionnons que la représentation de la fonction  $g$  peut aussi se faire à l'aide du système d'ondelettes complémentaires  $\{\tilde{\psi}_{j,k}\}_{k \in \mathbb{Z}}$ . Cette dernière est similaire à celle susmentionnée, les coefficients étant définis en termes du produit scalaire de  $g$  avec les  $\psi_{j,k}$  correspondants. Nous pouvons donc voir qu'une ondelette orthogonale est un cas particulier des ondelettes biorthogonales pour lequel  $\tilde{\phi} \equiv \phi$  et  $\tilde{\psi} \equiv \psi$ .

Rappelons que dans le cas des ondelettes orthogonales, les ondelettes  $\{\psi_{j,k}\}_{k \in \mathbb{Z}}$  forment une base orthonormée de  $W_j$  et les fonctions  $\{\phi_{j,k}\}_{k \in \mathbb{Z}}$  forment également une base orthonormée de  $V_j$ . En généralisant ce principe aux ondelettes biorthogonales, ces ensembles ne forment plus une base orthonormée pour leurs espaces respectifs, mais bien une base de Riesz dont voici la définition.

**Définition 1.3.3.** *L'ensemble  $\{\phi(\cdot - k), k \in \mathbb{Z}\}$  forme une base de Riesz pour l'espace  $V_0$  si les fonctions  $\{\phi(\cdot - k), k \in \mathbb{Z}\}$  sont éléments de  $V_0$  et qu'il existe des constantes  $0 < A \leq B < \infty$  telles que, pour toutes les séquences  $\{p_k\}_{k \in \mathbb{Z}} \in l^2(\mathbb{Z})$ ,*

nous avons

$$A \sum_k |p_k|^2 \leq \left\| \sum_k p_k \phi(\cdot - k) \right\|^2 \leq B \sum_k |p_k|^2.$$

Avec ce survol des ondelettes biorthogonales, nous sommes en mesure d'expliquer la première transformation que nous faisons à partir des ondelettes de Daubechies afin d'obtenir des fonctions qui sont toujours positives ou nulles. La prochaine sous-section est dédiée à cette procédure.

### 1.3.5. Les ondelettes non négatives

Il n'existe pas de fonction d'échelle non négative et continue qui génère une analyse multirésolution orthogonale (voir Vidakovic, 1999). Par contre, en s'intéressant au problème de l'estimation de densités, Walter et Shen (1999) ont proposé la construction d'un système biorthogonal dont les fonctions d'échelle primaires sont des densités. En voici un résumé.

Choisissons  $\phi(t)$ , une fonction d'échelle orthogonale qui a un support compact, et  $r$ , une valeur comprise entre 0 et 1 qui est suffisamment grande pour nous assurer que la fonction

$$\rho_r(t) = \sum_m r^{|m|} \phi(t - m), \quad (1.3.5)$$

soit non négative. Posons également

$$\tilde{\rho}_r(t) = \frac{(1 + r^2)\phi(t) - r[\phi(t + 1) + \phi(t - 1)]}{2\pi(1 - r^2)}.$$

Nous obtenons alors une base de fonctions d'échelle biorthogonales

$$\{2^{j/2}\rho_r(2^j t - k), k \in \mathbb{Z}\} \text{ et } \{2^{j/2}\tilde{\rho}_r(2^j t - k), k \in \mathbb{Z}\},$$

associées aux espaces  $V_j$  et  $\tilde{V}_j = V_j$  respectivement. Remarquons qu'il existe toujours une certaine valeur de  $r$  pour laquelle la fonction (1.3.5) est positive ou nulle. En effet, comme le mentionne Vidakovic (1999), puisque

$$\sum_{i=-\infty}^{\infty} \phi(t - i) = 1, \quad t \in \mathbb{R},$$

il s'ensuit que  $\rho_r(t) \rightarrow 1$  quand  $r \rightarrow 1$ . De plus, comme la convergence est uniforme sur des ensembles compacts, nous pouvons toujours trouver une valeur critique  $r^*$  telle que pour tout  $t$ ,  $\rho_r(t) \geq 0$  lorsque  $r \geq r^*$ .

Pour construire leur estimateur, les auteurs appliquent alors la théorie usuelle portant sur l'utilisation de noyaux lors de l'approximation de fonctions. Exposons-la brièvement à la façon de Vidakovic (1999).

Commençons par donner les propriétés associées au noyau  $\mathbb{K}$ .

**Définition 1.3.4.** Une fonction de deux variables  $x$  et  $y$ ,  $\mathbb{K}(x, y)$ , est appelée noyau reproducteur de la fonction pour l'espace fonctionnel  $\mathcal{H}$  si

(1) pour un  $y$  fixé,  $\mathbb{K}(x, y)$  est une fonction dans  $\mathcal{H}$ .

(2) Pour toute fonction  $f \in \mathcal{H}$  et tout  $y$ ,  $\mathbb{K}$  a la propriété de reproduction, c'est-à-dire

$$f(y) = \langle f(x), \mathbb{K}(x, y) \rangle.$$

La construction d'un tel noyau est facilement réalisable en présence d'une base orthonormée de fonctions, ce qui est expliqué dans le théorème qui suit.

**Théorème 1.3.1.** Soient  $V$ , un sous-espace de  $\mathcal{L}_2$  et  $\{e_1, e_2, \dots\}$ , une base orthonormée complète de  $V$ . Alors,  $V$  est un espace du noyau reproducteur de Hilbert et il est associé au noyau

$$\mathbb{K}(x, y) = \sum_l e_l(x)e_l(y). \quad (1.3.6)$$

De plus, pour toute fonction  $f \in V$ ,

$$f(y) = \int \mathbb{K}(x, y)f(x)dx. \quad (1.3.7)$$

Walter et Shen (1999) ont donc adapté ce principe aux fonctions  $\rho_r(t)$ . En fait, nous savons que toute fonction d'échelle ayant un support compact respecte la propriété suivante (voir Chui, 1992) :

$$\int_X \phi(x)dx = 1.$$

En utilisant cette propriété, nous obtenons ainsi que

$$\int \sum_n r^{|n|} \phi(t-n)dt = 2 \sum_{i=0}^{\infty} r^i - 1 = \frac{1+r}{1-r}, \quad (1.3.8)$$

et le noyau associé à  $V_0$  et introduit à l'équation (1.3.6) peut être défini par

$$\mathbb{K}_r(t, s) = \left( \frac{1-r}{1+r} \right)^2 \sum_{n \in \mathbb{Z}} \rho_r(t-n)\rho_r(s-n),$$

alors que celui associé à  $V_j$  est

$$\mathbb{K}_{r,j}(s, t) = 2^j \mathbb{K}_r(2^j s, 2^j t), \quad j \in \mathbb{Z}. \quad (1.3.9)$$

Notons que ce sont les fonctions  $\left(\frac{1-r}{1+r}\right) \rho_r(t)$  qui sont des densités.

Dans le but de construire une base de fonctions monotones croissantes, nous créons ce que nous appelons des ondelettes intégrées en intégrant la fonction  $\rho_r(t)$  et nous utilisons alors l'équivalent du noyau donné à l'équation (1.3.9). La procédure exacte sera donnée au second chapitre, qui décrit les modèles que nous supposons pour les coefficients  $\lambda_k$  définis en (1.0.2) et les estimateurs trouvés à l'aide de ces derniers.

## Chapitre 2

---

# MODÉLISATION D'UNE COURBE DE CROISSANCE DE RÉFÉRENCE POUR UNE POPULATION

Au début du premier chapitre, nous avons considéré le modèle

$$y_{ij} = g_i(x_{ij}) + \varepsilon_{ij}, \quad i = 1, \dots, N, \quad j = 1, \dots, n, \quad (2.0.1)$$

où  $g_i$  sont des fonctions inconnues à estimer, les  $x_{ij}$  appartiennent à un ensemble borné  $\mathcal{X} \subset \mathbb{R}$  et les  $\varepsilon_{ij}$  représentent les termes d'erreurs. Mentionnons que, bien qu'il soit possible de faire le contraire, nous supposons que la variance associée aux erreurs est la même pour tous les enfants de la population étudiée.

Notons cependant que nous ne nous intéressons pas à toutes ces fonctions. En effet, pour une population d'enfants donnée, nous nous attardons plutôt à la construction d'une courbe de croissance qui soit représentative de l'ensemble des  $N$  courbes observées. C'est cette courbe de référence que nous estimons alors.

Dans ce second chapitre, nous commençons donc par décrire la méthode utilisée pour définir la courbe de référence avant de fournir les détails quant à l'estimation de celle-ci.

### 2.1. SYNCHRONISATION ÉVÉNEMENTIELLE

Il n'est pas rare de vouloir résumer l'information contenue dans un jeu de données en ne conservant que quelques statistiques que nous pensons représentatives de la population. Lorsque nous travaillons avec des courbes, l'important est de

retrouver tous les attributs communs à la majorité de celles-ci dans la courbe finale. Pour les courbes de croissance, ces particularités sont dues aux poussées de croissance parmi lesquelles nous trouvons celle au début de la puberté qui est généralement la plus évoquée dans la littérature (voir Gasser *et al.*, 1984). L'étalonnage de ces caractéristiques saillantes à l'aide de transformations monotones constitue donc un prérequis essentiel au développement d'une courbe de référence. Il découle d'une décomposition en deux parties de la variation entre les fonctions observées : une phase (âge auquel l'événement se produit) et une amplitude (taille des enfants) (voir Ramsay et Silverman, 2002). La figure 2.1 illustre bien ce principe avec la fonction sinus. En présence de telles fonctions, il est raisonnable de les comparer aux temps  $x_1$  et  $x_2$  auxquels deux sujets sont dans des états similaires plutôt qu'à un temps fixé  $x$ . L'étalonnage vise donc à éliminer la première de ces deux sources de variation, ce qui permet de concentrer l'analyse sur la seconde par la suite, et ainsi, créer une courbe de référence dépeignant bien les courbes observées.

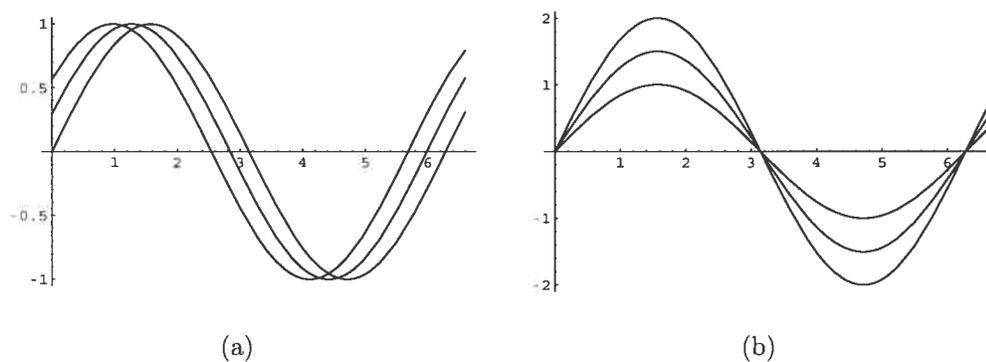


FIG. 2.1. Exemples de variation avec la fonction sinus (a) Phase (b) Amplitude

Expliquons donc, en termes plus mathématiques, le principe de la synchronisation étudié par plusieurs auteurs dont Sakoe et Chiba (1978) en ingénierie, mais surtout dans un contexte statistique par Kneip et Gasser (1988, 1992). Nous avons un échantillon de  $N$  fonctions  $g_i$  définies sur un même intervalle  $[T_0, T_n]$ . Cette procédure consiste à passer, pour chaque courbe, de l'âge original  $x$  au temps synchronisé  $\tau$ , en utilisant une fonction de transformation  $h_i(x)$ . En d'autres mots,

nous passons des fonctions  $g_i(x)$  aux fonctions

$$g_i(h_i(x)) = g_i(\tau),$$

où  $g_i(h_i(x))$  représente la courbe de croissance synchronisée pour l'enfant  $i$ .

Pour que les transformations soient bijectives, et donc que, pour tout  $y \in [T_0, T_n]$ , il existe une unique valeur de  $x$  pour laquelle  $h_i(x) = y$ , les fonctions  $h_i$  doivent être monotones croissantes. Ainsi, nous désirons que

$$h_i(x_1) > h_i(x_2) \iff x_1 > x_2.$$

Pour ce faire, nous respectons une série de contraintes prenant la forme

$$h_i(x_\nu) = \tau_\nu,$$

où  $x_\nu$  est le temps auquel l'événement  $\nu$  se produit pour l'enfant  $i$  et  $\tau_\nu$  est le temps étalonné auquel il survient pour tous les enfants dans le domaine transformé. C'est ce que nous appelons la synchronisation événementielle qui consiste en l'alignement des courbes en se basant sur une série d'événements marquants. Bien sûr, certaines méthodes alternatives ont été élaborées afin de synchroniser différentes courbes. Par exemple, Silverman (1995) a développé une technique qui ne requiert pas la spécification d'événements marquants. Il optimise plutôt un critère d'ajustement global en tenant compte d'une famille paramétrique de transformations du temps. Ramsay (1998) a également développé une approche de synchronisation non paramétrique où le degré de lissage de la transformation est contrôlé par un terme de pénalité. Ces techniques, plus complexes que celle que nous décrivons, peuvent être d'une grande utilité quand les événements marquants ne se retrouvent pas (ou sont très faiblement visibles) sur une grande partie des courbes que nous désirons synchroniser. Cependant, puisque nous ne nous trouvons pas dans cette situation, nous préférons utiliser la technique qui offre la meilleure combinaison entre simplicité et efficacité dans notre cas, c'est-à-dire la synchronisation événementielle.

La fonction  $h_i$  peut être modélisée de différentes façons. Mentionnons notamment l'utilisation d'une approximation en séries de Taylor (voir Angers *et al.*, 2004), de splines d'interpolation (voir Kneip et Gasser, 1992) ou d'une approche

telle que celle proposée par Ramsay et Li (1998) qui repose sur la minimisation d'un critère d'ajustement. Ici, nous considérons une technique très simple employée par Merleau *et al.* (2007), qui consiste à exprimer la fonction de synchronisation  $h_i(x)$  en une série de fonctions linéaires  $L_p(x)$ . Nous obtenons ainsi

$$h_i(x) = \sum_{p=1}^P L_p(x) I_{D_p}(x) = \sum_{p=1}^P (c_{p,0} + c_{p,1}x) I_{D_p}(x),$$

où  $D_p$  est le domaine de chacune des fonctions linéaires  $L_p(x) = c_{p,0} + c_{p,1}x$ ,  $P$  est le nombre de parties de la fonction de synchronisation et

$$I_{D_p}(x) = \begin{cases} 1, & \text{si } x \in D_p, \\ 0, & \text{sinon.} \end{cases}$$

Ce modèle possède une solution exacte quand une condition de continuité est imposée sur la fonction de synchronisation. Il satisfait également l'hypothèse de monotonie tant que l'ordre des événements est le même pour tous les enfants.

Pour obtenir la courbe de référence à la suite de ces transformations, il suffit de faire la moyenne, à chacun des temps, de toutes les hauteurs pour chacun des enfants. Cependant, puisque les transformations par rapport aux temps sont faites pour chacun des enfants séparément, les temps transformés sont maintenant différents d'un enfant à l'autre. Ainsi, il nous faut procéder par interpolation afin de retrouver la hauteur moyenne approximative pour chacun des temps originaux  $x_i$ ,  $i = 1, \dots, n$  (voir Merleau *et al.*, 2007). Bien sûr, ceci nous fait perdre un peu de précision, mais c'est une façon simple d'établir la courbe moyenne pour une population donnée. Nous détenons alors les couples de données suivants :

$$(x_i, \bar{y}_i), \quad i = 1, \dots, n, \quad (2.1.1)$$

où  $x_i$  représente l'âge à la  $i^e$  mesure,  $\bar{y}_i$  est la taille moyenne des enfants à ce moment et  $n$  est le nombre d'observations associé à la courbe de référence. Notons que la taille moyenne dont nous parlons ici est celle obtenue après interpolation linéaire. Pour ces données, le modèle (2.0.1) s'écrit sous la forme

$$\bar{y}_i = g(x_i) + \varepsilon_i, \quad (2.1.2)$$

où  $\varepsilon_i \sim N(0, \sigma^2)$ ,  $i = 1, \dots, n$ . C'est cette dernière fonction  $g$  que nous exprimons à l'aide d'une base de fonctions.

Soulignons finalement qu'à partir de maintenant et jusqu'à la fin de ce mémoire, nous présumons que les fonctions sont déjà synchronisées. Rappelons aussi que nous pourrions tenter d'estimer la courbe de référence à l'aide de toute base de fonctions connue telle qu'une base polynomiale qui fait partie des plus simples. Par contre, nous utilisons deux des caractéristiques des courbes de croissance pour nous guider dans le choix des bases à utiliser, c'est-à-dire le fait qu'elles soient positives et monotones croissantes. Il est alors préférable d'employer des bases qui présentent ces mêmes propriétés. C'est pourquoi notre choix s'est arrêté sur les I-splines ainsi que sur une modification de la base d'ondelettes non négatives. Dans les prochaines sections, nous développons en premier lieu le modèle et les estimations dans un cadre plus général, c'est-à-dire sans l'imposition de contraintes nous certifiant que les fonctions estimées soient bien croissantes. Nous établissons par la suite les estimations en rajoutant ces contraintes, et ce, pour les I-splines d'abord et pour les ondelettes intégrées ensuite. Remarquons que deux modèles ont été développés. Pour le premier, nous ignorons les courbes associées à chacun des enfants et nous travaillons directement avec la courbe moyenne trouvée, tandis que le second tient compte des courbes de chaque enfant. Donc, bien que le but principal ne soit pas d'estimer la courbe d'un enfant en particulier, le deuxième modèle nous permet néanmoins de le faire. De plus, puisqu'ils constituent tous les deux des modèles hiérarchiques bayésiens, il importe d'introduire ce type de modèle avant tout.

## 2.2. MODÈLE HIÉRARCHIQUE DE BAYES ET DÉCOMPOSITIONS CONDITIONNELLES

Débutons d'abord en éclaircissant la notion de modèle statistique de Bayes qui nous servira à comprendre celle de modèle hiérarchique.

**Définition 2.2.1.** *Un modèle statistique de Bayes consiste en un modèle statistique paramétrique  $f(x|\theta)$ , c'est-à-dire une série d'observations qui proviennent d'une variable aléatoire  $X$ , distribuée selon  $f(x|\theta)$ , où le seul paramètre inconnu*

est  $\theta$  qui appartient à un espace fini  $\Theta$ , auquel nous rajoutons une distribution a priori  $\pi(\theta)$  sur le paramètre.

En d'autres mots, nous partons avec un certain niveau d'information sur le paramètre  $\theta$ , que nous incorporons dans la distribution  $\pi(\theta)$ , et nous utilisons le théorème de Bayes donné par

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta},$$

afin d'actualiser nos connaissances sur  $\theta$  en exploitant celles contenues dans les observations  $x$ .

En sachant cela, nous sommes maintenant en mesure de définir un modèle hiérarchique.

**Définition 2.2.2.** *Un modèle de Bayes hiérarchique est un modèle statistique de Bayes  $(f(x|\theta), \pi(\theta))$  dans lequel la distribution a priori  $\pi(\theta)$  est décomposée en distributions conditionnelles*

$$\pi_1(\theta|\theta_1), \pi_2(\theta_1|\theta_2), \dots, \pi_n(\theta_{n-1}|\theta_n),$$

et une distribution marginale  $\pi_{n+1}(\theta_n)$  telle que

$$\pi(\theta) = \int_{\Theta_1 \times \dots \times \Theta_n} \pi_1(\theta|\theta_1)\pi_2(\theta_1|\theta_2) \cdots \pi_n(\theta_{n-1}|\theta_n)\pi_{n+1}(\theta_n)d\theta_1 \cdots d\theta_n. \quad (2.2.1)$$

Les paramètres  $\theta_i$  sont appelés hyperparamètres de niveau  $i$ ,  $i = 1, \dots, n$ .

Ce n'est donc qu'un modèle de Bayes qui prend une forme particulière. En fait, si

$$x \sim f(x|\theta), \theta \sim \pi_1(\theta|\theta_1), \dots, \theta_n \sim \pi_{n+1}(\theta_n),$$

nous retrouvons le modèle bayésien habituel

$$x \sim f(x|\theta), \theta \sim \pi(\theta),$$

ayant pour distribution a priori la densité donnée à l'équation (2.2.1).

Un des aspects les plus attrayants de ce genre de modèle vient du fait qu'il permet le conditionnement sur tous les niveaux, ce qui peut compenser pour l'apparente complexité occasionnée par ces derniers. Illustrons ceci par un exemple.

Si

$$x \sim f(x|\theta), \theta|\theta_1 \sim \pi_1(\theta|\theta_1) \text{ et } \theta_1 \sim \pi_2(\theta_1),$$

alors le résultat suivant s'applique.

**Lemme 2.2.1.** *La distribution a posteriori de  $\theta$  est*

$$\pi(\theta|x) = \int_{\Theta_1} \pi_1(\theta|\theta_1, x)\pi_2(\theta_1|x)d\theta_1,$$

où

$$\pi_1(\theta|\theta_1, x) = \frac{f(x|\theta)\pi_1(\theta|\theta_1)}{m_1(x|\theta_1)},$$

$$m_1(x|\theta_1) = \int_{\Theta} f(x|\theta)\pi_1(\theta|\theta_1)d\theta, \quad (2.2.2)$$

$$\pi_2(\theta_1|x) = \frac{m_1(x|\theta_1)\pi_2(\theta_1)}{m(x)}, \quad (2.2.3)$$

$$m(x) = \int_{\Theta_1} m_1(x|\theta_1)\pi_2(\theta_1)d\theta_1.$$

Ce qui est d'autant plus intéressant, c'est que cette décomposition est valable également pour les moments *a posteriori*, c'est-à-dire que pour toute fonction  $h$ , nous avons

$$\mathbb{E}^\pi[h(\theta)|x] = \mathbb{E}^{\pi_2(\theta_1|x)}[\mathbb{E}^{\pi_1}(h(\theta)|\theta_1, x)], \quad (2.2.4)$$

où  $\mathbb{E}^{\pi_1}(h(\theta)|\theta_1, x) = \int_{\Theta} h(\theta)\pi_1(\theta|\theta_1, x)d\theta$ .

Ce résultat est une conséquence directe du théorème de Bayes, la dernière égalité étant aisément établie à l'aide du théorème de Fubini. Nous le rapportons, car il est d'une grande importance quant au calcul des estimateurs de Bayes, puisqu'il montre que la distribution  $\pi(\theta|x)$  peut être simulée en générant premièrement,  $\theta_1$  à partir de  $\pi_2(\theta_1|x)$  et ensuite,  $\theta$  à l'aide de  $\pi_1(\theta|\theta_1, x)$ , s'il est bien sûr, facile de travailler avec ces deux distributions conditionnelles. Notons que ce type d'argumentation sera utilisé ultérieurement pour trouver l'estimateur de Bayes associé au second modèle que nous employons. Terminons la parenthèse ici et revenons maintenant aux modèles *a priori*.

## 2.3. MODÉLISATION PAR BASE DE SPLINES

### 2.3.1. Modèle *a priori*

#### 2.3.1.1. Modèle 1 : sur la moyenne des observations

Dans la présente sous-section, nous résumons la méthode d'estimation proposée par Merleau *et al.* (2007) qui, à l'origine, a été développée afin de modéliser différents hydrogrammes à l'aide de M-splines ainsi que de I-splines. Dans cette publication, la fonction  $g(x)$  de l'équation (2.1.2) est décomposée de la façon suivante :

$$g(x) = \sum_{k=1}^p \beta_k b_k(x), \quad (2.3.1)$$

où  $\beta_k$  représente le coefficient relié à l'élément  $b_k(\cdot)$  de la base. Rappelons que nous employons une base de I-splines dans ce mémoire.

Maintenant, si nous définissons le vecteur des coefficients  $\boldsymbol{\beta}$  par

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top,$$

et la matrice des splines  $X$  par

$$X = \begin{bmatrix} b_1(x_1) & b_2(x_1) & \dots & b_p(x_1) \\ b_1(x_2) & b_2(x_2) & \dots & b_p(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ b_1(x_n) & b_2(x_n) & \dots & b_p(x_n) \end{bmatrix},$$

alors nous pouvons écrire le modèle (2.1.2) sous la forme matricielle

$$\bar{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

où  $\bar{\mathbf{y}} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_n)^\top$ , constitue l'ensemble des moyennes calculées en (2.1.1) et  $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^\top$ , est le vecteur qui comprend les erreurs. Sous cette forme linéaire, il est facile de voir que

$$\bar{\mathbf{y}}|\boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{\Sigma}_\varepsilon),$$

si nous supposons que  $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}_\varepsilon)$ . Dans le présent ouvrage, nous faisons l'hypothèse que la structure de covariance  $\boldsymbol{\Sigma}_\varepsilon$  est connue et nous posons  $\boldsymbol{\Sigma}_\varepsilon = \mathbf{I}_n$ .

De cette façon, ce sont les coefficients  $\beta_j$ ,  $j = 1, \dots, p$ , que nous cherchons à estimer. Puisque nous sommes dans un contexte bayésien, nous présumons que ce sont des variables aléatoires normalement distribuées, ou plus précisément

$$\boldsymbol{\beta} | \sigma^2 \sim \mathcal{N}_p(\boldsymbol{\beta}^0, \sigma^2 \boldsymbol{\Sigma}_\beta).$$

De plus, puisque la quantité  $\sigma^2$  reste inconnue, nous présumons que

$$\sigma^2 \sim \mathcal{IG}\left(\frac{\alpha}{2}, \frac{\gamma}{2}\right),$$

et nous nous retrouvons ainsi avec une distribution *a priori* conjuguée. Le choix des hyperparamètres  $\boldsymbol{\beta}^0$ ,  $\boldsymbol{\Sigma}_\beta$ ,  $\alpha$  et  $\gamma$  dépend des connaissances *a priori* que nous possédons.

Dans un premier temps, puisque nous n'avons aucune idée de l'ordre de grandeur des coefficients  $\beta_j$ ,  $j = 1, \dots, p$ , nous tentons de spécifier un ensemble de valeurs moyennes plausibles pour  $\boldsymbol{\beta}^0$ . En fait, nous faisons passer une droite par les observations minimale  $(x_{(1)}, \bar{y}_{(1)})$  et maximale  $(x_{(n)}, \bar{y}_{(n)})$ . À partir de celle-ci, nous calculons les points  $p_i$  associés aux temps  $x_i$ ,  $i = 1, \dots, n$ , pour cette droite. Le vecteur  $\boldsymbol{\beta}^0$  est alors l'estimateur des moindres carrés résultant, c'est-à-dire

$$\boldsymbol{\beta}^0 = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{p},$$

où  $\mathbf{p} = (p_1, p_2, \dots, p_n)^\top$ , est le vecteur des points obtenus. En ce qui a trait à la matrice de covariance  $\boldsymbol{\Sigma}_\beta$ , nous supposons qu'elle prend la forme

$$\begin{aligned} \boldsymbol{\Sigma}_\beta &= \frac{1}{n_0} (\mathbf{X}^\top \boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{X})^{-1} \\ &= \frac{1}{n_0} (\mathbf{X}^\top \mathbf{X})^{-1}. \end{aligned}$$

Ce type de structure a été suggéré par Zellner (1986) et peut être justifié de différentes façons (voir Robert, 2001). Par manque d'information *a priori*, nous limitons les valeurs possibles de  $n_0$  à l'intervalle  $(0, 1)$ . La valeur conservée est ensuite celle qui maximise la loi marginale des observations  $\bar{\mathbf{y}}$ . Pour les mêmes raisons, nous prenons  $\alpha < 2$  et  $\gamma < 2$ , ce qui mène à l'obtention d'une distribution *a priori* diffuse pour  $\sigma^2$ . Nous incorporons tout de même une certaine quantité d'information en choisissant, comme le font Merleau *et al.* (2007),  $\alpha = \gamma = \exp(-2p)$ , une fonction décroissante en  $p$ , qui représente le nombre de paramètres.

Cette hypothèse permet, en quelque sorte, d'inclure un nouveau terme pénalisant dans la procédure de sélection du modèle. En effet, la vitesse avec laquelle la fonction  $\exp(-2p)$  décroît influence celle avec laquelle la pénalité augmente selon la dimension  $p$ .

À présent, avant de définir l'estimateur de  $\beta$ , donnons les lois *a priori* associées au second modèle.

### 2.3.1.2. Modèle 2 : sur les observations directement

Rappelons que la différence majeure entre ce modèle et le précédent vient du fait que nous précisons une loi *a priori* pour le vecteur des données observées pour le  $i^e$  enfant  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in})^\top$ ,  $i = 1, \dots, N$ , plutôt que sur le vecteur moyen  $\bar{\mathbf{y}} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_n)^\top$ . Si  $\beta_i$  est le vecteur des coefficients pour l'enfant  $i$ , nous présumons que

$$\mathbf{y}_i | \beta_i \sim \mathcal{N}_n(\mathbf{X}\beta_i, \sigma^2 \mathbf{I}_n).$$

Nous supposons également que

$$\begin{aligned} \beta_i | \beta &\sim \mathcal{N}_p\left(\beta, \frac{\sigma^2}{n_0} \mathbf{I}_p\right), \\ \beta | \sigma^2 &\sim \mathcal{N}_p\left(\beta^0, \frac{\sigma^2}{n_1} \mathbf{I}_p\right), \\ \sigma^2 &\sim \text{IG}\left(\frac{\alpha}{2}, \frac{\gamma}{2}\right), \end{aligned}$$

où  $\beta$ ,  $\beta^0$ ,  $\alpha$  et  $\gamma$  sont définis de la même façon qu'à la section précédente afin qu'il y ait un certain accord entre les deux modèles. Puisque nos connaissances par rapport aux structures de covariance de  $\beta_i$  et  $\beta$  sont plutôt limitées, nous posons  $0 < n_1 \leq n_0 < 1$  et nous les estimons en utilisant la même procédure qu'auparavant, en supposant que  $n_0 = n_1$  afin de simplifier les calculs.

L'adoption d'un modèle hiérarchique semblable à celui que nous proposons a été fait par Lindley et Smith (1972), à l'exception du fait que les matrices de variance-covariance sont supposées entièrement connues dans cet article. Comme nous ne les connaissons que partiellement, nous devons rajouter une loi *a priori* pour  $\sigma^2$ . Puisque nous ne l'avons donc pas directement tiré d'une publication, ce second modèle sera élaboré plus en détails que le premier dans la section qui suit.

### 2.3.2. Estimateur de la courbe de référence

Souvenons-nous que nous désirons obtenir un estimateur pour la fonction  $g$  décrite en (2.3.1). Pour le calculer, comme seuls les coefficients des splines sont inconnus, il suffit de trouver l'estimateur de ces coefficients qui correspond à l'espérance de la densité *a posteriori* de  $\beta|\bar{\mathbf{y}}$  sous la fonction de perte quadratique. Voici donc les quelques étapes qui permettent de retrouver cette densité.

#### 2.3.2.1. Modèle 1 : sur la moyenne des observations

Si nous posons  $\beta^{MC} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \bar{\mathbf{y}}$ , l'emploi de calculs bayésiens standards pour les modèles linéaires (voir Robert, 2001) nous permet d'écrire

$$\begin{aligned} \beta|\sigma^2, \bar{\mathbf{y}} &\sim \mathcal{N}_p(\beta^*, \sigma^2 \Sigma_\beta^*), \\ \sigma^2|\bar{\mathbf{y}} &\sim \mathcal{IG}\left(\frac{\alpha^*}{2}, \frac{\gamma^*}{2}\right), \end{aligned}$$

où

$$\begin{aligned} \beta^* &= \frac{1}{n_0 + 1} [\beta^{MC} + n_0 \beta^0], \\ \Sigma_\beta^* &= \frac{1}{n_0 + 1} (\mathbf{X}^\top \mathbf{X})^{-1}, \\ \alpha^* &= n + \alpha, \\ \gamma^* &= (\bar{\mathbf{y}} - \mathbf{X} \beta^{MC})^\top (\bar{\mathbf{y}} - \mathbf{X} \beta^{MC}) + \frac{n_0}{n_0 + 1} (\beta^{MC} - \beta^0)^\top \mathbf{X}^\top \mathbf{X} (\beta^{MC} - \beta^0) \\ &\quad + \gamma. \end{aligned}$$

Nous décomposons ensuite la distribution de la manière suivante :

$$\pi(\beta, \sigma^2|\bar{\mathbf{y}}) = \pi_1(\beta|\sigma^2, \bar{\mathbf{y}}) \pi_2(\sigma^2|\bar{\mathbf{y}}). \quad (2.3.2)$$

Donc, en intégrant l'égalité (2.3.2) par rapport à  $\sigma^2$ , nous obtenons la distribution *a posteriori* de  $\beta$  qui consiste en une distribution t de Student multidimensionnelle

$$\beta|\bar{\mathbf{y}} \sim T_p\left(\alpha^*, \beta^*, \frac{\gamma^*}{\alpha^*} \Sigma_\beta^*\right), \quad (2.3.3)$$

où la notation  $T_p(v, \theta, \Sigma)$  se réfère à une distribution t de Student multidimensionnelle ayant  $v$  degrés de liberté, de moyenne  $\theta$  et de matrice d'échelle  $\Sigma$ .

Notons que la densité de cette loi ainsi que de toutes celles qui apparaissent dans ce chapitre sont présentées dans l'annexe A.

Sous la fonction de perte quadratique, l'estimateur obtenu pour les coefficients est donc la moyenne *a posteriori*  $\beta^*$ , qui correspond ici à une moyenne pondérée entre l'estimateur des moindres carrés ordinaires  $\beta^{MC}$  et la moyenne *a priori*  $\beta^0$ .

Pour nous aider lors du choix de  $n_0$ , mentionnons finalement que la densité marginale des observations est donnée par

$$\begin{aligned} m(\bar{\mathbf{y}}) &= \left( \frac{|\Sigma_{\beta}^*|}{|\Sigma_{\epsilon}| |\Sigma_{\beta}|} \right)^{1/2} \left( \frac{\Gamma(\alpha^*/2)}{\pi^{n/2} \Gamma(\alpha/2)} \right) \left( \frac{\gamma^{\alpha/2}}{(\gamma^*)^{\alpha^*/2}} \right) \\ &= \left( \frac{n_0}{n_0 + 1} \right)^{p/2} \left( \frac{\Gamma(\alpha^*/2)}{\pi^{n/2} \Gamma(\alpha/2)} \right) \left( \frac{\gamma^{\alpha/2}}{(\gamma^*)^{\alpha^*/2}} \right) \\ &\propto \left( \frac{n_0}{n_0 + 1} \right)^{p/2} \left[ \frac{n_0}{n_0 + 1} (\beta^{MC} - \beta^0)^\top \mathbf{X}^\top \mathbf{X} (\beta^{MC} - \beta^0) + \gamma \right. \\ &\quad \left. + (\bar{\mathbf{y}} - \mathbf{X} \beta^{MC})^\top (\bar{\mathbf{y}} - \mathbf{X} \beta^{MC}) \right]^{-\frac{(n+\alpha)}{2}}. \end{aligned}$$

### 2.3.2.2. Modèle 2 : sur les observations directement

Dans cette sous-section, nous avons recours aux décompositions conditionnelles dont nous avons discuté à la section 2.2 afin de retrouver l'estimateur de Bayes recherché.

Effectivement, l'application de l'idée vue à l'équation (2.2.4) au second modèle nous permet de considérer que l'estimateur du vecteur des coefficients  $\beta$  vaut

$$\hat{\beta} = \mathbb{E}^\pi[\beta | \mathbf{Y}] = \mathbb{E}^{\pi(\sigma^2 | \mathbf{Y})}[\mathbb{E}^{\pi(\beta | \mathbf{Y}, \sigma^2)}(\beta | \mathbf{Y}, \sigma^2)], \quad (2.3.4)$$

où  $\mathbf{Y}$  est la matrice de dimensions  $N \times n$  dont la  $j^e$  ligne est le vecteur des observations pour l'enfant  $j$ ,  $j = 1, \dots, N$ . Pour le calculer, il suffit donc d'évaluer les distributions  $\pi_1(\beta | \mathbf{Y}, \sigma^2)$  et  $\pi_2(\sigma^2 | \mathbf{Y})$ .

En nous référant à l'équation (2.2.3), nous savons que

$$\pi_1(\beta | \mathbf{Y}, \sigma^2) \propto \left[ \prod_{i=1}^N m_i(\mathbf{y}_i | \beta, \sigma^2) \right] \times \pi_1(\beta | \sigma^2).$$

De plus, l'équation (2.2.2) nous dit que

$$\begin{aligned}
 m_i(\mathbf{y}_i|\boldsymbol{\beta}, \sigma^2) &= \int_{-\infty}^{\infty} f(\mathbf{y}_i|\boldsymbol{\beta}_i)\pi_1(\boldsymbol{\beta}_i|\boldsymbol{\beta}, \sigma^2)d\boldsymbol{\beta}_i \\
 &= \int_{-\infty}^{\infty} \frac{n_0^{p/2}}{(2\pi\sigma^2)^{\frac{n+p}{2}}} \exp \left\{ \frac{-1}{2\sigma^2} \left[ (\mathbf{y}_i - \mathbf{X}\boldsymbol{\beta}_i)^\top (\mathbf{y}_i - \mathbf{X}\boldsymbol{\beta}_i) \right. \right. \\
 &\quad \left. \left. + n_0(\boldsymbol{\beta}_i - \boldsymbol{\beta})^\top (\boldsymbol{\beta}_i - \boldsymbol{\beta}) \right] \right\} d\boldsymbol{\beta}_i. \tag{2.3.5}
 \end{aligned}$$

Réécrivons maintenant cette expression de sorte qu'il soit facile de retrouver la densité marginale et, par conséquent, la loi *a posteriori* de  $\boldsymbol{\beta}|\mathbf{Y}, \sigma^2$ . Pour nous faciliter la tâche, voici deux égalités qui nous seront d'une grande utilité tout au long de l'élaboration de ce modèle. La première sert à réexprimer la somme de deux formes quadratiques telle que celle que nous retrouvons dans l'équation (2.3.5) et la seconde concerne l'inverse de matrices.

**Proposition 2.3.1.** (voir Lindley et Smith, 1972)

*La somme de deux formes quadratiques prenant la forme*

$$(\mathbf{y} - \mathbf{A}_1\boldsymbol{\theta}_1)^\top \mathbf{C}_1^{-1}(\mathbf{y} - \mathbf{A}_1\boldsymbol{\theta}_1) + (\boldsymbol{\theta}_1 - \mathbf{A}_2\boldsymbol{\theta}_2)^\top \mathbf{C}_2^{-1}(\boldsymbol{\theta}_1 - \mathbf{A}_2\boldsymbol{\theta}_2),$$

*peut s'écrire comme*

$$(\boldsymbol{\theta}_1 - \mathbf{B}\mathbf{b})^\top \mathbf{B}^{-1}(\boldsymbol{\theta}_1 - \mathbf{B}\mathbf{b}) + [\mathbf{y}^\top \mathbf{C}_1^{-1}\mathbf{y} + \boldsymbol{\theta}_2^\top \mathbf{A}_2^\top \mathbf{C}_2^{-1}\mathbf{A}_2\boldsymbol{\theta}_2 - \mathbf{b}^\top \mathbf{B}\mathbf{b}], \tag{2.3.6}$$

où

$$\begin{aligned}
 \mathbf{B}^{-1} &= \mathbf{A}_1^\top \mathbf{C}_1^{-1}\mathbf{A}_1 + \mathbf{C}_2^{-1}, \\
 \mathbf{b} &= \mathbf{A}_1^\top \mathbf{C}_1^{-1}\mathbf{y} + \mathbf{C}_2^{-1}\mathbf{A}_2\boldsymbol{\theta}_2.
 \end{aligned}$$

Spécifions que cette équivalence est très pratique lorsque vient le temps de trouver la loi de  $\boldsymbol{\theta}_1$ , puisqu'elle permet de restructurer les termes afin d'obtenir une forme quadratique en ce paramètre. Par la suite, il suffit de réorganiser le terme restant pour avoir la loi marginale des observations et c'est là où l'identité suivante entre en jeu.

**Proposition 2.3.2.** (voir Searle, 1982)

*Soient  $\mathbf{A}$  et  $\mathbf{B}$ , deux matrices. Si  $\mathbf{A}$  est inversible, alors nous avons*

$$(\mathbf{A} + \mathbf{B}\mathbf{B}^\top)^{-1}\mathbf{B} = \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{B}^\top \mathbf{A}^{-1}\mathbf{B})^{-1}. \tag{2.3.7}$$

Ainsi, en nous servant de l'équation (2.3.6) pour réorganiser les termes dans la fonction exponentielle et en posant  $\mathbf{A} = n_0 \mathbf{I}_p$  et  $\mathbf{B} = \mathbf{X}^\top$  dans l'équation (2.3.7), nous pouvons montrer que

$$\begin{aligned} m_i(\mathbf{y}_i | \boldsymbol{\beta}, \sigma^2) &= c_1^* \exp \left\{ \frac{-n_0}{2\sigma^2} \left[ (\mathbf{y}_i - \mathbf{X}\boldsymbol{\beta})^\top (n_0 \mathbf{I}_n + \mathbf{X}\mathbf{X}^\top)^{-1} (\mathbf{y}_i - \mathbf{X}\boldsymbol{\beta}) \right] \right\} \\ &\times \int_{-\infty}^{\infty} c_2^* \exp \left\{ \frac{-1}{2\sigma^2} \left[ (\boldsymbol{\beta}_i - (\mathbf{X}^\top \mathbf{X} + n_0 \mathbf{I}_p)^{-1} (\mathbf{X}^\top \mathbf{y}_i + n_0 \boldsymbol{\beta}))^\top \right. \right. \\ &\times \left. \left. (\mathbf{X}^\top \mathbf{X} + n_0 \mathbf{I}_p) (\boldsymbol{\beta}_i - (\mathbf{X}^\top \mathbf{X} + n_0 \mathbf{I}_p)^{-1} (\mathbf{X}^\top \mathbf{y}_i + n_0 \boldsymbol{\beta})) \right] \right\} d\boldsymbol{\beta}_i, \end{aligned}$$

$$\text{où } c_1^* = \frac{1}{(2\pi\sigma^2)^{n/2}} \text{ et } c_2^* = \frac{n_0^{p/2}}{(2\pi\sigma^2)^{p/2}}.$$

Pour finir, puisque l'intégrand correspond, à une constante près, à la densité d'une loi

$$\mathcal{N}_p \left( (\mathbf{X}^\top \mathbf{X} + n_0 \mathbf{I}_p)^{-1} (\mathbf{X}^\top \mathbf{y}_i + n_0 \boldsymbol{\beta}), \sigma^2 (\mathbf{X}^\top \mathbf{X} + n_0 \mathbf{I}_p)^{-1} \right),$$

où  $\mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  désigne la loi normale multidimensionnelle de vecteur moyen  $\boldsymbol{\mu}$  et de matrice de variance-covariance  $\boldsymbol{\Sigma}$  associée à un vecteur de dimension  $k$ , et que (voir Searle, 1982)

$$|\mathbf{X}^\top \mathbf{X} + n_0 \mathbf{I}_p| = \frac{1}{n_0^{n-p}} |\mathbf{X}\mathbf{X}^\top + n_0 \mathbf{I}_n|,$$

nous avons

$$m_i(\mathbf{y}_i | \boldsymbol{\beta}, \sigma^2) \sim \mathcal{N}_n \left( \mathbf{X}\boldsymbol{\beta}, \frac{\sigma^2}{n_0} (n_0 \mathbf{I}_n + \mathbf{X}\mathbf{X}^\top) \right). \quad (2.3.8)$$

Avant de poursuivre, précisons que plutôt que de travailler directement avec la matrice  $\mathbf{Y}$ , nous préférons utiliser les statistiques exhaustives  $\mathbf{S}$  et  $\bar{\mathbf{y}}$ , c'est-à-dire la matrice de variance-covariance et la moyenne échantillonales, qui sont indépendantes lorsque conditionnées par rapport à  $\sigma^2$ . Cela simplifie grandement les calculs subséquents. Nous réexprimons donc la fonction de vraisemblance associée aux lois marginales de la façon que voici (voir Anderson, 2003).

En posant  $\Sigma = n_0 \mathbf{I}_n + \mathbf{X}\mathbf{X}^\top$ , nous avons

$$\begin{aligned} \prod_{i=1}^N m_i(\mathbf{y}_i | \boldsymbol{\beta}, \sigma^2) &= c^* \exp \left\{ \frac{-n_0}{2\sigma^2} \sum_{i=1}^N (\mathbf{y}_i - \mathbf{X}\boldsymbol{\beta})^\top \Sigma^{-1} (\mathbf{y}_i - \mathbf{X}\boldsymbol{\beta}) \right\} \\ &= c^* \exp \left\{ \frac{-n_0}{2\sigma^2} \left[ \sum_{i=1}^N (\mathbf{y}_i - \bar{\mathbf{y}})^\top \Sigma^{-1} (\mathbf{y}_i - \bar{\mathbf{y}}) \right. \right. \\ &\quad \left. \left. + N(\bar{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})^\top \Sigma^{-1} (\bar{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}) \right] \right\} \\ &= c^* \exp \left\{ \frac{-n_0}{2\sigma^2} [\text{tr}(\Sigma^{-1} \mathbf{S}) + N(\bar{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})^\top \Sigma^{-1} (\bar{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})] \right\}, \end{aligned}$$

où

$$\begin{aligned} c^* &= \left( \frac{n_0}{2\pi\sigma^2} \right)^{nN/2} \frac{1}{|\Sigma|^{N/2}}, \\ \mathbf{S} &= \sum_{i=1}^N (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^\top, \\ \bar{\mathbf{y}} &= (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_n)^\top. \end{aligned}$$

Rappelons que  $\bar{\mathbf{y}}$  est le vecteur des moyennes des tailles observées pour tous les enfants à chacun des temps  $x_i$ ,  $i = 1, \dots, n$ .

En employant cette forme, nous obtenons

$$\begin{aligned} \pi_1(\boldsymbol{\beta} | \mathbf{Y}, \sigma^2) &\propto \frac{n_0^{nN/2} n_1^{p/2}}{(2\pi\sigma^2)^{(nN+p)/2}} \frac{1}{|\Sigma|^{N/2}} \exp \left\{ \frac{-n_0}{2\sigma^2} \text{tr}(\Sigma^{-1} \mathbf{S}) \right\} \\ &\times \exp \left\{ \frac{-1}{2\sigma^2} \left[ n_0 N(\bar{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})^\top \Sigma^{-1} (\bar{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}) \right. \right. \\ &\quad \left. \left. + n_1 (\boldsymbol{\beta} - \boldsymbol{\beta}^0)^\top (\boldsymbol{\beta} - \boldsymbol{\beta}^0) \right] \right\}. \end{aligned}$$

Pour retrouver la loi *a posteriori* de  $\boldsymbol{\beta} | \mathbf{Y}, \sigma^2$ , nous appliquons alors les mêmes étapes que précédemment, c'est-à-dire que nous réarrangeons les termes dans la deuxième fonction exponentielle afin d'obtenir une forme quadratique en  $\boldsymbol{\beta}$  et nous manipulons ensuite le terme restant en nous aidant de l'égalité (2.3.7) pour trouver

$$m(\mathbf{Y} | \boldsymbol{\beta}^0, \sigma^2) = \int_{-\infty}^{\infty} \left[ \prod_{i=1}^N m_i(\mathbf{y}_i | \boldsymbol{\beta}, \sigma^2) \right] \pi_1(\boldsymbol{\beta} | \sigma^2) d\boldsymbol{\beta},$$

ce qui donne

$$\boldsymbol{\beta} | \mathbf{Y}, \sigma^2 \sim \mathcal{N}_p(\mathbf{A}\mathbf{a}, \sigma^2 \mathbf{A}), \quad (2.3.9)$$

où

$$\mathbf{A} = \left( n_0 N \mathbf{X}^\top (n_0 \mathbf{I}_n + \mathbf{X} \mathbf{X}^\top)^{-1} \mathbf{X} + n_1 \mathbf{I}_p \right)^{-1}, \quad (2.3.10)$$

$$\mathbf{a} = n_0 N \mathbf{X}^\top \Sigma^{-1} \bar{\mathbf{y}} + n_1 \boldsymbol{\beta}^0. \quad (2.3.11)$$

Établissons dès lors la densité de  $\sigma^2 | \mathbf{Y}$ . Nous avons

$$\begin{aligned} \pi_2(\sigma^2 | \mathbf{Y}) &\propto m(\mathbf{Y} | \boldsymbol{\beta}^0, \sigma^2) \pi_2(\sigma^2) \\ &\propto \frac{1}{(\sigma^2)^{(nN+\alpha)/2}} \exp \left\{ \frac{-1}{2\sigma^2} \left[ n_0 \text{tr}[\Sigma^{-1} \mathbf{S}] + \gamma \right. \right. \\ &\quad \left. \left. + N(\bar{\mathbf{y}} - \mathbf{X} \boldsymbol{\beta}^0)^\top \left( \mathbf{I}_n + \left( \frac{Nn_0 + n_1}{n_0 n_1} \right) \mathbf{X} \mathbf{X}^\top \right)^{-1} (\bar{\mathbf{y}} - \mathbf{X} \boldsymbol{\beta}^0) \right] \right\}. \end{aligned} \quad (2.3.12)$$

À partir de l'équation (2.3.12), nous remarquons tout de suite que

$$\sigma^2 | \mathbf{Y} \sim \mathcal{IG} \left( \frac{nN + \alpha}{2}, \frac{\gamma'}{2} \right), \quad (2.3.13)$$

où

$$\begin{aligned} \gamma' &= n_0 \text{tr}(\Sigma^{-1} \mathbf{S}) + N(\bar{\mathbf{y}} - \mathbf{X} \boldsymbol{\beta}^0)^\top \left( \mathbf{I}_n + \left( \frac{Nn_0 + n_1}{n_0 n_1} \right) \mathbf{X} \mathbf{X}^\top \right)^{-1} (\bar{\mathbf{y}} - \mathbf{X} \boldsymbol{\beta}^0) \\ &\quad + \gamma. \end{aligned} \quad (2.3.14)$$

Maintenant que les densités *a posteriori*  $\boldsymbol{\beta} | \mathbf{Y}, \sigma^2$  et  $\sigma^2 | \mathbf{Y}$  ont été trouvées, déterminons l'estimateur de  $\boldsymbol{\beta}$ . En nous fiant à l'égalité (2.3.4), nous avons

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \int_0^\infty \int_{-\infty}^\infty \left[ \boldsymbol{\beta} \times \pi_1(\boldsymbol{\beta} | \mathbf{Y}, \sigma^2) d\boldsymbol{\beta} \right] \pi_2(\sigma^2 | \mathbf{Y}) d\sigma^2 \\ &= \int_0^\infty \mathbf{A} \mathbf{a} \times \pi_2(\sigma^2 | \mathbf{Y}) d\sigma^2 \\ &= \mathbf{A} \mathbf{a}, \end{aligned} \quad (2.3.15)$$

où  $\mathbf{A}$  et  $\mathbf{a}$  sont respectivement donnés aux équations (2.3.10) et (2.3.11).

Donc, comme pour le premier modèle, l'estimateur obtenu, bien qu'un peu plus complexe, est également un mélange entre un estimateur dépendant des observations  $\bar{\mathbf{y}}$  et la moyenne *a priori*  $\boldsymbol{\beta}^0$ .

Pour terminer, il ne reste plus qu'à trouver la densité marginale des observations qui sert lors du choix des paramètres  $n_0$  et  $n_1$ . Pour ce faire, appliquons le

résultat suivant (voir Robert, 2001) :

$$\pi(\boldsymbol{\beta}, \sigma^2 | \mathbf{Y}) = \frac{\pi(\sigma^2)\pi(\boldsymbol{\beta}|\sigma^2) \left[ \prod_{i=1}^N m_i(\mathbf{y}_i | \boldsymbol{\beta}, \sigma^2) \right]}{m(\mathbf{Y})}.$$

Les densités de  $\boldsymbol{\beta}|\sigma^2$  et de  $\sigma^2$  sont connues *a priori*. De plus, la densité marginale  $m_i(\mathbf{y}_i | \boldsymbol{\beta}, \sigma^2)$  est donnée à l'équation (2.3.8) et la distribution  $\pi(\boldsymbol{\beta}, \sigma^2 | \mathbf{Y})$  peut s'exprimer par  $\pi(\boldsymbol{\beta} | \mathbf{Y}, \sigma^2)\pi(\sigma^2 | \mathbf{Y})$ , qui est le produit de deux densités que nous avons trouvées aux équations (2.3.9) et (2.3.13) respectivement. La seule quantité inconnue est donc la densité marginale des observations que nous évaluons en calculant

$$m(\mathbf{Y}) = \frac{\pi(\sigma^2)\pi(\boldsymbol{\beta}|\sigma^2) \left[ \prod_{i=1}^N m_i(\mathbf{y}_i | \boldsymbol{\beta}, \sigma^2) \right]}{\pi(\boldsymbol{\beta} | \mathbf{Y}, \sigma^2)\pi(\sigma^2 | \mathbf{Y})}.$$

Après plusieurs manipulations qui ressemblent à celles effectuées préalablement, nous obtenons

$$\begin{aligned} m(\mathbf{Y}) &= \frac{n_0^{\frac{n}{2}(N-1)} \Gamma([nN + \alpha]/2)}{\pi^{\frac{nN}{2}} \Gamma(\alpha/2)} \frac{\gamma^{\alpha/2}}{|n_0 \mathbf{I}_n + \mathbf{X} \mathbf{X}^\top|^{\frac{N-1}{2}}} \\ &\times \frac{(\gamma')^{-\left(\frac{nN+\alpha}{2}\right)}}{\left| \mathbf{I}_n + \left( \frac{Nn_0 + n_1}{n_0 n_1} \right) \mathbf{X} \mathbf{X}^\top \right|^{\frac{1}{2}}}, \end{aligned}$$

où  $\gamma'$  est défini de la même manière qu'en (2.3.14).

## 2.4. MODÉLISATION À L'AIDE DE FONCTIONS D'ÉCHELLE POSITIVES

Nous désirons estimer la courbe de croissance de référence à l'aide des ondelettes non négatives décrites dans la section 1.3.5 du premier chapitre. Pour ce faire, si nous suivons la théorie présentée dans l'article de Walter et Shen (1999), dont le but est d'estimer une fonction de densité  $f(s)$ , nous avons

$$f_j^r(x) = \int_{-\infty}^{\infty} \mathbb{K}_{r,j}(x, s) f(s) ds, \quad (2.4.1)$$

où  $f(s)$  est la fonction à estimer. Nous remarquons que l'expression (2.4.1) correspond à la forme définie à l'équation (1.3.7), qui découle de la théorie des noyaux avec, comme noyau spécifique, celui décrit à l'équation (1.3.9).

En développant les termes qui se trouvent dans l'expression (2.4.1), pour une valeur de  $j$  fixée, nous obtenons

$$\begin{aligned} f_j^r(x) &= \int_{-\infty}^{\infty} 2^j \left( \frac{1-r}{1+r} \right)^2 \sum_l \rho_r(2^j x - l) \rho_r(2^j s - l) f(s) ds \\ &= 2^j \left( \frac{1-r}{1+r} \right)^2 \sum_l \rho_r(2^j x - l) \int_{-\infty}^{\infty} \rho_r(2^j s - l) f(s) ds, \end{aligned}$$

la permutation de la somme et de l'intégrale étant permise grâce au théorème de Fubini.

Puisque la fonction  $f$  est inconnue, nous ne pouvons pas calculer l'intégrale précédente. Ainsi, nous posons

$$\beta_l = \int_{-\infty}^{\infty} \rho_r(2^j s - l) f(s) ds.$$

Il nous reste alors

$$f_j^r(x) = 2^j \left( \frac{1-r}{1+r} \right)^2 \sum_l \rho_r(2^j x - l) \beta_l. \quad (2.4.2)$$

Avec ceci, nous serions normalement en mesure de développer un modèle afin d'estimer les paramètres  $\beta_l$  et de retrouver l'estimateur de la courbe. Remarquons cependant que nous ne procédons pas exactement de cette façon. En effet, rappelons que dans la section 2.3, nous avons choisi les I-splines parmi toutes les bases de splines, puisque ces fonctions sont monotones croissantes, propriété qu'elles partagent avec la courbe de référence. Par contre, les fonctions  $\rho_r$  ne le sont pas. Mais, comme elles sont toujours positives ou nulles, il est possible de construire, en les intégrant, une base de fonctions monotones croissantes que nous surnomons « ondelettes intégrées ».

Nous définissons donc, toujours pour une valeur de  $j$  fixée,

$$\rho_{r,l}^*(x) = \int_{-\infty}^x \rho_r(2^j t - l) dt, \quad (2.4.3)$$

ce qui mène, lorsque nous appliquons le résultat à notre problème, à la décomposition

$$g(x) = 2^j \left( \frac{1-r}{1+r} \right)^2 \sum_l \rho_{r,l}^*(x) \times \beta_l. \quad (2.4.4)$$

C'est cette dernière expression que nous employons afin de retracer l'estimateur de la courbe de référence. Notons qu'elle ressemble énormément à celle de l'équation (2.3.1), mais que contrairement à celle-ci, elle comprend un nombre infini d'éléments. Bien sûr, pour obtenir une bonne approximation, il est possible de n'en utiliser qu'une partie, les autres étant inclus dans le terme d'erreur (voir Angers et Delampady, 2001).

Avant de discuter des modèles *a priori*, donnons quelques exemples de la fonction  $\rho_{r,l}^*(x)$  définie à l'équation (2.4.3).

#### 2.4.1. Développement analytique de la fonction $\rho_{r,l}^*(x)$ pour les ondelettes de Haar

Dans cette section, nous développons la fonction  $\rho_{r,l}^*(x) = \int_{-\infty}^x \rho_r(2^j t - l) dt$  pour le cas particulier des ondelettes de Haar. Notons que ce n'est que pour cette base qu'il est possible de le faire, puisqu'il n'existe pas de représentation explicite de la fonction d'échelle des ondelettes de Daubechies d'ordre supérieur ou égal à deux.

Tout d'abord, rappelons la définition de la fonction d'échelle de Haar. Nous avons

$$\phi(x) = \begin{cases} 1, & 0 \leq x < 1, \\ 0, & \text{sinon.} \end{cases} \quad (2.4.5)$$

De plus, d'après l'égalité (1.3.5), nous retrouvons

$$\rho_r(2^j t - l) = \sum_m r^{|m|} \phi(\{2^j t - l\} - m). \quad (2.4.6)$$

En combinant les équations (2.4.5) et (2.4.6), nous nous apercevons que les seules valeurs pour lesquelles la fonction  $\phi$  est non nulle sont celles où  $m = [2^j t] - l$ , où  $[\cdot]$  dénote la partie entière d'un nombre. Ainsi,

$$\rho_r(2^j t - l) = r^{|[2^j t] - l|}.$$

Le problème se réduit donc à trouver une expression pour

$$\rho_{r,l}^*(x) = \int_{-\infty}^x r^{|[2^j t] - l|} dt. \quad (2.4.7)$$

Pour ce faire, supposons que la partie entière  $[2^j t]$  corresponde à une valeur  $k \in \mathbb{Z}$ , ce qui signifie que  $k \leq 2^j t < k + 1$ , d'où  $2^{-j}(k) \leq t < 2^{-j}(k + 1)$ . Nous

pouvons alors montrer que, lorsque  $2^j x$  est un entier, l'équation (2.4.7) équivaut à

$$\rho_{r,l}^*(x) = \sum_{k=-\infty}^{\lfloor 2^j x \rfloor - 1} \int_{2^{-j}(k)}^{2^{-j}(k+1)} r^{|k-l|} dt. \quad (2.4.8)$$

Dans le cas contraire, le terme

$$2^{-j} (2^j - \lfloor 2^j x \rfloor) r^{|\lfloor 2^j x \rfloor - l|},$$

doit être rajouté à l'expression (2.4.8).

Posons maintenant  $q = k - l$ . De cette façon, l'énoncé (2.4.8) peut s'écrire comme

$$S_0 = 2^{-j} \sum_{q=-\infty}^{\lfloor 2^j x \rfloor - l - 1} r^{|q|},$$

qui n'est simplement qu'une série géométrique. Il existe alors 3 cas possibles.

Donnons les résultats de ces séries pour chacun de ceux-ci.

$$(1) \lfloor 2^j x \rfloor - l - 1 = 0 \implies S_0 = \frac{2^{-j}}{1 - r},$$

$$(2) \lfloor 2^j x \rfloor - l - 1 > 0 \implies S_0 = 2^{-j} \left( \frac{2 - r^{\lfloor 2^j x \rfloor - l - 1}}{1 - r} + (r^{\lfloor 2^j x \rfloor - l - 1} - 1) \right),$$

$$(3) \lfloor 2^j x \rfloor - l - 1 < 0 \implies S_0 = \frac{2^{-j} (r^{|\lfloor 2^j x \rfloor - l - 1|})}{1 - r}.$$

En résumé, nous obtenons

$$\rho_{r,l}^*(x) = \begin{cases} S_0, & \text{si } 2^j x \text{ est un entier,} \\ S_0 + 2^{-j} (2^j - \lfloor 2^j x \rfloor) r^{|\lfloor 2^j x \rfloor - l|}, & \text{sinon.} \end{cases}$$

Pour terminer, illustrons les fonctions  $\rho_r(2^j x - l)$  et  $\rho_{r,l}^*(x)$  respectivement définies aux équations (2.4.6) et (2.4.8). La figure 2.2 présente trois exemples de la fonction  $\rho_r$  pour différentes valeurs de  $j$  et de  $l$  sur le domaine  $[-10, 10]$ , quand  $r$  est fixée à 0,3. De plus, la figure 2.3 présente les fonctions  $\rho_{r,l}^*(x)$  associées à ces exemples.

Pour comprendre ces figures, exprimons l'équation (2.4.6) sous la forme suivante

$$\begin{aligned} \rho_r(2^j x - l) &= \sum_m r^{|m|} \phi(2^j x - l - m) \\ &= \sum_m r^{|m|} \phi \left( 2^j \left( x - \frac{l}{2^j} \right) - m \right). \end{aligned}$$

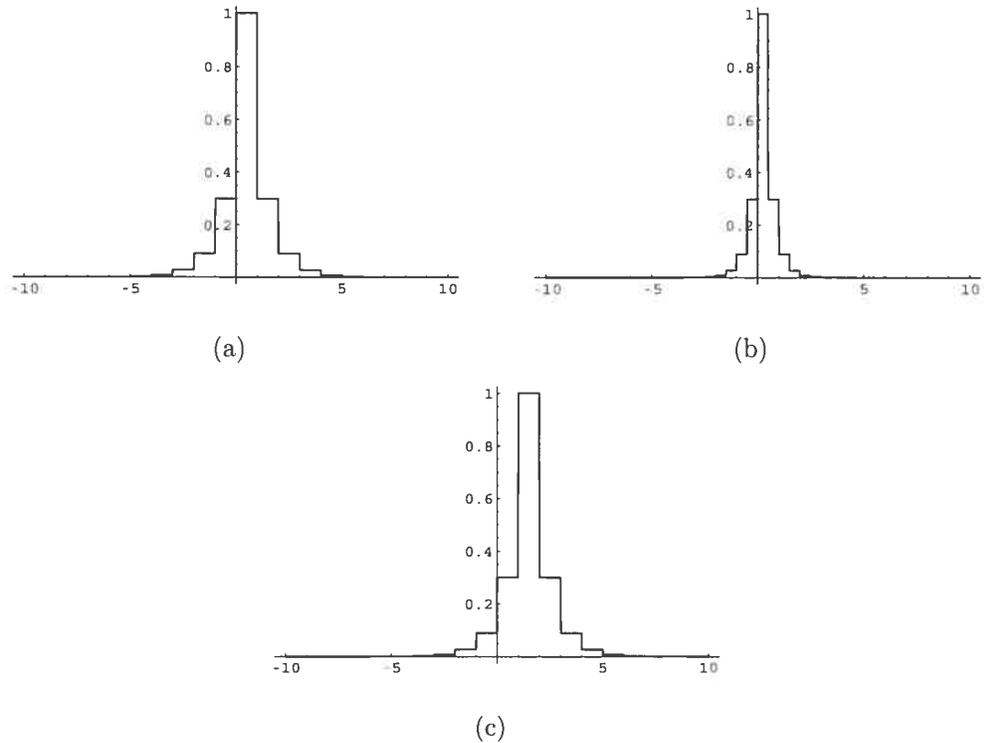


FIG. 2.2. Exemples de la fonction  $\rho_r(2^j x - l)$  (a) Pour  $j = l = 0$   
 (b) Pour  $j = 1$  et  $l = 0$  (c) Pour  $j = 0$  et  $l = 1$

En observant les deux figures, nous remarquons que le premier terme rencontré dans cette équation, soit  $2^j$ , est un facteur d'échelle. Il a pour rôle la contraction du domaine sur lequel les fonctions sont définies. Ensuite, le second facteur, ou  $\frac{l}{2^j}$ , est appelé facteur de translation et, comme son nom l'indique, il a pour rôle la translation des fonctions le long de l'axe horizontal. Mentionnons également que le facteur  $2^j$  qui se retrouve dans l'équation (2.4.2) est un facteur de normalisation et qu'il est responsable de l'étirement vertical des fonctions. Pour plus de renseignements sur ces facteurs dans le cas des ondelettes de Haar, voir Ouellette (2004).

Notons enfin que, bien qu'il nous soit impossible de trouver une expression pour les fonctions  $\rho_r(2^j x - l)$  et  $\rho_{r,l}^*(x)$  pour les fonctions d'échelle de Daubechies d'ordre supérieur à un, nous pouvons quand même les évaluer, le calcul de la seconde fonction requérant par contre certaines méthodes d'approximation des

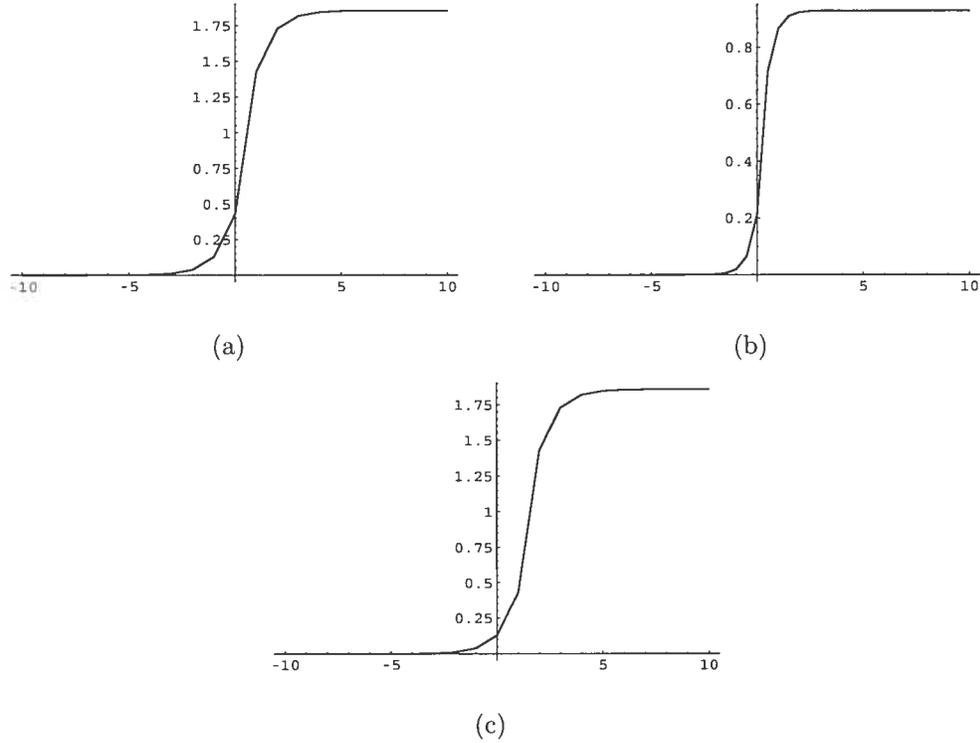


FIG. 2.3. Exemples de la fonction  $\rho_{r,l}^*(x)$  (a) Pour  $j = l = 0$  (b) Pour  $j = 1$  et  $l = 0$  (c) Pour  $j = 0$  et  $l = 1$

intégrales telles que la méthode du trapèze ou celle de Simpson (voir Conte et De Boor, 1980).

Maintenant que nous connaissons bien les fonctions avec lesquelles nous travaillons, définissons les modèles *a priori* choisis.

#### 2.4.2. Modèle *a priori*

Dans un premier temps, répétons que l'estimation de la courbe ne nécessite pas l'utilisation de toutes les fonctions de la base employée. Nous décomposons donc la fonction  $g(x)$  de l'équation (2.1.2) à l'aide de l'égalité (2.4.4) de la façon suivante

$$\begin{aligned}
 g(x) &= 2^j \left( \frac{1-r}{1+r} \right)^2 \sum_l \rho_{r,l}^*(x) \times \beta_l \\
 &= 2^j \left( \frac{1-r}{1+r} \right)^2 \left\{ \sum_{l=L}^U \rho_{r,l}^*(x) \times \beta_l + \left[ \sum_{l=-\infty}^{L-1} \rho_{r,l}^*(x) \times \beta_l + \sum_{l=U+1}^{\infty} \rho_{r,l}^*(x) \times \beta_l \right] \right\} \\
 &= g_{L,U}(x) + R_{L,U}(x), \tag{2.4.9}
 \end{aligned}$$

où

$$g_{L,U}(x) = 2^j \left( \frac{1-r}{1+r} \right)^2 \sum_{l=L}^U \rho_{r,l}^*(x) \times \beta_l,$$

$$R_{L,U}(x) = 2^j \left( \frac{1-r}{1+r} \right)^2 \left[ \sum_{l=-\infty}^{L-1} \rho_{r,l}^*(x) \times \beta_l + \sum_{l=U+1}^{\infty} \rho_{r,l}^*(x) \times \beta_l \right].$$

Soulignons que la technique que nous utilisons pour sélectionner les valeurs de  $L$  et de  $U$  ressemble énormément à celle décrite à la section 1.2.3 pour le choix des noeuds avec les splines. Nous définissons donc un très grand intervalle pour les valeurs de  $l$  au départ et nous en enlevons ensuite une à la fois. Notons toutefois que les seules valeurs qu'il nous est possible d'éliminer se limitent toujours aux extrémités de l'intervalle à cause de la forme de la fonction  $g_{L,U}(x)$  introduite à l'équation (2.4.9).

Regardons maintenant la définition des deux modèles *a priori*.

#### 2.4.2.1. Modèle 1 : sur la moyenne des observations

L'égalité (2.4.9) nous permet d'écrire l'équation (2.1.2) comme suit

$$\bar{y}_i = g_{L,U}(x_i) + \eta_i + \varepsilon_i, \quad (2.4.10)$$

où  $\eta_i = R_{L,U}(x_i)$ ,  $i = 1, \dots, n$ .

À présent, définissons le vecteur des coefficients  $\beta$  par

$$\beta = (\beta_L, \beta_{L+1}, \dots, \beta_U)^\top,$$

et la matrice de design  $X$  par

$$X = 2^j \left( \frac{1-r}{1+r} \right)^2 \begin{bmatrix} \rho_{r,L}^*(x_1) & \rho_{r,L+1}^*(x_1) & \dots & \rho_{r,U}^*(x_1) \\ \rho_{r,L}^*(x_2) & \rho_{r,L+1}^*(x_2) & \dots & \rho_{r,U}^*(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{r,L}^*(x_n) & \rho_{r,L+1}^*(x_n) & \dots & \rho_{r,U}^*(x_n) \end{bmatrix}.$$

Nous pouvons alors exprimer le modèle (2.4.10) sous la forme matricielle

$$\bar{\mathbf{y}} = \mathbf{X}\beta + \mathbf{u},$$

où  $\mathbf{u} = \boldsymbol{\eta} + \boldsymbol{\varepsilon}$ , et  $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)^\top$ , est le vecteur des erreurs découlant du fait que nous n'employons qu'une partie de la base pour estimer la courbe de référence.

Nous supposons que la distribution *a priori* de  $\boldsymbol{\eta}$  est

$$\boldsymbol{\eta} | \sigma^2 \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{Q}_n),$$

où la matrice  $\mathbf{Q}_n$  dépend de la structure de variance des  $(\beta_l)_{l \leq L-1, l \geq U+1}$ . Par souci de simplicité, nous supposons que ces coefficients sont indépendants et que leur variance est proportionnelle à  $\frac{\sigma^2}{n_0} \left(\frac{1}{k}\right)^{|l|}$ , en choisissant  $k > 1$ . Cette forme est similaire à celle suggérée par Zellner (1986). Cependant, comme nous le démontrerons un peu plus tard, l'ajout de la constante  $\left(\frac{1}{k}\right)^{|l|}$  nous assure que la variance est toujours finie. De cette façon, l'élément  $(i, s)$  de  $\mathbf{Q}_n$  est donné par

$$\begin{aligned} (\mathbf{Q}_n)_{i,s} &= \sigma^{-2} \text{Cov}(\eta_i, \eta_s) \\ &= \sigma^{-2} \text{Cov} \left( 2^j \left( \frac{1-r}{1+r} \right)^2 \left\{ \sum_{p=-\infty}^{L-1} \rho_{r,p}^*(x_i) \beta_p + \sum_{p=U+1}^{\infty} \rho_{r,p}^*(x_i) \beta_p \right\}, \right. \\ &\quad \left. 2^j \left( \frac{1-r}{1+r} \right)^2 \left\{ \sum_{q=-\infty}^{L-1} \rho_{r,q}^*(x_s) \beta_q + \sum_{q=U+1}^{\infty} \rho_{r,q}^*(x_s) \beta_q \right\} \right) \\ &= d^* \left\{ \sum_{p=-\infty}^{L-1} \sum_{q=-\infty}^{L-1} \rho_{r,p}^*(x_i) \rho_{r,q}^*(x_s) \text{Cov}(\beta_p, \beta_q) \right. \\ &\quad \left. + \sum_{p=U+1}^{\infty} \sum_{q=U+1}^{\infty} \rho_{r,p}^*(x_i) \rho_{r,q}^*(x_s) \text{Cov}(\beta_p, \beta_q) \right\} \\ &= d^* \left\{ \sum_{p=-\infty}^{L-1} \rho_{r,p}^*(x_i) \rho_{r,p}^*(x_s) \text{Var}(\beta_p) + \sum_{p=U+1}^{\infty} \rho_{r,p}^*(x_i) \rho_{r,p}^*(x_s) \text{Var}(\beta_p) \right\} \\ &= d_1^* \left\{ \sum_{p=-\infty}^{L-1} \rho_{r,p}^*(x_i) \rho_{r,p}^*(x_s) \left(\frac{1}{k}\right)^{|p|} + \sum_{p=U+1}^{\infty} \rho_{r,p}^*(x_i) \rho_{r,p}^*(x_s) \left(\frac{1}{k}\right)^{|p|} \right\}, \end{aligned}$$

$$\text{où } d^* = \sigma^{-2} 2^{2j} \left( \frac{1-r}{1+r} \right)^4 \text{ et } d_1^* = \frac{2^{2j}}{n_0} \left( \frac{1-r}{1+r} \right)^4.$$

Puisque la fonction  $\rho_{r,l}^*(x)$  est bornée, nous sommes assurés que la matrice  $\mathbf{Q}_n$

est bien définie. Effectivement, d'après la propriété (1.3.8) et l'équation (1.3.9),

$$\rho_{r,p}^*(x_i)\rho_{r,p}^*(x_s) < 2^{-2j} \left( \frac{1+r}{1-r} \right)^2,$$

et, par conséquent,

$$\begin{aligned} |(\mathbf{Q}_n)_{i,s}| &< \frac{1}{n_0} \left( \frac{1-r}{1+r} \right)^2 \left\{ \sum_{p=-\infty}^{\infty} \left( \frac{1}{k} \right)^{|p|} - \sum_{p=L+1}^{U-1} \left( \frac{1}{k} \right)^{|p|} \right\} \\ &= \frac{1}{n_0} \left( \frac{1-r}{1+r} \right)^2 \left\{ 2 \sum_{p=0}^{\infty} \left( \frac{1}{k} \right)^p - 1 - \sum_{p=L+1}^{U-1} \left( \frac{1}{k} \right)^{|p|} \right\} \\ &< \frac{1}{n_0} \left( \frac{1-r}{1+r} \right)^2 \left\{ 2 \sum_{p=0}^{\infty} \left( \frac{1}{k} \right)^p \right\} \\ &= \frac{1}{n_0} \left( \frac{1-r}{1+r} \right)^2 \left( \frac{2}{1-1/k} \right) \\ &< +\infty. \end{aligned}$$

En sachant cela, nous pouvons développer le modèle complet. De façon similaire à ce qui a été proposé pour le modèle utilisant les splines à la section 2.3.1.1, nous présumons que

$$\bar{\mathbf{y}}|\boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2[\mathbf{I}_n + \mathbf{Q}_n]),$$

si  $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$ .

En ce qui concerne les hypothèses par rapport au vecteur des coefficients  $\boldsymbol{\beta}$  et à la variance des erreurs  $\sigma^2$ , les présupposés faits sont les mêmes qu'auparavant, c'est-à-dire que

$$\begin{aligned} \boldsymbol{\beta}|\sigma^2 &\sim \mathcal{N}_{U-L} \left( \boldsymbol{\beta}^0, \frac{\sigma^2}{n_0} \left( \mathbf{X}^\top [\mathbf{I}_n + \mathbf{Q}_n]^{-1} \mathbf{X} \right)^{-1} \right), \\ \sigma^2 &\sim \mathcal{IG} \left( \frac{\alpha}{2}, \frac{\gamma}{2} \right), \end{aligned}$$

et les paramètres  $\boldsymbol{\beta}^0$ ,  $n_0$ ,  $\alpha$  et  $\gamma$  sont également choisis comme précédemment.

Mentionnons qu'il peut paraître curieux de choisir une matrice de covariance qui suppose une certaine dépendance entre les coefficients  $\beta_l$ ,  $l = L, L+1, \dots, U$ , que nous utilisons lors de l'estimation, alors que nous supposons l'indépendance pour les autres. En fait, si nous définissons

$$\boldsymbol{\beta}_c = (\boldsymbol{\beta}^\top, \boldsymbol{\beta}_\eta^\top)^\top,$$

où  $\beta_\eta$  est le vecteur des coefficients qui se retrouvent dans le terme d'erreur, nous avons

$$\text{Var}(\beta_c) = \frac{\sigma^2}{n_0} \begin{bmatrix} (\mathbf{X}^\top [\mathbf{I}_n + \mathbf{Q}_n]^{-1} \mathbf{X})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

Il aurait été beaucoup plus naturel de remplacer la matrice  $(\mathbf{X}^\top [\mathbf{I}_n + \mathbf{Q}_n]^{-1} \mathbf{X})^{-1}$  par l'identité, mais nous avons néanmoins convenu d'utiliser ce modèle, non seulement pour qu'il ressemble à celui utilisé avec les splines, mais aussi car il simplifie grandement l'expression des estimateurs obtenus. Soulignons que l'impact de cette décision n'est pas majeur quant au modèle obtenu. En effet, comme l'explique Goel (1983), plus nous nous éloignons des données, c'est-à-dire plus le niveau  $i$  d'un hyperparamètre  $\theta_i$  augmente, plus l'information à son sujet décroît. Ainsi, selon certaines mesures de divergence décrites dans cet article, l'importance de la différence entre les distributions *a priori* et *a posteriori* de cet hyperparamètre diminue. Donc, même si ce modèle peut sembler moins réaliste, ses conséquences sur l'inférence sont plus ou moins négligeables.

#### 2.4.2.2. Modèle 2 : sur les observations directement

Ce second modèle ressemble également énormément à celui développé avec les splines à la section 2.3.1.2. Le seul changement effectué est dû à la présence de la matrice  $\mathbf{Q}_n$  qui influence la structure de variance des observations. Le modèle adopté est donc

$$\begin{aligned} \mathbf{y}_i | \beta_i &\sim \mathcal{N}_n(\mathbf{X}\beta_i, \sigma^2[\mathbf{I}_n + \mathbf{Q}_n]), \\ \beta_i | \beta &\sim \mathcal{N}_{U-L}\left(\beta, \frac{\sigma^2}{n_0} \mathbf{I}_{U-L}\right), \\ \beta | \sigma^2 &\sim \mathcal{N}_{U-L}\left(\beta^0, \frac{\sigma^2}{n_1} \mathbf{I}_{U-L}\right), \\ \sigma^2 &\sim \mathcal{IG}\left(\frac{\alpha}{2}, \frac{\gamma}{2}\right). \end{aligned}$$

De plus, puisque les définitions de  $\beta^0$ ,  $n_0$ ,  $n_1$ ,  $\alpha$  et  $\gamma$  sont encore une fois les mêmes qu'avant, le lecteur peut se référer à la section 2.3.1.1 pour obtenir plus de détails quant à ces derniers.

Les modèles étant maintenant précisés, nous pouvons nous attarder au problème principal.

### 2.4.3. Estimateur de la courbe de référence

Nous cherchons donc à retrouver l'estimateur du vecteur  $\beta$  qui, sous la fonction de perte quadratique, correspond à l'espérance de la distribution *a posteriori*  $\pi(\beta|\bar{\mathbf{y}})$ . Pour ce faire, les mêmes raisonnements que ceux élaborés dans les sections 2.3.2.1 et 2.3.2.2 s'appliquent. C'est pourquoi toutes les techniques utilisées et les résultats intermédiaires trouvés sont laissés de côté. Nous ne présentons ainsi que les résultats pertinents au calcul de la densité marginale des observations  $m(\bar{\mathbf{y}})$  et de l'estimateur  $\hat{\beta}$ .

#### 2.4.3.1. Modèle 1 : sur la moyenne des observations

Si nous posons  $\beta^{MCP} = (\mathbf{X}^\top(\mathbf{I}_n + \mathbf{Q}_n)^{-1}\mathbf{X})^{-1}\mathbf{X}^\top(\mathbf{I}_n + \mathbf{Q}_n)^{-1}\bar{\mathbf{y}}$ , nous obtenons

$$\begin{aligned}\beta|\sigma^2, \bar{\mathbf{y}} &\sim \mathcal{N}_{U-L}(\beta^*, \sigma^2 \Sigma_\beta^*), \\ \sigma^2|\bar{\mathbf{y}} &\sim \mathcal{IG}\left(\frac{\alpha^*}{2}, \frac{\gamma^*}{2}\right),\end{aligned}$$

où

$$\begin{aligned}\beta^* &= \frac{1}{n_0 + 1} (\beta^{MCP} + n_0 \beta^0), \\ \Sigma_\beta^* &= \frac{1}{n_0 + 1} [\mathbf{X}^\top(\mathbf{I}_n + \mathbf{Q}_n)^{-1}\mathbf{X}]^{-1}, \\ \alpha^* &= n + \alpha, \\ \gamma^* &= \gamma + (\bar{\mathbf{y}} - \mathbf{X}\beta^{MCP})^\top(\mathbf{I}_n + \mathbf{Q}_n)^{-1}(\bar{\mathbf{y}} - \mathbf{X}\beta^{MCP}) \\ &\quad + \frac{n_0}{n_0 + 1} (\beta^{MCP} - \beta^0)^\top \mathbf{X}^\top(\mathbf{I}_n + \mathbf{Q}_n)^{-1}\mathbf{X}(\beta^{MCP} - \beta^0).\end{aligned}$$

La distribution *a posteriori* de  $\beta$  consiste alors en une distribution t de Student multidimensionnelle

$$\beta|\bar{\mathbf{y}} \sim T_{U-L}\left(\alpha^*, \beta^*, \frac{\gamma^*}{\alpha^*} \Sigma_\beta^*\right). \quad (2.4.11)$$

Cette fois, l'estimateur  $\beta^*$  obtenu correspond à une moyenne pondérée entre l'estimateur des moindres carrés pondérés  $\beta^{MCP}$  et la moyenne *a priori*  $\beta^0$ .

Mentionnons aussi que la distribution marginale des observations est donnée par

$$\left(\frac{n_0}{n_0+1}\right)^{\frac{U-L}{2}} \frac{1}{|\mathbf{I}_n + \mathbf{Q}_n|^{1/2}} \left(\frac{\Gamma(\alpha^*/2)}{\pi^{n/2}\Gamma(\alpha/2)}\right) \left(\frac{\gamma^{\alpha/2}}{(\gamma^*)^{\alpha^*/2}}\right).$$

Passons maintenant au second modèle.

#### 2.4.3.2. Modèle 2 : sur les observations directement

Rappelons que le calcul de l'estimateur  $\hat{\beta}$  à l'aide des distributions conditionnelles tel que décrit à l'équation (2.3.4) nécessite la connaissance des lois de  $\beta|\mathbf{Y}, \sigma^2$  et de  $\sigma^2|\mathbf{Y}$ . Les voici :

$$\beta|\mathbf{Y}, \sigma^2 \sim \mathcal{N}_{U-L}(\mathbf{C}\mathbf{c}, \sigma^2\mathbf{C}), \quad (2.4.12)$$

$$\sigma^2|\mathbf{Y} \sim \mathcal{IG}\left(\frac{nN + \alpha}{2}, \frac{\gamma'}{2}\right), \quad (2.4.13)$$

où

$$\mathbf{C} = (Nn_0\mathbf{X}^\top(n_0(\mathbf{I}_n + \mathbf{Q}_n) + \mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{X} + n_1\mathbf{I}_{U-L})^{-1}, \quad (2.4.14)$$

$$\mathbf{c} = Nn_0\mathbf{X}^\top(n_0(\mathbf{I}_n + \mathbf{Q}_n) + \mathbf{X}\mathbf{X}^\top)^{-1}\bar{\mathbf{y}} + n_1\beta^0, \quad (2.4.15)$$

$$\begin{aligned} \gamma' &= N(\bar{\mathbf{y}} - \mathbf{X}\beta^0)^\top \left( \mathbf{I}_n + \mathbf{Q}_n + \left(\frac{n_0N + n_1}{n_0n_1}\right) \mathbf{X}\mathbf{X}^\top \right) (\bar{\mathbf{y}} - \mathbf{X}\beta^0) \\ &+ n_0\text{tr}(\{n_0(\mathbf{I}_n + \mathbf{Q}_n) + \mathbf{X}\mathbf{X}^\top\}^{-1}\mathbf{S}) + \gamma. \end{aligned} \quad (2.4.16)$$

Pour trouver l'estimateur des coefficients, nous appliquons alors l'équation (2.3.15) qui, dans ce cas, devient

$$\hat{\beta} = \mathbf{C}\mathbf{c}, \quad (2.4.17)$$

où la matrice  $\mathbf{C}$  est définie à l'équation (2.4.14) et le vecteur  $\mathbf{c}$  se trouve à l'équation (2.4.15).

Finalement, la densité marginale des observations est donnée par

$$\begin{aligned} m(\mathbf{Y}) &= \frac{n_0^{\frac{n}{2}(N-1)}\Gamma([nN + \alpha]/2)}{\pi^{\frac{nN}{2}}\Gamma(\alpha/2)} \frac{\gamma^{\alpha/2}}{|n_0(\mathbf{I}_n + \mathbf{Q}_n) + \mathbf{X}\mathbf{X}^\top|^{\frac{N-1}{2}}} \\ &\times \frac{(\gamma')^{-\frac{nN+\alpha}{2}}}{\left|\mathbf{I}_n + \mathbf{Q}_n + \left(\frac{Nn_0 + n_1}{n_0n_1}\right) \mathbf{X}\mathbf{X}^\top\right|^{\frac{1}{2}}}, \end{aligned}$$

où  $\gamma'$  est défini de la même façon qu'à l'équation (2.4.16).

Le développement de tous les modèles est maintenant terminé. Puisqu'ils ont été élaborés dans un contexte général, il suffit seulement de les modifier légèrement afin de tenir compte des contraintes de monotonie. Notons que nous aurions pu imposer ces contraintes dès la spécification du modèle *a priori*. Ceci a d'ailleurs déjà été fait avec les splines par quelques auteurs dont Bennachmouch (1992) qui utilise la méthode de Monte Carlo avec fonction d'importance afin de retrouver l'estimateur qui a été élaboré à l'aide d'un lisseur sous ce type de contraintes. Par contre, les calculs analytiques deviennent impossibles à ce moment et nous avons plutôt opté pour une procédure qui se trouve à être beaucoup plus facile à utiliser et qui permet également l'estimation d'autres fonctions pour lesquelles ces contraintes ne doivent pas nécessairement être respectées.

## 2.5. ESTIMATION AVEC CONTRAINTES

Comme mentionné précédemment, nous voulons tenir compte du fait que nos estimations doivent absolument être non décroissantes, c'est-à-dire que  $\hat{f}(x') \geq \hat{f}(x)$  si  $x' \geq x$ . Pour ce faire, il existe plusieurs moyens d'y arriver dont les deux techniques décrites dans cette section.

### 2.5.1. Coefficients positifs

Selon nos modèles, l'estimation de la fonction en un point  $x$  est donnée par

$$\hat{g}(x) = \sum_i \hat{\beta}_i b_i(x),$$

où  $b_i(x)$  représente le  $i^e$  élément de la base évalué au point  $x$ . Évidemment, nous avons choisi de travailler avec des bases pour lesquelles

$$b_i(x') \geq b_i(x), \quad \forall x' \geq x.$$

Pour que l'estimation soit non décroissante, une condition suffisante, mais non nécessaire, est donc que

$$\hat{\beta}_i \geq 0, \quad \forall i. \tag{2.5.1}$$

Une technique pour parvenir à n'obtenir que des coefficients positifs consiste à utiliser la méthode de Monte Carlo pour générer des vecteurs de coefficients selon la loi *a posteriori* trouvée et à ne conserver que ceux dont tous les éléments sont

positifs. Cette méthode est une technique classique de calcul bayésien à la fois simple et efficace qui permet d'approcher de façon numérique des intégrales non calculables analytiquement. Le principe de la méthode est le suivant (voir Robert, 2001).

Si nous dénotons par  $D(\theta)$ , l'espace du paramètre  $\theta$  lorsque nous incluons les contraintes qui nous intéressent, nous voulons calculer une intégrale de la forme :

$$\int_{\Theta} g(\theta) f(x|\theta) \pi(\theta) I_D(\theta) d\theta, \quad (2.5.2)$$

que nous pouvons aussi écrire comme

$$\int_{\Theta} \frac{g(\theta) f(x|\theta) \pi(\theta) I_D(\theta)}{h(\theta)} h(\theta) d\theta,$$

où  $h$  est une densité de probabilité sur  $\Theta$  qui vérifie les trois propriétés suivantes :

- (1) le support de  $f(x|\theta) \pi(\theta) I_D(\theta)$  est inclus dans le support de  $h$ ,
- (2)  $\lim_{|\theta| \rightarrow \infty} \frac{f(x|\theta) \pi(\theta) I_D(\theta)}{h(|\theta|)} < \infty$ ,
- (3) il doit être relativement facile, d'un point de vue numérique, de générer  $\theta$  à partir de  $h$ .

Nous appelons méthode de Monte Carlo avec fonction d'importance  $h$  l'algorithme suivant :

- (1) générer  $\theta_1, \dots, \theta_m$  selon la densité  $h$ ,
- (2) faire l'approximation de l'intégrale (2.5.2) par :

$$\frac{1}{m_0} \sum_{i=1}^{m_0} g(\theta_i) w(\theta_i),$$

$$\text{où } m_0 = \sum_{i=1}^m I_D(\theta_i) \text{ et } w(\theta_i) = \frac{f(x|\theta_i) \pi(\theta_i) I_D(\theta_i)}{h(\theta_i)}.$$

La loi forte des grands nombres nous assure que, sous quelques conditions de régularité comme le fait que l'intégrale soit finie, cette approximation converge vers l'équation (2.5.2).

Avant de donner les estimateurs finaux, vérifions d'abord pour chacun des modèles, à partir de quelle distribution nous pouvons générer le vecteur de coefficients  $\beta$ . Autrement dit, trouvons en partie la fonction  $h$ .

Pour le premier modèle, nous connaissons la distribution *a posteriori* de  $\beta|\bar{y}$ . Nous utilisons donc cette loi de Student multidimensionnelle, définie en (2.3.3) pour les splines et en (2.4.11) pour les ondelettes intégrées, pour générer  $\beta$  directement. Par contre, nous ne l'avons pas calculée pour le second modèle. Malgré le fait qu'il serait possible de le faire en appliquant la même procédure qu'à l'équation (2.3.2) et en intégrant par la suite par rapport à  $\sigma^2$ , nous préférons utiliser la théorie sur les décompositions conditionnelles et donc, procéder par étapes :

- (1) Nous générons d'abord  $\sigma^2$  à l'aide de la loi  $\sigma^2|\mathbf{Y}$ . Cette dernière correspond à une inverse gamma et nous la retrouvons aux équations (2.3.13) et (2.4.13) pour les splines et les ondelettes respectivement.
- (2) Nous générons ensuite  $\beta$  à partir de la distribution  $\beta|\mathbf{Y}, \sigma^2$ . Celle-ci consiste en une loi normale multidimensionnelle et est définie à l'équation (2.3.9) pour les splines et à l'équation (2.4.12) pour les ondelettes.

À partir de ceci, il est facile de voir que si nous remplaçons la fonction  $g(\beta)$  par  $\beta$  et la fonction  $h$  par les lois décrites précédemment, en prenant soin de tenir compte de l'espace  $D(\beta)$  qui se trouve à être l'espace des vecteurs dont tous les éléments sont non négatifs, nous obtenons l'estimateur désiré. Notons qu'en pratique, ceci signifie que nous générons le vecteur  $\beta$  avec les techniques mentionnées ci-dessus et que nous ne conservons que ceux qui respectent la contrainte (2.5.1). Nous obtenons ainsi un ensemble de vecteurs de coefficients  $\beta_1^+, \dots, \beta_{m_0}^+$ , et l'estimateur recherché devient

$$\beta^+ = \frac{1}{m_0} \sum_{i=1}^{m_0} \beta_i^+. \quad (2.5.3)$$

### 2.5.2. Dérivées positives

Cette condition de non négativité donne les résultats escomptés, mais elle est assez restrictive et peut parfois s'avérer trop sévère pour certaines bases. Cependant, rappelons le lien entre la monotonie d'une fonction et sa dérivée : une fonction  $g(x)$  est non décroissante si sa dérivée  $g'(x)$  est positive ou nulle pour tout  $x$ . Pour obtenir l'estimateur, nous pouvons donc de nouveau utiliser la méthode de

Monte Carlo en changeant cette fois, la contrainte (2.5.1) par la suivante :

$$(\hat{g}(x))' = \sum_i \hat{\beta}_i b'_i(x) \geq 0. \quad (2.5.4)$$

L'espace  $D(\beta)$  est donc celui comprenant tous les vecteurs dont les coefficients font en sorte que la dérivée de la fonction estimée soit positive pour toute valeur de  $x$  comprise dans le domaine de la courbe de référence.

Notons qu'il est très simple de calculer la dérivée de la fonction estimée en un point  $x$  pour les bases que nous utilisons, car

$$\begin{aligned} \frac{\partial I_i(x|k, \mathbf{t})}{\partial x} &= M_i(x|k, \mathbf{t}), \\ \frac{\partial \rho_{r,l}^*(x)}{\partial x} &= \rho_r(2^j x - l), \end{aligned}$$

c'est-à-dire que la dérivée d'une fonction I-spline est une fonction M-spline et que celle de la fonction  $\rho_{r,l}^*(x)$  correspond à la fonction  $\rho_r(2^j x - l)$ .

En pratique, pour chacun des vecteurs générés, nous calculons les dérivées définies à l'équation (2.5.4) pour un grand nombre de valeurs du domaine de la courbe de référence et, si le minimum de ces dérivées est positif, alors nous concluons que le vecteur de coefficients généré respecte la contrainte et nous le conservons. L'estimateur se calcule alors de la même façon qu'à l'équation (2.5.3). Notons que, contrairement au critère (2.5.1), la condition (2.5.4) est nécessaire à l'obtention d'estimations croissantes. En fait, ce second critère englobe le premier, c'est-à-dire que si tous les coefficients sont positifs, les dérivées le sont automatiquement, puisque les bases de fonctions avec lesquelles nous travaillons sont nécessairement positives et croissantes. Par contre, lorsque ce premier critère n'est pas trop sévère, la vérification de ce dernier devient très avantageuse par rapport au temps d'exécution, puisque l'évaluation des dérivées doit être faite pour un très grand nombre de valeurs du domaine de la courbe, ce qui devient souvent très long.

Ces deux estimateurs tiennent ainsi compte des contraintes que nous avons imposées au début de ce chapitre. Nous avons donc complété le développement de quatre modèles similaires, en utilisant deux bases de fonctions différentes, à raison de deux modèles par base. Dans le chapitre qui suit, nous les appliquerons en utilisant un jeu de données réelles.

# Chapitre 3

---

## APPLICATIONS

Ce chapitre est consacré à l'application de la théorie développée dans le chapitre précédent. Nous l'illustrons d'abord à l'aide d'un jeu de données réelles, ce qui nous permet également d'établir certaines comparaisons quant à la performance des deux modèles lorsque chacune des deux bases est utilisée. Nous poursuivons ensuite cette analyse en vérifiant la validité de nos estimateurs pour des fonctions qui sont beaucoup plus variables. Commençons tout de suite avec un exemple portant sur les courbes de croissance.

### 3.1. EXEMPLE AVEC DONNÉES RÉELLES

Dans cet exemple, les données proviennent de l'étude de croissance de Berkeley (voir Tuddenham et Snyder, 1954). Cette dernière comprend les courbes de 39 garçons et de 54 filles. La taille de chacun des enfants a été prise à 31 reprises entre 1 et 18 ans. De plus, les mesures ne sont pas équidistantes : elles sont trimestrielles jusqu'à l'âge de deux ans, annuelles de deux à huit ans et, finalement, biannuelles jusqu'à dix-huit ans. Plus d'information a donc été amassée aux périodes où la croissance est la plus rapide, soient durant la tendre enfance et à l'adolescence. Rajoutons finalement que l'erreur type de ces mesures est d'environ 3 millimètres et qu'elle a tendance à être un peu plus grande pour les premières mesures. Notons que ces données sont plutôt difficiles à obtenir, puisqu'elles nécessitent un suivi d'une vingtaine d'années. Donc, comme toute étude de type longitudinal, ceci engendre des coûts élevés et comporte des risques d'attrition. Il existe quand même certaines études plus vastes qui ont été complétées. Pensons notamment

aux données de Fels (voir Roche, 1992) ou de Zurich (voir Falkner, 1960). Par contre, l'étude de Berkeley a déjà été publiée et elle est disponible gratuitement. C'est pourquoi nous avons choisi de travailler avec celle-ci. Bien sûr, nous ne sommes pas les premiers à tenter d'estimer ces courbes. Ramsay, Bock et Gasser (1995) comparent, par exemple, les courbes d'accélération associées à la hauteur pour les trois jeux de données cités antérieurement. La figure 3.1 présente un sous-ensemble de courbes pour dix garçons choisis aléatoirement dans l'échantillon.

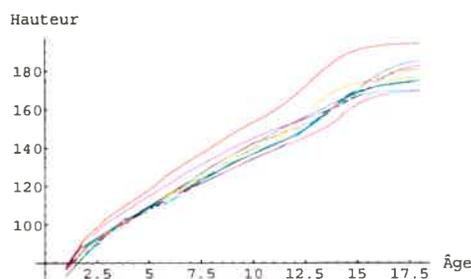


FIG. 3.1. Courbes de croissance de dix garçons choisis aléatoirement

Nous remarquons qu'il y a un garçon qui est grand à tout âge, mais que certains autres sont grands durant l'enfance, bien qu'ils finissent avec une stature relativement petite à l'âge adulte et vice-versa. Plus généralement parlant, nous observons une tendance commune : la croissance la plus rapide se retrouve dans les premières années, mais il y a augmentation de la pente durant la poussée de croissance de la puberté, c'est-à-dire entre 11 et 16 ans. Nous faisons d'ailleurs l'étalonnage des courbes par rapport à cette poussée, car elle ne survient pas au même moment pour tous les enfants. Regardons la méthode utilisée pour y parvenir d'un peu plus près. Soulignons seulement que les estimations pour les garçons et les filles sont effectuées séparément.

### 3.1.1. Synchronisation événementielle

Premièrement, il nous faut déterminer avec exactitude le temps auquel l'événement marquant, qui consiste en cette poussée, se produit pour chacun des enfants. Ceci n'est pas trivial, puisqu'il n'est pas évident de retrouver, à l'oeil, l'âge auquel la poussée de croissance débute ou se termine. Pour contrer ce problème, nous calculons la vitesse de croissance approximative pour les temps 2 à 30, ce qui nous

aide à mieux distinguer les détails par rapport aux changements dans chacune des courbes. Pour ce faire, nous appliquons la formule d'approximation de la dérivée que voici. Au temps  $x_i$ ,  $i = 2, \dots, 30$ , nous avons

$$V_{x_i} = \frac{g(x_i) - g(x_{i-1})}{x_i - x_{i-1}},$$

où  $g(x_i)$  représente la valeur de la mesure observée au temps  $x_i$ . Le temps auquel la vitesse maximale est observée lors de la puberté (l'endroit où nous retrouvons un pic) est alors considéré, pour chacun des enfants, comme le temps de référence servant à l'alignement des courbes. En résumé, les trois conditions devant être respectées dont nous avons parlé à la section 2.1 sont

$$h_i(T_0) = T_0, \quad h_i(x_a) = \tau_a, \quad h_i(T_n) = T_n,$$

où  $T_0$  et  $T_n$  sont, respectivement, les bornes inférieure et supérieure du domaine de  $x$ ,  $x_a$  est, pour chacune des courbes, l'âge auquel la vitesse atteint son seuil maximal et  $\tau_a$  est la médiane de ces âges pour toutes les courbes. Ainsi, nos fonctions de synchronisation sont faites de deux parties linéaires séparées par la médiane  $\tau_a$ . La figure 3.2 illustre bien l'effet de la synchronisation chez deux des garçons faisant partie de notre échantillon. De plus, la figure 3.3 montre les transformations associées à l'étalonnage de ces deux fonctions.

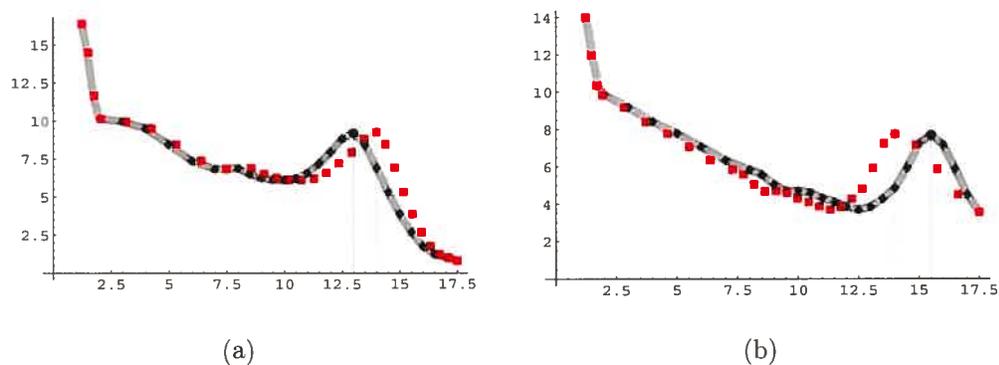


FIG. 3.2. Effet de la synchronisation : Données originales (trait plein gris), Données transformées (carrés rouges) (a) Garçon 1 (b) Garçon 2

Pour l'ensemble des garçons, l'âge médian  $\tau_a$  obtenu est de 14 ans. Dans le premier exemple, la vitesse maximale observée est plutôt atteinte à 13,5 ans. Pour

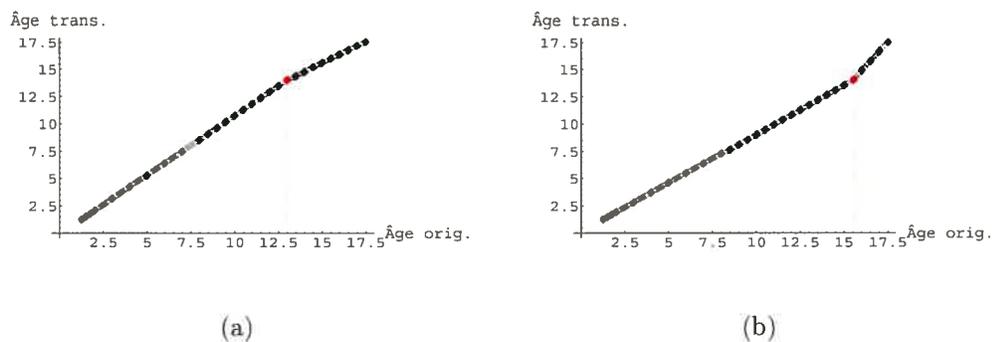


FIG. 3.3. Temps transformés en fonction de ceux observés (a) Garçon 1 (b) Garçon 2

compenser par rapport à l'avance qu'avait cet enfant par rapport à la croissance de l'ensemble des garçons, le domaine associé à la première partie des observations, donc celles avant 13,5 ans, a été élargi, alors qu'il y a eu contraction des données entre cet âge et dix-huit ans. Cet effet est également visible en regardant la pente de la droite représentant les temps transformés en fonction des temps originaux. En fait, lorsqu'il y a contraction, cette pente est inférieure à l'unité et, dans le cas contraire, elle est supérieure à l'unité. Finalement, le cas du second enfant est l'opposé du premier. En effet, il accusait un certain retard, car sa vitesse maximale a plutôt été observée à l'âge de 15,5 ans, ce qui explique le fait qu'il y a d'abord eu contraction des données. La synchronisation pour l'ensemble des garçons et des filles est illustrée aux figures 3.4 et 3.5 respectivement.

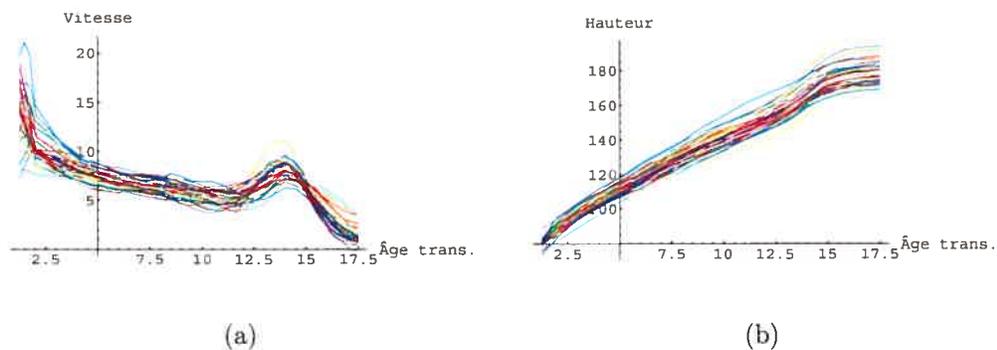


FIG. 3.4. Fonctions synchronisées pour les garçons (a) Vitesses (b) Hauteurs

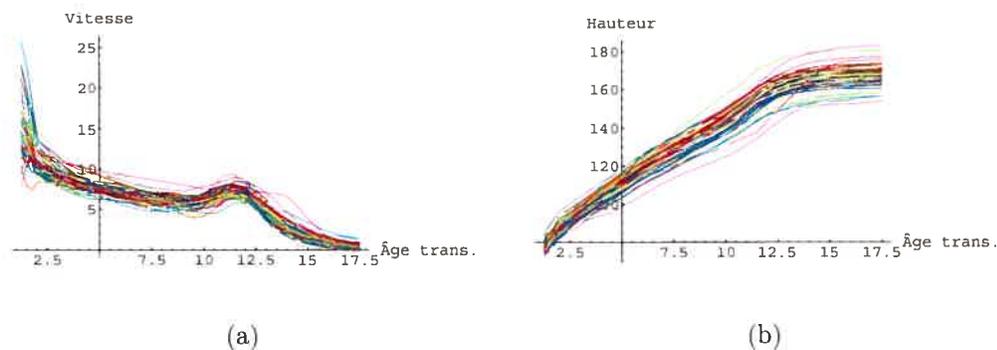


FIG. 3.5. Fonctions synchronisées pour les filles (a) Vitesses (b) Hauteurs

Notons que l'âge médian auquel la vitesse maximale lors de la puberté a été observée pour les filles est de 11,5 ans. Mentionnons aussi qu'il y a une fille pour laquelle la synchronisation n'a pas très bien fonctionné. Ceci est dû au fait que cette enfant ne présentait pas une poussée de croissance très accentuée et qu'il a donc été difficile de trouver la valeur pour laquelle l'alignement serait fait. Ceci est d'ailleurs l'un des principaux reproches dont fait l'objet la synchronisation événementielle (voir Ramsay et Li, 1998). En fait, il arrive que certaines courbes ne présentent pas les particularités désirées et, à ce moment, il devient difficile de synchroniser les courbes. Par contre, dans notre cas, rappelons qu'il y a 54 filles dans ce jeu de données et qu'il est donc concevable qu'une seule de ces courbes ne soit pas très bien alignée par rapport aux autres.

Une fois cette étape franchie, il suffit de faire la moyenne, à chacun des temps, de toutes les hauteurs pour chacun des enfants pour obtenir la courbe de référence que nous estimons. Cependant, comme nous l'avons déjà expliqué à la première section du second chapitre, puisque les transformations par rapport aux temps sont faites pour chacun des enfants séparément, il nous faut procéder par interpolation linéaire afin de retrouver la hauteur moyenne approximative pour chacun des temps originaux  $x_i$ ,  $i = 1, \dots, 31$ .

La figure 3.6 rapporte les courbes moyennes obtenues pour les deux groupes étudiés. Notons que les données ont été standardisées entre 0 et 1 afin de faciliter l'obtention de coefficients positifs lors de nos estimations.

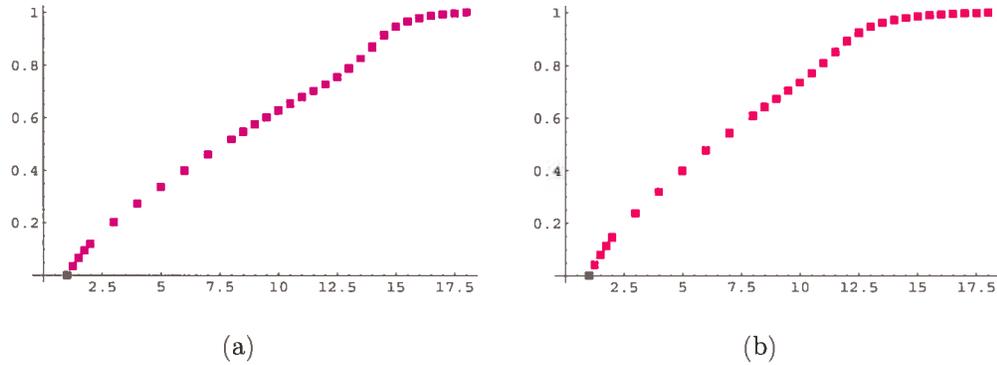


FIG. 3.6. Courbes moyennes obtenues après la standardisation (a) Garçons (b) Filles

Nous remarquons que pour les garçons, la poussée de croissance est plus accentuée que pour les filles et, comme mentionné précédemment, qu'elle arrive un peu plus tard (14 ans contre 11,5 ans). La prochaine étape consiste maintenant à estimer ces courbes, et ce, au moyen des deux modèles développés ainsi que pour chacune des bases utilisées.

### 3.1.2. Modélisation

Indiquons, avant tout, les étapes à suivre afin d'obtenir la modélisation voulue pour chacun des différents modèles.

- (1) Nous commençons par fixer un ensemble de valeurs possibles variant entre 0 et 1 pour l'hyperparamètre  $n_0$ . Pour chacune de ces valeurs, nous choisissons les éléments de la base qui font en sorte que le critère  $BIC$  introduit à l'équation (1.2.2) est minimisé. Dans le cas des splines, ces éléments sont totalement déterminés par l'ensemble des noeuds intérieurs que nous gardons. Dans le cas des ondelettes, ce sont les fonctions  $\rho_{r,l}^*(x)$  qui constituent cet ensemble. Il suffit donc de trouver l'intervalle  $[L, U]$  défini au début de la section 2.4.2 qui répond aux exigences voulues. Les fonctions utilisées pour nos estimations seront alors  $\rho_{r,L}^*(x), \dots, \rho_{r,U}^*(x)$ . Notons que, pour une valeur de  $n_0$  particulière, l'algorithme nous permettant de réaliser ce choix est le suivant :

- (a) Pour les splines :

- (i) nous fixons un ensemble de noeuds en prenant soin d'en inclure un entre chaque observation.
  - (ii) Après avoir éliminé le premier noeud intérieur, nous calculons l'estimateur  $\beta^*$  décrit à l'équation (2.3.3) pour le premier modèle, ou l'estimateur  $\hat{\beta}$  que nous retrouvons à l'équation (2.3.15) pour le second, ainsi que les valeurs prédites  $\hat{y}_i$ ,  $i = 1, \dots, n$ , correspondantes. Nous évaluons ensuite le critère *BIC*. Cette étape est répétée pour tous les noeuds intérieurs. Nous conservons alors l'ensemble de noeuds pour lequel la valeur minimale du critère est observée.
  - (iii) Nous répétons l'étape précédente jusqu'à ce que la valeur du critère *BIC* ne puisse plus être diminuée.
- (b) Pour les ondelettes intégrées :
- (i) nous spécifions un intervalle relativement grand pour les valeurs possibles de  $l$ . Au départ, nous pouvons par exemple supposer que toute valeur variant entre -20 et 20 est plausible pour  $l$ .
  - (ii) Nous appliquons alors la procédure décrite aux étapes (ii) et (iii) pour les splines en utilisant plutôt les estimateurs  $\beta^*$  et  $\hat{\beta}$  introduits aux équations (2.4.11) et (2.4.17) respectivement, non pas pour toutes les valeurs de  $l$ , mais bien uniquement pour celles qui se situent aux bornes de l'intervalle.
- (2) Nous calculons ensuite la valeur de la loi marginale en considérant chacune des combinaisons  $n_0$ /éléments de la base obtenues à l'étape précédente. La valeur de  $n_0$  choisie est alors celle qui maximise la valeur de la loi marginale. De plus, les éléments de la base conservés sont ceux qui ont été associés à ce  $n_0$  à la première étape.
- (3) Nous générons 10 000 vecteurs de coefficients de la façon décrite à la section 2.5.
- (4) Pour chacun de ces vecteurs, nous vérifions s'il respecte les conditions imposées afin de nous assurer que les estimations soient non décroissantes.

Si c'est le cas, nous le conservons. Sinon, nous le laissons de côté. Notons que pour les splines, il a été possible de satisfaire la première condition décrite dans la section 2.5.1, c'est-à-dire que tous les coefficients devaient être positifs. Par contre, cette restriction s'est avérée être trop sévère dans le cas des ondelettes intégrées, puisque le premier coefficient était, pour la plupart du temps, négatif. Dans ce cas, c'est plutôt la seconde condition décrite à la section 2.5.2 qui a été vérifiée.

- (5) L'estimateur final est calculé en faisant la moyenne des vecteurs de coefficients que nous avons gardés à l'étape 4. Notons que nous nous sommes assurés de toujours faire les estimations à partir d'au moins 2000 vecteurs, c'est-à-dire 20% des vecteurs générés. Ainsi, si le nombre de vecteurs respectant la condition était inférieur à 2000, nous générions un peu plus de 10 000 vecteurs afin d'en conserver suffisamment.

Discutons maintenant des résultats obtenus pour chacun des modèles. Mentionnons que, pour ces exemples, les splines utilisés sont des I-splines. En outre, pour construire les fonctions  $\rho_{r,l}(x)$ , les ondelettes de Haar ont été privilégiées et nous avons fixé les valeurs de  $j$  et de  $r$  dans l'équation (2.4.3) à zéro et 0,3.

Arrêtons-nous d'abord sur le choix de  $n_0$ . Notons que nous montrons les résultats trouvés à partir du critère *BIC*, mais que le critère de He et Shi décrit à la section 1.2.3 aurait très bien pu être utilisé également. En fait, comme les résultats associés aux deux critères étaient très semblables, nous avons décidé de ne présenter que ceux obtenus à partir du premier critère. La figure 3.7 présente deux exemples de graphiques représentant le logarithme de la densité marginale en fonction de différentes valeurs possibles pour l'hyperparamètre  $n_0$ . Remarquons que les sauts observés dans le premier cas sont dus au changement de configuration des noeuds, c'est-à-dire que le nombre de noeuds total ainsi que l'emplacement des noeuds conservés diffèrent d'une valeur de  $n_0$  à une autre. En effet, comme nous pouvons le constater dans le second exemple, lorsque les éléments de la base demeurent les mêmes peu importe la valeur de  $n_0$ , la fonction résultante ne contient aucune discontinuité.

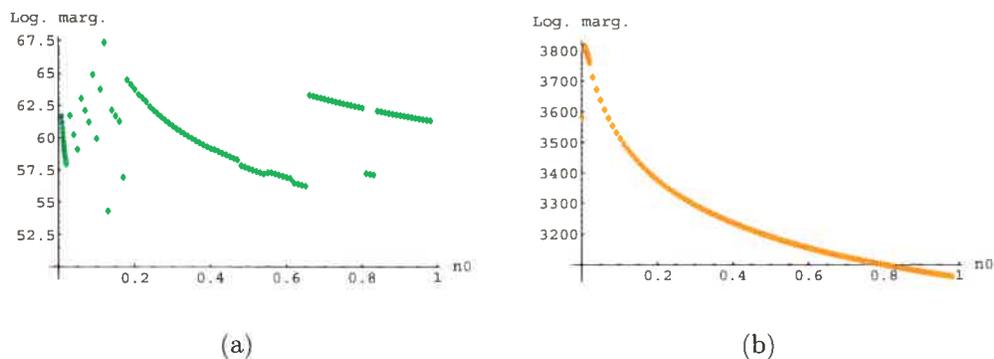


FIG. 3.7. Exemples du logarithme de la marginale en fonction de  $n_0$  pour les garçons (a) Modèle 1 - splines (b) Modèle 2 - ondelettes intégrées

Les tableaux 3.1 et 3.2 résument l'information qu'il est nécessaire de connaître avant de réaliser l'estimation des courbes moyennes, à savoir les valeurs de  $n_0$  et des éléments de la base retenus en utilisant les techniques décrites aux étapes 1 et 2, et ce, pour les splines et les ondelettes intégrées respectivement. Nous constatons que, pour les ondelettes intégrées, les valeurs de  $n_0$  choisies pour les différents modèles sont très près l'une de l'autre (environ 0,01). Par contre, pour les splines, elles fluctuent un peu plus. De plus, il est intéressant de relever que le nombre de coefficients associé au vecteur  $\beta$  est toujours inférieur dans le cas des splines. En fait, avec le premier modèle pour les garçons, il est précisément égal à la moitié de celui que nous retrouvons avec les ondelettes intégrées, ce qui constitue un atout considérable, puisque le but de toute estimation est généralement d'obtenir le meilleur ajustement possible en restreignant le nombre de paramètres pour que ce dernier soit minimal. Par contre, ajoutons ici un bémol quant au temps d'exécution relié au choix de ces valeurs, qui est beaucoup plus élevé dans le cas des splines. Ceci s'explique facilement en regardant l'algorithme décrit précédemment. En effet, à l'itération (ii) de l'étape (1), le critère *BIC* doit être déterminé pour tous les noeuds intérieurs dans le cas des splines, alors qu'à cette même étape, l'utilisation des ondelettes intégrées ne nécessite que deux évaluations de ce critère.

TAB. 3.1. Résultats en utilisant les splines.

Groupe	Modèle	$n_0$	Noeuds intérieurs	Nombre de coeff.	% des vecteurs respectant la contrainte 1
Filles	1	0,141	3,5; 10,25; 10,75; 11,25; 14,75;	8	17,2
	2	0,019	3,5; 4,5; 5,5; 6,5; 7,5; 8,25; 10,75; 11,25; 12,25; 13,75,	13	21,2
Garçons	1	0,121	2,5; 6,5; 10,75; 13,25; 14,25	8	28,6
	2	0,014	2,5; 6,5; 7,5; 11,25; 13,25; 14,75; 15,25	10	29,7

TAB. 3.2. Résultats en utilisant les ondelettes intégrées.

Groupe	Modèle	$n_0$	Intervalle pour $l$	Nombre de coeff.	% des vecteurs respectant la contrainte 2
Filles	1	0,008	[0, 14]	15	68,4
	2	0,01	[0, 14]	15	18,6
Garçons	1	0,007	[0, 15]	16	74,6
	2	0,008	[0, 15]	16	97,5

Ces tableaux indiquent également le nombre total de vecteurs de coefficients, parmi les 10 000 ayant été générés, qui satisfont les contraintes imposées afin de respecter la non décroissance des estimations. Sur ce point, les deux modèles sont difficilement comparables, car la condition imposée à ces vecteurs est différente d'une base à l'autre. Les résultats sont donc compréhensibles, car il est plus restrictif d'exiger que tous les coefficients soient positifs que d'imposer l'obligation que les dérivées le soient. En observant les résultats, nous constatons toutefois qu'il semble que les restrictions soient plus facilement satisfaites pour les garçons, ce qui dépend certainement de l'allure de la courbe estimée. Ceci devient particulièrement évident en constatant qu'il n'y a que 18,6% des vecteurs générés qui

respectent la contrainte observée pour le second modèle associé aux ondelettes dans le cas des filles. En réalité, ce résultat est annonciateur de ce qui suivra lors des simulations, c'est-à-dire que rien ne nous assure qu'il existe un grand nombre de vecteurs qui respecteront la contrainte imposée, ce qui est une conséquence de la méthode utilisée. Notons que ceci se produit surtout à cause des extrémités de la courbe où la hauteur est assez stable.

Finalement, la modélisation des courbes moyennes est illustrée aux figures 3.8 à 3.11. Les points représentent les vraies données, alors que les estimations sont en couleur. Soulignons que les ajustements sont tous excellents pour le premier modèle. Par contre, certaines estimations, bien qu'acceptables, sont un peu moins bonnes pour le second modèle. En effet, nous observons que, pour les splines, peu importe le groupe dont nous modélisons la courbe, il y a une légère surestimation des fonctions vers la fin de celles-ci.

Pour fins de comparaison, examinons les résultats associés aux estimateurs usuels habituellement utilisés en régression. Pour les splines, nous comparons alors les estimations que nous avons obtenues avec celles générées à l'aide de l'estimateur des moindres carrés ordinaires défini à la section 2.3.2. Pour les ondelettes intégrées, c'est plutôt l'estimateur des moindres carrés pondérés introduit à la section 2.4.3 que nous employons. En général, les estimateurs usuels performant aussi bien que ceux obtenus à l'aide du premier modèle et même parfois mieux que ceux obtenus à l'aide du second. La figure 3.12 illustre d'ailleurs la modélisation obtenue pour les filles en utilisant la même matrice de design que pour le second modèle employant les ondelettes intégrées (voir la figure 3.11 (b)). Rappelons toutefois qu'il est possible que ces estimations soient décroissantes, ce que nous devons tenter d'éviter. Bien sûr, certains estimateurs ont déjà été développés afin d'obtenir des estimations croissantes pour une fonction. Nous en avons déjà mentionné quelques-uns dans l'introduction de ce mémoire. Mais, comme notre but était surtout de vérifier la différence entre les deux modèles proposés, c'est-à-dire l'effet de suggérer un modèle *a priori* pour chacune des courbes individuelles comparativement à un modèle pour la moyenne des observations, nous

n'avons pas jugé nécessaire de comparer nos résultats à ceux de différentes méthodes déjà publiées. Le lecteur intéressé peut toutefois se référer à l'article de Angers et Delampady (2001) pour obtenir plus de détails quant à ces différentes comparaisons.

Notons enfin que les données avec lesquelles nous travaillons ne comportent que très peu de bruit, ce qui est souvent le cas des courbes de croissance où les mesures sont prises au millimètre près. Pour vérifier comment nos estimateurs se comportent lorsque nous sommes en présence de courbes qui varient un peu plus, nous tentons d'estimer de telles courbes dans la prochaine section en n'appliquant qu'un seul critère, défini par la présence de dérivées positives ou nulles sur tout le domaine qui nous intéresse.

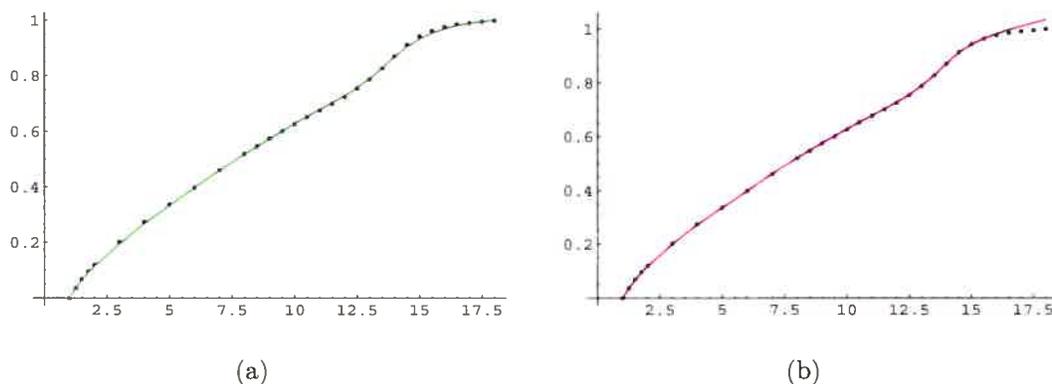


FIG. 3.8. Estimation (couleur) de la courbe moyenne (points) avec les splines pour les garçons (a) Modèle 1 (b) Modèle 2

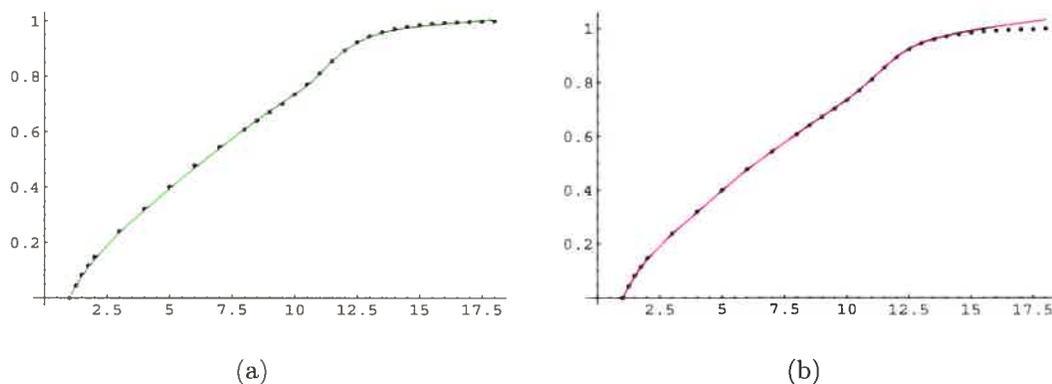


FIG. 3.9. Estimation (couleur) de la courbe moyenne (points) avec les splines pour les filles (a) Modèle 1 (b) Modèle 2

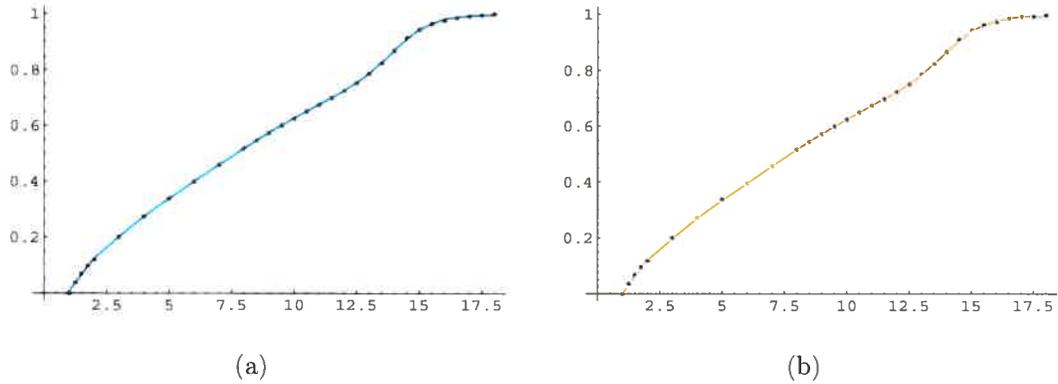


FIG. 3.10. Estimation (couleur) de la courbe moyenne (points) avec les ondelettes intégrées pour les garçons (a) Modèle 1 (b) Modèle 2

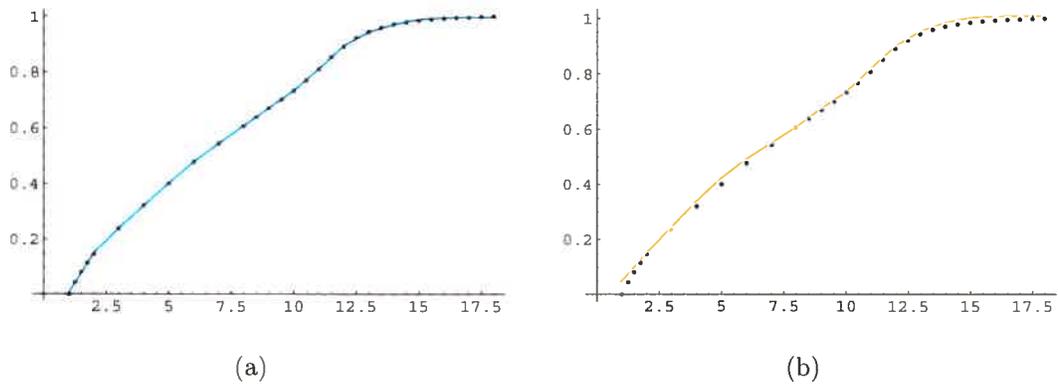


FIG. 3.11. Estimation (couleur) de la courbe moyenne (points) avec les ondelettes intégrées pour les filles (a) Modèle 1 (b) Modèle 2

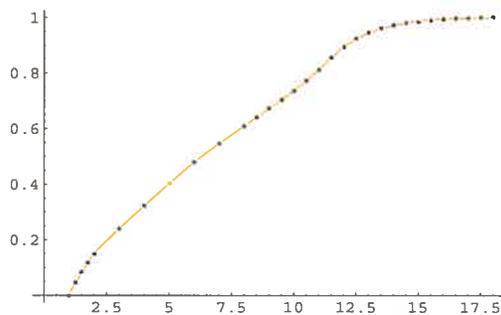


FIG. 3.12. Estimation (couleur) de la courbe moyenne (points) pour les filles avec les ondelettes intégrées en utilisant l'estimateur des moindres carrés pondérés

### 3.2. APPLICATION DES MODÈLES DÉVELOPPÉS À DES FONCTIONS CONTENANT UN CERTAIN DEGRÉ DE VARIATION

Dans cette section, trois fonctions sont considérées : les deux premières sont strictement croissantes, alors que la troisième l'est sur la majorité de son domaine,

mais devient légèrement décroissante vers la fin de ce dernier. Sa création nous permet de vérifier ce qui se passe quand il est faux de croire que les estimations doivent être croissantes. Ces fonctions ainsi que la notation que nous utilisons pour leur appellation sont décrites dans le tableau 3.3.

TAB. 3.3. Description des fonctions utilisées.

Notation	Fonction	Écart type $\sigma$ pour l'erreur
A	$g(x) = x^2, 1 \leq x \leq 18$	7,5
B	$g(x) = x/5, 1 \leq x \leq 18$	0,5
C	$g(x) = \begin{cases} x, & 1 \leq x \leq 16, \\ 24 - (1/2)x, & 16 < x \leq 18. \end{cases}$	0,5

Pour générer des observations provenant de ces fonctions, nous procédons de cette façon :

- (1) nous fixons d'abord les valeurs de  $x_i, i = 1, \dots, 69$ . Dans nos exemples, les données sont équidistantes et comprises entre 1 et 18.
- (2) Nous générons ensuite des erreurs  $\epsilon_i$  normalement distribuées centrées à 0 et d'écart type  $\sigma$  présenté au tableau 3.3.
- (3) Les valeurs des observations deviennent alors

$$y_i = g(x_i) + \epsilon_i, \quad i = 1, \dots, 69.$$

Remarquons que le choix des éléments des bases à utiliser n'a pas été fait exactement de la même façon que pour l'exemple avec les courbes de croissance. En effet, pour accélérer le processus, nous avons décidé de fixer les noeuds ainsi que l'intervalle comprenant les valeurs de  $l$  à l'avance. Ainsi, pour les deux modèles, quatre noeuds intérieurs équidistants donnés par 4,4 ; 7,8 ; 11,2 et 14,6 sont employés. De plus, pour les deux modèles toujours, l'intervalle  $[0, 16]$  a été choisi pour définir les fonctions  $\rho_{r,l}^*(x)$ . Soulignons que, dans le cas des ondelettes intégrées, nous travaillons toujours avec les ondelettes de Haar et les valeurs de  $j$  et de  $r$  sont respectivement 0 et 0,3. Par la suite, les valeurs de  $n_0$  choisies ont été trouvées en maximisant la densité marginale des observations pour chacune des

fonctions en incluant certaines erreurs aléatoires. Bien sûr, puisque ces erreurs varient d'une fois à l'autre, les valeurs de  $n_0$  changent également, mais très peu, puisque la fonction de base demeure la même. Une valeur approximative a donc été retenue, ce qui ne cause aucun problème, puisque nous n'observons pas de différence majeure entre une estimation faite à partir de  $n_0 = 0,02$  ou de  $n_0 = 0,05$ , par exemple. Le tableau 3.4 présente les valeurs de  $n_0$  sélectionnées pour chacune des fonctions.

TAB. 3.4. Résultats associés aux fonctions A, B et C.

Fonction	Base	Modèle	$n_0$	$SSE$ moyen
A	Splines	1	0,002	3 462,95
		2	0,2	8 174,24
	Ondelettes	1	0,025	3 362,55
		2	0,15	4 915,63
B	Splines	1	0,08	16,21
		2	0,95	15,54
	Ondelettes	1	0,35	947,77 <sup>a</sup>
		2	0,9	15,63
C	Splines	1	0,02	25,75
		2	0,35	25,29
	Ondelettes	1	0,1	378,13
		2	0,2	23,37

<sup>a</sup>Utilisation de la méthode itérative

Nous avons répété l'estimation de chacune des fonctions 100 fois. Pour chacune des estimations, la somme des carrés des résidus a été évaluée. Rappelons sa définition.

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

où  $\hat{y}_i$  représente la valeur prédite par le modèle pour l'observation  $y_i$ . La valeur moyenne de cette somme est présentée dans le tableau 3.4. Remarquons qu'il est plutôt difficile de juger de l'efficacité de l'ensemble des modèles en n'observant que ces valeurs, puisqu'il n'y en a pas un qui se distingue des autres, sauf bien

sûr, le premier modèle avec les ondelettes qui a donné un très mauvais ajustement pour les fonctions B et C. Par contre, si nous assignons un rang allant de 1 à 4 à chacun des modèles en fonction de la somme obtenue (le premier étant associé au modèle ayant obtenu la plus petite somme des carrés), nous constatons que c'est le second modèle utilisant les ondelettes qui a, en moyenne, le plus petit rang. Ainsi, nous pouvons dire que ce modèle performe plutôt bien, en gardant cependant en tête le fait que les estimations sont faites avec plus de paramètres pour les ondelettes que pour les splines.

Mentionnons que, pour la troisième fonction, nous ne parvenions jamais à obtenir des estimations croissantes, ce qui est souhaitable, puisque la fonction ne l'était pas totalement. Nous l'avons néanmoins estimée en remplaçant tout coefficient négatif par une fraction de sa valeur, et ce, à titre informatif uniquement. Ceci nous donne une idée du type d'ajustement pouvant être obtenu de cette façon. Éclaircissons l'idée derrière cette technique. S'il existe des dérivées négatives, c'est que certains coefficients le sont également. Ainsi, si nous remplaçons tous ces coefficients par zéro, nous serions certains d'obtenir des estimations non décroissantes. Par contre, s'il y en a plusieurs, ces estimations risquent d'être très éloignées de la vraie fonction. Donc, pour tenter de minimiser les erreurs commises à l'aide de cette méthode, plutôt que de remplacer tous ces coefficients négatifs par zéro, nous les multiplions par une certaine constante, inférieure à l'unité. Bien sûr, plus sa valeur est près de un, moins l'estimation obtenue s'en trouvera affectée. Nous commençons donc par vérifier s'il est suffisant d'opter pour une constante égale à 0,99 afin d'obtenir des estimations dont la dérivée est toujours supérieure ou égale à zéro. Si ce n'est pas le cas, nous multiplions cette constante par elle-même jusqu'à ce que la condition imposée soit respectée. En répétant cette procédure pour chacun des vecteurs générés, nous nous assurons que la condition sera bien satisfaite pour l'ensemble des estimations. L'utilisation de cette méthode itérative est indiquée en rouge dans le tableau 3.4. Soulignons que nous avons également dû l'appliquer au premier modèle utilisant les ondelettes intégrées pour la fonction B. En effet, dans ce cas, il a été impossible d'obtenir des estimations croissantes pour la même raison que précédemment, c'est-à-dire

que plusieurs coefficients étaient négatifs et, par conséquent, aucun vecteur généré ne satisfaisait la condition imposée. Ce problème représente un défaut de la méthode employée. En effet, nous nous attendons normalement à ce qu'une bonne partie des vecteurs de coefficients obtenus fasse en sorte que les dérivées retrouvées soient positives ou nulles. Par contre, rien ne nous garantit que ce sera toujours le cas, même si les fonctions sont croissantes. Pour être bien certain de ne pas rencontrer un tel problème, il aurait fallu imposer les conditions dès le départ, ce qui ne sera cependant pas fait dans ce mémoire pour les raisons énumérées à la fin de la section 2.4.

Finalement, les figures 3.13 à 3.15 illustrent la modélisation des courbes obtenues pour chacune des fonctions. La vraie fonction est en noir, alors que la moyenne des valeurs prédites pour les 100 évaluations est en couleur. Les points représentent également un exemple du bruit généré. Nous avons déjà discuté des défauts du premier modèle avec les ondelettes intégrées pour la fonction B. Ceci se reflète dans l'ajustement obtenu, puisqu'il y a une surestimation flagrante de la fonction. Pour la fonction C, le même phénomène se produit. Il y a donc trop de coefficients négatifs qui sont modifiés et c'est pourquoi la modélisation est mauvaise dans ce cas. De plus, de manière plus globale, les résultats associés au premier modèle employant les splines ainsi qu'au second modèle utilisant les ondelettes intégrées sont plutôt comparables, même si le premier performe parfois un peu mieux. En ce qui concerne le second modèle évalué à l'aide des splines, il performe relativement bien, malgré le fait qu'il surestime légèrement la fonction A.

En conclusion, en combinant les résultats trouvés à l'exemple avec des données réelles et aux petites simulations dont nous venons de discuter, nous pouvons dire que c'est le premier modèle construit à partir des I-splines qui semble être le plus efficace.

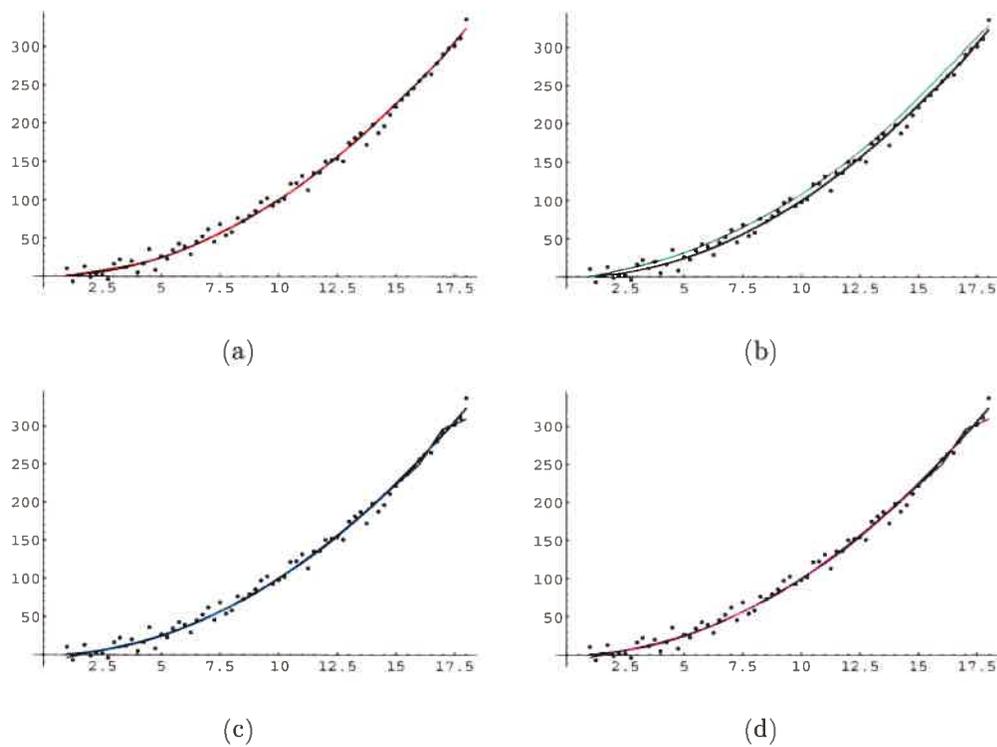


FIG. 3.13. Estimation de la fonction A : fonction réelle (trait plein noir), estimation moyenne (trait plein en couleur) et bruit (points noirs) (a) Modèle 1 - splines (b) Modèle 2 - splines (c) Modèle 1 - ondelettes intégrées (d) Modèle 2 - ondelettes intégrées

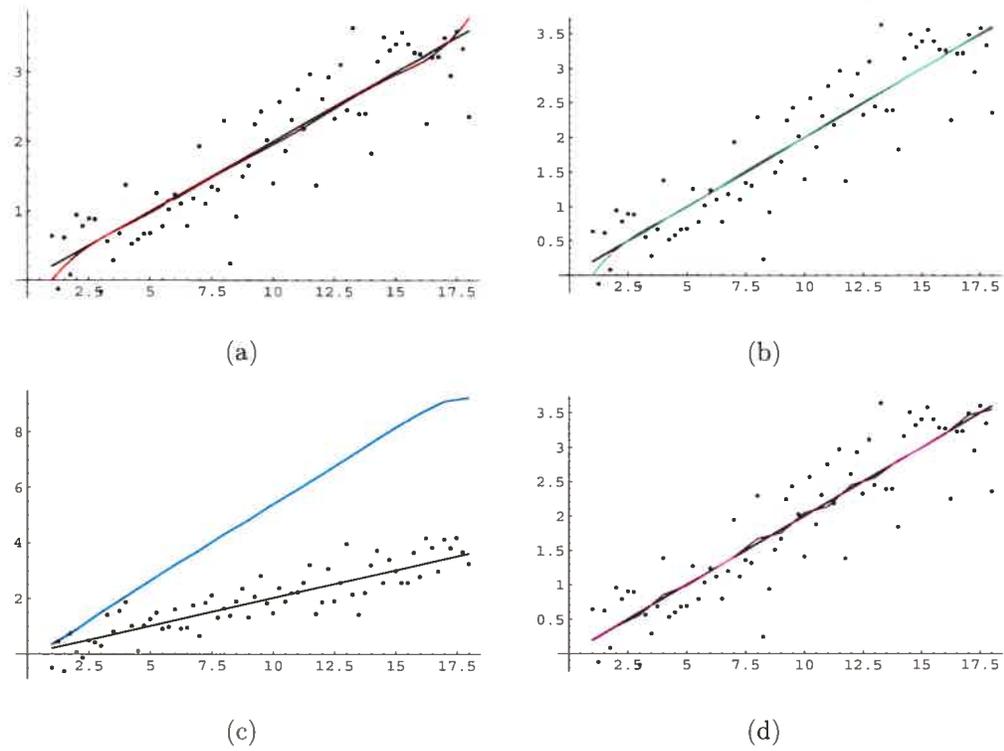


FIG. 3.14. Estimation de la fonction B : fonction réelle (trait plein noir), estimation moyenne (trait plein en couleur) et bruit (points noirs) (a) Modèle 1 - splines (b) Modèle 2 - splines (c) Modèle 1 - ondelettes intégrées (d) Modèle 2 - ondelettes intégrées

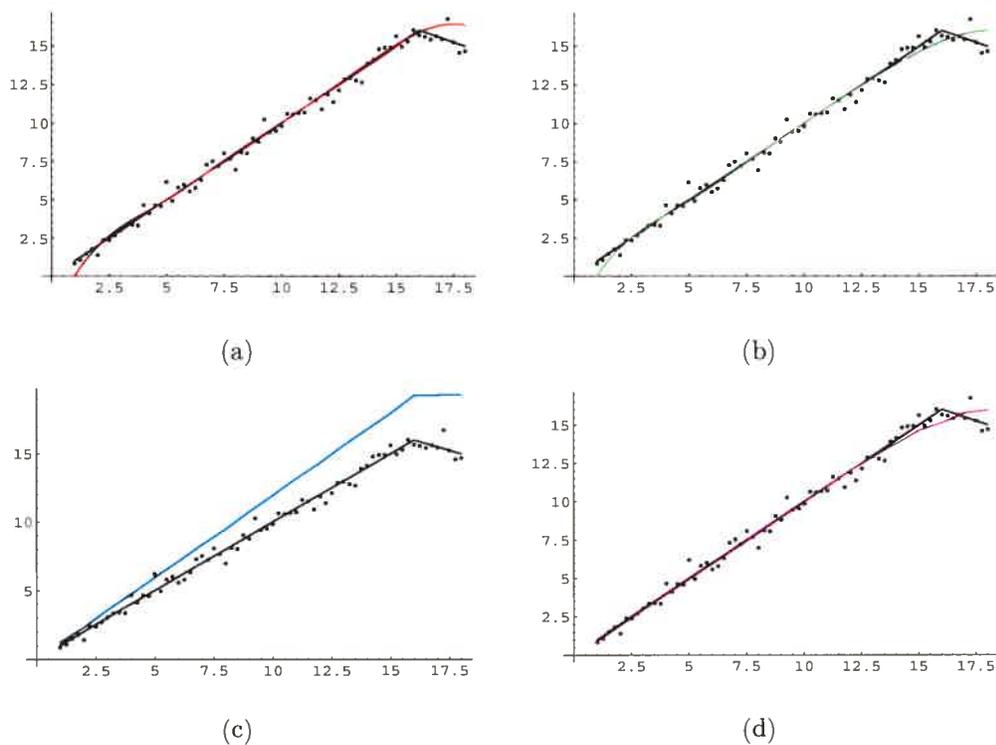


FIG. 3.15. Estimation de la fonction  $C$  : fonction réelle (trait plein noir), estimation moyenne (trait plein en couleur) et bruit (points noirs) (a) Modèle 1 - splines (b) Modèle 2 - splines (c) Modèle 1 - ondelettes intégrées (d) Modèle 2 - ondelettes intégrées

## CONCLUSION

---

Dans ce mémoire, nous avons étudié l'estimation de fonctions monotones non décroissantes dans un contexte de régression bayésienne non paramétrique, en nous préoccupant particulièrement du cas des courbes de croissance humaine qui est un exemple très fréquemment rencontré dans la littérature. Pour ce faire, nous nous sommes inspirés d'un article dans lequel les auteurs proposent l'estimation d'hydrogrammes cumulatifs à partir de splines monotones.

La première étape de nos analyses était de trouver une courbe de référence qui représentait la tendance moyenne de la croissance des enfants provenant d'une certaine population. Pour ce faire, nous devions d'abord tenir compte du fait que nous étions en présence d'un phénomène naturel et que, par conséquent, l'âge auquel certains événements importants se produisaient variait d'un enfant à l'autre. C'est pourquoi nous avons commencé par synchroniser les courbes par rapport aux événements qui nous intéressaient à l'aide de ce que nous avons appelé la synchronisation événementielle avant de faire la moyenne des observations. Ceci consistait en l'étalonnage des caractéristiques saillantes à l'aide de transformations monotones du temps.

Par la suite, l'objectif que nous voulions atteindre était d'estimer correctement la courbe moyenne résultante. Nous avons donc présenté deux modèles bayésiens relativement semblables nous permettant d'obtenir une approximation de cette fonction. La différence entre ces deux modèles venait du fait que le premier n'utilisait que les observations associées à la courbe moyenne, alors que le second employait également celles de chacun des enfants faisant partie de notre échantillon. Ces deux modèles ont été établis à l'aide de deux bases de fonctions différentes ayant cependant toutes deux la propriété de monotonie. La première,

déjà connue, était une base de splines intégrés, alors que nous avons construit la seconde en intégrant une base d'ondelettes biorthogonales développée par Walter et Shen (1999).

Finalement, le comportement des estimateurs obtenus à partir des deux modèles utilisant ces deux bases de fonctions a été étudié à l'aide du jeu de données de Berkeley qui représente un ensemble de courbes de croissance. Par la suite, une courte simulation nous a permis de vérifier l'allure des estimations de certaines fonctions qui étaient croissantes sur au moins une partie de leur domaine et qui comprenaient plus de bruit que les courbes de croissance.

Globalement, les estimateurs ont bien performé, le premier modèle étant sensiblement meilleur que le second. En outre, nous pouvons affirmer que les deux modèles développés avec les splines donnaient, en général, de meilleurs résultats que ceux développés avec les ondelettes intégrées.

En conclusion, le premier modèle utilisant les I-splines se trouve à être non seulement le plus simple des quatre, mais également celui qui fournissait les meilleures estimations dans la plupart des circonstances. C'était également le seul qui se comparait facilement, en termes d'efficacité, aux estimateurs classiques généralement rencontrés avec, comme avantage, le fait qu'il nous assurait l'obtention d'estimations non décroissantes.

Par contre, nous avons également rencontré certains problèmes. Ceux-ci étaient dus au fait que notre technique consistait à générer des vecteurs de coefficients qui ne respectaient pas nécessairement les conditions désirées avant d'imposer les contraintes nous assurant que nos estimations étaient bien croissantes. Rien ne nous garantissait alors qu'il existait des coefficients qui satisfaisaient ces restrictions. Ainsi, pour de futures recherches, il serait intéressant de trouver de nouvelles techniques d'estimation des paramètres qui permettraient l'imposition de contraintes dès le départ. De plus, la construction d'intervalles de confiance pourrait être utile pour être en mesure de savoir si un enfant a une croissance hors de la normale.

# Annexe A

---

## DENSITÉS

Dans cette annexe, nous rappelons les fonctions de densité de toutes les distributions que nous avons utilisées aux deux premiers chapitres.

### A.1. DISTRIBUTIONS NORMALES

#### A.1.1. Distribution normale $\mathcal{N}(\mu, \sigma^2)$

Si  $x \in \mathbb{R}$  a pour densité

$$f(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\},$$

où  $\mu \in \mathbb{R}$  et  $\sigma^2 \in \mathbb{R}$ , alors  $x$  suit une loi normale, notée  $\mathcal{N}(\mu, \sigma^2)$ , de moyenne  $\mu$  et de variance  $\sigma^2$ .

#### A.1.2. Distribution normale multidimensionnelle $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Si  $\boldsymbol{x} = (x_1, \dots, x_p)^\top$ , un vecteur aléatoire défini dans  $\mathbb{R}^p$ , a pour densité

$$f(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{p/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp \left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \right\},$$

où  $\boldsymbol{\mu} \in \mathbb{R}^p$  et  $\boldsymbol{\Sigma}$  est une matrice  $p \times p$  symétrique définie positive, alors  $\boldsymbol{x}$  suit une loi normale multidimensionnelle, notée  $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , d'espérance  $\boldsymbol{\mu}$  et de matrice de variance-covariance  $\boldsymbol{\Sigma}$ .

## A.2. DISTRIBUTION INVERSE GAMMA $\mathcal{IG}(\alpha, \beta)$

Si  $x \in \mathbb{R}^+$  a pour densité

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\exp(-\beta/x)}{x^{\alpha+1}} I_{[0, \infty)}(x),$$

où  $\alpha > 0$  et  $\beta > 0$ , alors  $x$  suit une loi inverse gamma.

## A.3. DISTRIBUTION T DE STUDENT MULTIDIMENSIONNELLE

$$T_p(\nu, \boldsymbol{\theta}, \boldsymbol{\Sigma})$$

Si  $\mathbf{x} = (x_1, \dots, x_p)^\top$ , un vecteur aléatoire défini dans  $\mathbb{R}^p$ , a pour densité

$$f(\mathbf{x}|\nu, \boldsymbol{\theta}, \boldsymbol{\Sigma}) = \frac{\Gamma([\nu + p]/2)}{\Gamma(\nu/2)(\nu\pi)^{\frac{p}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \left[ 1 + \frac{(\mathbf{x} - \boldsymbol{\theta})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\theta})}{\nu} \right]^{-\left(\frac{\nu+p}{2}\right)},$$

où  $\nu > 0$ ,  $\boldsymbol{\theta} \in \mathbb{R}^p$  et  $\boldsymbol{\Sigma}$  est une matrice  $p \times p$  symétrique définie positive, alors  $\mathbf{x}$  suit une loi t de Student multidimensionnelle de moyenne  $\boldsymbol{\theta}$  et de matrice d'échelle  $\boldsymbol{\Sigma}$ . De plus,  $\nu$  degrés de liberté sont associés à cette distribution. Sa matrice de covariance est donnée par  $\frac{\nu}{\nu - 2} \boldsymbol{\Sigma}$  lorsque  $\nu > 2$ .

# Annexe B

---

## PROGRAMMATION

Cette annexe présente la programmation reliée à la modélisation faite à l'aide des splines. Ce programme a servi à retrouver les ajustements pour les garçons avec les données provenant de l'exemple des courbes de croissance. Il a été construit à partir de la version 5.2 du logiciel Mathematica.

```
(***** Chargement des bibliothèques nécessaires *****)
```

```
<< "Statistics`MultinormalDistribution`";  
<< "Statistics`NormalDistribution`";  
<< "Statistics`ContinuousDistributions`";  
<< "LinearAlgebra`MatrixManipulation`";  
<< "Graphics`MultipleListPlot`";  
<< "Statistics`DataManipulation`";  
<< "Statistics`DataSmoothing`";
```

```
(*****)
```

```
(***** Lecture des données *****)
```

```
(*****)
```

```
hgtmb = Import["hgtm.txt", "List"];
```

```
(***** Nombre de garçons dans le fichier de données *****)
```

```
ncasm = 39;
```

```
(***** Entrée des temps observés pour chacun des enfants *****)
age = Join[Range[1, 2, 0.25], Range[3, 8], Range[8.5, 18, 0.5]];
nage = Length[age];
```

```
(***** Restructuration des données sous forme matricielle *****)
hgtm = Table[0, {i, ncasm}, {j, nage}];
For[i = 1, i <= ncasm, i++, {n = 31*(i - 1); For[j = 1, j <= nage,
j++, hgtm[[i,j]] = hgtmb[[n + j]]}]];
```

```
(*****
***** Calcul des vitesses de croissance approximatives *****)
***** pour les temps 2 à 30 *****)
*****
```

```
vitessem = Table[0, {i, ncasm}, {j, nage - 2}];
```

```
(** Fonction permettant de calculer la vitesse à un temps donné **)
```

```
calculvitesse1[vecteur_, pos_, temps_] := (vecteur[[pos + 1]]
- vecteur[[pos - 1]])/(temps[[pos + 1]] - temps[[pos - 1]]);
```

```
(***** Calcul des vitesses à l'aide de la fonction précédente *****)
For[i = 1, i <= ncasm, i++, For[j = 2, j < nage, j++,
vitessem[[i,j - 1]] = calculvitesse1[hgtm[[i]], j, age]]];
```

```
(***** Lissage des vitesses calculées *****)
agev = Take[age, {2, nage - 1}];
```

```
vitessemlisse = Table[0, {i, ncasm}, {j, nage - 2}];
```

```

For[i = 1, i <= ncasm, i++,
vitessemlisse[[i]] = ExponentialSmoothing[vitesseml[[i]], 0.5]];

(***** Synchronisation par rapport à la puberté *****)
(***** Recherche de la vitesse maximale observée à l'adolescence *****)
modesm=Table[0,{i,ncasm}];

tm=Range[10,29];

(***** Fonction permettant de trouver ce maximum *****)
modes[vecteur_,te_]:=Position[vecteur,Max[vecteur[[te]]]][[1]][[1]];

(***** Calcul du maximum à l'aide de la fonction précédente *****)
For[i=1,i<=ncasm,i++,modesm[[i]]=modes[vitessemlisse[[i]],tm]];

(***** Calcul de la médiane de ces maximums *****)
medm=Median[modesm];

(***** Points servant à la synchronisation *****)
t1=agev[[1]]
t2m=agev[[medm]]
t3=agev[[29]]

(** Calcul des pentes pour chacune des transformations linéaires **)
beta[x1_,x2_,y1_,y2_]:=Module[{}
,
beta1=(y2-y1)/(x2-x1);
beta0=y2-beta1*x2;
]

```

```
betam=Table[0,{i,ncasm},{j,4}];
```

```
For[i=1,i<=ncasm,i++,
{beta[t1,agev[[modesm[[i]]]],t1,t2m];
betam[[i,1]]=beta0;
betam[[i,2]]=beta1;
beta[agev[[modesm[[i]]]],t3,t2m,t3];
betam[[i,3]]=beta0;
betam[[i,4]]=beta1;
}];
```

```
(***** Transformations linéaires pour les temps *****)
```

```
agetrans[i_,j_,k_,bet_]:=bet[[i,k]] + bet[[i,k+1]]*agev[[j]];
```

```
agetm=Table[0,{i,ncasm},{j,nage-2}];
```

```
For[i=1,i<=ncasm,i++,
{For[j=1,j<=modesm[[i]],j++,agetm[[i,j]]=agetrans[i,j,1,betam]];
For[j=(modesm[[i]]+1),j<=Length[agev],j++,
agetm[[i,j]]=agetrans[i,j,3,betam]];
}
];
```

```
(*****)
```

```
(** Calcul des courbes moyennes pour les données standardisées **)
```

```
(*****)
```

```
v1m=Table[1,{i,ncasm}];
```

```
v2m=Table[18,{i,ncasm}];
```

(\* Âges synchronisés auxquels nous rajoutons les bornes 1 et 18 \*)

```
agesynchrom=Table[0,{i,ncasm},{j,nage}];
```

```
agesynchrom[[All,1]]=v1m;
```

```
agesynchrom[[All,nage]]=v2m;
```

```
For[i=2,i<=(nage-1),i++,agesynchrom[[All,i]]=agetm[[All,i-1]]];
```

(\*\*\*\*\* Standardisation des fonctions entre 0 et 1 \*\*\*\*\*)

```
hgtmstand=Table[0,{i,ncasm},{j,nage}];
```

```
For[i=1,i<=ncasm,i++,
```

```
{
```

```
min=Min[hgtm[[i]]];
```

```
max=Max[hgtm[[i]]];
```

```
hgtmstand[[i]]=(hgtm[[i]]-min)/(max-min);
```

```
}]
```

```
moyhautmi2=Table[0,{i,ncasm},{j,nage}];
```

(\*\* Fonction faisant l'interpolation linéaire afin de trouver \*\*)

(\*\*\*\*\* la hauteur approximative pour un temps donné \*\*\*\*\*)

```
ptmoyenne[donnees_,tsynchro_,temps_,i_,j_,k_]:=
```

```
donnees[[i,k]]-(donnees[[i,k]]-donnees[[i,k-1]])*
```

```
(tsynchro[[i,k]]-temps[[j]])/(tsynchro[[i,k]]-tsynchro[[i,k-1]]);
```

(\* Calcul des hauteurs approximatives à l'aide de la  
fonction précédente \*)

```
For[i=1,i<=ncasm,i++,
```

```
{For[j=1,j<=nage,j++,If[age[[j]]==agesynchrom[[i,j]],
moyhautmi2[[i,j]]=hgtmstand[[i,j]],For[k=2,k<=nage,k++,
If[agesynchrom[[i,k]]>=age[[j]] && agesynchrom[[i,k-1]]<age[[j]],
moyhautmi2[[i,j]]=ptmoyenne[hgtmstand,agesynchrom,age,i,j,k]]]]];
```

```
(**** Calcul de la hauteur moyenne à chacun des temps observés ****)
hautmoyennem2=Table[0,{i,nage}];
```

```
For[i=1,i<=nage,i++,hautmoyennem2[[i]]=Mean[moyhautmi2[[All,i]]]]];
```

```
(*****
(***** Définition des fonctions splines *****)
(*****
```

```
(***** M-spline *****)
Mspline[i_, k_, t_, x_] := Module[{rep},
If[IntervalMemberQ[Interval[{t[[i]], t[[i + k]]}], x] == True
&& t[[i + k]] - t[[i]] > 0, If[k == 1,
rep = 1/(t[[i + k]] - t[[i]]),
rep = k*((x - t[[i]])*Mspline[i, k - 1, t, x]
+ (t[[i + k]] - x)*Mspline[i + 1, k - 1, t, x])/((k - 1)*
(t[[i + k]] - t[[i]]))], rep = 0];
rep];
```

```
(***** I-spline *****)
Ispline[m_, k_, t_, x_] := Module[{somme},
If[x < t[[m]], rep = 0, If[x >= t[[m + k]], rep = 1,
{somme = 0; For[j = m, j <= m + k - 1, j++, If[x > t[[j]]
&& x <= t[[j + 1]] && t[[j + 1]] - t[[j]] > 0,
{For[i = m, i <= j, i++, somme += (t[[i + k + 1]] - t[[i]])*
(Mspline[i, k + 1, t, x]/(k + 1))}; }]; ]; rep = somme; }]; rep];
```

```

(*****)
(** Fonctions permettant le calcul des éléments nécessaires **)
(***** à l'obtention des estimateurs finaux *****)
(*****)

(***** Fonction calculant la matrice de design *****)
Bw[k_,t_,don_,Kw_]:=Module[{q},
matsplines=Table[0,{i,Length[don]},{j,Kw}];
For[l=1,l<=Length[don],l++,For[q=1,q<=Kw,q++,
matsplines[[l,q]]=Ispline[q,k,t,don[[l]]]];];
matsplines
]

(** Fonction calculant le nombre de fonctions splines utilisées *)
Kw[ordre_,m_]:=ordre+m;

(***** Ordre des fonctions splines *****)
ordre=3;

(***** Paramètre a priori de la variance *****)
alphaw[Kw_]:=Exp[-2*Kw];

(***** Moyenne a priori des coefficients *****)
betaw0[Bw_,hbar_,temps_]:=Module[{pente,beta0,points},
pente1=(hbar[[11]]-Min[hbar])/(temps[[11]]-temps[[1]]);
beta1=Min[hbar]-pente1*temps[[1]];
points1=Table[0,{i,11}];
For[i=1,i<=11,i++,points1[[i]]=pente1*temps[[i]]+beta1];

pente2=(hbar[[22]]-hbar[[11]])/(temps[[22]]-temps[[11]]);

```

```

beta2=hbar[[11]]-pente2*temps[[11]];
points2=Table[0,{i,11}];
For[i=12,i<=22,i++,points2[[i-11]]=pente2*temps[[i]]+beta2];

pente3=(Max[hbar]-hbar[[22]])/(temps[[Length[temps]]]-temps[[22]]);
beta3=hbar[[22]]-pente3*temps[[22]];
points3=Table[0,{i,Length[hbar]-22}];
For[i=23,i<=Length[hbar],i++,
points3[[i-22]]=pente3*temps[[i]]+beta3];

points=Join[points1,points2,points3];

bet=Inverse[Transpose[Bw].Bw].Transpose[Bw].points
];

(** Paramètres nécessaires pour le calcul de la loi marginale **)

alphawstar[vectemps_,Kw_]:=Length[vectemps]+alphaw[Kw];

Sw[hbar_,Bw_,betaL_]:=({hbar-Bw.betaL}).
Transpose[({hbar-Bw.betaL})][[1]][[1]];

Tw[n0_,betawL_,betaw0_,Bw_]:=((n0/(n0+1))*({Bw.(betawL-betaw0)}).
Transpose[({Bw.(betawL-betaw0})})][[1]][[1]];

gammawstar[Sw_,Tw_,Kw_]:=Sw+Tw+alphaw[Kw];

(***** Estimateur des moindres carrés ordinaires *****)
betawL[Bw_,hbar_]:=Inverse[Transpose[Bw].Bw].Transpose[Bw].hbar;

(***** Paramètre utile lors de la méthode de Monte Carlo *****)

```

```
sigmawstar[Bw_,n0_] := ((n0+1)^(-1))*Inverse[Transpose[Bw].Bw];
```

```
(***** Estimateur associé au second modèle *****)
```

```
tetaghat[m_,n0_,n1_,X_,ybar_,beta0_] := Module[{mat},
mat=m*n0*Transpose[X].Inverse[n0*IdentityMatrix[Length[ybar]]
+X.Transpose[X]];
```

```
tetahat=Inverse[mat.X+n1*IdentityMatrix[Length[beta0]]].
```

```
(mat.ybar+n1*beta0)
```

```
];
```

```
(*****)
```

```
(***** Choix des noeuds et de n0 *****)
```

```
(*****)
```

```
(* Fonction qui permet de choisir les noeuds pour un n0 fixé *)
```

```
enlevenoeud[noeuds_,k_,temps_,n0_,vecmoy_,Modele_] := Module[{diff},
```

```
(* Initialisation de vecteurs dont la longueur est égale au nombre
de noeuds intérieurs *)
```

```
diff=Table[0,{i,Length[noeuds]-7}];
```

```
bic=Table[0,{i,Length[noeuds]-7}];
```

```
(* Nous enlevons un noeud intérieur à la fois *)
```

```
For[s=4,s<=(Length[noeuds]-4),s++,
```

```
{noeuds1=Drop[noeuds,{s]}];
```

```
(* Nombre de fonctions splines à utiliser *)
```

```
(* Pour le calculer, nous enlevons les 7 noeuds aux extrémités *)
```

```
Kw1=Kw[k,Length[noeuds1]-7];
```

```

(* Matrice de design *)
Bw1=Bw[k,noeuds1,temps,Kw1];

(* Moyenne a priori des coefficients *)
betaw10=betaw0[Bw1,vecmoy,temps];

If[Modele==1,{
alphaw1=alphaw[Kw1];
alphaw1star=alphawstar[temps,Kw1];
gammaw1=alphaw1;
betaw1L=betawL[Bw1,vecmoy];
Sw1=Sw[vecmoy,Bw1,betaw1L];
Tw1=Tw[n0,betaw1L,betaw10,Bw1];
gammaw1star=gammawstar[Sw1,Tw1,Kw1];

(* Coefficients estimés *)
betaw1star=(1/(n0+1))*betaw1L+(n0/(n0+1))*betaw10;

For[i=1,i<=Length[betaw1star],i++,
If[betaw1star[[i]]<0,betaw1star[[i]]=0]];

(* Valeurs prédites *)
hbarw1star=Bw1.betaw1star;
}];

If[Modele==2,{
betaw1star=tetaghat[ncasm,n0,n0,Bw1,vecmoy,betaw10];
hbarw1star=Bw1.betaw1star;
}];

```

```
(* Calcul des critères *)
diff[[s-3]]=Log[Sum[Abs[vecmoy[[i]]-hbarw1star[[i]]],
{i,Length[vecmoy]}]]+2*(Length[noeuds1]+2)/Length[vecmoy];

bic[[s-3]]=Length[vecmoy]*
  Log[Sum[(vecmoy[[j]]-hbarw1star[[j]])^2,{j,Length[vecmoy]}]/
  Length[vecmoy]] + Length[betaw1star]*Log[Length[vecmoy]];

}];
```

```
(* Enregistrement de la valeur minimale observée
et de la position de celle-ci *)
```

```
no=Range[1,Length[noeuds1]-7];
critheshi=Min[diff];
numheshi=Position[diff,critheshi][[1]][[1]] + 3;
valnoeud=noeuds[[numheshi]];
noeuds2=Drop[noeuds,{numheshi}];
```

```
nobic=Range[1,Length[noeuds1]-7];
critbic=Min[bic];
numbic=Position[bic,critbic][[1]][[1]] + 3;
valnoeudbic=noeuds[[numbic]];
noeuds2bic=Drop[noeuds,{numbic}];
hbarw1star=Bw1.betaw1star;
```

```
]
```

```
(***** Calcul de la marginale pour diverses valeurs de n0 *****)
```

```
vecn0=Sort[Join[Range[0.001,0.99,0.01],Range[0.005,0.02,0.001]]];
```

(\*\*\*\*\* Pour le premier modèle \*\*\*\*\*)

```
marg=Table[0,{i,Length[vecn0]}];
```

```
For[u=1,u<=Length[vecn0],u++,
```

```
{noeuds=Join[Table[1,{i,3}],{1.38,1.63},Range[2.5,7.5,1],
Range[8.25,17.25,0.5],Table[18,{i,4}]];
```

```
criteresmbic=Table[0,{i,Length[noeuds]+2}];
```

```
numerosmbic=Table[0,{i,Length[noeuds]+2}];
```

```
noeudsotembic=Table[0,{i,Length[noeuds]+2}];
```

(\*\*\*\*\* Nous prenons une grande valeur au départ \*\*\*\*\*)

```
criteresmbic[[1]]=10000;
```

```
enlevenoeud[noeuds,ordre,age,vecn0[[u]],hautmoyennem2,1];
```

```
criteresmbic[[2]]=critbic;
```

```
numerosmbic[[1]]=numbic;
```

```
noeudsotembic[[1]]=valnoeudbic;
```

```
n=2;
```

```
While[criteresmbic[[n]]<criteresmbic[[n-1]],{n++;
```

```
enlevenoeud[noeuds2bic,ordre,age,vecn0[[u]],hautmoyennem2,1];
```

```
criteresmbic[[n]]=critbic;
```

```
numerosmbic[[n-1]]=numbic;
```

```
noeudsotembic[[n-1]]=valnoeudbic;};];
```

```
noeudsg=Sort[Join[noeuds2bic,{noeudsotembic[[n-1]]}]];
```

```
Print[noeudsg];
```

```
Kwg=Kw[ordre,Length[noeudsg]-7];
```

```

Bwg=Bw[ordre,noeudsg,age,Kwg];
betawg0=betaw0[Bwg,hautmoyennem2,age];
alphawg=alphaw[Kwg];
alphawgstar=alphawstar[age,Kwg];
gammawg=alphawg;
betawgL=betawL[Bwg,hautmoyennem2];
Swg=Sw[hautmoyennem2,Bwg,betawgL];
Twg=Tw[vecn0[[u]],betawgL,betawg0,Bwg];
gammawgstar=gammawstar[Swg,Twg,Kwg];

marg[[u]]=((vecn0[[u]]/(vecn0[[u]]+1))^(Kwg))*(Gamma[alphawgstar/2]
/(Pi^(Length[age]/2)*Gamma[alphawg/2]))*((gammawg^(alphawg/2))/
(gammawgstar^(alphawgstar/2)));
};];

(***** Pour le second modèle *****)

marg=Table[0,{i,Length[vecn0]}];

For[u=1,u<=Length[vecn0],u++,
{noeuds = Join[Table[1, {i, 3}], {1.38, 1.63}, Range[2.5, 7.5, 1],
  Range[8.25, 17.25, 0.5], Table[18, {i, 4}]];
criteresmbic = Table[0, {i, Length[noeuds] + 2}];
numerosmbic = Table[0, {i, Length[noeuds] + 2}];
noeudsotembic = Table[0, {i, Length[noeuds] + 2}];

(***** Nous prenons une grande valeur au départ *****)
criteresmbic[[1]] = 10000;
enlevenoed[noeuds, ordre, age, vecn0[[u]], hautmoyennem2,2];
criteresmbic[[2]] = critbic;
numerosmbic[[1]] = numbic;

```

```

noeudsotembic[[1]] = valnoeudbic;

p = 2;
While[criteresmbic[[p]] < criteresmbic[[p - 1]],
  {p++;
  enlevenoeud[noeuds2bic, ordre, age, vecn0[[u]],hautmoyennem2,2];
  criteresmbic[[p]] = critbic;
  numerosmbic[[p - 1]] = numbic;
  noeudsotembic[[p - 1]] = valnoeudbic;
  }; ];

noeudsg=Sort[Join[noeuds2bic,{noeudsotembic[[p-1]]}]];
Print[noeudsg];

(* Pour calculer le nombre de paramètres, nous ne conservons que
les noeuds intérieurs *)
Kwg= Kw[ordre, Length[noeudsg] - 7];
Bwg = Bw[ordre, noeudsg, age, Kwg];
betawg0 = betaw0[Bwg, hautmoyennem2, age];
alphawg=alphaw[Kwg];

Sg= Sum[({hgtmstand[[i]]-hautmoyennem2}).Inverse[vecn0[[u]]
*IdentityMatrix[nage]+Bwg.Transpose[Bwg]].
Transpose[({hgtmstand[[i]]-hautmoyennem2}),{i,ncasm}][[1]][[1]];

milg=ncasm*({hautmoyennem2-Bwg.betawg0}.
Inverse[IdentityMatrix[nage] +((ncasm*vecn0[[u]]+vecn0[[u]])/
(vecn0[[u]]*vecn0[[u]])]*Bwg.Transpose[Bwg]].
Transpose[{hautmoyennem2-Bwg.betawg0}][[1]][[1]]);

betag=vecn0[[u]]*Sg + milg + alphawg;

```

```

constante = ((alphawg^(alphawg/2))*Gamma[(nage*ncasm+alphawg)/2])
/(Gamma[alphawg/2]*(Pi^((nage*ncasm)/2)));
tern0=vecn0[[u]]^((nage/2)*(ncasm-1));

```

```

det1=Det[vecn0[[u]]*IdentityMatrix[nage]+Bwg.
Transpose[Bwg]]^((ncasm-1)/2);
det2=Det[IdentityMatrix[nage]+((ncasm*vecn0[[u]]+vecn0[[u]])/
(vecn0[[u]]^2))*Bwg.Transpose[Bwg]]^(1/2);
marg[[u]]=((constante*tern0)/(det1*det2))*((vecn0[[u]]*
Sg+milg+alphawg)^(-(nage*ncasm+alphawg)/2));
}]

```

```

(*****
(***** Estimation de la fonction avec le premier modèle *****)
(*****

```

```

(***** Noeuds et valeurs de n0 trouvés précédemment *****)
n0g=0.121;
noeudsg= {1,1,1,2.5,6.5,10.75,13.25,14.25,18,18,18,18};

```

```

(***** Calcul du nombre de fonctions splines à utiliser *****)
Kwg=Kw[ordre,Length[noeudsg]-7];

```

```

(***** Calcul de la matrice de design *****)
Bwg=Bw[ordre,noeudsg,age,Kwg];

```

```

(***** Calcul de la moyenne a priori *****)
betawg0=betaw0[Bwg,hautmoyennem2,age];

```

```

(***** Calcul des paramètres a priori pour la variance *****)
alphawg=alphaw[Kwg];

```

```
gammawg=alphawg;
```

```
(***** Calcul de l'estimateur des moindres carrés ordinaires *****)
betawgL=betawL[Bwg,hautmoyennem2];
```

```
(* Calcul de l'estimateur a posteriori sans les
contraintes de non décroissance *)
betawgstar=(1/(n0g+1))*betawgL+(n0g/(n0g+1))*betawg0;
```

```
(***** Méthode de Monte Carlo *****)
```

```
(***** Décomposition de la matrice de covariance a posteriori *****)
```

```
(***** Calcul de cette matrice *****)
sigmag=sigmawstar[Bwg,n0g];
consm=gammawgstar/alphawgstar;
```

```
(* Avant de la décomposer, nous nous assurons
que la matrice soit bien symétrique *)
New1g=Table[If[i<j,sigmatg[[i,j]],0],
{i,1,Length[betawgstar]},{j,1,Length[betawgstar]}];
New2g=Transpose[New1g];
new3g=Table[sigmatg[[i,i]},{i,1,Length[betawgstar]}];
sigg=New1g+New2g+DiagonalMatrix[new3g];
```

```
(***** Décomposition en valeurs singulières *****)
{ug,vg,wg}=SingularValueDecomposition[sigg];
cholg=ug.(vg^(1/2));
```

```
decm=(consm^(1/2))*cholg;
```

```
(*** Nous générons maintenant 10 000 vecteurs de coefficients ***)
betamr=Transpose[Table[betawgstar,{i,10000}]];
```

```
yg=Table[RandomArray[StudentTDistribution[alphawgstar],
Length[betawgstar]],{i,10000}];
```

```
matm=decn.Transpose[yg]+betamr;
```

```
(** Nous ne retenons que les vecteurs de coefficients positifs **)
```

```
For[i=1,i<=Kwg,i++,For[j=1,j<=10000,j++,
If[matm[[i,j]]<0,matm[[All,j]]=0]]];
```

```
sommem=0;
```

```
For[i=1,i<=10000,i++,If[matm[[1,i]]!=0,sommem+=1]];
```

```
sommem1=(10000-sommem)
```

```
(* Calcul du nouveau vecteur de coefficients obtenu
avec la méthode de Monte Carlo *)
```

```
betampos=Table[0,{i,Kwg}];
```

```
For[i=1,i<=Kwg,i++,betampos[[i]]=Total[matm[[i]]]];
```

```
(***** Coefficients finaux *****)
```

```
betampos=betampos/sommem1;
```

```
(***** Valeurs prédites finales *****)
```

```
hbarposm=Bwg.betampos;
```

```

(*****)
(***** Estimation de la fonction avec le second modèle *****)
(*****)

(***** Noeuds et valeurs de n0=n1 trouvés précédemment *****)
n0 = 0.014;
n1 = 0.014;
noeudsg = {1,1,1,2.5,6.5,7.5,11.25,13.25,14.75,15.25,18,18,18,18};

(***** Calcul du nombre de fonctions splines à utiliser *****)
Kwg = Kw[ordre, Length[noeudsg] - 7];

(***** Calcul de la matrice de design *****)
Bwg = Bw[ordre, noeudsg, age, Kwg];

(***** Calcul de la moyenne a priori *****)
betawg0 = betaw0[Bwg, hautmoyennem2, age]

(***** Calcul des paramètres a priori pour la variance *****)
alphawg=alphaw[Kwg];
gammawg=alphawg;

(* Calcul de l'estimateur obtenu sans les
contraintes de non décroissance *)
betagstar = tetaghat[ncasm, n0, n1, Bwg, hautmoyennem2, betawg0];

(***** Méthode de Monte Carlo *****)

(** Calcul des valeurs nécessaires pour générer la variance **)
Sg= Sum[({hgtmstand[[i]]-hautmoyennem2}).

```

```

Inverse[n0*IdentityMatrix[nage]+Bwg.Transpose[Bwg]].
Transpose[({hgtmstand[[i]]-hautmoyennem2}),{i,ncasm}][[1]][[1]];
milg=ncasm*({hautmoyennem2-Bwg.betawg0}.
Inverse[IdentityMatrix[nage]+((ncasm*n0+n1)/(n0*n1))*
Bwg.Transpose[Bwg]].Transpose[{hautmoyennem2-Bwg.betawg0}]]
[[1]][[1]]);
betag=(n0*Sg + milg + gammawg)/2;

(***** Nous générons 10 000 valeurs pour les variances *****)
vecsigmag=RandomArray[GammaDistribution[(alphawg+nage)/2,1/betag],
10000];
vecsigmag=1/vecsigmag;

mattetag=Table[0,{i,10000},{j,Kwg}];

matg1=Inverse[ncasm*n0*Transpose[Bwg].
Inverse[n0*IdentityMatrix[nage]+Bwg.Transpose[Bwg]].Bwg
+ n1*IdentityMatrix[Kwg]];

(* Nous nous assurons que la matrice de covariance est
symétrique *)

mat1g=Table[If[i<j,matg1[[i,j]],0],
{i,1,Length[betagstar]},{j,1,Length[betagstar]};
mat2g=Transpose[mat1g];
mat3g=Table[matg1[[i,i]},{i,1,Length[betagstar]};
matg=mat1g+mat2g+DiagonalMatrix[mat3g];

(***** Calcul du vecteur moyen *****)
moyg=matg.(ncasm*n0*(Transpose[Bwg].

```

```

Inverse[n0*IdentityMatrix[nage]+Bwg.Transpose[Bwg]].
hautmoyennem2) + n1*betawg0);

(***** Nous générons 10 000 vecteurs de coefficients *****)

For[i=1,i<=10000,i++,mattetag[[i]]=
Random[MultinormalDistribution[moyg,vecsigmag[[i]]*matg]]]

(***) Nous ne conservons que les vecteurs complètement positifs (***)

For[i=1,i<=10000,i++,For[j=1,j<=Kwg,j++,
If[mattetag[[i,j]]<0,mattetag[[i]]=Table[0,{i,Kwg}]]]]]

sommeg = 0;
For[i = 1, i <= 10000, i++, If[mattetag[[i,1]] == 0, sommeg += 1]];
sommeg1 = 10000 - sommeg

(* Calcul du nouveau vecteur de coefficients obtenu
avec la méthode de Monte Carlo *)

For[i = 1, i <= 10000, i++, If[mattetag[[i,1]] == 0, sommeg += 1]];
sommeg1 = 10000 - sommeg
betagpos = Table[0, {i, Kwg}];
For[i = 1, i <= Kwg, i++,
betagpos[[i]] = Total[mattetag[[All,i]]]];

(***** Coefficients finaux *****)
betagpos1 = betagpos/sommeg1;

(***** Valeurs prédites finales *****)
hbarposg= Bwg.betagpos1;

```

## BIBLIOGRAPHIE

---

- [1] AHLBERG, J. H., NILSON, E. N. ET WALSH, J. L. (1967), Complex Cubic Splines, *Transactions of the American Mathematical Society*, **129**, 391-413.
- [2] ALPERT, B.K. (1992), Wavelets and other bases for fast numerical linear algebra, *Wavelets-A tutorial in theory and applications (C.K. Chui ed.)*, 181-216, Academic Press, Boston.
- [3] ANDERSON, T.W. (2003), *An introduction to Multivariate Statistical Analysis*, Troisième édition, John Wiley & Sons Inc.
- [4] ANGERS, J.-F. ET DELAMPADY, M. (1992), Hierarchical Bayesian Curve Fitting and Smoothing, *The Canadian Journal of Statistics*, **20**, 35-49.
- [5] ANGERS, J.-F. ET DELAMPADY, M. (2001), Bayesian Nonparametric Regression Using Wavelets, *Sankhya, Series B*, **63**, 287-308.
- [6] ANGERS, J.-F., MERLEAU, J. ET PERREAULT, L. (2004), Landmark Registration of Hydrographs and Bayesian Estimation of a Mean Hydrograph, Centre de recherches mathématiques, CRM-3023, mai 2004.
- [7] BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. ET BRUNK, H. D. (1972), *Statistical Inference under Order Restrictions : The Theory and Application of Isotonic Regression*, Wiley, New York.
- [8] BENNACHMOUCH, Z. (1992), Estimation bayésienne d'une fonction avec contraintes, Mémoire de maîtrise, Université de Montréal.
- [9] BURDEN, R. L. ET FAIRES, J. D. (2001), *Numerical Analysis*, 7<sup>e</sup> édition, Australia ; Pacific Grove, CA : Brooks/Cole.
- [10] CHAMBERS, J. M. (1977), *Computational Methods for Data Analysis*, Wiley, New York.

- [11] CHIPMAN, H. A., KOLACZIK, E. D. ET MCCULLOCH, R. E. (1997), Adaptive Bayesian Wavelet Shrinkage, *Journal of The American Statistical Association*, **92**, 1413-1421.
- [12] CHUI, C. K. (1992), *An Introduction to Wavelets*, Boston, Academic Press.
- [13] COHEN, A., DAUBECHIES, I. ET FEAUVEAU, J. C. (1992), Biorthogonal bases of compactly supported wavelets, *Communications in Pure and Applied Mathematics*, **45**, 485-560.
- [14] CONTE, S. D. ET DE BOOR, C. (1980), *Elementary Numerical Analysis : An Algorithmic Approach*, Troisième édition, McGraw-Hill Book Company.
- [15] CURRY, H. B. ET SCHOENBERG, I. J. (1966), On Polya frequency functions, IV : The fundamental spline functions and their limits, *Journal d'analyse mathématique*, **17**, 71-107.
- [16] DAUBECHIES, I. (1992), *Ten lectures on wavelets*, SIAM, Philadelphie.
- [17] DE BOOR, C. (1963), Best Approximation Properties of Spline Functions of Odd Degree, *Journal of Applied Mathematics and Mechanics*, **12**, 747-750.
- [18] DE BOOR, C. (1978), *A practical guide to splines*, Springer-Verlag, New York.
- [19] DONOHO, D. L. ET JOHNSTONE, I. M. (1994), Ideal Spatial Adaptation via Wavelet Shrinkage, *Biometrika*, **81**, 425-455.
- [20] DONOHO, D. L. ET JOHNSTONE, I. M. (1995), Adapting to Unknown Smoothness via Wavelet Shrinkage, *Journal of the American Statistical Association*, **90**, 1200-1224.
- [21] DONOHO, D. L. ET JOHNSTONE, I. M. (1998), Minimax Estimation via Wavelet Shrinkage, *The Annals of Statistics*, **26**, 879-921.
- [22] EUBANK, R. L. (1988), *Spline smoothing and nonparametric regression, Statistics, textbooks and monographs*, New York.
- [23] FALKNER, F. T. (1960), *Child Development : An International Method of Study*, Basel : Karger.
- [24] GASSER, T., MULLER, H.-G., KOHLER, W., MOLINARI, L. ET PRADER, A. (1984), Nonparametric Regression Analysis of Growth Curves, *The Annals of Statistics*, **12**, 210-229.

- [25] GREEN, P. J. (1995), Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination, *Biometrika*, **82**, 711-732.
- [26] GROSSMAN, A. ET MORLET, J. (1984), Decomposition of Hardy functions into square integrable wavelets of constant shape, *Siam Journal on Mathematical Analysis*, **15**, 723-736.
- [27] HAAR, A. (1910), Zur theorie der orthogonalen funktionensysteme, *Mathematische Annalen*, **69**, 331-371.
- [28] HÄRDLE, W., KERKYACHARIAN, G., PICARD, D. ET TSYBAKOV, A. B. (1998), *Wavelets, approximations, and statistical applications*, Springer-Verlag, New York.
- [29] HE, X. ET SHI, P. (1998), Monotone B-Spline Smoothing, *Journal of the American Statistical Association*, **93**, 643-650.
- [30] KNEIP, A. ET GASSER, T. (1988), Convergence and consistency results for self-modeling nonlinear regression, *The Annals of Statistics*, **16**, 82-112.
- [31] KNEIP, A. ET GASSER, T. (1992), Statistical tools to analyze data representing a sample of curves, *The Annals of Statistics*, **20**, 1266-1305.
- [32] LEBLANC, A. (2001), *Utilisation des ondelettes de Haar en estimation bayésienne*, Thèse de doctorat, Université de Montréal.
- [33] LINDLEY, D. V. ET SMITH, A. F. (1972), Bayes Estimates for the Linear Model, *Journal of the Royal Statistical Society, Series B*, **34**, 1-41.
- [34] MALLAT, S. (1989), Multiresolution approximations and wavelet orthonormal bases of  $\mathcal{L}^2(\mathbb{R})$ , *Transactions of the American Mathematical Society*, **315**, 69-88.
- [35] MERLEAU, J., PERREAULT, L., ANGERS, J.-F. ET FAVRE, A.-C. (2007), Bayesian Modelling of Hydrographs, à paraître dans *Water Research Resources*.
- [36] MEYER, Y. (1990), *Ondelettes et opérateurs I : Ondelettes*, Hermann, Paris.
- [37] MEYER, Y. (1992), *Wavelets and operators*, Cambridge University Press, Cambridge.
- [38] MEYER, Y. (1993), *Wavelets : algorithms and applications*, Philadelphia : Society for Industrial and Applied Mathematics.

- [39] MUKERJEE, H. (1988), Monotone Nonparametric Regression, *The Annals of Statistics*, **16**, 741-750.
- [40] MÜLLER, P. ET VIDA KOVIC, B. (1999A), *Bayesian Inference in Wavelet-Based Models*, Springer-Verlag, New York.
- [41] MÜLLER, P. ET VIDA KOVIC, B. (1999B), MCMC Methods in Wavelet Shrinkage : Non-Equally Spaced Regression, Density and Spectral Density Estimation, *Lecture Notes in Statistics*, **141**, 187-202.
- [42] OGDEN, R. T. (1997), *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhäuser, Boston.
- [43] OUELLETTE, N. (2004), *Approximation de la distribution de la distance entre deux courbes empiriques*, Mémoire de maîtrise, Université de Montréal.
- [44] PREECE, M. A. ET BAINES, M. J. (1978), A New Family of Mathematical Models Describing the Human Growth Curves, *Annals of Human Biology*, **5**, 1-24.
- [45] RAMSAY, J. O. (1988), Monotone Regression Splines in Action, *Statistical Science*, **3**, 425-461.
- [46] RAMSAY, J. O., BOCK, R. D. ET GASSER, T. (1995), Comparison of Height Acceleration Curves in the Fels, Zurich, and Berkeley Growth Data, *Annals of Human Biology*, **22**, 413-426.
- [47] RAMSAY, J. O. ET LI, X. (1998), Curve Registration, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **60**, 351-363.
- [48] RAMSAY, J. O. ET SILVERMAN, B. W. (2002), *Applied Functional Data Analysis : Methods and Case Studies*, New-York : Springer-Verlag.
- [49] RAMSAY, J. O. ET SILVERMAN, B. W. (2005), *Functional Data Analysis*, New-York : Springer-Verlag.
- [50] RAO, C. R. (1965), The Theory of Least Squares when the Parameters are Stochastic and its Application to the Analysis of Growth Curves, *Biometrika*, **52**, 447-458.
- [51] RAO, R. M. ET BOPARDIKAR, A. S. (1998), *Wavelet transforms : Introduction to theory and applications*, Addison-Wesley.
- [52] ROBERT, C. P. (2001), *The Bayesian Choice*, 2<sup>e</sup> édition, Springer-Verlag, New York.

- [53] ROCHE, A. F. (1992), *Growth, Maturation and Body Composition : The Fels Longitudinal Study 1929-1991*, Cambridge : Cambridge University Press.
- [54] SAKOE, H. ET CHIBA, S. (1978), Dynamic programming algorithm optimization for spoken word recognition, *IEEE Transactions on Acoustics, Speech & Signal Processing*, **26**, 43-49.
- [55] SCHOENBERG, I. J. (1946), Contributions to the problem of approximation of equidistant data by analytic functions, *Quarterly of Applied Mathematics*, **4**, 45-99 et 112-141.
- [56] SCHOENBERG, I. J. (1964), Spline functions and the problem of graduation, *Proceedings of the American Mathematical Society*, **52**, 947-950.
- [57] SCHUMAKER, L. L. (1981), *Spline functions : basic theory*, Wiley, New York.
- [58] SCHWARZ, G. (1978), Estimating the Dimension of a Model, *The Annals of Statistics*, **6**, 461-464.
- [59] SEARLE, S. (1982), *Matrix Algebra Useful for Statistics*, Wiley, New York.
- [60] SILVERMAN, B. W. (1995), Incorporating parametric effects into functional principal component analysis, *Journal of the Royal Statistical Society, Series B*, **57**, 673-689.
- [61] SMITH, P. (1979), Splines as a Useful and Convenient Statistical Tool, *The American Statistician*, **33**, 57-62.
- [62] TCHAMITCHIAN, P. (1987), Biorthogonalité et théorie des opérateurs, *Revista Mathematica Iberoamericana*, **3**, 163-189.
- [63] TUDDENHAM, R. D. ET SNYDER, M. M. (1954), Physical growth of California boys and girls from birth to eighteen years, *University of California Publications in Child Development*, **1**, 183-364.
- [64] VIDAKOVIC, B. (1999), *Statistical modeling by wavelets*, John Wiley, New York.
- [65] WAHBA, G. (1978), Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression, *Journal of The Royal Statistical Society, Series B*, **40**, 364-372.

- [66] WAHBA, G. ET CRAVEN, P. (1979), Smoothing noisy data with spline functions : Estimating the correct degree of smoothing by the method of generalized cross-validation, *Numerical Mathematics*, **31**, 377-403.
- [67] WAHBA, G. ET KIMELDORF, G. (1971), Some results on Tchebycheffian spline functions, *Journal of Mathematical Analysis and Applications*, **33**, 82-95.
- [68] WAHBA, G. ET WOLD, S. (1975), A Completely Automatic French Curve : Fitting Spline Functions by Cross Validation, *Communications in Statistics, Theory and Methods*, **4**, 1-17.
- [69] WALNUT, D.F. (2002), *An introduction to wavelet analysis*, Birkhäuser, Boston.
- [70] WALTER, G. ET SHEN, X. (1999), Continuous Non-Negative Wavelets and Their Use in Density Estimation, *Communications in Statistics, Theory and Methods*, **28**, 1-17.
- [71] WEGMAN, E. J., ET WRIGHT, I. W. (1983), Splines in Statistics, *Journal of the American Statistical Association*, **78**, 351-365.
- [72] YOUNG, A. S. (1977), A Bayesian Approach to Prediction Using Polynomials, *Biometrika*, **64**, 309-317.
- [73] ZELLNER, A. (1986), Bayesian Estimation and Prediction using Asymmetric Loss Function, *Journal of the American Statistical Association*, **81**, 446-451.

**AUTORISATION DE REPRODUIRE ET DE DIFFUSER  
LE MÉMOIRE DE MAÎTRISE OU LA THÈSE DE DOCTORAT**

*Prière de dactylographier ou d'écrire en caractères d'imprimerie*

Nom du candidat	<u>UBARTAS CINDY</u>	Code permanent	
Grade postulé	<u>M. Sc.</u>	Programme	<u>STATISTIQUE</u>
Option	_____		
Département/Faculté/École	<u>MATHÉMATIQUES ET STATISTIQUE</u>		
Directeur de recherche	<u>JEAN-FRANÇOIS ANGERS</u>	Codirecteur(s)	_____
Titre du mémoire de maîtrise ou de la thèse de doctorat <u>ESTIMATION NON PARAMÉTRIQUE</u>			
<u>BAYÉSIENNE DE COURBES DE CROISSANCE</u>			

1. Par la présente, j'accorde à l'Université de Montréal l'autorisation de mettre ce mémoire ou cette thèse à la disposition des lecteurs dans les bibliothèques de l'Université ou dans d'autres bibliothèques, soit sous sa forme actuelle, soit sous forme d'une reproduction que l'Université est, par la présente, autorisée à faire.

2. J'autorise l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

Sauf entente contraire, je conserve, avec les coauteurs le cas échéant, la liberté reconnue au titulaire du droit d'auteur de diffuser, éditer et utiliser commercialement ou non ce travail, dont je possède un exemplaire original.

L'Université ne sera aucunement responsable d'une utilisation commerciale, industrielle ou autre du mémoire ou de la thèse par un tiers, y compris les professeurs.

3. Je déclare avoir accordé à la Bibliothèque nationale du Canada une licence non exclusive de reproduire ma thèse ou mon mémoire. L'Université prend acte de mon engagement en vertu de cette licence selon lequel l'Université fera remise du mémoire ou de la thèse à cet organisme. Aux fins de cette licence, l'Université accepte de remettre au Directeur général de la Bibliothèque nationale du Canada ou à ses agents une copie du mémoire ou de la thèse.

\_\_\_\_\_  
*Signature de l'auteur*

29/05/07  
*Date*

4. J'accepte que l'Université divulgue mon nom, mon adresse personnelle et mon numéro de téléphone aux organismes qui en feraient officiellement la demande dans le seul but de me faire participer à des concours de bourses ou à des prix reliés à mon mémoire ou à ma thèse.

\_\_\_\_\_  
*Signature de l'auteur*

29/05/07  
*Date*

La présente autorisation entre en vigueur à la date ci-dessus à moins que le Comité exécutif de la Faculté des études supérieures n'ait déclaré le mémoire ou la thèse confidentiel. L'autorisation prend alors effet à l'expiration de la période de confidentialité fixée par le Comité exécutif.

National Library  
of CanadaBibliothèque nationale  
du Canada

Canada

NON-EXCLUSIVE LICENSE  
TO REPRODUCE THESESLICENCE NON EXCLUSIVE  
DE REPRODUIRE DES THÈSES

Surname / Nom de famille UBARTAS	Given Names / Prénoms CINDY
Date of Birth / Date de naissance [REDACTED]	Canadian Citizen / Citoyen canadien Yes / Oui <input checked="" type="checkbox"/> No / Non <input type="checkbox"/>
Telephone / Téléphone [REDACTED]	E-mail / Courriel [REDACTED]
[REDACTED]	If Current Address is temporary, please provide a secondary permanent address Si l'adresse actuelle est temporaire, veuillez fournir une adresse secondaire permanente
Full Name of University / Nom complet de l'université UNIVERSITÉ DE MONTRÉAL	
Faculty, Department, School / Faculté, département, école ARTS ET SCIENCES / MATHÉMATIQUES ET STATISTIQUE	
Degree for which thesis was presented / Grade pour lequel cette thèse a été présentée M.Sc.	Date Degree Awarded / Date d'obtention du grade
Supervisor / Directeur de thèse JEAN-FRANÇOIS ANGERS	Language of Text / Langue du texte FRANÇAIS
Thesis Title / Titre de la thèse ESTIMATION NON PARAMÉTRIQUE BAYÉSIENNE DE COURBES DE CROISSANCE	

<p>In consideration of the National Librarian making copies of my thesis available to interested persons, I,</p> <p>hereby grant a non-exclusive, royalty free license to Her Majesty the Queen in right of Canada as represented by the National Librarian:</p> <p>(a) to reproduce, loan, distribute or sell copies of my thesis (the title of which is set forth above) anywhere in the world, for commercial or non-commercial purposes, in microform, paper or electronic formats, including the Internet;</p> <p>(b) to authorize, sub-license, sub-contract or procure any of the acts mentioned in paragraph (a).</p> <p>I undertake to submit a copy of my thesis, through my university, to the National Librarian. Any abstract submitted with the thesis will be considered to form part of the thesis.</p> <p>I represent and promise that my thesis is my original work, does not infringe any rights of others, and that I have the right to make the grant conferred by this non-exclusive license.</p> <p>I retain ownership of the copyright in my thesis, and may deal with the copyright in my thesis, in any way consistent with rights granted by me to Her Majesty in this non-exclusive license.</p> <p>I further promise to inform any person to whom I may hereafter assign or license my copyright in my thesis of the rights granted by me to Her Majesty in this non-exclusive license.</p> <p>I retain moral rights in my thesis in accordance with Section 14 of the <u>Copyright Act</u>.</p>	<p>Considérant le fait que l'Administrateur général de la Bibliothèque nationale du Canada désire mettre des exemplaires de ma thèse à la disposition des personnes intéressées, je,</p> <p><u>CINDY UBARTAS</u></p> <p>accorde par les présentes à Sa Majesté la Reine du Canada, représentée par l'Administrateur général de la Bibliothèque nationale, une licence non exclusive, sans frais d'auteurs, pour :</p> <p>(a) reproduire, prêter, distribuer ou vendre des copies de ma thèse (dont le titre est mentionné ci-haut) à travers le monde entier, à des fins commerciales ou non commerciales, sur les supports suivants : microforme, papier ou électronique, y compris sur l'Internet;</p> <p>(b) autoriser, accorder une sous-licence ou une sous-traitance, ou engager toute mesure mentionnée à l'alinéa (a).</p> <p>Je m'engage à ce qu'une copie de ma thèse soit remise à l'Administrateur général de la Bibliothèque nationale par mon université. Tout résumé analytique soumis avec la thèse sera considéré comme faisant partie de la thèse.</p> <p>Je déclare sur mon honneur que ma thèse est mon œuvre originale, qu'elle n'empiète pas sur les droits de quiconque et que j'agis de plein droit en accordant cette licence non exclusive.</p> <p>Je conserve la propriété du droit d'auteur qui protège ma thèse et peux disposer de ce droit de toute manière compatible avec les droits accordés à Sa Majesté par les présentes.</p> <p>Je promets également d'informer toute personne à qui je pourrais ultérieurement céder mon droit d'auteur sur ma thèse ou à qui je pourrais accorder une licence, des droits accordés à Sa Majesté par les présentes.</p> <p>Je conserve des droits moraux sur ma thèse conformément à l'article 14 de la <u>Loi sur le droit d'auteur</u>.</p>
Signature [REDACTED]	Date 29/05/07