

Université de Montréal

Utilisation des ondelettes de Haar en estimation  
bayésienne

par

Alexandre Leblanc

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

Thèse présentée à la Faculté des études supérieures  
en vue de l'obtention du grade de  
Philosophiæ Doctor (Ph.D.)  
en statistique

décembre 2001

© Alexandre Leblanc, MMI



QA

3

U54

7002

1,013

U

U

**Université de Montréal**

Faculté des études supérieures

Cette thèse intitulée

**Utilisation des ondelettes de Haar en estimation  
bayésienne**

présentée par

**Alexandre Leblanc**

a été évaluée par un jury composé des personnes suivantes :

*Roch Roy*

---

(président-rapporteur)

*Jean-François Angers*

---

(directeur de recherche)

*François Perron*

---

(membre du jury)

*Bertrand Clarke*

---

(examineur externe)

*Jean-Marc Lina*

---

(représentant du doyen de la FES)

Thèse acceptée le :

*3 juillet 2002*

---

## Résumé

---

Depuis maintenant quelques années, les statisticiens ont pu découvrir l'extraordinaire potentiel des ondelettes dans le domaine de la statistique, plus particulièrement pour traiter des problèmes d'estimation de densité, de lissage et de régression non paramétriques. Cependant, elles peuvent aussi fournir des approximations simples de densités tout comme le font les bases orthonormales classiques (Hermite, Fourier, Laplace, etc.). Ce type d'approximation est très utile en intégration numérique, et par le fait même, peut jouer un rôle important en calcul bayésien.

Dans le cadre de cette thèse, nous étudions, dans un premier temps, comment la base d'ondelettes de Haar peut être employée efficacement afin de construire des approximations des moments *a posteriori* d'un paramètre de position, noté  $\theta$ , et de la densité marginale des observations. Ces résultats sont obtenus dans le contexte de l'estimation bayésienne de  $\theta$ . Les techniques développées sont aussi utilisées afin de traiter du problème bayésien empirique de l'estimation de la densité *a priori* de  $\theta$  lorsque celle-ci est non spécifiée. Les méthodologies résultantes suggèrent toute une série de nouvelles applications des bases d'ondelettes à divers problèmes bayésiens typiques.

**Mots clés :** paramètre de position, moments *a posteriori*, calcul bayésien, densités *a priori* non informatives, méthodes bayésiennes empiriques non paramétriques, déconvolution.



## Abstract

---

A few years ago, statisticians have discovered the extraordinary potential of wavelets as a tool for statistical inference, especially in the fields of density estimation, nonparametric regression and smoothing. However, wavelets certainly offer many other possibilities, among which is the construction of simple approximations to known densities, just as the more classical orthonormal bases of functions (Hermite, Fourier, Laplace, etc.) do. This kind of approximation can be used to perform numerical integration and therefore, Bayesian calculation.

The main goal of this thesis is to study how the Haar basis of wavelets can be used efficiently to construct approximations to the posterior moments of a location parameter, denoted  $\theta$ , and to the marginal density of the observations. The results are obtained using the Bayesian paradigm. The proposed technique is also used to get a solution to the empirical Bayes problem of nonparametrically estimating the prior of  $\theta$  when it is unspecified. The resulting methodologies suggest a wide range of new interesting applications of wavelet bases to typical Bayesian problems.

**Kew words :** location parameter, posterior moments, Bayesian calculation, non-informative priors, nonparametric empirical Bayes methods, deconvolution.

## Remerciements

---

D'abord, je remercie sincèrement tous mes parents et amis pour leurs nombreux encouragements, leur support moral et leur aide.

Je remercie également les organismes subventionnaires CRSNG et FCAR pour leur support financier, sans lequel mes études de doctorat auraient été beaucoup plus difficiles.

Je remercie aussi tous les membres du jury chargé de l'évaluation de cette thèse, un travail certainement long et exigeant, et plus spécifiquement l'examinateur externe, Bertrand Clarke, pour l'intérêt qu'il a porté à mes travaux et pour ses précieux conseils.

Ensuite, je tiens à remercier mon directeur de recherche, Jean-François Angers, pour l'énorme investissement de temps et d'efforts qu'il a su faire. Côtéyer quelqu'un d'aussi ouvert et généreux intellectuellement m'a non seulement beaucoup aidé dans la production de cette thèse, mais a aussi contribué en grande partie à la réussite de mes études graduées en général.

Finalement et tout spécialement, merci à

- mes amis Alain et Mathieu, pour m'avoir donné l'exemple en me précédant dans le grand monde des docteurs;
- ma grand-mère Jeannette, pour toutes ses prières et son support spirituel soutenu;
- mon père, pour son aide financière et philosophique;
- ma mère, qui m'a plongé dans l'univers de la statistique avant même ma naissance;
- mon frère, qui combat le crime sans relâche;
- ma belle Julie, pour son amour, sa patience, ses encouragements parfois enflammés et toutes ces soirées passées sans moi.

# Table des matières

---

Résumé .....	iii
Abstract .....	iv
Remerciements .....	v
Table des figures .....	ix
Liste des tableaux .....	xii
Introduction .....	1
<b>Chapitre 1. Les ondelettes</b> .....	<b>4</b>
1.1. L'analyse multirésolution .....	5
1.2. L'équation de dilatation .....	10
1.3. Les ondelettes de Haar .....	18
1.4. Les ondelettes de Daubechies .....	23
1.5. Décomposition d'une densité à l'aide des bases d'ondelettes .....	38
1.6. Les ondelettes en statistique .....	43
<b>Chapitre 2. Estimation bayésienne d'un paramètre de position à l'aide de la base de Haar : cas d'une seule observation</b> .....	<b>46</b>
2.1. Estimation de $\theta$ .....	46
2.2. Approximation de $\mathbb{I}_l(x)$ .....	49

2.3.	Approximation de $\widehat{\theta}_B(x)$ et de $\rho(\pi, \widehat{\theta}_B x)$ .....	62
2.4.	Alternatives à $\widetilde{\theta}_B(x)$ et $\widetilde{\rho}(\pi, \widehat{\theta}_B x)$ .....	66
2.5.	Alternatives à $\widetilde{\theta}_B(x)$ et $\widetilde{\rho}(\pi, \widehat{\theta}_B x)$ : formes simples .....	76
2.6.	Considérations pratiques .....	83
2.7.	Quelques exemples .....	98
2.7.1.	Le modèle normale/double exponentielle .....	98
2.7.2.	Le modèle normale/Cauchy .....	112
2.7.3.	Discussion .....	124
<b>Chapitre 3.</b>	<b>Généralisations à plusieurs observations .....</b>	<b>126</b>
3.1.	Estimation de $\theta$ : cas de plusieurs observations .....	127
3.2.	Approximation de $\mathbb{I}_l(\mathbf{x})$ .....	130
3.3.	Approximation de $\widehat{\theta}_B(\mathbf{x})$ et de $\rho(\pi, \widehat{\theta}_B \mathbf{x})$ .....	136
3.4.	Cas de la loi <i>a priori</i> uniforme sur $\mathbb{R}$ .....	140
<b>Chapitre 4.</b>	<b>Estimation de la densité <i>a priori</i> d'un paramètre de position .....</b>	<b>148</b>
4.1.	Introduction au problème .....	149
4.2.	Expression alternative de la densité marginale .....	151
4.3.	Estimation non paramétrique de $\pi$ .....	157
4.4.	Propriétés de $\widehat{\mathbf{b}}$ .....	164
4.5.	Propriétés de $\widehat{\pi}_J$ .....	171
4.6.	Considérations pratiques .....	182
4.7.	Exemple simulé .....	191

4.7.1. Cas d'un mélange de deux normales.....	191
4.7.2. Discussion.....	208
<b>Conclusion</b> .....	210
<b>Annexe A</b> .....	215
<b>Annexe B</b> .....	228
<b>Annexe C</b> .....	234
<b>Bibliographie</b> .....	240

## Table des figures

---

1.1	Fonction d'échelle et mère des ondelettes de Daubechies d'ordre 2 (a) ${}_2\phi$ (b) ${}_2\psi$ .....	33
1.2	Fonction d'échelle et mère des ondelettes de Daubechies d'ordre 3 (a) ${}_3\phi$ (b) ${}_3\psi$ .....	35
1.3	Fonctions d'échelle de Daubechies d'ordre 5 et 7 (a) ${}_5\phi$ (b) ${}_7\phi$ .....	36
2.1	$m(x)$ (ligne pleine) et $S_0^T(x)$ (tirets courts) au niveau $J = 1$ pour $C = 6$ et $x \in [-8, 8]$ .....	102
2.2	Erreur relative de $S_0^T(x)$ pour $x \in [0, 10]$ , pour plusieurs valeurs de $C$ et aux niveaux (a) $J = 1$ , (b) $J = 2$ .....	103
2.3	$\hat{\theta}_B(x)$ (ligne pleine), $\tilde{\theta}_1(x)$ (tirets longs) et $\tilde{\theta}_2(x)$ (tirets courts) au niveau $J = 1$ pour (a) $C = 6$ et $x \in [-6, 6]$ , (b) $C = 8$ et $x \in [0, 8]$ ....	104
2.4	$\hat{\theta}_B(x)$ (ligne pleine), $\tilde{\theta}_1(x)$ (tirets longs) et $\tilde{\theta}_2(x)$ (tirets courts) au niveau $J = 2$ pour (a) $C = 6$ et $x \in [-6, 6]$ , (b) $C = 8$ et $x \in [0, 8]$ ....	105
2.5	$\hat{\theta}_B(x)$ (ligne pleine), $\tilde{\theta}_1(x)$ (tirets longs) et $\tilde{\theta}_2(x)$ (tirets courts) au niveau $J = 2$ pour $C = 4$ et $x \in [-6, 6]$ .....	106
2.6	Erreur relative de $\tilde{\theta}_2(x)$ pour $x \in [0, 10]$ , pour plusieurs valeurs de $C$ et aux niveaux (a) $J = 1$ , (b) $J = 2$ .....	107
2.7	$\rho(\pi, \hat{\theta}_B x)$ (ligne pleine), $\tilde{\rho}_1(\pi, \hat{\theta}_B x)$ (tirets longs) et $\tilde{\rho}_2(\pi, \hat{\theta}_B x)$ (tirets courts) au niveau $J = 1$ pour (a) $C = 6$ et $x \in [-6, 6]$ , (b) $C = 8$ et $x \in [0, 8]$ .....	108
2.8	$\rho(\pi, \hat{\theta}_B x)$ (ligne pleine), $\tilde{\rho}_1(\pi, \hat{\theta}_B x)$ (tirets longs) et $\tilde{\rho}_2(\pi, \hat{\theta}_B x)$ (tirets courts) au niveau $J = 2$ pour (a) $C = 6$ et $x \in [-6, 6]$ , (b) $C = 8$ et $x \in [0, 8]$ .....	109

2.9	$\rho(\pi, \hat{\theta}_B x)$ (ligne pleine), $\tilde{\rho}_1(\pi, \hat{\theta}_B x)$ (tirets longs) et $\tilde{\rho}_2(\pi, \hat{\theta}_B x)$ (tirets courts) au niveau $J = 3$ pour $C = 8$ et $x \in [0, 8]$ .....	110
2.10	Erreur relative de $\tilde{\rho}_1(\pi, \hat{\theta}_B x)$ pour $x \in [0, 10]$ , pour plusieurs valeurs de $C$ et aux niveaux (a) $J = 1$ , (b) $J = 2$ , (c) $J = 3$ .....	111
2.11	$m_{gm}(x)$ (ligne pleine) et $S_0^T(x)$ (tirets) au niveau $J = 1$ , pour $C = 4$ et $x \in [-10, 10]$ .....	116
2.12	Erreur relative approximative de $S_0^T(x)$ pour $x \in [0, 15]$ , pour plusieurs valeurs de $C$ et aux niveaux (a) $J = 1$ , (b) $J = 2$ .....	117
2.13	$\hat{\theta}_{gm}(x)$ (ligne pleine), $\tilde{\theta}_1(x)$ (tirets longs) et $\tilde{\theta}_2(x)$ (tirets courts) au niveau $J = 1$ pour $C = 4$ et $x \in [-10, 10]$ .....	118
2.14	Erreur relative approximative de $\tilde{\theta}_2(x)$ pour plusieurs valeurs de $C$ , pour $x \in [0, 15]$ et aux niveaux (a) $J = 1$ , (b) $J = 2$ .....	119
2.15	$\rho_{gm}(x)$ (ligne pleine), $\tilde{\rho}_1(\pi, \hat{\theta}_B x)$ (tirets longs) et $\tilde{\rho}_2(\pi, \hat{\theta}_B x)$ (tirets courts) au niveau $J = 1$ , pour $x \in [-10, 10]$ , et pour (a) $C = 3$ , (b) $C = 4$ .....	121
2.16	$\rho_{gm}(x)$ (ligne pleine), $\tilde{\rho}_1(\pi, \hat{\theta}_B x)$ (tirets longs) et $\tilde{\rho}_2(\pi, \hat{\theta}_B x)$ (tirets courts) pour $C = 4$ , pour $x \in [0, 15]$ , et aux niveaux (a) $J = 1$ , (b) $J = 2$ .....	122
2.17	Erreur relative de $\tilde{\rho}_1(\pi, \hat{\theta}_B x)$ pour $x \in [-15, 15]$ , pour plusieurs valeurs de $C$ et aux niveaux (a) $J = 1$ , (b) $J = 2$ , (c) $J = 3$ .....	123
4.1	Densité <i>a priori</i> réelle $\pi_1$ et densité reconstruite à partir de $m_1$ (a) au niveau $J = 0$ (b) au niveau $J = 1$ (c) au niveau $J = 2$ .....	195
4.2	Densité <i>a priori</i> réelle $\pi_2$ et densité reconstruite à partir de $m_2$ (a) au niveau $J = 0$ (b) au niveau $J = 1$ (c) au niveau $J = 2$ .....	196
4.3	Densité <i>a priori</i> réelle $\pi_3$ et densité reconstruite à partir de $m_3$ (a) au niveau $J = 0$ (b) au niveau $J = 1$ (c) au niveau $J = 2$ .....	197
4.4	Densité marginale réelle $m_j$ et son estimation par histogramme $\hat{m}_{1(j)}$ obtenue à partir du $j^{\text{ième}}$ échantillon simulé (a) $m_1$ et $\hat{m}_{1(1)}$ (b) $m_2$ et $\hat{m}_{1(2)}$ (c) $m_3$ et $\hat{m}_{1(3)}$ .....	198

- 4.5 Densité *a priori* réelle  $\pi_1$  et densité estimée à partir du premier échantillon simulé  $\hat{\pi}_{J(1)}$  (a) au niveau  $J = 0$  (b) au niveau  $J = 1$  (c) au niveau  $J = 2$ . . . . . 200
- 4.6 Densité *a priori* réelle  $\pi_2$  et densité estimée à partir du deuxième échantillon simulé  $\hat{\pi}_{J(2)}$  (a) au niveau  $J = 0$  (b) au niveau  $J = 1$  (c) au niveau  $J = 2$ . . . . . 202
- 4.7 Densité *a priori* réelle  $\pi_3$  et densité estimée à partir du troisième échantillon simulé  $\hat{\pi}_{J(3)}$  (a) au niveau  $J = 0$  (b) au niveau  $J = 1$  (c) au niveau  $J = 2$ . . . . . 203
- 4.8 Densité *a priori* réelle  $\pi_2$  et densité estimée à partir du deuxième échantillon simulé  $\hat{\pi}_{0(2)}$  pour plusieurs valeurs de  $n$  (a)  $n = 8$  (b)  $n = 9$  (c)  $n = 10$ . . . . . 205
- 4.9 Densité *a priori* réelle  $\pi_2$  et densité estimée à partir du deuxième échantillon simulé  $\hat{\pi}_{0(2)}$  pour plusieurs valeurs de  $n$  (a)  $n = 11$  (b)  $n = 12$  (c)  $n = 13$ . . . . . 206



## Liste des tableaux

---

4.1	Composition des modèles utilisés et choix de $N$ associé à chaque modèle.....	162
4.2	Coefficients de détermination $R^2$ et estimations de $\delta$ obtenus pour chaque modèle.....	162
4.3	Coefficients de détermination $R^2$ et estimations de $\delta$ obtenus pour chaque modèle.....	193
4.4	Quelques statistiques sur les trois échantillons simulés.....	194
4.5	Indices de conditionnement $\text{Cond}(\mathbf{B}_n)$ pour $J = 0, 1, 2$ .....	207

# Introduction

---

Depuis déjà quelques années, les ondelettes ont pris une place de choix parmi les différents outils utilisés par les statisticiens. Bien qu'au départ elles n'apparaissent que dans des travaux portant sur le lissage, la régression et l'estimation de densité non paramétriques, les ondelettes sont maintenant employées à plusieurs autres fins (une brève revue de travaux de statistique utilisant les ondelettes est présentée à la section 1.6). Le but premier de cette thèse est d'explorer une autre avenue où les ondelettes peuvent prendre un rôle intéressant. En effet, le calcul bayésien est un domaine où les ondelettes n'ont pas encore fait leur marque, mais où elles offrent certainement plusieurs possibilités.

Dans le but d'illustrer certaines de ces possibilités, nous nous concentrons sur les modèles bayésiens où la vraisemblance appartient à une famille de paramètres de position et d'échelle quelconque. À la base, notre projet consiste à établir une méthode d'approximation numérique des moments *a posteriori* du paramètre de position, noté  $\theta$ . Cette approximation numérique est obtenue à l'aide de la base de Haar, qui est une base d'ondelettes due à Alfred Haar (1910), mais seulement reconnue comme telle depuis les travaux de Daubechies (1988), Mallat (1989a) et Meyer (1990). De là, nous portons notre attention plus particulièrement sur l'approximation de la densité marginale des observations et, en considérant la fonction de perte quadratique, de l'estimateur de Bayes de  $\theta$  et de la perte espérée *a posteriori* de l'estimateur bayésien.

L'étude des estimateurs bayésiens de paramètres de position obtenus à partir de la fonction de perte quadratique a déjà été faite pour plusieurs modèles particuliers. Nous pensons ici entre autres aux célèbres travaux de Stein (1981), ainsi qu'à ceux de Spiegelhalter (1985), Fan et Berger (1990) et Angers (1996a, 1996b, 1997). Ces travaux ont été généralisés aux familles exponentielles par Robert (2001, section 4.2). De plus, Angers et Berger (1991) et Angers (1992) se sont aussi attaqués à ce problème, mais dans le cas de modèles hiérarchiques. Tous ces travaux permettent d'obtenir, dans divers contextes, les formes explicites de l'estimateur de Bayes d'un paramètre de position ainsi que d'autres quantités d'intérêt souvent reliées aux moments *a posteriori* de ce paramètre de position.

Il y a par ailleurs plusieurs autres contextes où les moments *a posteriori* d'un paramètre de position ne peuvent être obtenus de façon explicite. C'est habituellement le cas, entre autres, lorsque les densités du modèle bayésien ne sont pas conjuguées, ou lorsque la densité *a priori* du paramètre de position est modélisée de façon hiérarchique. Dans de telles conditions, il est nécessaire d'approximer les moments *a posteriori* et les quantités d'intérêt qui en dépendent à l'aide d'une méthode de calcul bayésien, comme par exemple :

- l'approximation gaussienne de Naylor et Smith (1982),
- les simulations de Monte-Carlo avec fonction d'importance,
- l'approximation laplacienne de Tierney et Kadane (1986),
- l'échantillonneur de Gibbs (voir Gelfand et Smith, 1990),
- les simulations de Monte-Carlo avec chaînes de Markov,
- l'approximation à l'aide de l'algorithme de transformation de Fourier rapide de Leblanc et Angers (1995).

Tanner (1993) présente en détail les diverses méthodes de simulations de Monte-Carlo, dont le populaire algorithme de Metropolis-Hastings.

Comme nous l'avons mentionné précédemment, nous proposons plutôt d'employer la théorie des ondelettes pour obtenir une forme d'approximation intéressante des moments *a posteriori*. Ainsi, lorsque les densités impliquées dans le modèle bayésien satisfont certaines conditions de régularité, les moments *a posteriori* pourront être approximés lorsqu'ils ne possèdent pas de forme explicite, ou de forme implicite aisément calculable. La technique d'approximation est développée dans le cadre du problème de l'estimation bayésienne du paramètre de position. Elle peut cependant être employée à d'autres fins. Nous illustrons l'une de ces nombreuses possibilités en l'appliquant à un problème bayésien empirique d'estimation de la densité *a priori*.

Pour résumer le tout, nous commencerons, au chapitre 1, par présenter au lecteur une introduction détaillée aux bases d'ondelettes. Pour ce faire, nous emprunterons l'approche due à Mallat (1989a) et Meyer (1990) et basée sur le concept d'analyse multirésolution. Nous présenterons les ondelettes de Haar ainsi que la famille d'ondelettes possiblement la plus populaire, celle due à Daubechies (1988) et qui est vue par certains comme une généralisation de la base de Haar. Nous nous intéresserons spécialement au développement en série d'une fonction de densité à l'aide des bases d'ondelettes et discuterons de l'emploi de telles séries en estimation non paramétrique. Nous terminerons cette introduction par un survol de différents travaux du domaine de la statistique et utilisant les bases d'ondelettes.

Au chapitre 2, nous indiquerons comment utiliser les ondelettes de Haar, dans le contexte d'un modèle bayésien simple composé d'une seule observation  $X$ , afin d'approximer la densité marginale de cette observation, l'estimateur bayésien du paramètre de position  $\theta$ , ainsi que la perte espérée *a posteriori* de cet estimateur. La qualité des approximations obtenues sera illustrée à l'aide de deux exemples où nous étudierons en détail le comportement des approximations utilisées. Le deuxième exemple mettra également en valeur l'utilité de la technique lorsqu'il n'existe pas d'expression explicite pour les diverses quantités d'intérêt.

Puis, au chapitre 3, nous verrons comment généraliser les résultats vus au chapitre 2 au cas d'un modèle bayésien à plusieurs observations  $X_1, X_2, \dots, X_p$ . Pour cela, nous discuterons brièvement de l'existence et du rôle éventuel d'une statistique exhaustive pour  $\theta$  dans la construction des approximations. Ensuite, et principalement, nous nous intéresserons au cas des modèles bayésiens pour lesquels il est impossible d'identifier une statistique exhaustive pour le paramètre de position  $\theta$ . De nouvelles approximations de l'estimateur bayésien de  $\theta$  ainsi que la perte espérée *a posteriori* de cet estimateur seront obtenues. De plus, nous nous intéresserons au cas des densités *a priori* impropres et, en particulier, au cas de la loi *a priori* uniforme sur  $\mathbb{R}$ .

Enfin, le chapitre 4 présente l'une des nombreuses possibilités d'application de la technique d'approximation développée au chapitre 2. Le problème que nous y examinerons est celui de l'estimation de la densité *a priori* du paramètre de position  $\theta$ , lorsqu'elle est non spécifiée. En effet, nous indiquerons comment construire, à partir de l'approximation de la densité marginale d'une observation  $X$  présentée au chapitre 2, une estimation non paramétrique de la densité *a priori* de  $\theta$ . Diverses caractéristiques théoriques de l'estimateur de la densité *a priori* ainsi obtenu seront étudiées : l'absence de biais au niveau asymptotique, la convergence en moyenne quadratique, etc. Finalement, la procédure d'estimation résultante sera illustrée à l'aide d'un exemple simulé.

# Chapitre 1

---

## Les ondelettes

Nous nous intéressons ici particulièrement aux paramètres de position et notre but est double. D'abord, développer une technique de calcul bayésien pour l'estimation d'un paramètre de position, puis proposer une méthode d'estimation de la densité *a priori* d'un paramètre de position. Dans les deux cas, nous proposons de faire le travail à l'aide des bases d'ondelettes.

Les bases d'ondelettes sont devenues, depuis quelques années, bien à la mode pour la simple raison qu'elles sont en quelque sorte une version améliorée des bases orthogonales classiques (Fourier, L'Hermite, Legendre, etc.). En effet, quoique habituellement plus difficile d'application que les bases classiques (puisqu'elles demandent une analyse plus approfondie des problèmes ainsi que du calcul intensif), les bases d'ondelettes possèdent certaines propriétés fort intéressantes qu'elles ne partagent pas avec celles-ci.

Par exemple, les bases d'ondelettes peuvent être construites de façon à obtenir différents niveaux de régularité (continuité, dérivabilité), alors que les bases classiques ne sont généralement pas si flexibles. Également, il est possible de construire des bases d'ondelettes dont tous les éléments ont un support compact. Ce n'est pas le cas pour plusieurs bases classiques qui sont composées de fonctions définies sur  $\mathbb{R}$ .

Finalement, les bases d'ondelettes permettent de saisir le comportement local (et global, bien sûr) des fonctions. Un mode très prononcé, difficilement identifiable en travaillant à l'aide des bases classiques, devient souvent évident lorsque le problème est étudié avec les ondelettes. Les travaux de régression et d'estimation de densité (dans un contexte non paramétrique) de David Donoho et de ses collaborateurs l'ont clairement démontré (voir section 1.6). En fait, l'impact de ces travaux a été tel que la réputation des ondelettes n'est plus à faire parmi les statisticiens. Plusieurs ont même adopté les ondelettes comme outil principal en estimation non paramétrique. Nous discutons brièvement de quelques-uns des travaux conférant ce rôle aux ondelettes à la section 1.6.

Auparavant, nous présentons au lecteur une introduction aux ondelettes, lui permettant ainsi de se familiariser avec cet outil qui gagne sans cesse en popularité. En cas de besoin, le lecteur peut consulter Meyer (1990, 1992), Daubechies (1992), Chui (1992), Mallat (1998) et Walnut (2002), qui sont des ouvrages beaucoup plus complets sur la théorie des ondelettes. Le lecteur intéressé plus particulièrement à un historique de la création des bases d'ondelettes peut consulter Daubechies (1988). Finalement, pour une introduction aux ondelettes centrée sur leur application en statistique, le lecteur peut consulter Ogden (1997), Härdle *et al.* (1998) et Vidakovic (1999).

### 1.1. L'analyse multirésolution

Depuis leur création, les ondelettes ont été l'objet de plusieurs travaux. Plusieurs auteurs se sont intéressés spécialement à la façon de construire une base d'ondelettes. Nous adoptons l'approche due à Mallat (1989a) et Meyer (1990), soit l'utilisation du concept d'*analyse multirésolution*, autour duquel une base d'ondelettes peut être construite. (Le lecteur peut aussi consulter les ouvrages cités précédemment.)

**Définition 1.1.** Une analyse multirésolution de  $\mathcal{L}^2(\mathbb{R})$  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)  $\bigcup_{j \in \mathbb{Z}} V_j$  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)  $f(x) \in V_j \implies f(x - 2^{-j}k) \in V_j \quad \forall j, k \in \mathbb{Z}$ ,
- (vi)  $\exists h \in V_0$  telle que  $\{h(x - k)\}_{k \in \mathbb{Z}}$  est une base orthonormée de  $V_0$ , c'est-à-dire  $V_0$  peut s'écrire comme

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

Notons immédiatement que le "0" de la propriété (iii) correspond à la fonction identiquement nulle. Notons aussi que la propriété (iv) est présentée à la manière de Mallat (1989a). Une forme équivalente de cette propriété est donnée par

$$f(x) \in V_0 \iff f(2^j x) \in V_j \quad \forall j \in \mathbb{Z}.$$

Ce dernier résultat sera employé plus tard, à la section 1.3, pour plus de simplicité.

Maintenant, la notion d'analyse multirésolution nous amène à définir une fonction bien particulière : la *fonction d'échelle* (ou "scaling function"). En effet, la fonction d'échelle aura ici un rôle central puisque c'est autour d'elle que pourra être construite une analyse multirésolution.

**Définition 1.2.** Une fonction  $\phi \in \mathcal{L}^2(\mathbb{R})$  est appelée *fonction d'échelle* si

(i) les sous-espaces  $V_j$  de  $\mathcal{L}^2(\mathbb{R})$  définis par

$$V_j = \left\{ f \in \mathcal{L}^2(\mathbb{R}) \mid f(x) = \sum_{k \in \mathbb{Z}} c_{jk} \phi_{jk}(x) \right\},$$

où

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

forment une analyse multirésolution de  $\mathcal{L}^2(\mathbb{R})$ ;

(ii) la suite  $\{\phi(x - k)\}_{k \in \mathbb{Z}} = \{\phi_{0k}\}_{k \in \mathbb{Z}}$  forme une base orthonormée de  $V_0$ , c'est-à-dire  $\phi$  satisfait la propriété (vi) de la définition 1.1.

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

Examinons rapidement les implications de la définition 1.2. D'abord, remarquons que la propriété (ii) affirme que tous les translatés de  $\phi$  forment une base orthonormée de  $V_0$ . Ceci nous permet de déduire, pour  $j \neq 0$  fixé, l'orthonormalité des suites  $\{\phi_{jk}\}_{k \in \mathbb{Z}}$  engendrant les sous-espaces  $V_j$ . En effet, posons d'abord

$$\delta_{kl} = \begin{cases} 1 & \text{si } k = l, \\ 0 & \text{sinon,} \end{cases}$$

(c'est-à-dire  $\delta_{kl}$  correspond au delta de Kronecker), alors

$$\begin{aligned} \langle \phi_{0k}, \phi_{0l} \rangle &= \int_{\mathbb{R}} \phi_{0k}(x) \phi_{0l}(x) dx \\ &= \int_{\mathbb{R}} \phi(x - k) \phi(x - l) dx = \delta_{kl}, \end{aligned}$$

et, pour  $j \in \mathbb{Z}$ ,

$$\begin{aligned}
 \langle \phi_{jk}, \phi_{jl} \rangle &= \int_{\mathbb{R}} \phi_{jk}(x) \phi_{jl}(x) dx \\
 &= 2^j \int_{\mathbb{R}} \phi(2^j x - k) \phi(2^j x - l) dx \\
 &= \int_{\mathbb{R}} \phi(x - k) \phi(x - l) dx \\
 &= \langle \phi_{0k}, \phi_{0l} \rangle = \delta_{kl},
 \end{aligned} \tag{1.1}$$

et la suite  $\{\phi_{jk}\}_{k \in \mathbb{Z}}$  est bien orthonormée. Nous pouvons, par le fait même, conclure que  $\{\phi_{jk}\}_{k \in \mathbb{Z}}$  forme une base de  $V_j$ .

Remarquons également qu'une fonction d'échelle permet de construire une décomposition orthogonale de  $\mathcal{L}^2(\mathbb{R})$ . En effet, notons  $W_j$  le complément orthogonal de  $V_j$  dans  $V_{j+1}$ . Nous avons alors, pour  $j \in \mathbb{Z}$ ,

$$V_j \oplus W_j = V_{j+1}.$$

Ainsi, nous pouvons écrire  $V_j$  comme la somme directe de sous-espaces orthogonaux entre eux, soit

$$\begin{aligned}
 V_j &= V_{j-1} \oplus W_{j-1} \\
 &= (V_{j-2} \oplus W_{j-2}) \oplus W_{j-1} \\
 &= \dots = \bigoplus_{i < j} W_i,
 \end{aligned} \tag{1.2}$$

puisque la propriété (iii) des analyses multirésolutions (voir définition 1.1) amène

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\}.$$

De là, et par la propriété (ii) des analyses multirésolutions, une première décomposition orthogonale de  $\mathcal{L}^2(\mathbb{R})$  est donnée par

$$\mathcal{L}^2(\mathbb{R}) = \overline{\bigcup_{j \in \mathbb{Z}} V_j} = \overline{\bigoplus_{j \in \mathbb{Z}} W_j}. \tag{1.3}$$

Notons que l'équation (1.2) permet aussi d'écrire, pour  $j \in \mathbb{Z}$ ,

$$\mathcal{L}^2(\mathbb{R}) = \overline{V_j \oplus \bigoplus_{i \geq j} W_i}, \tag{1.4}$$

et qu'encore une fois, la décomposition de  $\mathcal{L}^2(\mathbb{R})$  est orthogonale,  $V_j$  étant orthogonal à  $W_i$  ( $\forall i \geq j$ ).



Maintenant, Mallat (1989a) montre, à l'aide d'arguments de la théorie des groupes, que  $W_0$ , le complément orthogonal de  $V_0$  dans  $V_1$ , possède une base similaire à celle de  $V_0$ . En effet, Mallat démontre qu'il existe une fonction  $\psi \in W_0$  telle que  $\{\psi(x - k)\}_{k \in \mathbb{Z}}$  forme une base orthonormale de  $W_0$ . Ainsi, en posant

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

il est possible d'exprimer  $W_j$ , le complément orthogonal de  $V_j$  dans  $V_{j+1}$ , comme

$$W_j = \left\{ f \in \mathcal{L}^2(\mathbb{R}) \mid f(x) = \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk}(x) \right\}.$$

De plus, la suite  $\{\psi_{jk}\}_{k \in \mathbb{Z}}$  forme une base orthonormale de  $W_j$ . En effet, comme dans le cas de  $\{\phi_{jk}\}_{k \in \mathbb{Z}}$ , l'orthonormalité de la suite  $\{\psi_{jk}\}_{k \in \mathbb{Z}}$  peut être obtenue facilement à partir de la base de  $W_0$ . D'abord, nous savons que

$$\langle \psi_{0k}, \psi_{0l} \rangle = \int_{\mathbb{R}} \psi(x - k) \psi(x - l) dx = \delta_{kl},$$

la suite  $\{\psi_{0k}\}_{k \in \mathbb{Z}}$  formant une base orthonormée de  $W_0$ . Ainsi, comme précédemment, pour  $j \in \mathbb{Z}$ ,

$$\begin{aligned} \langle \psi_{jk}, \psi_{jl} \rangle &= 2^j \int_{\mathbb{R}} \psi(2^j x - k) \psi(2^j x - l) dx \\ &= \int_{\mathbb{R}} \psi(x - k) \psi(x - l) dx \\ &= \langle \psi_{0k}, \psi_{0l} \rangle = \delta_{kl}, \end{aligned} \tag{1.5}$$

et la suite  $\{\psi_{jk}\}_{k \in \mathbb{Z}}$  est bien orthonormée (et forme donc une base de  $W_j$ ).

Finalement, en combinant la décomposition orthogonale de  $\mathcal{L}^2(\mathbb{R})$  donnée par l'équation (1.3) au fait que, pour  $j \in \mathbb{Z}$ ,  $\{\psi_{jk}\}_{k \in \mathbb{Z}}$  est une base orthonormée de  $W_j$ , nous obtenons que la suite  $\{\psi_{jk}\}_{j, k \in \mathbb{Z}}$  forme une base orthonormée de  $\mathcal{L}^2(\mathbb{R})$ . Nous pouvons donc écrire toute fonction  $f \in \mathcal{L}^2(\mathbb{R})$  comme

$$f(x) \stackrel{\mathcal{L}^2}{=} \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk}(x),$$

l'égalité précédente étant vérifiée au sens de  $\mathcal{L}^2(\mathbb{R})$ , c'est-à-dire signifiant que

$$\left\| f - \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk}(\cdot) \right\|_2 = 0.$$

Remarquons que les coefficients  $d_{jk}$  respectent ici

$$\sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} d_{jk}^2 = \|f\|_2^2 < +\infty,$$

ce dernier résultat provenant de l'égalité de Parseval (voir Rudin, 1987, sections 4.13 à 4.18).

Ainsi, il est possible d'obtenir, à partir de toute fonction d'échelle  $\phi$ , une nouvelle fonction  $\psi$  telle que la suite  $\{\psi_{jk}\}_{j,k \in \mathbb{Z}}$  forme une base orthonormée de  $\mathcal{L}^2(\mathbb{R})$ . Notons que ce sont les fonctions  $\psi_{jk}$  qui forment les *ondelettes*. Elles sont ainsi nommées depuis les travaux de Grossman et Morlet (1984). Il est important de remarquer que toutes les ondelettes sont issues de la même fonction, soit  $\psi$ . Dans le but de mettre de l'emphase sur ce phénomène de filiation, Meyer (1990) a désigné  $\psi$  comme la *mère* des ondelettes. Dans la même veine, il a aussi proclamé  $\phi$ , la fonction d'échelle, comme le *père* des ondelettes. Ce dernier lien de filiation, quoique *a priori* moins apparent que le lien unissant les ondelettes à leur mère, deviendra évident à la section suivante.

Notons qu'une autre base orthonormée de  $\mathcal{L}^2(\mathbb{R})$  peut être obtenue à l'aide des ondelettes. L'équation (1.4) permet d'exprimer toute fonction de  $\mathcal{L}^2(\mathbb{R})$  comme

$$f(x) \stackrel{\mathcal{L}^2}{=} \sum_{k \in \mathbb{Z}} c_{Jk} \phi_{Jk}(x) + \sum_{j \geq J} \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk}(x),$$

quelque soit  $J \in \mathbb{Z}$ , et où les coefficients  $c_{Jk}$  et  $d_{jk}$  respectent

$$\sum_{k \in \mathbb{Z}} c_{Jk}^2 + \sum_{j \geq J} \sum_{k \in \mathbb{Z}} d_{jk}^2 = \|f\|_2^2 < +\infty.$$

En effet, ceci est possible puisque  $\{\phi_{Jk}\}_{k \in \mathbb{Z}}$  est une base orthonormée de  $V_J$ , que  $\{\psi_{jk}\}_{k \in \mathbb{Z}}$  est une base orthonormée de  $W_j$  (pour  $j \in \mathbb{Z}$ ), et que finalement,  $V_J$  est orthogonal à  $W_j$ , pour tout  $j \geq J$ . Ainsi, quelque soit  $J \in \mathbb{Z}$ , la suite  $\{\phi_{Jk}, \psi_{jk}\}_{j \geq J, k \in \mathbb{Z}}$  forme aussi une base orthonormée de  $\mathcal{L}^2(\mathbb{R})$ .

Grâce aux ondelettes, nous avons maintenant accès à deux types de bases orthonormales de  $\mathcal{L}^2(\mathbb{R})$  : le premier composé uniquement d'ondelettes et le second composé d'ondelettes, mais faisant aussi appel à la fonction d'échelle. Ce résultat est fort intéressant, quoique pour le moment, limité. En effet, une ombre demeure au tableau : deux questions importantes restent sans réponse.

D'abord, comment construire  $\psi$ ? Si l'analyse multirésolution est construite autour de  $\phi$ , il faut tout de même pouvoir obtenir  $\psi$ . N'oublions pas que la mère des ondelettes joue un rôle central, puisque ce sont les dilatés et translatés de  $\psi$  qui forment une base orthonormée de  $\mathcal{L}^2(\mathbb{R})$ .

Ensuite, comment trouver une fonction d'échelle? Une preuve de l'existence de telles fonctions ne suffit pas. Il faut pouvoir exhiber une (ou plusieurs) fonction(s) d'échelle pour traiter des problèmes concrets.

Nous répondrons successivement à ces deux questions aux sections suivantes. Plus particulièrement, nous examinerons, à la section 1.2, les liens à faire entre le père et la mère des ondelettes, et aux sections 1.3 et 1.4, nous discuterons de quelques fonctions d'échelle importantes.

## 1.2. L'équation de dilatation

Le problème qui nous intéresse ici est celui de trouver une façon, à partir d'une fonction d'échelle donnée, de construire  $\psi$ , la mère des ondelettes correspondante. Pour cela, remarquons que si  $\{V_j\}_{j \in \mathbb{Z}}$  est une analyse multirésolution engendrée par  $\phi$ , une fonction d'échelle donnée, alors

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

et  $\phi$  peut être développée en série à partir de la base de  $V_1$ , soit  $\{\phi_{1k}\}_{k \in \mathbb{Z}}$ . Ainsi, il existe une suite  $\{A_k\}_{k \in \mathbb{Z}}$  de coefficients telle que

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

et

$$\sum_{k \in \mathbb{Z}} A_k^2 < +\infty.$$

Plus directement, comme

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

il existe une suite de carré sommable  $\{a_k\}_{k \in \mathbb{Z}}$  telle que

$$\phi(x) = \sum_{k \in \mathbb{Z}} a_k \phi(2x - k), \quad (1.6)$$

où  $a_k = \sqrt{2} A_k$ , et il devient alors clair que l'expression obtenue est construite à partir de dilatés de  $\phi$ . C'est pour cette raison que cette dernière équation est appelée *équation de dilatation*. Nous le verrons plus loin, cette équation joue un rôle capital dans la construction de  $\psi$ , la mère des ondelettes.

Maintenant, la réflexion menant à l'équation de dilatation, donnée par l'équation (1.6), peut également être faite dans la cas de  $\psi$ . Effectivement, comme

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

nous pouvons aussi développer  $\psi$  en série à l'aide de la base de  $V_1$ . Il existe donc une suite de coefficients  $\{b_k\}_{k \in \mathbb{Z}}$  de carré sommable telle que

$$\psi(x) = \sum_{k \in \mathbb{Z}} b_k \phi(2x - k). \quad (1.7)$$

Notons la grande similitude entre l'équation de dilatation et l'équation précédente.

Les coefficients  $a_k$  et  $b_k$  des équations (1.6) et (1.7) sont appelés des *filtres*, terminologie empruntée à la théorie de l'analyse des signaux. Ces filtres permettent d'obtenir  $\psi$  à partir de  $\phi$ . En effet, si  $\phi$  est connue (les filtres  $a_k$  le sont alors aussi), il suffit de choisir les filtres  $b_k$  de façon à ce que  $\psi$  ait les propriétés souhaitées.

En fait, la structure même de l'analyse multirésolution, présentée à la section précédente, nous permet de fixer les valeurs des filtres  $b_k$ . Pour cela, il faut simplement se rappeler que, pour  $j \in \mathbb{Z}$ ,

$$V_j \perp W_j.$$

Ainsi, nous devons avoir, pour  $j \in \mathbb{Z}$ ,

$$\langle \phi_{jk}, \psi_{jl} \rangle = 0 \quad \forall k, l \in \mathbb{Z}. \quad (1.8)$$

La proposition suivante permet d'étudier cette égalité.

**Proposition 1.1.** *Si  $\phi \in \mathcal{L}^2(\mathbb{R})$  est une fonction d'échelle et  $\psi$  est la mère des ondelettes qui lui est associée, et si  $\{a_k\}_{k \in \mathbb{Z}}$  et  $\{b_k\}_{k \in \mathbb{Z}}$  sont leurs suites de filtres respectives, alors pour  $j \in \mathbb{Z}$ ,*

$$\langle \phi_{jk}, \psi_{jl} \rangle = \frac{1}{2} \sum_{m \in \mathbb{Z}} a_m b_{m+2(k-l)} \quad \forall k, l \in \mathbb{Z}.$$

**Démonstration.** D'abord, nous savons que pour  $j \in \mathbb{Z}$ ,

$$\langle \phi_{jk}, \psi_{jl} \rangle = 2^j \int_{\mathbb{R}} \phi(2^j x - k) \psi(2^j x - l) dx,$$

pour  $k, l \in \mathbb{Z}$ . En employant les équations (1.6) et (1.7) ainsi que le théorème de Fubini pour inverser sommes et intégrales, nous avons donc

$$\begin{aligned} \langle \phi_{jk}, \psi_{jl} \rangle &= 2^j \int_{\mathbb{R}} \left( \sum_{m \in \mathbb{Z}} a_m \phi(2(2^j x - k) - m) \right) \left( \sum_{n \in \mathbb{Z}} b_n \phi(2(2^j x - l) - n) \right) dx \\ &= 2^j \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} a_m b_n \int_{\mathbb{R}} \phi(2^{j+1}x - (2k + m)) \phi(2^{j+1}x - (2l + n)) dx \\ &= \frac{1}{2} \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} a_m b_n \langle \phi_{(j+1)(2k+m)}, \phi_{(j+1)(2l+n)} \rangle, \end{aligned}$$

ce qui est équivalent, d'après l'équation (1.1), à

$$\begin{aligned} \langle \phi_{jk}, \psi_{jl} \rangle &= \frac{1}{2} \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} a_m b_n \delta_{(2k+m)(2l+n)} \\ &= \frac{1}{2} \sum_{m \in \mathbb{Z}} a_m b_{m+2(k-l)}, \end{aligned}$$

pour  $j \in \mathbb{Z}$  et  $k, l \in \mathbb{Z}$ . □

Notons qu'à travers cette thèse, sommes et intégrales seront inversées très fréquemment. Nous ne spécifierons donc pas à chaque fois que l'inversion utilisée est valide puisque les conditions d'application du théorème de Fubini sont toujours satisfaites. Pour plus de détails, le lecteur peut consulter Rudin (1987, sections 8.8 et 8.9).

Maintenant, la proposition précédente et l'équation (1.8) suggèrent la solution suivante au problème de la construction de  $\psi$ . Il suffit de trouver une suite  $\{b_k\}_{k \in \mathbb{Z}}$  de filtres telle que

$$\frac{1}{2} \sum_{m \in \mathbb{Z}} a_m b_{m+2(k-l)} = 0 \quad \forall k, l \in \mathbb{Z},$$

ou, plus simplement,

$$\sum_{m \in \mathbb{Z}} a_m b_{m+2k} = 0 \quad \forall k \in \mathbb{Z}, \quad (1.9)$$

et d'utiliser l'équation (1.7) pour obtenir l'expression de  $\psi$  qui lui est associée. La mère des ondelettes est ainsi construite à partir des filtres de la fonction d'échelle, et ce, de façon à respecter les contraintes d'orthogonalité imposées par l'analyse multirésolution. Clairement, le choix de la suite  $\{b_k\}_{k \in \mathbb{Z}}$  n'est pas unique. Il est donc possible de construire, à partir d'une seule fonction d'échelle, plusieurs

fonctions différentes pouvant agir comme mère des ondelettes, l'équation (1.9) permettant de les caractériser toutes.

Par ailleurs, il est intéressant de remarquer qu'en posant

$$b_k = (-1)^k a_{1-k} \quad \forall k \in \mathbb{Z}, \quad (1.10)$$

l'équation (1.9) est satisfaite, et ce quelque soit la suite  $\{a_k\}_{k \in \mathbb{Z}}$ . En effet, nous obtenons alors

$$\begin{aligned} \sum_{m \in \mathbb{Z}} a_m b_{m+2k} &= \sum_{m \in \mathbb{Z}} (-1)^{m+2k} a_m a_{1-m-2k} \\ &= \sum_{m \in \mathbb{Z}} (-1)^m a_m a_{1-m-2k} \\ &= 0 \quad \forall k \in \mathbb{Z}, \end{aligned}$$

sans égard à la suite  $\{a_k\}_{k \in \mathbb{Z}}$ , chaque terme de la forme  $a_i a_j$  étant présent exactement deux fois dans la somme, mais avec des signes opposés! De là, la proposition 1.1 permet de conclure que l'équation (1.8) est bien respectée pour ce choix particulier de la suite  $\{b_k\}_{k \in \mathbb{Z}}$ .

Ce dernier résultat suggère donc une solution générale à la construction de  $\psi$ . Cette solution, qui s'avère fort simple, consiste à choisir les filtres  $b_k$  selon l'équation (1.10). L'expression générale de  $\psi$ , obtenue à partir de l'équation (1.7), devient alors

$$\psi(x) = \sum_{k \in \mathbb{Z}} (-1)^k a_{1-k} \phi(2x - k). \quad (1.11)$$

Cette expression est la réponse à la première des deux questions soulevées à la section précédente. En effet, si nous disposons d'une fonction d'échelle  $\phi$ , alors l'équation (1.11) donne la forme de  $\psi$  suffisante pour qu'elle agisse comme mère des ondelettes et, par le fait même, que  $\{\psi_{jk}\}_{j,k \in \mathbb{Z}}$  forme une base orthonormale de  $\mathcal{L}^2(\mathbb{R})$ .

La deuxième question, quant à elle, reste toujours sans réponse. Effectivement, nous sommes toujours confrontés à la difficulté d'identifier une (ou plusieurs) fonction(s) d'échelle. Sans le passage par cette étape cruciale, la théorie présentée jusqu'ici demeure inutile. Cependant, nous n'examinerons ce problème qu'aux sections suivantes. Pour le moment, nous allons plutôt tourner notre attention vers quelques propriétés des suites de filtres découlant des équations (1.6) et (1.7).

D'abord, l'orthonormalité des translatés de  $\phi$  permet d'établir une première égalité intéressante.

**Proposition 1.2.** *Si  $\phi \in \mathcal{L}^2(\mathbb{R})$  est une fonction d'échelle et si  $\{a_k\}_{k \in \mathbb{Z}}$  est la suite de filtres qui lui est associée, alors*

$$\frac{1}{2} \sum_{m \in \mathbb{Z}} a_m a_{m+2(k-l)} = \delta_{kl} \quad \forall k, l \in \mathbb{Z}. \quad (1.12)$$

**Démonstration.** D'abord, d'après l'équation (1.1), pour  $j \in \mathbb{Z}$ ,

$$\langle \phi_{jk}, \phi_{jl} \rangle = \langle \phi_{0k}, \phi_{0l} \rangle = \delta_{kl} \quad \forall k, l \in \mathbb{Z}.$$

Nous pouvons donc écrire, en employant l'équation (1.6),

$$\begin{aligned} \delta_{kl} &= \int_{\mathbb{R}} \phi(x-k) \phi(x-l) dx \\ &= \int_{\mathbb{R}} \left( \sum_{m \in \mathbb{Z}} a_m \phi(2(x-k) - m) \right) \left( \sum_{n \in \mathbb{Z}} a_n \phi(2(x-l) - n) \right) dx \\ &= \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} a_m a_n \int_{\mathbb{R}} \phi(2x - (2k+m)) \phi(2x - (2l+n)) dx \\ &= \frac{1}{2} \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} a_m a_n \langle \phi_{2(2k+m)}, \phi_{2(2l+n)} \rangle, \end{aligned}$$

et, toujours d'après l'équation (1.1),

$$\begin{aligned} \delta_{kl} &= \frac{1}{2} \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} a_m a_n \delta_{(2k+m)(2l+n)} \\ &= \frac{1}{2} \sum_{m \in \mathbb{Z}} a_m a_{m+2(k-l)}, \end{aligned}$$

pour  $k, l \in \mathbb{Z}$ . □

Il est intéressant de constater qu'aucune information supplémentaire ne peut être tirée de l'orthonormalité des translatés de  $\psi$ , la mère des ondelettes. En effet, de l'équation (1.5), nous avons, pour  $j \in \mathbb{Z}$ ,

$$\langle \psi_{jk}, \psi_{jl} \rangle = \langle \psi_{0k}, \psi_{0l} \rangle = \delta_{kl} \quad \forall k, l \in \mathbb{Z}.$$

Ainsi, grâce à l'équation (1.7), nous obtenons

$$\begin{aligned}
\delta_{kl} &= \int_{\mathbb{R}} \psi(x-k)\psi(x-l) dx \\
&= \int_{\mathbb{R}} \left( \sum_{m \in \mathbb{Z}} b_m \phi(2(x-k) - m) \right) \left( \sum_{n \in \mathbb{Z}} b_n \phi(2(x-l) - n) \right) dx \\
&= \frac{1}{2} \sum_{m \in \mathbb{Z}} b_m b_{m+2(k-l)},
\end{aligned}$$

comme précédemment. Or, nous avons déjà fixé les valeurs prises par la suite de filtres  $\{b_k\}_{k \in \mathbb{Z}}$ . En remplaçant ces valeurs, données par l'équation (1.10), dans l'équation précédente, nous trouvons,

$$\begin{aligned}
\delta_{kl} &= \frac{1}{2} \sum_{m \in \mathbb{Z}} (-1)^m a_{1-m} (-1)^{m+2(k-l)} a_{1-m-2(k-l)} \\
&= \frac{1}{2} \sum_{m \in \mathbb{Z}} a_{1-m} a_{1-m-2(k-l)} \\
&= \frac{1}{2} \sum_{n \in \mathbb{Z}} a_n a_{n+2(k-l)},
\end{aligned}$$

en posant finalement  $n = 1 - m - 2(k - l)$ . Ceci nous ramène donc directement à l'équation (1.12), encore une fois. Ainsi, cette équation s'avère très importante puisqu'elle a un rôle double. Elle assure non seulement l'orthonormalité des translatés de  $\phi$ , mais aussi celle des translatés de  $\psi$ .

Deux autres propriétés des suites de filtres peuvent être obtenues à partir des équations (1.6) et (1.7). La première est obtenue en intégrant l'équation de dilatation, et elle fixe la valeur de  $\sum_{k \in \mathbb{Z}} a_k$ .

**Proposition 1.3.** *Sous les conditions de la proposition 1.2,*

$$\sum_{k \in \mathbb{Z}} a_k = 2.$$



**Démonstration.** D'abord, l'équation (1.6) permet d'écrire

$$\begin{aligned} \int_{\mathbb{R}} \phi(x) dx &= \int_{\mathbb{R}} \sum_{k \in \mathbb{Z}} a_k \phi(2x - k) dx \\ &= \sum_{k \in \mathbb{Z}} a_k \int_{\mathbb{R}} \phi(2x - k) dx \\ &= \frac{1}{2} \sum_{k \in \mathbb{Z}} a_k \int_{\mathbb{R}} \phi(x) dx, \end{aligned}$$

c'est-à-dire,

$$\int_{\mathbb{R}} \phi(x) dx = \left( \int_{\mathbb{R}} \phi(x) dx \right) \left( \frac{1}{2} \sum_{k \in \mathbb{Z}} a_k \right). \quad (1.13)$$

Maintenant, il est possible de montrer (voir théorème A.2, annexe A) que

$$\int_{\mathbb{R}} \phi(x) dx \neq 0.$$

Nous pouvons donc simplifier l'équation (1.13) pour obtenir

$$1 = \frac{1}{2} \sum_{k \in \mathbb{Z}} a_k,$$

ou encore,

$$\sum_{k \in \mathbb{Z}} a_k = 2,$$

qui est le résultat voulu. □

La dernière caractéristique des suites de filtres que nous obtenons ici provient de l'équation (1.7). En effet, nous procédons ici comme à la proposition précédente, mais en intégrant plutôt l'équation (1.7), afin d'arriver à la propriété recherchée.

**Proposition 1.4.** *Toujours sous les conditions de la proposition 1.2,*

$$\sum_{k \in \mathbb{Z}} (-1)^k a_k = 0.$$

**Démonstration.** D'abord, d'après l'équation (1.7), et en procédant comme à la proposition 1.3, nous avons

$$\begin{aligned} \int_{\mathbb{R}} \psi(x) dx &= \int_{\mathbb{R}} \sum_{k \in \mathbb{Z}} b_k \phi(2x - k) dx \\ &= \left( \int_{\mathbb{R}} \phi(x) dx \right) \left( \frac{1}{2} \sum_{k \in \mathbb{Z}} b_k \right). \end{aligned} \quad (1.14)$$

Par ailleurs, nous pouvons montrer que contrairement à  $\phi$ , la mère des ondelettes respecte (voir théorème A.2, annexe A)

$$\int_{\mathbb{R}} \psi(x) dx = 0.$$

L'équation (1.14) peut donc être simplifiée pour donner

$$\sum_{k \in \mathbb{Z}} b_k = 0.$$

De là, l'équation (1.10) permet d'écrire

$$\sum_{k \in \mathbb{Z}} (-1)^k a_{1-k} = 0.$$

Cette dernière équation peut être ramenée à une forme plus simple. En effet, nous pouvons aussi écrire

$$\sum_{k \in \mathbb{Z}} (-1)^{1-k} a_k = 0,$$

ou encore,

$$\sum_{k \in \mathbb{Z}} (-1)^k a_k = 0,$$

soit, le résultat souhaité.  $\square$

Malgré le fait que nous ayons pu identifier trois propriétés importantes des suites de filtres, le problème de l'identification d'une (ou plusieurs) fonction(s) d'échelle demeure. En effet, les propriétés données par les propositions 1.2, 1.3 et 1.4 ne caractérisent pas une unique suite  $\{a_k\}_{k \in \mathbb{Z}}$ . Il existe donc possiblement une infinité de fonctions d'échelle et le problème d'identification demeure entier : comment en trouver au moins une ?

Pour répondre à cette question, nous étudions d'abord, à la section suivante, la base de Haar. Cette base d'ondelettes est obtenue à partir de la fonction

d'échelle toute simple donnée par

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

Nous étudions également, mais à la section 1.4, la classe des fonctions d'échelle due à Daubechies (1988), et les ondelettes qui leurs sont associées. Cette classe de fonctions d'échelle est très populaire à cause de plusieurs propriétés qui lui sont propres, et qui s'avèrent très intéressantes pour l'estimation non paramétrique.

### 1.3. Les ondelettes de Haar

Nous avons déjà mentionné le fait que toute la théorie présentée jusqu'ici reste inutile s'il nous est impossible d'exhiber ne serait-ce qu'au moins une fonction d'échelle. Nous nous arrêtons donc dès maintenant (ainsi qu'à la section suivante) à une courte étude des fonctions d'échelle les plus populaires et des ondelettes qui leurs sont associées.

Nous nous intéressons d'abord plus particulièrement à la fonction d'échelle la plus simple qu'il soit possible de trouver, soit

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

Cette fonction d'échelle a été identifiée par Haar (1910), bien avant l'émergence du concept même de fonction d'échelle. En fait, Haar a indiqué comment construire, à partir de la fonction  $\phi$  mentionnée plus haut, une base orthonormée de  $\mathcal{L}^2(\mathbb{R})$ . Cette construction innovatrice était le précurseur des concepts d'analyse multirésolution et de fonction d'échelle.

Dans le but d'illustrer la théorie introduite jusqu'ici, nous consacrons la présente section à l'étude de  $\phi$  et de la base d'ondelettes qu'elle génère. Pour cela, nous examinons d'abord la première question qui vient tout naturellement à l'esprit :  $\phi$  est-elle bien une fonction d'échelle?

Pour s'en convaincre, il suffit en fait de s'assurer que  $\phi$  engendre une analyse multirésolution. D'abord, notons, comme précédemment,

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

et remarquons que, dans le cas bien particulier de la base de Haar,

$$\phi_{jk}(x) = \begin{cases} 2^{j/2} & \text{si } x \in \left[\frac{k}{2^j}, \frac{k+1}{2^j}\right), \\ 0 & \text{sinon.} \end{cases} \quad (1.15)$$

Nous débutons notre réflexion en considérant le sous-espace de  $\mathcal{L}^2(\mathbb{R})$ , noté  $V_0$ , composé des fonctions constantes sur les intervalles unitaires, soit

$$V_0 = \left\{ f \in \mathcal{L}^2(\mathbb{R}) \mid f \text{ est constante sur } [k, k+1), k \in \mathbb{Z} \right\}.$$

La structure même de  $V_0$  amène

$$f \in V_0 \implies f(x-k) \in V_0 \quad \forall k \in \mathbb{Z},$$

la translation entière étant clairement une opération fermée sur  $V_0$ . D'autre part, pour toute fonction  $f \in \mathcal{L}^2(\mathbb{R})$ , nous avons

$$f \in V_0 \iff f(x) = c_k \quad \text{pour } x \in [k, k+1), k \in \mathbb{Z},$$

ou encore,

$$\begin{aligned} f \in V_0 \iff f(x) &= \sum_{k \in \mathbb{Z}} c_k \phi_{0k}(x) \\ &= \sum_{k \in \mathbb{Z}} c_k \phi(x-k), \end{aligned}$$

pour une certaine suite  $\{c_k\}_{k \in \mathbb{Z}}$ . Ceci permet d'écrire

$$V_0 = \left\{ f \mid f(x) = \sum_{k \in \mathbb{Z}} c_k \phi(x-k), \sum_{k \in \mathbb{Z}} c_k^2 < +\infty \right\},$$

et comme

$$\langle \phi_{0k}, \phi_{0l} \rangle = \delta_{kl} \quad \forall k, l \in \mathbb{Z},$$

alors  $\{\phi_{0k}\}_{k \in \mathbb{Z}}$  forme une base orthonormée de  $V_0$ .

Maintenant, considérons un nouveau sous-espace de  $\mathcal{L}^2(\mathbb{R})$ , cette fois noté  $V_1$ . Ce nouveau sous-espace est composé des fonctions constantes sur les intervalles de la forme  $\left[\frac{k}{2}, \frac{k+1}{2}\right)$ ,  $k \in \mathbb{Z}$ . Il est évident que

$$V_0 \subset V_1,$$

et que

$$f \in V_1 \implies f\left(x - \frac{k}{2}\right) \in V_1 \quad \forall k \in \mathbb{Z},$$

la translation par un multiple de  $\frac{1}{2}$  étant ici une opération fermée sur  $V_1$ . Aussi, comme précédemment, il est clair que pour  $f \in \mathcal{L}^2(\mathbb{R})$ ,

$$f \in V_1 \iff f(x) = \sum_{k \in \mathbb{Z}} c_k \phi_{1k}(x),$$

pour une certaine suite  $\{c_k\}_{k \in \mathbb{Z}}$ . Ainsi, nous pouvons écrire

$$V_1 = \left\{ f \mid f(x) = \sum_{k \in \mathbb{Z}} c_k \phi_{1k}(x), \sum_{k \in \mathbb{Z}} c_k^2 < +\infty \right\},$$

et, comme la suite de fonctions  $\{\phi_{1k}\}_{k \in \mathbb{Z}}$  est orthonormée, elle forme donc une base orthonormale de  $V_1$ . Remarquons finalement que

$$f \in V_1 \iff \exists h \in V_0 \text{ telle que } f(x) = h(2x),$$

c'est-à-dire qu'une fonction  $f$  est dans  $V_1$  si et seulement si elle correspond à une fonction de  $V_0$  que l'on contracte d'un facteur deux. Il est donc aussi possible d'écrire

$$V_1 = \left\{ f \mid f(x) = h(2x), h \in V_0 \right\}.$$

Ce procédé peut être répété, de façon à obtenir (pour  $j = 2, 3, \dots$ ) les sous-espaces

$$V_j = \left\{ f \mid f(x) = h(2^j x), h \in V_0 \right\}, \quad (1.16)$$

composés des fonctions constantes sur les intervalles de la forme  $[\frac{k}{2^j}, \frac{k+1}{2^j})$ ,  $k \in \mathbb{Z}$ . Chaque sous-espace  $V_j$  a pour base orthonormée  $\{\phi_{jk}\}_{k \in \mathbb{Z}}$  et tous respectent

$$V_0 \subset V_1 \subset V_2 \subset \dots,$$

ainsi que

$$f \in V_j \implies f(x - 2^{-j}k) \in V_j. \quad (1.17)$$

Le procédé peut aussi être élargi de façon à obtenir les sous-espaces  $V_j$ , pour  $j < 0$ , correspondants aux fonctions dilatées de  $V_0$ . Nous avons alors

$$\dots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \dots,$$

ou encore,

$$V_j \subset V_{j+1} \quad \forall j \in \mathbb{Z}. \quad (1.18)$$

Remarquons que

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\}, \quad (1.19)$$

la fonction identiquement nulle étant la seule fonction de  $\mathcal{L}^2(\mathbb{R})$  qui soit constante sur toute la droite réelle.

Enfin, en jumelant les équations (1.16) à (1.19) au fait que  $\{\phi_{0k}\}_{k \in \mathbb{Z}}$  est une base orthonormale de  $V_0$  et que  $\bigcup_{j \in \mathbb{Z}} V_j$  est dense dans  $\mathcal{L}^2(\mathbb{R})$  (puisque les fonctions simples sont denses dans  $\mathcal{L}^2(\mathbb{R})$ , voir Rudin, 1987, sections 3.6 à 3.13), il devient clair que  $\phi$  et la suite de sous-espaces  $\{V_j\}_{j \in \mathbb{Z}}$  respectent les définitions 1.1 et 1.2. De ce fait,  $\{V_j\}_{j \in \mathbb{Z}}$  est une analyse multirésolution engendrée par  $\phi$ , qui est bien une fonction d'échelle.

À ce titre,  $\phi$  respecte l'équation de dilatation. Ici, comme

$$\phi(2x - k) = \frac{1}{\sqrt{2}} \phi_{1k}(x) = \begin{cases} 1 & \text{si } x \in \left[\frac{k}{2}, \frac{k+1}{2}\right), \\ 0 & \text{sinon,} \end{cases}$$

pour  $k \in \mathbb{Z}$ , il est clair que l'équation de dilatation se réduit à

$$\phi(x) = \phi(2x) + \phi(2x - 1),$$

et donc, que la suite de filtres associée à la fonction d'échelle de Haar est donnée par

$$a_k = \begin{cases} 1 & \text{si } k = 0 \text{ ou } 1, \\ 0 & \text{sinon.} \end{cases}$$

Il est facile de vérifier que cette suite de filtres satisfait les propositions 1.2 à 1.4.

Grâce à l'équation (1.11), il est aussi possible d'obtenir la mère des ondelettes associée à la base de Haar. En effet, nous avons

$$\begin{aligned} \psi(x) &= a_1 \phi(2x) - a_0 \phi(2x - 1) \\ &= \phi(2x) - \phi(2x - 1). \end{aligned}$$

Cette dernière équation permet d'abord d'identifier clairement les filtres  $\{b_k\}_{k \in \mathbb{Z}}$  de la mère des ondelettes, donnés par

$$b_k = \begin{cases} 1 & \text{si } k = 0, \\ -1 & \text{si } k = 1, \\ 0 & \text{sinon,} \end{cases}$$

mais elle permet aussi de trouver la forme explicite de  $\psi$ , soit

$$\psi(x) = \begin{cases} 1 & \text{si } x \in [0, \frac{1}{2}), \\ -1 & \text{si } x \in [\frac{1}{2}, 1), \\ 0 & \text{sinon.} \end{cases} \quad (1.20)$$

Notons finalement que les ondelettes de Haar, formant une base orthonormale de  $\mathcal{L}^2(\mathbb{R})$ , sont données par

$$\psi_{jk}(x) = \begin{cases} 2^{j/2} & \text{si } x \in [\frac{k}{2^j}, \frac{2k+1}{2^{j+1}}), \\ -2^{j/2} & \text{si } x \in [\frac{2k+1}{2^{j+1}}, \frac{k+1}{2^j}), \\ 0 & \text{sinon.} \end{cases} \quad (1.21)$$

Nous avons déjà mentionné, à la section précédente, le fait que plusieurs mères des ondelettes peuvent être associées à une fonction d'échelle donnée. C'est le cas ici puisque l'équation (1.11) ne représente qu'un des choix possibles, soit celui que nous considérons, à l'instar de Daubechies (1988), du fait de son étonnante simplicité. Ce choix mène à l'équation (1.20), soit la mère des ondelettes adoptée par Haar (1910) lui-même. Notons que certains auteurs, comme Härdle *et al.* (1998) identifient plutôt

$$\psi^*(x) = -\psi(x) = \begin{cases} -1 & \text{si } x \in [0, \frac{1}{2}), \\ 1 & \text{si } x \in [\frac{1}{2}, 1), \\ 0 & \text{sinon,} \end{cases}$$

comme mère des ondelettes de Haar. Cette manière de faire ne va pas à l'encontre de l'équation (1.9) qui caractérise tous les choix possibles pour la mère des ondelettes. En effet, la suite de filtres  $\{b_k^*\}_{k \in \mathbb{Z}}$  associée à  $\psi^*$  est donnée par

$$b_k^* = \begin{cases} -1 & \text{si } k = 0, \\ 1 & \text{si } k = 1, \\ 0 & \text{sinon.} \end{cases}$$

D'où,

$$b_k^* = -b_k \quad \forall k \in \mathbb{Z},$$

et, comme  $\{b_k\}_{k \in \mathbb{Z}}$  satisfait l'équation (1.9) (puisque  $\psi$  est admissible comme mère des ondelettes),

$$\sum_{k \in \mathbb{Z}} a_m b_{m+2k}^* = - \sum_{k \in \mathbb{Z}} a_m b_{m+2k} = 0,$$

et ainsi,  $\{b_k^*\}_{k \in \mathbb{Z}}$  satisfait aussi l'équation (1.9). La fonction  $\psi^*$  est donc bien aussi admissible comme mère des ondelettes. En fait, il est facile de montrer que, de manière générale, si  $\psi$  est la mère des ondelettes associée à  $\phi$ , une fonction d'échelle, alors  $\psi^* = -\psi$  est aussi admissible comme mère des ondelettes.

Revenons maintenant à l'équation (1.21). De ce résultat, il apparaît évident que l'emploi des ondelettes de Haar, dans le but de décrire une fonction très chaotique, permet de bien saisir les caractéristiques locales de celle-ci. En effet, comme chaque élément  $\psi_{jk}$  de la base a un support de longueur  $2^{-j}$ , les éléments pour lesquels  $j$  est grand permettent d'exprimer les changements de comportement brusques d'une fonction. En particulier, les discontinuités pourront être bien rendues, ce qui n'est habituellement pas le cas avec les bases classiques, comme la base de Fourier. En plus de sa simplicité, la base de Haar présente donc des avantages intéressants.

Évidemment, cette simplicité a son prix, et l'emploi de la base de Haar comporte aussi certains désavantages. Par exemple, les constructions obtenues à partir de la base de Haar ne sont généralement pas continues. Ainsi, lorsque la régularité (continuité et dérivabilité) des constructions est un point d'intérêt, il peut arriver que la base de Haar devienne inadéquate. Il faut alors se tourner vers d'autres bases d'ondelettes. Pour obtenir des informations plus techniques à propos de la base de Haar et de son utilisation, le lecteur peut consulter Kashin et Saakyan (1989, chapitre 3) et Walnut (2002, chapitres 5 et 6).

À la prochaine section, nous présentons la classe des fonctions d'échelle due à Daubechies (1988), et les ondelettes qui leurs sont associées. Ces bases d'ondelettes ont été obtenues spécifiquement de façon à satisfaire certaines contraintes de régularité.

## 1.4. Les ondelettes de Daubechies

Nous avons discuté, à la section 1.2, du fait que les propriétés données par les propositions 1.2 à 1.4 ne caractérisent pas une unique suite de filtres  $\{a_k\}_{k \in \mathbb{Z}}$ . Elles peuvent cependant être le point de départ dans la recherche de fonctions d'échelle. En effet, Daubechies (1988) a proposé une approche qui consiste à rechercher des fonctions d'échelle à travers leur suite de filtres. Cette approche permet de construire de toute pièce des fonctions d'échelle ayant des propriétés désirées, et ce, à partir de contraintes imposées aux suites de filtres. La technique employée est présentée par Alpert (1992) d'une manière ingénieuse et avec une étonnante simplicité. Nous adoptons ici cette façon de voir les travaux de Daubechies.



Dans cette optique, où la recherche de fonctions d'échelle passe par la recherche des suites de filtres, les propositions 1.2 à 1.4 forment effectivement un bon point de départ. En particulier, la proposition 1.2 est primordiale puisqu'elle assure l'orthonormalité des divers translatés de la fonction d'échelle et de la mère des ondelettes.

Il nous suffit en fait maintenant de rajouter suffisamment de contraintes aux équations déjà disponibles pour obtenir un système d'équations ayant pour solution la suite de filtres d'une fonction d'échelle. La première des contraintes identifiées par Daubechies est toute simple : la suite de filtres doit contenir un nombre fini d'éléments. Ceci permet en fait à  $\phi$  d'avoir un support compact. Comme nous l'avons déjà mentionné à quelques occasions, ceci s'avère extrêmement précieux lors de l'approximation de fonctions à l'aide de bases d'ondelettes.

C'est en choisissant différemment les coefficients non nuls de la suite de filtres que Daubechies a identifié plusieurs fonctions d'échelle distinctes. En fait, c'est plus particulièrement en augmentant le nombre de coefficients utiles de la suite de filtres que toute une classe de fonctions d'échelle a pu être créée.

En effet, Daubechies a choisi sa fonction d'échelle d'ordre  $N$  ( $N \geq 1$ ), notée  ${}_N\phi$ , comme ayant  $2N$  coefficients non nuls dans sa suite de filtres. Ces coefficients sont notés  $a_0, a_1, \dots, a_{2N-1}$ . L'équation de dilatation associée à  ${}_N\phi$  peut donc s'écrire

$${}_N\phi(x) = \sum_{k=0}^{2N-1} a_k {}_N\phi(2x - k),$$

et, comme l'explique bien Alpert (1992), le support de  ${}_N\phi$  est donc  $[0, 2N - 1)$ . Aussi, la mère des ondelettes correspondante, notée  ${}_N\psi$ , est quant à elle donnée, d'après l'équation (1.11), par

$$\begin{aligned} {}_N\psi(x) &= \sum_{k \in \mathbb{Z}} (-1)^k a_{1-k} {}_N\phi(2x - k) \\ &= \sum_{k=2-2N}^1 (-1)^k a_{1-k} {}_N\phi(2x - k), \end{aligned} \quad (1.22)$$

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

Avant même d'examiner d'autres contraintes, remarquons qu'il nous est déjà possible de trouver  ${}_1\phi$ , la fonction d'échelle de Daubechies d'ordre un. En effet, cette fonction d'échelle possède seulement deux coefficients non nuls dans sa suite de filtres, soit  $a_0$  et  $a_1$ , qu'il est très facile d'obtenir. D'abord, la proposition 1.3

permet d'écrire

$$a_0 + a_1 = 2,$$

et la proposition 1.4 amène

$$a_0 - a_1 = 0.$$

Il est évident que l'unique solution de ce système d'équations est  $a_0 = a_1 = 1$ . L'équation de dilatation associée à  ${}_1\phi$  est donc

$${}_1\phi(x) = {}_1\phi(2x) + {}_1\phi(2x - 1),$$

ce qui nous ramène de facto à la base de Haar puisque chaque suite de filtres caractérise une unique fonction d'échelle. De même, la base d'ondelettes de Daubechies d'ordre 1 n'est en fait, tout simplement, que la base d'ondelettes de Haar. Ce fait, qui peut paraître au départ surprenant, est en réalité fort intéressant puisqu'il permet de voir les bases d'ondelettes de Daubechies comme une généralisation directe de la base de Haar.

Maintenant, pour obtenir les fonctions d'échelle d'ordre  $N > 1$ , les contraintes vues jusqu'ici ne sont pas suffisantes. La deuxième contrainte supplémentaire identifiée par Daubechies en est une qui a trait plutôt à la régularité des fonctions d'échelle, et par le fait même, des ondelettes. Cette nouvelle contrainte exige que les polynômes de degré inférieur à  $N$  puissent s'exprimer directement à l'aide des translatés de  ${}_N\phi$ . En d'autres termes, les polynômes de degré inférieur à  $N$  doivent être inclus dans

$$T_0^N = \left\{ f \mid f(x) = \sum_{k \in \mathbb{Z}} \beta_k {}_N\phi(x - k) \right\},$$

l'espace engendré par les translatés de  ${}_N\phi$ . Cette exigence amène les fonctions d'échelle  ${}_N\phi$  à être de plus en plus régulières au fur et à mesure que  $N$  augmente (voir Alpert, 1992).

La proposition suivante permet de traduire la contrainte identifiée par Daubechies en termes d'équations sur la suite des filtres  $\{a_k\}_{k \in \mathbb{Z}}$ .

**Proposition 1.5.** *Si  $\phi \in \mathcal{L}^2(\mathbb{R})$  est une fonction d'échelle, si  $\{a_k\}_{k \in \mathbb{Z}}$  est la suite de filtres associée à  $\phi$ , et s'il existe des suites  $\{\alpha_{pk}\}_{k \in \mathbb{Z}}$  telles que pour  $p = 0, 1, \dots, N - 1$ ,*

$$x^p = \sum_{k \in \mathbb{Z}} \alpha_{pk} \phi(x - k) \quad \forall x \in \mathbb{R}, \quad (1.23)$$

alors

$$\sum_{k \in \mathbb{Z}} (-1)^k k^p a_k = 0,$$

pour  $p = 0, 1, \dots, N - 1$ .

**Démonstration.** Posons d'abord

$$h_p(x) = x^p,$$

pour  $p = 0, 1, \dots, N - 1$ , et montrons que  $h_p$  est orthogonale à  $\psi$ , la mère des ondelettes associée à  $\phi$ . En effet, l'hypothèse donnée par l'équation (1.23) permet d'écrire

$$\begin{aligned} \langle h_p, \psi \rangle &= \int_{\mathbb{R}} h_p(x) \psi(x) dx \\ &= \int_{\mathbb{R}} \left( \sum_{k \in \mathbb{Z}} \alpha_{pk} \phi(x - k) \right) \psi(x) dx \\ &= \sum_{k \in \mathbb{Z}} \alpha_{pk} \int_{\mathbb{R}} \phi(x - k) \psi(x) dx, \end{aligned}$$

et donc, de par l'équation (1.8),

$$\begin{aligned} \langle h_p, \psi \rangle &= \sum_{k \in \mathbb{Z}} \alpha_{pk} \langle \phi_{0k}, \psi_{00} \rangle \\ &= 0, \end{aligned}$$

pour  $p = 0, 1, \dots, N - 1$ .

Examinons maintenant les conséquences de cette orthogonalité. D'abord, l'équation (1.7) amène

$$\begin{aligned} \langle h_p, \psi \rangle &= \int_{\mathbb{R}} h_p(x) \psi(x) dx \\ &= \int_{\mathbb{R}} x^p \left( \sum_{k \in \mathbb{Z}} b_k \phi(2x - k) \right) dx \\ &= \sum_{k \in \mathbb{Z}} b_k \int_{\mathbb{R}} x^p \phi(2x - k) dx. \end{aligned}$$

En faisant un simple changement de variable, nous avons

$$\begin{aligned} \langle h_p, \psi \rangle &= \sum_{k \in \mathbb{Z}} b_k \int_{\mathbb{R}} \left( \frac{x+k}{2} \right)^p \phi(x) \frac{dx}{2} \\ &= \frac{1}{2^{p+1}} \sum_{k \in \mathbb{Z}} b_k \int_{\mathbb{R}} (x+k)^p \phi(x) dx. \end{aligned}$$

Finalement, en développant à l'aide de la formule du binôme, nous obtenons

$$\begin{aligned} \langle h_p, \psi \rangle &= \frac{1}{2^{p+1}} \sum_{k \in \mathbb{Z}} b_k \int_{\mathbb{R}} \left( \sum_{n=0}^p \binom{p}{n} x^n k^{p-n} \right) \phi(x) dx \\ &= \frac{1}{2^{p+1}} \sum_{k \in \mathbb{Z}} b_k \sum_{n=0}^p \binom{p}{n} k^{p-n} \int_{\mathbb{R}} x^n \phi(x) dx \\ &= \frac{1}{2^{p+1}} \sum_{k \in \mathbb{Z}} b_k \sum_{n=0}^p \binom{p}{n} k^{p-n} \langle h_n, \phi \rangle, \end{aligned} \quad (1.24)$$

pour  $p = 0, 1, \dots, N-1$ . Notons que pour  $p = 0$ , cette expression peut être simplifiée pour obtenir

$$0 = \langle h_0, \psi \rangle = \frac{1}{2} \langle h_0, \phi \rangle \sum_{k \in \mathbb{Z}} b_k.$$

De là, comme (voir théorème A.2, annexe A)

$$\langle h_0, \phi \rangle = \int_{\mathbb{R}} \phi(x) dx \neq 0,$$

nous avons alors

$$\sum_{k \in \mathbb{Z}} b_k = 0, \quad (1.25)$$

un résultat déjà rencontré, quoique dans un contexte légèrement différent, à la proposition 1.4.

Revenons à l'équation (1.24). Prenons, cette fois, le cas où  $p > 0$ . Nous pouvons inverser l'ordre des sommations et écrire

$$\begin{aligned} \langle h_p, \psi \rangle &= \frac{1}{2^{p+1}} \sum_{n=0}^p \binom{p}{n} \langle h_n, \phi \rangle \sum_{k \in \mathbb{Z}} b_k k^{p-n} \\ &= \frac{1}{2^{p+1}} \left[ \langle h_0, \phi \rangle \sum_{k \in \mathbb{Z}} b_k k^p + \sum_{n=1}^p \binom{p}{n} \langle h_n, \phi \rangle \sum_{k \in \mathbb{Z}} b_k k^{p-n} \right], \end{aligned} \quad (1.26)$$

pour  $p = 1, 2, \dots, N-1$ . Dans le cas particulier où  $p = 1$ , nous avons donc

$$\begin{aligned} 0 = \langle h_1, \psi \rangle &= \frac{1}{4} \left[ \langle h_0, \phi \rangle \sum_{k \in \mathbb{Z}} b_k k + \langle h_1, \phi \rangle \sum_{k \in \mathbb{Z}} b_k \right] \\ &= \frac{1}{4} \langle h_0, \phi \rangle \sum_{k \in \mathbb{Z}} b_k k, \end{aligned}$$

d'après l'équation (1.25). Ainsi, nous avons aussi

$$\sum_{k \in \mathbb{Z}} b_k k = 0. \quad (1.27)$$

En revenant à l'équation (1.26), mais pour le cas de  $p = 2$ , nous pouvons écrire

$$\begin{aligned} 0 = \langle h_2, \psi \rangle &= \frac{1}{8} \left[ \langle h_0, \phi \rangle \sum_{k \in \mathbb{Z}} b_k k^2 + \sum_{n=1}^2 \binom{2}{n} \langle h_n, \phi \rangle \sum_{k \in \mathbb{Z}} b_k k^{2-n} \right] \\ &= \frac{1}{8} \langle h_0, \phi \rangle \sum_{k \in \mathbb{Z}} b_k k^2, \end{aligned}$$

puisque d'après les équations (1.25) et (1.27),

$$\sum_{k \in \mathbb{Z}} b_k k^{2-n} = 0,$$

pour  $n = 1, 2$ . Il nous est donc possible d'écrire

$$\sum_{k \in \mathbb{Z}} b_k k^2 = 0.$$

Il suffit maintenant de répéter ce raisonnement et de remplacer successivement les valeurs  $p = 3, 4, \dots, N - 1$  dans l'équation (1.26). Ceci permet finalement d'obtenir

$$\sum_{k \in \mathbb{Z}} b_k k^p = 0, \quad (1.28)$$

pour  $p = 0, 1, \dots, N - 1$ .

Maintenant, pour retrouver le résultat souhaité, nous devons employer encore une fois l'équation (1.10). Cette équation amène

$$\begin{aligned} \sum_{k \in \mathbb{Z}} b_k k^p &= \sum_{k \in \mathbb{Z}} (-1)^k a_{1-k} k^p \\ &= \sum_{k \in \mathbb{Z}} (-1)^{1-k} a_k (1-k)^p \\ &= \sum_{k \in \mathbb{Z}} (-1)^{k+1} a_k (1-k)^p. \end{aligned}$$

Par ailleurs, comme

$$(1-k)^p = (-1)^p (k-1)^p = (-1)^p \sum_{n=0}^p \binom{p}{n} (-1)^n k^{p-n},$$

nous avons donc

$$\sum_{k \in \mathbb{Z}} b_k k^p = (-1)^{p+1} \sum_{k \in \mathbb{Z}} (-1)^k a_k \sum_{n=0}^p \binom{p}{n} (-1)^n k^{p-n}, \quad (1.29)$$

pour  $p = 0, 1, \dots, N - 1$ . Remarquons que pour le cas de  $p = 0$ , cette expression prend une forme toute simple, soit

$$\sum_{k \in \mathbb{Z}} b_k = - \sum_{k \in \mathbb{Z}} (-1)^k a_k,$$

et donc, d'après l'équation (1.28),

$$\sum_{k \in \mathbb{Z}} (-1)^k a_k = 0,$$

qui n'est autre que le résultat vu à la proposition 1.4.

Pour les valeurs positives de  $p$ , l'équation (1.29) peut s'écrire sous une forme similaire à l'équation (1.26) vue plus tôt. En effet, en changeant l'ordre des sommations, nous obtenons

$$\begin{aligned} \sum_{k \in \mathbb{Z}} b_k k^p &= (-1)^{p+1} \sum_{n=0}^p \binom{p}{n} (-1)^n \sum_{k \in \mathbb{Z}} (-1)^k a_k k^{p-n} \\ &= (-1)^{p+1} \left[ \sum_{k \in \mathbb{Z}} (-1)^k a_k k^p + \sum_{n=1}^p \binom{p}{n} (-1)^n \sum_{k \in \mathbb{Z}} (-1)^k a_k k^{p-n} \right], \end{aligned}$$

pour  $p = 1, 2, \dots, N-1$ . Il ne reste plus qu'à faire comme auparavant en remplaçant successivement les valeurs  $p = 1, 2, \dots, N-1$  dans l'équation précédente et d'employer l'équation (1.28) pour voir apparaître le résultat espéré, soit

$$\sum_{k \in \mathbb{Z}} (-1)^k a_k k^p = 0,$$

pour  $p = 0, 1, \dots, N-1$ . □

La proposition précédente permet donc d'introduire  $N-1$  nouvelles équations que doivent satisfaire les filtres  $\{a_k\}_{k \in \mathbb{Z}}$  de la fonction d'échelle  ${}_N\phi$ . En jumelant ces équations aux propositions 1.2 à 1.4, il est possible de trouver les suites de filtres des fonctions d'échelle d'ordre  $N > 1$ . À titre d'exemple, nous obtenons ici la suite de filtres de  ${}_2\phi$ .

D'abord, nous tirons de la proposition 1.3,

$$\sum_{k=0}^3 a_k = 2,$$

et de la proposition 1.4,

$$\sum_{k=0}^3 (-1)^k a_k = 0.$$

De plus, comme les polynômes du premier degré s'expriment à l'aide des translatés de  ${}_2\phi$ , la proposition 1.5 permet d'ajouter, comme contrainte supplémentaire,

$$\sum_{k=0}^3 (-1)^k a_k k = 0.$$

Ces trois équations peuvent être maintenant regroupées afin d'écrire le système suivant,

$$\begin{cases} a_0 + a_1 + a_2 + a_3 = 2, \\ a_0 - a_1 + a_2 - a_3 = 0, \\ -a_1 + 2a_2 - 3a_3 = 0. \end{cases}$$

De là, nous avons

$$\begin{cases} a_1 = a_0 + \frac{1}{2}, \\ a_2 = 1 - a_0, \\ a_3 = \frac{1}{2} - a_0, \end{cases} \quad (1.30)$$

la valeur de  $a_0$  restant à être déterminée. Pour cela, notons que la proposition 1.2 nous permet d'écrire

$$a_0 a_2 + a_1 a_3 = 0. \quad (1.31)$$

Ainsi, en remplaçant les valeurs données par le système (1.30) dans l'égalité précédente, nous obtenons l'équation du second degré suivante,

$$a_0(1 - a_0) + \left(a_0 + \frac{1}{2}\right) \left(\frac{1}{2} - a_0\right) = 0,$$

ou encore,

$$2a_0^2 - a_0 - \frac{1}{4} = 0,$$

dont les solutions sont données par

$$a_0 = \frac{1 \pm \sqrt{3}}{4}.$$

Daubechies a choisi la valeur positive de  $a_0$  pour avoir que

$$\int_{\mathbb{R}} {}_2\phi(x) dx = 1,$$



(voir théorème A.2, annexe A) et ainsi obtenir, comme valeurs des filtres de  ${}_2\phi$ ,

$$\begin{aligned} a_0 &= \frac{1 + \sqrt{3}}{4} \simeq 0,683013, \\ a_1 &= \frac{3 + \sqrt{3}}{4} \simeq 1,183013, \\ a_2 &= \frac{3 - \sqrt{3}}{4} \simeq 0,316987, \\ a_3 &= \frac{1 - \sqrt{3}}{4} \simeq -0,183013. \end{aligned}$$

Notons que la proposition 1.2 permet aussi d'écrire

$$a_0^2 + a_1^2 + a_2^2 + a_3^2 = 2,$$

cette dernière équation pouvant être substituée, dans le raisonnement précédent, à l'équation (1.31). Ceci amène cependant les mêmes valeurs pour la suite de filtres de  ${}_2\phi$ .

Il nous est maintenant possible d'écrire l'équation de dilatation liée à la fonction d'échelle de Daubechies d'ordre 2 sous une forme plus explicite, c'est-à-dire

$$\begin{aligned} {}_2\phi(x) &= \left(\frac{1 + \sqrt{3}}{4}\right) {}_2\phi(2x) + \left(\frac{3 + \sqrt{3}}{4}\right) {}_2\phi(2x - 1) + \left(\frac{3 - \sqrt{3}}{4}\right) {}_2\phi(2x - 2) \\ &\quad + \left(\frac{1 - \sqrt{3}}{4}\right) {}_2\phi(2x - 3). \end{aligned}$$

Notons que  ${}_2\phi$  a comme support l'intervalle  $[0, 3)$ . Aussi, la mère des ondelettes de Daubechies d'ordre 2 est donnée, selon l'équation (1.22), par

$$\begin{aligned} {}_2\psi(x) &= \left(\frac{1 - \sqrt{3}}{4}\right) {}_2\phi(2x + 2) - \left(\frac{3 - \sqrt{3}}{4}\right) {}_2\phi(2x + 1) + \left(\frac{3 + \sqrt{3}}{4}\right) {}_2\phi(2x) \\ &\quad - \left(\frac{1 + \sqrt{3}}{4}\right) {}_2\phi(2x - 1), \end{aligned}$$

et son support est  $[-1, 2)$ .

La figure 1.1 donne les graphes de  ${}_2\phi$  et  ${}_2\psi$ . Ces graphes sont obtenus, à partir de la suite de filtres de  ${}_2\phi$ , à l'aide de l'algorithme pyramidal local dû à Daubechies et Lagarias (1991, 1992). Cet algorithme permet d'évaluer des fonctions d'échelle et la mère des ondelettes qui leurs sont associées à des points donnés, à partir de leur suite de filtres (voir aussi Vidakovic, 1999, section 3.5.4). Une autre technique,

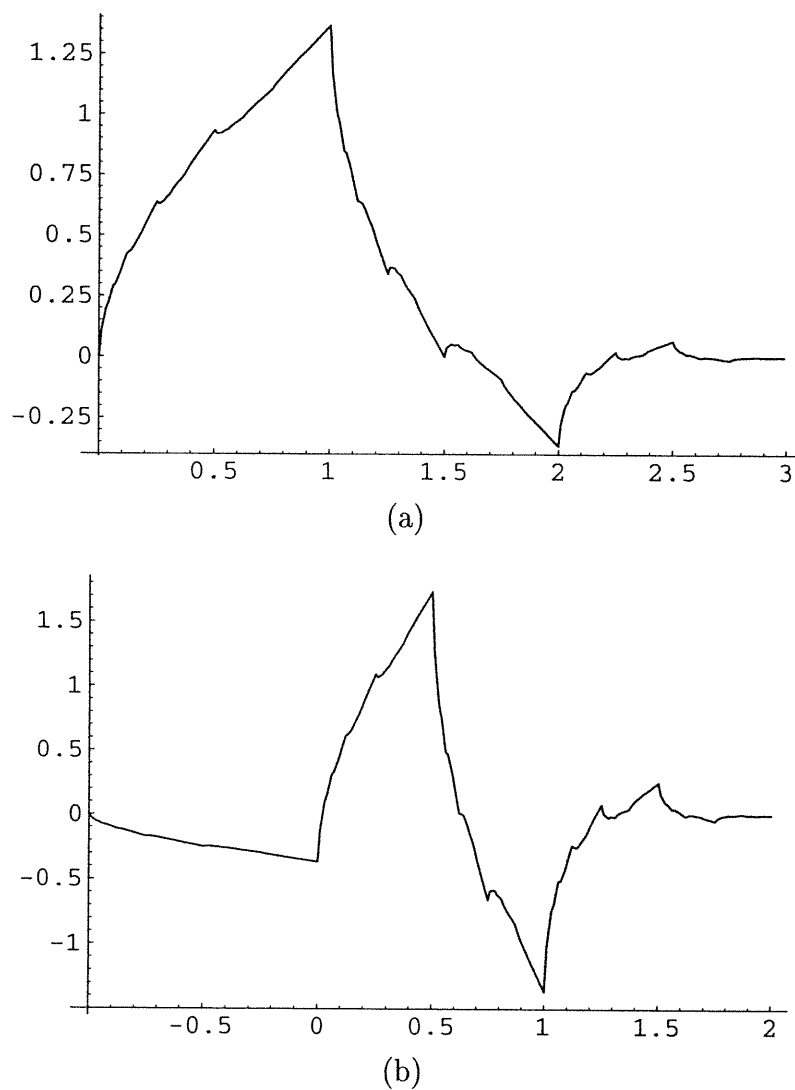


FIG. 1.1. Fonction d'échelle et mère des ondelettes de Daubechies d'ordre 2 (a)  ${}_2\phi$  (b)  ${}_2\psi$ .

due à Strang (1989) et clairement illustrée par Alpert (1992), peut aussi être employée.

En étudiant rapidement ces deux figures, il est évident que  ${}_2\phi$  et  ${}_2\psi$  sont continues. Il semble aussi que les deux fonctions ne sont pas continûment dérivables. Nous remarquons finalement que ces fonctions sont très asymétriques. En fait, Daubechies (1988) a pu démontrer que l'unique fonction d'échelle symétrique

possédant aussi un support compact est  ${}_1\phi$ , la fonction d'échelle associée à la base de Haar.

Aussi, rappelons nous que les polynômes de degré inférieur à deux peuvent s'écrire à l'aide des translatés de  ${}_2\phi$ . Plus précisément, il est possible de montrer que

$$1 = \sum_{k \in \mathbb{Z}} {}_2\phi(x - k),$$

et

$$x = \sum_{k \in \mathbb{Z}} \left( k + \frac{3 - \sqrt{3}}{2} \right) {}_2\phi(x - k),$$

et ce, quelque soit  $x \in \mathbb{R}$ . Ceci peut paraître très surprenant vue l'allure chaotique de  ${}_2\phi$  mise en lumière par la figure 1.1.

Voyons maintenant le comportement de  ${}_3\phi$  et de  ${}_3\psi$ . Comme précédemment, il suffit ici d'écrire un système de six équations à l'aide des propositions 1.2 à 1.5. Nous pouvons alors obtenir les coefficients  $a_0, a_1, \dots, a_5$  de la suite de filtres de  ${}_3\phi$ , et finalement, l'équation de dilatation associée à  ${}_3\phi$ , soit

$$\begin{aligned} {}_3\phi(x) = & \left( \frac{1 + \sqrt{10} + \sqrt{5 + 2\sqrt{10}}}{16} \right) {}_3\phi(2x) \\ & + \left( \frac{5 + \sqrt{10} + 3\sqrt{5 + 2\sqrt{10}}}{16} \right) {}_3\phi(2x - 1) \\ & + \left( \frac{10 - 2\sqrt{10} + 2\sqrt{5 + 2\sqrt{10}}}{16} \right) {}_3\phi(2x - 2) \\ & + \left( \frac{10 - 2\sqrt{10} - 2\sqrt{5 + 2\sqrt{10}}}{16} \right) {}_3\phi(2x - 3) \\ & + \left( \frac{5 + \sqrt{10} - 3\sqrt{5 + 2\sqrt{10}}}{16} \right) {}_3\phi(2x - 4) \\ & + \left( \frac{1 + \sqrt{10} - \sqrt{5 + 2\sqrt{10}}}{16} \right) {}_3\phi(2x - 5). \end{aligned}$$

Notons que  ${}_3\phi$  est non nulle uniquement sur l'intervalle  $[0, 5)$ . Quant à elle,  ${}_3\psi$  peut être facilement obtenue à partir de l'équation (1.22), et son support est  $[-2, 3)$ .

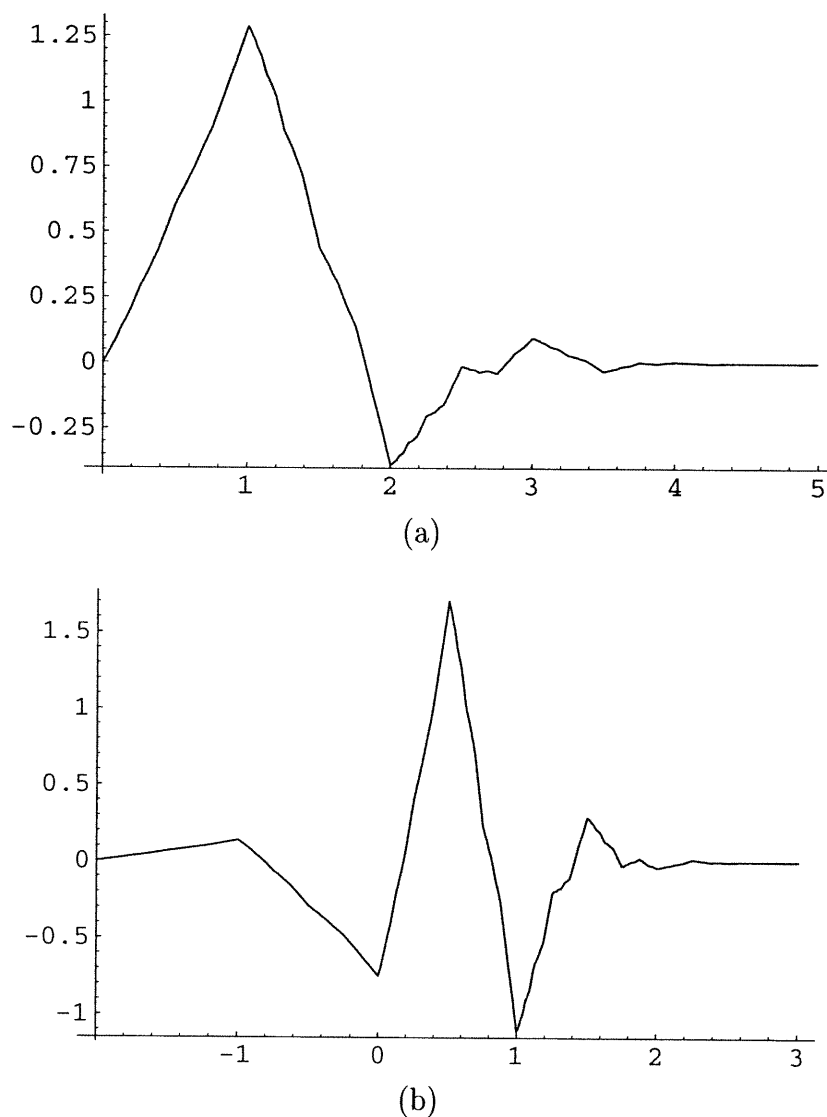


FIG. 1.2. *Fonction d'échelle et mère des ondelettes de Daubechies d'ordre 3 (a)  ${}_3\phi$  (b)  ${}_3\psi$ .*

La figure 1.2 présente les graphes de  ${}_3\phi$  et  ${}_3\psi$ . Comme  ${}_2\phi$  et  ${}_2\psi$ , ces fonctions sont très asymétriques. Par ailleurs, elles ne semblent, à première vue, que légèrement plus régulières que ces dernières. Ceci est trompeur, puisque Daubechies (1992) a réussi à démontrer que  ${}_3\phi$  et  ${}_3\psi$  sont continûment dérivables.

D'autre part, nous savons que les polynômes de degré inférieur à trois peuvent être exprimés à l'aide des translatés de  ${}_3\phi$ . Effectivement, il est possible d'obtenir,

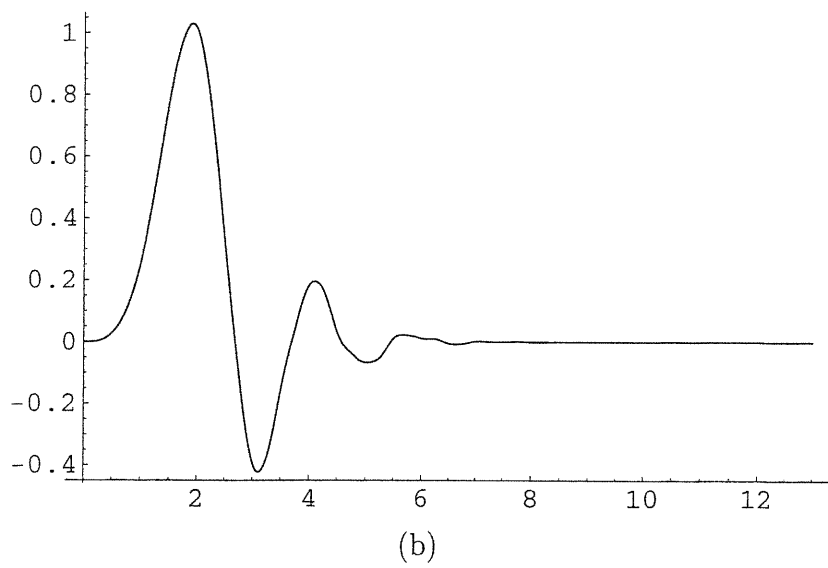
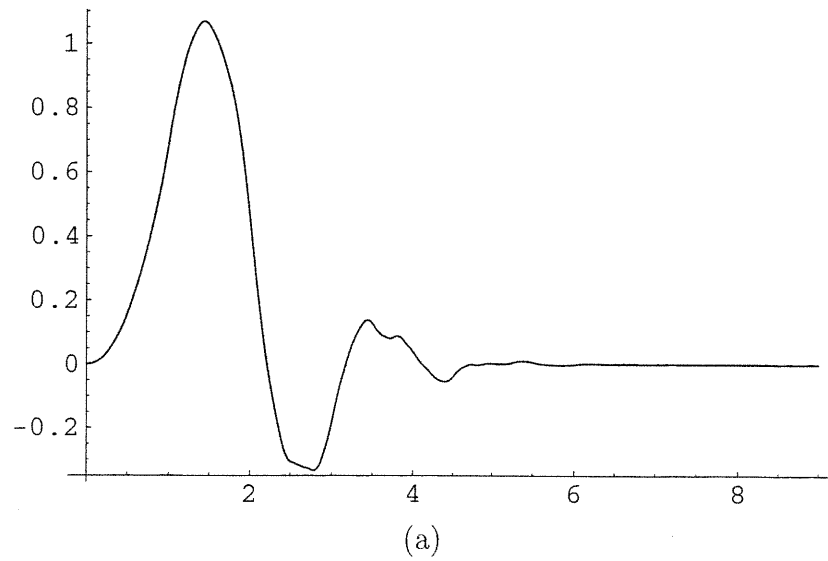


FIG. 1.3. *Fonctions d'échelle de Daubechies d'ordre 5 et 7 (a)  ${}_5\phi$  (b)  ${}_7\phi$ .*

pour tout  $x \in \mathbb{R}$ ,

$$1 = \sum_{k \in \mathbb{Z}} {}_3\phi(x - k),$$

$$x = \sum_{k \in \mathbb{Z}} \left( k + \frac{5 - \sqrt{5 + 2\sqrt{10}}}{2} \right) {}_3\phi(x - k),$$

et finalement,

$$x^2 = \sum_{k \in \mathbb{Z}} \left( k^2 + \left( 5 - \sqrt{5 + 2\sqrt{10}} \right) k + \frac{15 + \sqrt{10} - 5\sqrt{5 + 2\sqrt{10}}}{2} \right) {}_3\phi(x - k).$$

Encore une fois, ces propriétés peuvent paraître surprenantes puisque  ${}_3\phi$  ne semble pas non plus très régulière.

Dans le but de voir des fonctions d'échelle plus régulières, nous présentons à la figure 1.3, les fonctions  ${}_5\phi$  et  ${}_7\phi$ . Ceci permet au lecteur de voir l'augmentation de régularité des fonctions d'échelle de Daubechies avec  $N$ , l'ordre des fonctions d'échelle. Effectivement, il est évident que  ${}_5\phi$  et  ${}_7\phi$  sont beaucoup plus régulières que  ${}_2\phi$ , et même que  ${}_3\phi$ . En fait, tout comme  ${}_3\phi$ , les fonctions  ${}_5\phi$  et  ${}_7\phi$  sont continûment dérivables. Par ailleurs, les graphes de ces fonctions permettent ici de le voir d'une façon beaucoup plus claire. Daubechies (1992, chapitre 7) étudie en détail la régularité des fonctions d'échelle  ${}_N\phi$ , pour  $N \geq 1$ , et démontre que lorsque  $N$  est grand,  ${}_N\phi \in C^{\mu N}(\mathbb{R})$  où  $\mu \simeq 0,2$ . Notons finalement que la forte asymétrie des fonctions d'échelle est clairement maintenue avec l'augmentation de  $N$ .

Finalement, nous terminons cette introduction aux ondelettes de Daubechies en mentionnant que les suites de filtres des fonctions d'échelle  ${}_2\phi$  à  ${}_{10}\phi$  sont tabulées par Daubechies (1988). Nous tenons également à mentionner que plusieurs autres classes de fonctions d'échelle (et d'ondelettes associées) ont été créées depuis la découverte de Daubechies (1988), notamment les *coiflets* dues à Beylkin *et al.* (1991) et Daubechies (1993), et les *symmlets* dues à Daubechies (1992, section 6.4) et initialement nommées ondelettes à support compact les moins asymétriques ("least asymmetric compactly supported wavelets").

Nous portons maintenant notre attention sur l'emploi des ondelettes dans le domaine de la statistique. D'abord, nous verrons à la section suivante comment exprimer une fonction de densité à l'aide des bases d'ondelettes. Nous nous concentrerons particulièrement sur la base de Haar, puisqu'elle jouera un rôle central dans la suite de nos travaux. Enfin, nous terminerons le présent chapitre en soulignant quelques travaux de statistique intéressants et centrés sur les ondelettes.

## 1.5. Décomposition d'une densité à l'aide des bases d'ondelettes

Notre but ici est de discuter de l'emploi des bases d'ondelettes pour représenter une fonction de densité  $f \in \mathcal{L}^2(\mathbb{R})$ . Nous porterons une attention toute spéciale à la base de Haar.

D'abord, nous savons que, de façon générale, si  $\{w_k\}_{k \in \mathbb{Z}}$  est une suite orthonormée de fonctions formant une base de  $\mathcal{L}^2(\mathbb{R})$ , alors toute fonction  $f \in \mathcal{L}^2(\mathbb{R})$  peut être développée en série selon

$$f(x) \stackrel{\mathcal{L}^2}{=} \sum_{k \in \mathbb{Z}} c_k w_k(x),$$

où les coefficients  $c_k$  sont donnés par

$$\begin{aligned} c_k &= \langle f, w_k \rangle \\ &= \int_{\mathbb{R}} f(x) w_k(x) dx, \end{aligned}$$

pour  $k \in \mathbb{Z}$ . Ce résultat est bien sûr valable quelque soit la base orthonormale employée. Voyons comment l'appliquer aux bases d'ondelettes. Pour plus de détails sur les bases orthonormales de fonctions, le lecteur peut consulter Rudin (1987, chapitres 4 et 5) et Kashin et Saakyan (1989).

Comme nous l'avons vu à la section 1.1, deux choix s'offrent à nous. En effet, deux types de bases orthonormales de  $\mathcal{L}^2(\mathbb{R})$  peuvent être construites à partir d'une fonction d'échelle  $\phi$  et de la mère des ondelettes  $\psi$  qui lui est associée. D'abord, l'équation (1.3) amène à une base orthonormée composée uniquement d'ondelettes, soit  $\{\psi_{jk}\}_{j,k \in \mathbb{Z}}$ . En utilisant cette base, toute fonction  $f \in \mathcal{L}^2(\mathbb{R})$  peut être exprimée comme

$$f(x) \stackrel{\mathcal{L}^2}{=} \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk}(x), \quad (1.32)$$

où les coefficients des ondelettes sont donnés par

$$d_{jk} = \langle f, \psi_{jk} \rangle = \int_{\mathbb{R}} f(x) \psi_{jk}(x) dx, \quad (1.33)$$

pour  $j, k \in \mathbb{Z}$ . Malheureusement,  $\psi$  ne peut habituellement pas être obtenue de façon explicite. C'est le cas, par exemple, pour  ${}_N\psi$ , la mère des ondelettes de Daubechies d'ordre  $N$ , lorsque  $N > 1$ . Dans ce cas, il devient évidemment alors plus difficile d'obtenir la valeur des coefficients des ondelettes. Des valeurs approximatives peuvent quand même être obtenues en adaptant les techniques

dues à Delyon et Juditsky (1995, 1997). Cependant, remarquons que pour la base de Haar (voir l'équation (1.21)), les ondelettes ont une expression très simple, ce qui permet d'obtenir les coefficients  $d_{jk}$  de façon explicite. Nous y reviendrons plus tard.

Une deuxième approche est possible pour développer une fonction en série d'ondelettes. En effet, d'après l'équation (1.4), une deuxième base orthonormée de  $\mathcal{L}^2(\mathbb{R})$  peut être construite à l'aide des ondelettes. Ainsi, quelque soit  $J \in \mathbb{Z}$ ,  $\{\phi_{Jk}, \psi_{jk}\}_{j \geq J, k \in \mathbb{Z}}$  est une base orthonormée de  $\mathcal{L}^2(\mathbb{R})$ , et donc, toute fonction  $f$ , définie sur  $\mathbb{R}$  et de carré sommable, peut s'écrire

$$f(x) \stackrel{\mathcal{L}^2}{=} \sum_{k \in \mathbb{Z}} c_{Jk} \phi_{Jk}(x) + \sum_{j \geq J} \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk}(x), \quad (1.34)$$

où les coefficients  $d_{jk}$  sont toujours donnés par l'équation (1.33), et où les coefficients  $c_{Jk}$  sont définis par

$$c_{Jk} = \langle f, \phi_{Jk} \rangle = \int_{\mathbb{R}} f(x) \phi_{Jk}(x) dx, \quad (1.35)$$

pour  $k \in \mathbb{Z}$ . Encore une fois, comme pour les coefficients des ondelettes, il peut arriver qu'il soit impossible d'obtenir les coefficients  $c_{Jk}$  sous une forme explicite. Ce n'est pas le cas avec la base de Haar, comme nous le verrons plus tard.

Remarquons que c'est cette deuxième approche qui est généralement employée dans les travaux de statistique portant sur les ondelettes. C'est aussi sur cette représentation que nous allons nous concentrer.

D'abord, notons  $f_j$  la projection orthogonale de  $f$  dans  $V_j$ . Nous avons alors, pour  $j \in \mathbb{Z}$ ,

$$f_j(x) = \sum_{k \in \mathbb{Z}} c_{jk} \phi_{jk}(x),$$

et cette approximation est en quelque sorte la meilleure approximation disponible de  $f$  au niveau  $j$ . Il est très intéressant de constater que puisque

$$V_{j+1} = V_j \oplus W_j,$$

nous pouvons écrire  $f_{j+1}$ , l'approximation de  $f$  au niveau  $j+1$ , comme

$$\begin{aligned} f_{j+1}(x) &= \sum_{k \in \mathbb{Z}} c_{jk} \phi_{jk}(x) + \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk}(x) \\ &= f_j(x) + \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk}(x). \end{aligned} \quad (1.36)$$



Ainsi,  $f_{j+1}$  peut être obtenue en ajoutant certains détails à  $f_j$ , l'approximation de  $f$  au niveau précédent.

Intuitivement, il est assez clair qu'en rajoutant des couches successives de détails, nous accédons à une approximation beaucoup plus fine de  $f$ . En fait, nous pouvons améliorer notre approximation de façon substantielle, et ce, jusqu'à s'approcher de  $f$  autant que nécessaire. Ceci est possible puisque la convergence en norme quadratique de  $f_j$ , soit

$$\|f_j - f\|_2 \xrightarrow{j \rightarrow \infty} 0,$$

nous est assurée par les propriétés (i) et (ii) des analyses multirésolutions (voir définition 1.1). Dans ce contexte, nous jugeons naturel d'employer  $f_j$  comme approximation de  $f$ . En effet, si la valeur de  $j$  est assez grande, l'approximation obtenue devrait être adéquate. Nous utiliserons donc ce type de projection orthogonale dans la suite de nos travaux. Pour plus de détails sur la convergence de  $f_j$  dans un contexte général, voir Mallat (1989a) et Alpert (1992).

Tournons maintenant notre attention vers une autre propriété des bases d'ondelettes qui a contribué à les rendre si attrayantes. Il existe des relations récursives permettant de relier entre eux les divers coefficients de niveaux successifs. En effet, en combinant les équations (1.6) et (1.35), il est possible de montrer (voir théorème B.1, annexe B) que

$$c_{jk} = \frac{1}{\sqrt{2}} \sum_{m \in \mathbb{Z}} a_{m-2k} c_{(j+1)m},$$

pour  $j, k \in \mathbb{Z}$ , et où la suite  $\{a_k\}_{k \in \mathbb{Z}}$  correspond à la suite de filtres de la fonction d'échelle associée à la base d'ondelettes employée. En combinant plutôt les équations (1.7) et (1.33), il est possible de trouver (voir théorème B.1, annexe B)

$$d_{jk} = \frac{1}{\sqrt{2}} \sum_{m \in \mathbb{Z}} (-1)^m a_{1+2k-m} c_{(j+1)m},$$

pour  $j, k \in \mathbb{Z}$ . Ainsi, les coefficients des translatés de la fonction d'échelle d'un niveau donné permettent de retrouver tous les coefficients du niveau précédent.

À l'inverse, il est aussi possible de regrouper tous les coefficients d'un niveau donné pour obtenir les coefficients des translatés de  $\phi$  au niveau suivant. L'équation (1.36) permet effectivement d'arriver à montrer (voir théorème B.2,

annexe B) que

$$c_{(j+1)k} = \frac{1}{\sqrt{2}} \left[ \sum_{m \in \mathbb{Z}} a_{k-2m} c_{jm} + (-1)^k \sum_{m \in \mathbb{Z}} a_{1+2m-k} d_{jm} \right],$$

pour  $j, k \in \mathbb{Z}$ . Ces trois résultats sont fort importants puisqu'ils permettent, dans les faits, de décomposer rapidement une fonction pour l'exprimer en un ensemble de coefficients, ou encore, de reconstruire une fonction à partir de l'ensemble de ses coefficients. Ceci peut être fait à l'aide de l'algorithme pyramidal dû à Mallat (1989b). Notons que cet algorithme a grandement contribué à populariser les ondelettes car il permet de les voir comme une alternative efficace à plusieurs techniques classiques. Plus particulièrement dans les domaines de la statistique et de la reconnaissance d'images, l'utilisation des ondelettes s'est répandue à une vitesse étonnante.

Nous arrivons maintenant au but premier de cette section, soit la décomposition d'une fonction de densité à l'aide d'une base d'ondelettes, et en particulier, à l'aide de la base de Haar présentée à la section 1.3. Tout d'abord, quelques soient la fonction de densité  $f \in \mathcal{L}^2(\mathbb{R})$  considérée et la base d'ondelettes employée, les équations (1.32) à (1.35) demeurent valides. Il est cependant intéressant de constater que les coefficients  $c_{jk}$  et  $d_{jk}$  prennent ici un sens particulier. En effet, comme  $f$  est une densité, alors pour  $j, k \in \mathbb{Z}$ ,

$$c_{jk} = \mathbb{E}[\phi_{jk}(X)],$$

et

$$d_{jk} = \mathbb{E}[\psi_{jk}(X)],$$

où les espérances sont prises sur une variable aléatoire  $X$  de densité  $f$ . Cette interprétation théorique des coefficients est extrêmement intéressante puisqu'elle ouvre la voie à l'estimation de la densité  $f$  à travers l'estimation de ses coefficients. En effet, si  $X_1, X_2, \dots, X_n$  est un échantillon tiré de la densité  $f$ , alors il est facile de montrer que

$$\hat{c}_{jk} = \frac{1}{n} \sum_{i=1}^n \phi_{jk}(X_i)$$

et

$$\hat{d}_{jk} = \frac{1}{n} \sum_{i=1}^n \psi_{jk}(X_i)$$

sont des estimateurs sans biais de  $c_{jk}$  et  $d_{jk}$  (pour  $j, k \in \mathbb{Z}$ ). Ce résultat rend l'emploi des ondelettes tout à fait naturel pour l'estimation des densités. Malheureusement, il ne permet pas, en général, d'obtenir plus facilement les valeurs des coefficients d'une densité  $f$  connue.

Limitons nous donc dès maintenant au cas de la base de Haar. Dans ce cas bien particulier, tous les coefficients peuvent être obtenus facilement à cause de la simplicité de la fonction d'échelle et de la mère des ondelettes. L'équation (1.15) nous autorise effectivement à écrire

$$\begin{aligned}
 c_{jk} &= \int_{\mathbb{R}} f(x) \phi_{jk}(x) dx \\
 &= 2^{j/2} \int_{\frac{k}{2^j}}^{\frac{k+1}{2^j}} f(x) dx \\
 &= 2^{j/2} \mathbb{P} \left( \frac{k}{2^j} \leq X < \frac{k+1}{2^j} \right) \\
 &= 2^{j/2} \left[ F \left( \frac{k+1}{2^j} \right) - F \left( \frac{k}{2^j} \right) \right], \tag{1.37}
 \end{aligned}$$

pour  $j, k \in \mathbb{Z}$ , et où  $F$  est la fonction de répartition associée à la densité  $f$ . Pour les coefficients des ondelettes, il suffit d'employer l'expression de  $\psi_{jk}$  donnée par l'équation (1.21). Nous obtenons ainsi

$$\begin{aligned}
 d_{jk} &= \int_{\mathbb{R}} f(x) \psi_{jk}(x) dx \\
 &= 2^{j/2} \left[ \int_{\frac{k}{2^j}}^{\frac{2k+1}{2^{j+1}}} f(x) dx - \int_{\frac{2k+1}{2^{j+1}}}^{\frac{k+1}{2^j}} f(x) dx \right] \\
 &= 2^{j/2} \left[ \mathbb{P} \left( \frac{k}{2^j} \leq X < \frac{2k+1}{2^{j+1}} \right) - \mathbb{P} \left( \frac{2k+1}{2^{j+1}} \leq X < \frac{k+1}{2^j} \right) \right] \\
 &= 2^{j/2} \left[ 2F \left( \frac{2k+1}{2^{j+1}} \right) - F \left( \frac{k}{2^j} \right) - F \left( \frac{k+1}{2^j} \right) \right],
 \end{aligned}$$

pour  $j, k \in \mathbb{Z}$ . Les divers coefficients d'une densité  $f$  connue deviennent donc très faciles à obtenir lorsque la base de Haar est employée. Par le fait même, il est donc aussi très simple et très rapide de calculer  $f_j$ , l'approximation de  $f$  au niveau  $j$ .

Finalement, il est intéressant de noter qu'avec la base de Haar, l'approximation  $f_j$  d'une densité continue converge non seulement en norme quadratique

selon

$$\|f_j - f\|_2 \xrightarrow{j \rightarrow \infty} 0, \quad (1.38)$$

comme dans le cas de toutes les bases d'ondelettes, mais aussi de façon ponctuelle. Ainsi, lorsque la densité  $f$  est continue, alors il est aussi possible d'écrire

$$f_j(x) \xrightarrow{j \rightarrow \infty} f(x), \quad (1.39)$$

quelque soit  $x \in \mathbb{R}$ . Pour plus de détails à ce sujet, voir Kashin et Saakyan (1989, section 3.2) et Walnut (2002, sections 5.3 et 5.5).

À la lumière de ces derniers résultats, nous proposons d'utiliser, à titre d'approximation d'une densité  $f \in \mathcal{L}^2(\mathbb{R})$  connue,

$$f_j(x) = \sum_{k \in \mathbb{Z}} c_{jk} \phi_{jk}(x), \quad (1.40)$$

où les coefficients  $c_{jk}$  sont donnés par l'équation (1.37) et où les fonctions  $\phi_{jk}$  sont données par l'équation (1.15). Nous avons donc en fait

$$f_j(x) = 2^{j/2} c_{jk} \quad \text{pour } x \in \left[ \frac{k}{2^j}, \frac{k+1}{2^j} \right), k \in \mathbb{Z},$$

où  $j$  est choisi de façon à rendre l'approximation adéquate. Notons que  $f_j$  est aussi une densité. Nous discuterons plus en détail du choix du niveau d'approximation  $j$  au cours des prochains chapitres, puisque ce type d'approximation est à la base de nos résultats.

## 1.6. Les ondelettes en statistique

Dans le but de terminer adéquatement cette introduction aux ondelettes, nous jugeons utile de citer quelques travaux intéressants du domaine de la statistique faisant appel à la théorie des ondelettes.

Nous mentionnons d'abord les travaux de David Donoho et de ses divers collaborateurs. Il est important de souligner le rôle primordial qu'ont joué ces travaux. En effet, Donoho (1995), Donoho et Johnstone (1994, 1995, 1998), Donoho, Johnstone, Kerkyacharian et Picard (1995, 1996) et Kerkyacharian et Picard (1992, 1993) 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 est, sans aucun doute, la mise au point des techniques de rétrécissement ("shrinkage") et de seuillage ("thresholding"), aujourd'hui largement répandues. Ces

techniques constituent, en quelque sorte, les fondements de l'utilisation des ondelettes en statistique, en plus de soulever certaines questions qui n'ont trouvé de réponses qu'un peu plus tard.

D'abord, dans le contexte de la régression, Delyon et Juditsky (1995, 1997) discutent des façons adéquates d'obtenir les coefficients des ondelettes. Fan *et al.* (1999) discutent plutôt du bien fondé de l'utilisation des ondelettes et des techniques de seuillage. Tao (1996) et Tao et Vidakovic (1998) s'intéressent à la convergence des estimateurs obtenus par seuillage. Hall et Patil (1996a, 1996b) et Marron *et al.* (1998) comparent différentes techniques de seuillage à travers divers critères d'optimalité. Nason (1995, 1996) s'intéresse au choix des seuils par validation croisée. Ruggeri et Vidakovic (1999) et Ruggeri (1999) développent, à l'aide d'arguments empruntés à la théorie de la décision, des critères bayésiens permettant d'obtenir un seuil optimal. McCoy et Walden (1996), Wang (1996) et Abry *et al.* (1998) discutent de l'emploi des méthodes de seuillage dans le contexte plus particulier des séries chronologiques.

Certains auteurs préfèrent voir les techniques de seuillage d'une façon différente. Par exemple, Abramovich et Benjamini (1995, 1996) les présentent comme étant des tests d'hypothèses multiples. Vidakovic (1998) fait de même dans une approche bayésienne à la régression. Ogden et Parzen (1996) les présentent plutôt comme des tests du rapport de vraisemblance séquentiels. Zhang (1995) et Antoniadis *et al.* (1997) les considèrent similaires à un problème d'identification des meilleurs régresseurs.

Chipman *et al.* (1997), Abramovich *et al.* (1998), Clyde *et al.* (1998), Abramovich et Sapatinas (1999), Clyde et George (1999) et Holmes et Denison (1999) adoptent une approche radicalement différente. Ils construisent des modèles bayésiens où l'estimation des coefficients des ondelettes amène du rétrécissement et/ou du seuillage de façon naturelle. Dans ce même contexte, Chipman et Wolfson (1999) et Vannucci et Corradi (1999) discutent plus particulièrement du choix des densités *a priori*. Müller et Vidakovic (1998) adoptent également cette approche mais concentrent leurs efforts sur le seuillage dans le cadre d'un problème d'estimation de densité. Cette dernière technique est généralisée à divers types de problèmes par Müller et Vidakovic (1999).

Toujours dans un contexte d'estimation de densité, Vannucci et Vidakovic (1995) et Pinheiro et Vidakovic (1997) traitent de différents problèmes rencontrés avec l'emploi des bases d'ondelettes. Huang (1999) traite du biais et de la variance des estimateurs linéaires basés sur les ondelettes. Tribouley (1995a, 1995b) propose le choix des seuils par validation croisée.

L'impact de ces divers travaux portant sur la régression et l'estimation de densité est tel que les ondelettes sont maintenant vues par plusieurs comme un outil offrant un très grand nombre de possibilités, pour la plupart encore inexplorées. Entre autres, les problèmes reliés de près ou de loin à la régression et à l'estimation de densité peuvent être abordés sous un regard nouveau à l'aide des ondelettes. Par exemple, Fan (1996) construit des tests d'ajustements et Huang (1997) et Pensky (1999) étudient des problèmes d'estimation bayésienne empirique. Dans le contexte des séries chronologiques, Li et Xie (1997) s'intéressent à la détection de périodicité et Wang (1995), à la détection des sauts et changements brusques. Aussi, Neumann (1996), Priestley (1996) et Gao (1997) utilisent les ondelettes dans le but d'estimer des densités spectrales. Dans un tout autre ordre d'idée, Vidakovic (1996) explique comment construire, à l'aide des ondelettes et de façon aléatoire, des densités permettant d'évaluer le comportement de plusieurs types d'estimateurs.

Finalement, Ogden (1997), Härdle *et al.* (1998) et Vidakovic (1999) sont des ouvrages plus approfondis sur l'emploi des ondelettes dans le domaine de la statistique. Nous attirons particulièrement l'attention du lecteur sur le dernier ouvrage mentionné. À notre avis, il constitue à l'heure actuelle le répertoire le plus complet des méthodes statistiques basées sur les ondelettes.

## Chapitre 2

---

### Estimation bayésienne d'un paramètre de position à l'aide de la base de Haar : cas d'une seule observation

Nous nous intéressons, tout au long de ce chapitre, au cas du modèle simple dont nous avons discuté précédemment, soit

$$\begin{aligned} X|\theta &\sim L(\theta|x) = \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right), \\ \theta &\sim \pi(\theta), \end{aligned}$$

où  $f \in \mathcal{L}^2(\mathbb{R})$  est spécifiée, et où le paramètre d'échelle  $\sigma > 0$  est connu. Nous notons ici  $L(\theta|x)$  la densité conditionnelle de l'observation  $X$  pour accentuer le fait que cette densité correspond à la vraisemblance du paramètre de position  $\theta$ . Notre intérêt est tourné vers  $\theta$  et la question à laquelle nous nous proposons de répondre à travers ce chapitre est la suivante : si la densité *a priori* est complètement spécifiée, comment estimer  $\theta$  ?

#### 2.1. Estimation de $\theta$

D'abord, dans le but d'obtenir un estimateur bayésien de  $\theta$ , nous supposons également que  $\pi$  appartient aussi à une famille de paramètres de position et d'échelle. Plus particulièrement, nous admettons que

$$\pi(\theta) = \frac{1}{\tau} g\left(\frac{\theta - \mu}{\tau}\right),$$

où  $\mu \in \mathbb{R}$  et  $\tau > 0$  sont connus, et  $g \in \mathcal{L}^2(\mathbb{R})$  est spécifiée.

Dans ce contexte, nous pouvons exprimer la densité marginale de  $X$ , notée  $m$ , comme

$$\begin{aligned} m(x) &= \int_{\mathbb{R}} L(\theta|x)\pi(\theta) d\theta \\ &= \frac{1}{\sigma\tau} \int_{\mathbb{R}} f\left(\frac{x-\theta}{\sigma}\right) g\left(\frac{\theta-\mu}{\tau}\right) d\theta, \end{aligned}$$

ce qui permet d'écrire la densité *a posteriori* de  $\theta$  comme

$$\begin{aligned} \pi(\theta|x) &= \frac{L(\theta|x)\pi(\theta)}{m(x)} \\ &= \frac{f\left(\frac{x-\theta}{\sigma}\right) g\left(\frac{\theta-\mu}{\tau}\right)}{\int_{\mathbb{R}} f\left(\frac{x-\theta}{\sigma}\right) g\left(\frac{\theta-\mu}{\tau}\right) d\theta}. \end{aligned}$$

Ainsi, les moments *a posteriori* de  $\theta$  sont donnés, pour  $n > 0$ , par

$$\mathbb{E}[\theta^n|x] = \frac{\int_{\mathbb{R}} \theta^n f\left(\frac{x-\theta}{\sigma}\right) g\left(\frac{\theta-\mu}{\tau}\right) d\theta}{\int_{\mathbb{R}} f\left(\frac{x-\theta}{\sigma}\right) g\left(\frac{\theta-\mu}{\tau}\right) d\theta},$$

avec l'intéressante propriété que, pour  $n \geq 0$ ,

$$\int_{\mathbb{R}} \theta^n f\left(\frac{x-\theta}{\sigma}\right) g\left(\frac{\theta-\mu}{\tau}\right) d\theta = \sum_{l=0}^n \binom{n}{l} \mu^{n-l} \tau^{l+1} \mathbb{I}_l(x), \quad (2.1)$$

où

$$\mathbb{I}_l(x) = \int_{\mathbb{R}} s^l g(s) f\left(\frac{\tau}{\sigma} \left[\frac{x-\mu}{\tau} - s\right]\right) ds. \quad (2.2)$$

Ce résultat est obtenu à l'aide du changement de variable simple

$$\theta = \tau s + \mu.$$

Pour plus de détails, le lecteur peut consulter Leblanc (1995), Leblanc et Angers (1995), et Angers (1996a).

Maintenant, pour obtenir une estimation bayésienne de  $\theta$ , l'approche usuelle consiste à considérer l'estimateur de Bayes sous la perte quadratique, noté  $\hat{\theta}_B(x)$ , soit l'espérance *a posteriori* de  $\theta$  (voir Berger, 1985, section 4.4). Il est aussi d'usage de considérer la perte espérée *a posteriori* de l'estimateur de Bayes, notée  $\rho(\pi, \hat{\theta}_B|x)$ , comme mesure de précision de cet estimateur. Notons que sous la



perte quadratique, il est facile de montrer que  $\rho(\pi, \hat{\theta}_B|x)$  se réduit à la variance *a posteriori* du paramètre  $\theta$  (voir Berger, 1985, section 4.3).

En employant la notation introduite précédemment, il est d'abord possible d'écrire

$$m(x) = \frac{1}{\sigma} \mathbb{I}_0(x). \quad (2.3)$$

De là, les deux quantités qui monopolisent ici notre attention peuvent à leur tour s'exprimer (voir Angers, 1996a) comme

$$\hat{\theta}_B(x) = \mu + \tau \frac{\mathbb{I}_1(x)}{\mathbb{I}_0(x)}, \quad (2.4)$$

et

$$\rho(\pi, \hat{\theta}_B|x) = \tau^2 \left[ \frac{\mathbb{I}_2(x)}{\mathbb{I}_0(x)} - \left( \frac{\mathbb{I}_1(x)}{\mathbb{I}_0(x)} \right)^2 \right]. \quad (2.5)$$

Malheureusement, à moins de se restreindre au cas où la loi *a priori* est conjuguée à la vraisemblance, il est habituellement difficile, voire impossible, d'obtenir une expression simple pour les fonctions  $\mathbb{I}_l(x)$ . Ce problème se répercute clairement sur  $\hat{\theta}_B(x)$ ,  $\rho(\pi, \hat{\theta}_B|x)$  et même  $m(x)$ , qu'il devient alors aussi difficile (ou impossible) d'exprimer sous une forme explicite. Il est alors nécessaire d'employer diverses techniques, telles que l'approximation gaussienne de Naylor et Smith (1982), l'approximation laplacienne de Tierney et Kadane (1986) ou l'échantillonneur de Gibbs (voir Gelfand et Smith, 1990) pour obtenir des approximations de  $\hat{\theta}_B(x)$  et  $\rho(\pi, \hat{\theta}_B|x)$ .

Afin d'éviter d'avoir à traiter chaque problème de façon particulière et d'être éventuellement confronté à toutes les difficultés techniques que chaque modèle spécifique peut apporter, nous proposons une méthode générale d'approximation des intégrales  $\mathbb{I}_l(x)$  à l'aide de la base de Haar. Cette méthode, d'une étonnante simplicité, est présentée en détail à la prochaine section. Elle nous permettra plus tard d'approximer l'estimateur de Bayes et sa perte espérée *a posteriori*.

Finalement, remarquons tout de même que plusieurs se sont attardés à la recherche d'une forme simple pour l'estimateur de Bayes dans le cas de différents modèles bayésiens. Notons, en plus des travaux de Stein (1981), ceux de Spiegelhalter (1985), Fan et Berger (1990) et Angers (1996a, 1996b, 1997).

## 2.2. Approximation de $\mathbb{I}_l(x)$

La première étape, dans le développement de notre technique d'approximation, consiste à construire une approximation de  $\mathbb{I}_l(x)$  (pour  $l = 0, 1, 2$ ). Pour cela, nous employons la base de Haar et les approximations obtenues à l'aide des projections orthogonales vues à la section 1.5. Les approximations de  $\mathbb{I}_l(x)$  ainsi construites, que nous notons  $\tilde{\mathbb{I}}_l(x)$ , serviront plus tard à approximer  $\hat{\theta}_B(x)$  et  $\rho(\pi, \hat{\theta}_B|x)$ .

Nous débutons la tâche qui nous préoccupe ici en définissant, pour alléger la notation,

$$f_x(s) = f\left(\frac{\tau}{\sigma} \left[ \frac{x - \mu}{\tau} - s \right]\right),$$

et

$$g_l(s) = s^l g(s),$$

pour  $l \in \mathbb{N}$ . Remarquons qu'en employant cette notation, l'équation (2.2) devient

$$\mathbb{I}_l(x) = \int_{\mathbb{R}} g_l(s) f_x(s) ds. \quad (2.6)$$

Toujours par souci de clarté, nous notons l'intervalle

$$A_{jk} = \left[ \frac{k}{2^j}, \frac{k+1}{2^j} \right),$$

pour  $j, k \in \mathbb{Z}$ . Ainsi, d'après l'équation (1.40), nous pouvons approximer, en utilisant les ondelettes de Haar, toute densité  $h \in \mathcal{L}^2(\mathbb{R})$  associée à une variable aléatoire  $Y$  par

$$\tilde{h}(y) = \sum_{k \in \mathbb{Z}} c_{Jk} \phi_{Jk}(y), \quad (2.7)$$

où, pour  $k \in \mathbb{Z}$ ,

$$\phi_{Jk}(y) = \begin{cases} 2^{J/2} & \text{si } y \in A_{Jk}, \\ 0 & \text{sinon,} \end{cases}$$

et

$$c_{Jk} = \langle h, \phi_{Jk} \rangle = 2^{J/2} \mathbb{P}(Y \in A_{Jk}),$$

et où la valeur de  $J \in \mathbb{Z}$  est choisie assez grande. L'impact du choix de  $J$  sur le comportement des approximations de l'estimateur de Bayes et de sa perte espérée *a posteriori* construites à l'aide de l'approximation donnée par l'équation (2.7)

sera établi au corollaire 2.1. À la section 2.7, nous discuterons aussi, à travers quelques exemples, de l'effet du choix de  $J$  du point de vue pratique.

Le résultat qui suit nous permet dès maintenant d'établir l'approximation de  $\mathbb{I}_l(x)$  obtenue à l'aide de la base de Haar.

**Résultat 2.1.** *Si  $f \in \mathcal{L}^2(\mathbb{R})$ , si*

$$a_{jk}(x) = \langle f_x, \phi_{jk} \rangle \quad \forall j, k \in \mathbb{Z},$$

*et si*

$$b_{l,jk} = \langle g_l, \phi_{jk} \rangle \quad \forall j, k \in \mathbb{Z},$$

*alors l'approximation de  $\mathbb{I}_l(x)$  obtenue à partir de l'approximation de Haar de niveau  $J$  de  $f_x$  est donnée par*

$$\begin{aligned} \tilde{\mathbb{I}}_l(x) &= \int_{\mathbb{R}} g_l(s) \tilde{f}_x(s) ds \\ &= \sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{l,Jk}. \end{aligned}$$

**Démonstration.** D'abord, l'équation (2.7) nous permet d'approximer  $f_x$  par

$$\tilde{f}_x(s) = \sum_{k \in \mathbb{Z}} a_{Jk}(x) \phi_{Jk}(s), \quad (2.8)$$

si la valeur de  $J$  est suffisamment grande. Il nous suffit maintenant de remplacer  $f_x$  par son approximation dans l'équation (2.6). En effet, nous avons alors

$$\begin{aligned} \tilde{\mathbb{I}}_l(x) &= \int_{\mathbb{R}} g_l(s) \tilde{f}_x(s) ds \\ &= \int_{\mathbb{R}} g_l(s) \left( \sum_{k \in \mathbb{Z}} a_{Jk}(x) \phi_{Jk}(s) \right) ds \\ &= \sum_{k \in \mathbb{Z}} a_{Jk}(x) \int_{\mathbb{R}} g_l(s) \phi_{Jk}(s) ds \\ &= \sum_{k \in \mathbb{Z}} a_{Jk}(x) \langle g_l, \phi_{Jk} \rangle \\ &= \sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{l,Jk}, \end{aligned}$$

et cette approximation devrait être adéquate si la valeur de  $J$  est choisie avec soin.  $\square$

Attardons nous maintenant au calcul des divers coefficients nécessaires afin d'obtenir les approximations de  $\mathbb{I}_l(x)$  pour  $l = 0, 1, 2$ . D'abord, les coefficients  $a_{jk}(x)$  doivent être trouvés puisqu'ils interviennent dans le calcul de  $\tilde{\mathbb{I}}_l(x)$  pour toutes les valeurs de  $l \in \mathbb{N}$ . Ces coefficients sont obtenus à partir de l'équation (1.37), qu'il suffit de modifier légèrement. Nous avons en effet, pour  $j, k \in \mathbb{Z}$ ,

$$\begin{aligned} a_{jk}(x) &= \langle f_x, \phi_{jk} \rangle \\ &= 2^{j/2} \frac{\sigma}{\tau} \left[ F \left( \frac{\tau}{\sigma} \left[ \frac{x - \mu}{\tau} - \frac{k}{2^j} \right] \right) - F \left( \frac{\tau}{\sigma} \left[ \frac{x - \mu}{\tau} - \frac{(k+1)}{2^j} \right] \right) \right], \end{aligned} \quad (2.9)$$

où  $F$  est la fonction de répartition associée à la densité  $f$ .

Les coefficients  $b_{l,jk}$  doivent aussi être obtenus pour  $l = 0, 1, 2$ . Dans le cas où  $l = 0$ , l'équation (1.37) amène directement

$$\begin{aligned} b_{0,jk} &= \langle g_0, \phi_{jk} \rangle \\ &= \langle g, \phi_{jk} \rangle \\ &= 2^{j/2} \left[ G \left( \frac{k+1}{2^j} \right) - G \left( \frac{k}{2^j} \right) \right], \end{aligned} \quad (2.10)$$

pour  $j, k \in \mathbb{Z}$ , où  $G$  est la fonction de répartition associée à la densité  $g$ . Notons que ces coefficients peuvent s'interpréter comme

$$b_{0,jk} = 2^{j/2} \mathbb{P}_g(A_{jk}), \quad (2.11)$$

où  $\mathbb{P}_g(A_{jk})$  correspond à la probabilité associée à l'intervalle  $A_{jk}$  par la densité  $g$  et donc, dans le cas présent, à la probabilité *a priori* associée à l'intervalle  $A_{jk}$ . L'obtention des coefficients  $b_{1,jk}$  et  $b_{2,jk}$  demande un peu plus de travail. La proposition suivante donne la forme de ces coefficients.

**Proposition 2.1.** *Les coefficients  $b_{l,jk}$ , tels que définis par le résultat 2.1, sont donnés, pour  $l = 1$  et 2, par*

$$b_{l,jk} = \left( \frac{k}{2^j} \right)^l b_{0,jk} + e_{l,jk} \quad \forall j, k \in \mathbb{Z},$$

où

$$e_{l,jk} = \frac{1}{2^{j/2}} \left( \frac{2k+1}{2^j} \right)^{l-1} G \left( \frac{k+1}{2^j} \right) - 2^{j/2} \int_{A_{jk}} (2s)^{l-1} G(s) ds.$$

**Démonstration.** Tout d'abord, quelque soit la valeur de  $l \in \mathbb{N}$ , nous pouvons écrire

$$\begin{aligned} b_{l,jk} &= \langle g_l, \phi_{jk} \rangle \\ &= \int_{\mathbb{R}} g_l(s) \phi_{jk}(s) ds \\ &= 2^{j/2} \int_{A_{jk}} s^l g(s) ds. \end{aligned}$$

Dans le cas particulier où  $l = 1$ , l'intégration par parties amène

$$\begin{aligned} b_{1,jk} &= 2^{j/2} \int_{A_{jk}} s g(s) ds \\ &= 2^{j/2} \left[ \frac{(k+1)}{2^j} G\left(\frac{k+1}{2^j}\right) - \frac{k}{2^j} G\left(\frac{k}{2^j}\right) - \int_{A_{jk}} G(s) ds \right]. \end{aligned}$$

En employant la définition de  $b_{0,jk}$  donnée par l'équation (2.10), il est donc possible d'écrire

$$\begin{aligned} b_{1,jk} &= \frac{k}{2^{j/2}} \left[ G\left(\frac{k+1}{2^j}\right) - G\left(\frac{k}{2^j}\right) \right] + \frac{1}{2^{j/2}} G\left(\frac{k+1}{2^j}\right) - 2^{j/2} \int_{A_{jk}} G(s) ds \\ &= \frac{k}{2^j} b_{0,jk} + e_{1,jk}, \end{aligned}$$

pour  $j, k \in \mathbb{Z}$ .

En procédant de la même façon, nous avons, pour  $l = 2$ ,

$$\begin{aligned} b_{2,jk} &= 2^{j/2} \int_{A_{jk}} s^2 g(s) ds \\ &= 2^{j/2} \left[ \frac{(k+1)^2}{2^{2j}} G\left(\frac{k+1}{2^j}\right) - \frac{k^2}{2^{2j}} G\left(\frac{k}{2^j}\right) - \int_{A_{jk}} 2s G(s) ds \right], \end{aligned}$$

et, toujours d'après l'équation (2.10),

$$\begin{aligned} b_{2,jk} &= \frac{k^2}{2^{3j/2}} \left[ G\left(\frac{k+1}{2^j}\right) - G\left(\frac{k}{2^j}\right) \right] + \frac{2k+1}{2^{3j/2}} G\left(\frac{k+1}{2^j}\right) - 2^{j/2} \int_{A_{jk}} 2s G(s) ds \\ &= \frac{k^2}{2^{2j}} b_{0,jk} + e_{2,jk}, \end{aligned}$$

pour  $j, k \in \mathbb{Z}$ . □

Notons que le comportement de  $b_{1,jk}$  et  $b_{2,jk}$  (et par le fait même, de  $e_{1,jk}$  et  $e_{2,jk}$ ) sera étudié aux lemmes 2.3 et 2.4. Nous verrons alors aussi pourquoi

leur valeur exacte ne sera en fait jamais calculée. Pour l'instant, nous nous intéressons plutôt à la qualité de l'approximation introduite au résultat 2.1. Plus particulièrement, nous voulons examiner l'erreur d'approximation liée à  $\tilde{\mathbb{I}}_l(x)$ .

Auparavant, nous obtenons un lemme permettant de quantifier la qualité de l'approximation de  $f_x(s)$  donnée par  $\tilde{f}_x(s)$ . Ce résultat sera employé afin d'étudier les erreurs d'approximation associées à  $\tilde{\mathbb{I}}_l(x)$ , pour  $l \in \mathbb{N}$ .

**Lemme 2.1.** *Si  $f \in \mathcal{L}^2(\mathbb{R}) \cap C^1(\mathbb{R})$  et si  $f' \in \mathcal{L}^\infty(\mathbb{R})$ , alors*

$$\left| f_x(s) - \tilde{f}_x(s) \right| \leq \frac{\tau}{\sigma 2^J} \|f'\|_\infty \quad \forall x, s \in \mathbb{R},$$

où  $\|f'\|_\infty = \sup_{s \in \mathbb{R}} |f'(s)|$ , et donc

$$\left| f_x(s) - \tilde{f}_x(s) \right| = O(2^{-J}),$$

et ce résultat est uniforme en  $x$  et en  $s$ .

**Démonstration.** D'abord, remarquons que pour tout  $s \in \mathbb{R}$ , il existe un et un seul  $k_s \in \mathbb{Z}$  tel que

$$s \in A_{Jk_s} = \left[ \frac{k_s}{2^J}, \frac{k_s + 1}{2^J} \right).$$

Il est ainsi possible d'écrire, d'après les équations (2.8) et (2.9),

$$\begin{aligned} \tilde{f}_x(s) &= \sum_{k \in \mathbb{Z}} a_{Jk}(x) \phi_{Jk}(s) \\ &= 2^{J/2} a_{Jk_s}(x) \\ &= 2^J \int_{A_{Jk_s}} f_x(t) dt. \end{aligned}$$

De là, nous avons

$$\begin{aligned} \left| f_x(s) - \tilde{f}_x(s) \right| &= \left| f_x(s) - 2^J \int_{A_{Jk_s}} f_x(t) dt \right| \\ &= 2^J \left| \int_{A_{Jk_s}} (f_x(s) - f_x(t)) dt \right| \\ &\leq 2^J \int_{A_{Jk_s}} |f_x(s) - f_x(t)| dt. \end{aligned} \tag{2.12}$$

Maintenant, pour toute fonction  $p \in C^1(\mathbb{R})$ , un développement en série de Taylor permet d'écrire

$$p(y+h) = p(y) + hp'(\varepsilon),$$

où  $\varepsilon \in (y, y+h)$  si  $h > 0$ , et  $\varepsilon \in (y+h, y)$  si  $h < 0$ . Ainsi, pour  $h \in \mathbb{R}$ , nous avons

$$\begin{aligned} |p(y+h) - p(y)| &= |h| |p'(\varepsilon)| \\ &\leq |h| \|p'\|_\infty, \end{aligned}$$

si  $p' \in \mathcal{L}^\infty(\mathbb{R})$ . Comme ici

$$\frac{\partial f_x}{\partial s}(s) = -\frac{\tau}{\sigma} f' \left( \frac{\tau}{\sigma} \left[ \frac{x-\mu}{\tau} - s \right] \right),$$

alors ce résultat amène

$$\begin{aligned} |f_x(s) - f_x(t)| &\leq |s-t| \frac{\tau}{\sigma} \|f'\|_\infty \\ &\leq \frac{\tau}{\sigma 2^J} \|f'\|_\infty, \end{aligned}$$

pour tout  $t \in A_{Jk_s}$ .

Finalement, en employant cette inégalité avec l'équation (2.12), nous pouvons borner l'erreur d'approximation associée à  $\tilde{f}_x(s)$  par

$$\begin{aligned} |f_x(s) - \tilde{f}_x(s)| &\leq \frac{\tau}{\sigma} \|f'\|_\infty \int_{A_{Jk_s}} dt \\ &= \frac{\tau}{\sigma 2^J} \|f'\|_\infty, \end{aligned}$$

pour toutes les valeurs de  $x, s \in \mathbb{R}$ . □

Il nous est enfin possible d'étudier l'erreur faite lors de l'approximation de  $\mathbb{I}_l(x)$  par  $\tilde{\mathbb{I}}_l(x)$ . Le prochain théorème présente notre résultat général.

**Théorème 2.1.** *Si  $f \in \mathcal{L}^2(\mathbb{R}) \cap C^1(\mathbb{R})$ , si  $f' \in \mathcal{L}^\infty(\mathbb{R})$  et si  $g_l \in \mathcal{L}^1(\mathbb{R})$ , alors l'erreur d'approximation liée à  $\tilde{\mathbb{I}}_l(x)$  est bornée, pour  $l \in \mathbb{N}$ , par*

$$\left| \mathbb{I}_l(x) - \tilde{\mathbb{I}}_l(x) \right| \leq \frac{\tau}{\sigma 2^J} \|f'\|_\infty \|g_l\|_1 \quad \forall x \in \mathbb{R}.$$

Ainsi, nous avons

$$\left| \mathbb{I}_l(x) - \tilde{\mathbb{I}}_l(x) \right| = O(2^{-J}),$$

quelque soit la valeur de  $l \in \mathbb{N}$ , ce résultat étant uniforme en  $x$ .

**Démonstration.** Pour obtenir le résultat, il suffit d'abord de remarquer qu'en employant l'équation (2.6) et le résultat 2.1, nous avons

$$\begin{aligned} \left| \mathbb{I}_l(x) - \tilde{\mathbb{I}}_l(x) \right| &= \left| \int_{\mathbb{R}} g_l(s) f_x(s) ds - \int_{\mathbb{R}} g_l(s) \tilde{f}_x(s) ds \right| \\ &= \left| \int_{\mathbb{R}} g_l(s) (f_x(s) - \tilde{f}_x(s)) ds \right| \\ &\leq \int_{\mathbb{R}} |g_l(s)| |f_x(s) - \tilde{f}_x(s)| ds. \end{aligned}$$

Par ailleurs, d'après le lemme 2.1,

$$\left| f_x(s) - \tilde{f}_x(s) \right| \leq \frac{\tau}{\sigma 2^J} \|f'\|_{\infty},$$

pour  $x, s \in \mathbb{R}$ . L'erreur d'approximation que nous cherchions à caractériser ici satisfait donc

$$\begin{aligned} \left| \mathbb{I}_l(x) - \tilde{\mathbb{I}}_l(x) \right| &\leq \frac{\tau}{\sigma 2^J} \|f'\|_{\infty} \int_{\mathbb{R}} |g_l(s)| ds \\ &= \frac{\tau}{\sigma 2^J} \|f'\|_{\infty} \|g_l\|_1, \end{aligned}$$

pour tout  $x \in \mathbb{R}$ . □

La première étape du développement de notre technique d'approximation est maintenant complétée. Nous avons en effet réussi à construire des approximations intéressantes des intégrales  $\mathbb{I}_l(x)$ , soit  $\tilde{\mathbb{I}}_l(x)$ . Il ne nous reste plus qu'à étudier l'emploi de ces approximations dans le but d'arriver à approximer efficacement l'estimateur de Bayes et sa perte espérée *a posteriori*. Les autres étapes nécessaires à la construction de nos approximations sont effectuées aux sections suivantes.

Auparavant, nous faisons remarquer au lecteur que pour le cas où  $l = 0$ , il est possible de faire mieux qu'au théorème précédent. Pour cela, nous obtenons d'abord un lemme donnant des expressions alternatives des coefficients  $a_{jk}(x)$  et  $b_{0,jk}$ . Ces résultats seront employés afin de quantifier diverses erreurs d'approximation.

**Lemme 2.2.** *Si  $f \in C^1(\mathbb{R})$ , alors les coefficients  $a_{jk}(x)$ , tels que définis par l'équation (2.9), peuvent aussi s'exprimer comme*

$$a_{jk}(x) = \frac{1}{2^{j/2}} \left[ f_x \left( \frac{k}{2^j} \right) - \frac{\tau}{\sigma 2^{j+1}} f'(x, \varepsilon_{1jk}) \right] \quad \forall j, k \in \mathbb{Z},$$



où  $\varepsilon_{1jk} \in A_{jk}$ , et où la fonction  $f'(x, s)$  est définie par

$$f'(x, s) = f' \left( \frac{\tau}{\sigma} \left[ \frac{x - \mu}{\tau} - s \right] \right).$$

De même, si  $g \in C^1(\mathbb{R})$ , alors les coefficients  $b_{0,jk}$ , donnés par l'équation (2.10), peuvent être réécrits comme

$$b_{0,jk} = \frac{1}{2^{j/2}} \left[ g \left( \frac{k}{2^j} \right) + \frac{1}{2^{j+1}} g'(\varepsilon_{2jk}) \right] \quad \forall j, k \in \mathbb{Z},$$

où  $\varepsilon_{2jk} \in A_{jk}$ .

**Démonstration.** D'abord, il est bien connu que toute fonction  $p \in C^2(\mathbb{R})$  peut être développée en série de Taylor selon

$$p(y + h) = p(y) + hp'(y) + \frac{h^2}{2} p''(\varepsilon),$$

où  $h > 0$  et  $\varepsilon \in (y, y + h)$ . Cette relation permet aussi d'écrire

$$\frac{p(y + h) - p(y)}{h} = p'(y) + \frac{h}{2} p''(\varepsilon), \quad (2.13)$$

cette dernière forme nous étant plus utile.

Maintenant, notons que comme  $f \in C^1(\mathbb{R})$ , alors  $F \in C^2(\mathbb{R})$ , et posons

$$F_x(s) = F \left( \frac{\tau}{\sigma} \left[ \frac{x - \mu}{\tau} - s \right] \right).$$

Remarquons ensuite que

$$\frac{\partial F_x}{\partial s}(s) = -\frac{\tau}{\sigma} f_x(s),$$

et que

$$\frac{\partial^2 F_x}{\partial s^2}(s) = \frac{\tau^2}{\sigma^2} f'(x, s).$$

Remarquons finalement que l'équation (2.9) s'écrit, avec cette nouvelle notation, comme

$$a_{jk}(x) = 2^{j/2} \frac{\sigma}{\tau} \left[ F_x \left( \frac{k}{2^j} \right) - F_x \left( \frac{k+1}{2^j} \right) \right].$$

En employant l'équation (2.13), nous pouvons donc écrire

$$\begin{aligned} a_{jk}(x) &= -\frac{1}{2^{j/2}} \frac{\sigma}{\tau} \left[ \frac{F_x\left(\frac{k+1}{2^j}\right) - F_x\left(\frac{k}{2^j}\right)}{1/2^j} \right] \\ &= \frac{1}{2^{j/2}} \left[ f_x\left(\frac{k}{2^j}\right) - \frac{\tau}{\sigma 2^{j+1}} f'(x, \varepsilon_{1jk}) \right], \end{aligned}$$

où  $\varepsilon_{1jk} \in A_{jk}$ , pour  $j, k \in \mathbb{Z}$ .

Pour les coefficients  $b_{0,jk}$ , nous avons directement

$$\begin{aligned} b_{0,jk} &= 2^{j/2} \left[ G\left(\frac{k+1}{2^j}\right) - G\left(\frac{k}{2^j}\right) \right] \\ &= \frac{1}{2^{j/2}} \left[ \frac{G\left(\frac{k+1}{2^j}\right) - G\left(\frac{k}{2^j}\right)}{1/2^j} \right] \\ &= \frac{1}{2^{j/2}} \left[ g\left(\frac{k}{2^j}\right) + \frac{1}{2^{j+1}} g'(\varepsilon_{2jk}) \right], \end{aligned}$$

où  $\varepsilon_{2jk} \in A_{jk}$ , et pour  $j, k \in \mathbb{Z}$ . □

Comme nous l'avons dit plus haut, le résultat qui suit n'est pas directement relié au développement de la technique d'approximation qui nous intéresse ici. L'importance de ce résultat sera mise en lumière au chapitre 4.

**Théorème 2.2.** *Si  $f \in \mathcal{L}^2(\mathbb{R}) \cap C^2(\mathbb{R})$ ,  $g \in C^2(\mathbb{R})$ , et si, pour  $i = 1$  et  $2$ ,  $f^{(i)} \in \mathcal{L}^\infty(\mathbb{R})$  et  $g^{(i)} \in \mathcal{L}^1(\mathbb{R})$ , alors*

$$\left| \mathbb{I}_0(x) - \tilde{\mathbb{I}}_0(x) \right| = O(2^{-2J}) \quad \forall x \in \mathbb{R},$$

et ce résultat est uniforme en  $x$ .

**Démonstration.** D'abord, nous savons que

$$\left| \mathbb{I}_0(x) - \tilde{\mathbb{I}}_0(x) \right| = \left| \mathbb{I}_0(x) - \sum_{k \in \mathbb{Z}} a_{jk}(x) b_{0,jk} \right|.$$

Par ailleurs, comme  $f$  et  $g$  sont continûment dérivables, le lemme 2.2 amène

$$\begin{aligned} \sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk} &= \frac{1}{2^J} \sum_{k \in \mathbb{Z}} \left[ f_x \left( \frac{k}{2^J} \right) - \frac{\tau}{\sigma 2^{J+1}} f'(x, \varepsilon_{1Jk}) \right] \left[ g \left( \frac{k}{2^J} \right) + \frac{1}{2^{J+1}} g'(\varepsilon_{2Jk}) \right] \\ &= \frac{1}{2^J} \sum_{k \in \mathbb{Z}} f_x \left( \frac{k}{2^J} \right) g \left( \frac{k}{2^J} \right) \\ &\quad + \frac{1}{2^{2J+1}} \sum_{k \in \mathbb{Z}} \left[ f_x \left( \frac{k}{2^J} \right) g'(\varepsilon_{2Jk}) - \frac{\tau}{\sigma} f'(x, \varepsilon_{1Jk}) g \left( \frac{k}{2^J} \right) \right] \\ &\quad - \frac{\tau}{\sigma 2^{3J+2}} \sum_{k \in \mathbb{Z}} f'(x, \varepsilon_{1Jk}) g'(\varepsilon_{2Jk}), \end{aligned}$$

où  $\varepsilon_{1Jk}, \varepsilon_{2Jk} \in A_{Jk}$ . En posant

$$S_{R_1} = \frac{1}{2^J} \sum_{k \in \mathbb{Z}} \left[ f_x \left( \frac{k}{2^J} \right) g'(\varepsilon_{2Jk}) - \frac{\tau}{\sigma} f'(x, \varepsilon_{1Jk}) g \left( \frac{k}{2^J} \right) \right],$$

et

$$S_{R_2} = \frac{1}{2^J} \sum_{k \in \mathbb{Z}} |f'(x, \varepsilon_{1Jk}) g'(\varepsilon_{2Jk})|,$$

nous avons donc

$$\begin{aligned} \left| \mathbb{I}_0(x) - \tilde{\mathbb{I}}_0(x) \right| &= \left| \mathbb{I}_0(x) - \frac{1}{2^J} \sum_{k \in \mathbb{Z}} f_x \left( \frac{k}{2^J} \right) g \left( \frac{k}{2^J} \right) \right. \\ &\quad \left. - \frac{1}{2^{J+1}} S_{R_1} + \frac{\tau}{\sigma 2^{2J+2}} S_{R_2} \right| \\ &\leq \left| \mathbb{I}_0(x) - \frac{1}{2^J} \sum_{k \in \mathbb{Z}} f_x \left( \frac{k}{2^J} \right) g \left( \frac{k}{2^J} \right) \right| \\ &\quad + \frac{1}{2^{J+1}} |S_{R_1}| + \frac{\tau}{\sigma 2^{2J+2}} S_{R_2}. \end{aligned} \quad (2.14)$$

Il nous suffit maintenant de procéder à une étude terme à terme de l'équation précédente.

Commençons par considérer le premier terme. Pour cela, notons d'abord que comme  $f \in \mathcal{L}^1(\mathbb{R})$  (puisque c'est une densité) et  $f$  est continue, alors  $f \in \mathcal{L}^\infty(\mathbb{R})$ . Donc, il est possible de vérifier que

$$\int_{\mathbb{R}} \left| \frac{\partial^2}{\partial t^2} (f_x(t) g(t)) \Big|_{t=s} \right| ds \leq M_1 < +\infty, \quad (2.15)$$

où

$$M_1 = \frac{\tau^2}{\sigma^2} \|f''\|_\infty \|g\|_1 + 2 \frac{\tau}{\sigma} \|f'\|_\infty \|g'\|_1 + \|f\|_\infty \|g''\|_1,$$

et que

$$\sup_{x, s \in \mathbb{R}} \left| \frac{\partial}{\partial t} (f_x(t)g(t)) \Big|_{t=s} \right| \leq M_2 < +\infty,$$

où  $M_2$  est donné par

$$M_2 = \|f\|_\infty \|g'\|_\infty + \frac{\tau}{\sigma} \|f'\|_\infty \|g\|_\infty,$$

car  $g$  et  $g'$  sont toutes deux bornées (puisque continues et absolument intégrables sur  $\mathbb{R}$ ). Ceci permet l'utilisation du lemme C.2 (voir annexe C) pour obtenir le comportement du premier terme de l'équation (2.14). Nous avons ainsi

$$\begin{aligned} \left| \mathbb{I}_0(x) - \frac{1}{2^J} \sum_{k \in \mathbb{Z}} f_x \left( \frac{k}{2^J} \right) g \left( \frac{k}{2^J} \right) \right| &= \left| \int_{\mathbb{R}} f_x(s)g(s) ds - \frac{1}{2^J} \sum_{k \in \mathbb{Z}} f_x \left( \frac{k}{2^J} \right) g \left( \frac{k}{2^J} \right) \right| \\ &\leq \frac{M_2}{6 \times 2^{2J}} + \frac{M_1}{12 \times 2^{2J}} \\ &= \frac{1}{2^{2J}} \left( \frac{M_2}{6} + \frac{M_1}{12} \right), \end{aligned}$$

pour  $x \in \mathbb{R}$ , c'est-à-dire

$$\left| \mathbb{I}_0(x) - \frac{1}{2^J} \sum_{k \in \mathbb{Z}} f_x \left( \frac{k}{2^J} \right) g \left( \frac{k}{2^J} \right) \right| = O(2^{-2J}), \quad (2.16)$$

de façon uniforme en  $x$ , puisque  $M_1$  et  $M_2$  sont des constantes finies indépendantes de  $x$  et  $J$ . Remarquons que le lemme C.2 ne peut être employé pour décrire le comportement des deux derniers termes de l'équation (2.14). Ceci vient du fait que  $f'(x, \cdot)$  et  $g'$  sont évaluées aux points  $\varepsilon_{1Jk}$  et  $\varepsilon_{2Jk}$  (pour  $k \in \mathbb{Z}$ ) et que ces points ne sont pas connus explicitement. Le comportement de ces deux termes doit donc être obtenu autrement.

Dans cette optique, notons que  $S_{R_1}$  est une série apparentée aux sommes de Riemann. Dans le but de faire apparaître des sommes de Riemann dans l'expression de  $S_{R_1}$ , nous utilisons encore une fois des développements en série de Taylor. Ainsi, comme  $f' \in C^1(\mathbb{R})$  et  $g' \in C^1(\mathbb{R})$ , nous écrivons

$$f'(x, \varepsilon_{1Jk}) = f' \left( x, \frac{k}{2^J} \right) - \frac{\tau}{\sigma} \left( \varepsilon_{1Jk} - \frac{k}{2^J} \right) f''(x, \varepsilon_{3Jk}),$$

où

$$\varepsilon_{3Jk} \in \left[ \frac{k}{2^J}, \varepsilon_{1Jk} \right) \subset A_{Jk},$$

et

$$g'(\varepsilon_{2Jk}) = g' \left( \frac{k}{2^J} \right) + \left( \varepsilon_{2Jk} - \frac{k}{2^J} \right) g''(\varepsilon_{4Jk}),$$

où

$$\varepsilon_{4Jk} \in \left[ \frac{k}{2^J}, \varepsilon_{2Jk} \right) \subset A_{Jk}.$$

Ces deux résultats permettent d'exprimer  $S_{R_1}$  comme

$$\begin{aligned} S_{R_1} &= \frac{1}{2^J} \sum_{k \in \mathbb{Z}} \left[ f_x \left( \frac{k}{2^J} \right) g' \left( \frac{k}{2^J} \right) - \frac{\tau}{\sigma} f' \left( x, \frac{k}{2^J} \right) g \left( \frac{k}{2^J} \right) \right] \\ &\quad + \frac{1}{2^J} \sum_{k \in \mathbb{Z}} \left( \varepsilon_{2Jk} - \frac{k}{2^J} \right) f_x \left( \frac{k}{2^J} \right) g''(\varepsilon_{4Jk}) \\ &\quad + \frac{\tau^2}{\sigma^2 2^J} \sum_{k \in \mathbb{Z}} \left( \varepsilon_{1Jk} - \frac{k}{2^J} \right) f''(x, \varepsilon_{3Jk}) g \left( \frac{k}{2^J} \right). \end{aligned}$$

De là, comme

$$\left| \varepsilon_{iJk} - \frac{k}{2^J} \right| \leq \frac{1}{2^J},$$

pour  $i = 1, 2$  et pour  $k \in \mathbb{Z}$ , alors il est possible de borner cette somme par

$$|S_{R_1}| \leq |S_{R_{1,1}}| + \frac{1}{2^J} \|f\|_\infty S_{R_{1,2}} + \frac{\tau^2}{\sigma^2 2^J} \|f''\|_\infty S_{R_{1,3}}, \quad (2.17)$$

où

$$S_{R_{1,1}} = \frac{1}{2^J} \sum_{k \in \mathbb{Z}} \left[ f_x \left( \frac{k}{2^J} \right) g' \left( \frac{k}{2^J} \right) - \frac{\tau}{\sigma} f' \left( x, \frac{k}{2^J} \right) g \left( \frac{k}{2^J} \right) \right],$$

$$S_{R_{1,2}} = \frac{1}{2^J} \sum_{k \in \mathbb{Z}} |g''(\varepsilon_{4Jk})|,$$

et

$$S_{R_{1,3}} = \frac{1}{2^J} \sum_{k \in \mathbb{Z}} g \left( \frac{k}{2^J} \right),$$

et où les sommes  $S_{R_{1,i}}$  sont maintenant des sommes de Riemann pour  $i = 1, 2, 3$ .

Notons que la première de ces nouvelles sommes converge vers

$$\int_{\mathbb{R}} \left( f_x(s)g'(s) - \frac{\tau}{\sigma} f'(x, s)g(s) \right) ds = f_x(s)g(s) \Big|_{s=-\infty}^{s=+\infty} = 0.$$

Notons aussi que

$$\begin{aligned} \int_{\mathbb{R}} \left| \frac{\partial}{\partial t} \left( f_x(t)g'(t) - \frac{\tau}{\sigma} f'(x, t)g(t) \right) \Big|_{t=s} \right| ds &= \int_{\mathbb{R}} \left| \frac{\partial^2}{\partial t^2} (f_x(t)g(t)) \Big|_{t=s} \right| ds \\ &\leq M_1 < +\infty, \end{aligned}$$

d'après l'équation (2.15). Ainsi, en utilisant le lemme C.1 (voir annexe C), nous pouvons ici écrire

$$|S_{R_{1,1}} - 0| \leq \frac{M_1}{2^{J+1}},$$

et donc,

$$S_{R_{1,1}} = O(2^{-J}), \quad (2.18)$$

la constante  $M_1$  étant indépendante de  $x$  et  $J$ .

Les autres sommes de Riemann de l'équation (2.17), soient  $S_{R_{1,2}}$  et  $S_{R_{1,3}}$ , sont indépendantes de  $x$  et convergent respectivement vers  $\|g''\|_1$  et  $\|g\|_1 = 1$ . Ainsi, elles représentent toutes les deux des quantités bornées et il est possible d'écrire

$$S_{R_{1,2}} = O(1),$$

et

$$S_{R_{1,3}} = O(1),$$

dans les deux cas de manière uniforme en  $x$ . En jumelant ceci au résultat donné par l'équation (2.18), nous pouvons alors revenir à l'équation (2.17) et obtenir

$$\begin{aligned} |S_{R_1}| &\leq O(2^{-J}) + \frac{1}{2^J} \|f\|_{\infty} O(1) + \frac{\tau^2}{\sigma^2 2^J} \|f''\|_{\infty} O(1) \\ &= O(2^{-J}), \end{aligned} \quad (2.19)$$

ce résultat étant uniforme en  $x$ , ce qui complète l'étude du comportement de  $S_{R_1}$ .

Étudions finalement le dernier terme de l'équation (2.14). Il est facile de vérifier que

$$S_{R_2} \leq \|f'\|_{\infty} S_{R_{2,1}},$$

où

$$S_{R_{2,1}} = \frac{1}{2^J} \sum_{k \in \mathbb{Z}} |g'(\varepsilon_{2^J k})|,$$

et correspond ainsi à une somme de Riemann convergeant vers  $\|g'\|_1$ . En reprenant le même argument que précédemment, il est clair que

$$S_{R_{2,1}} = O(1),$$

et donc que

$$S_{R_2} \leq \|f'\|_\infty O(1) = O(1), \quad (2.20)$$

indépendamment de  $x \in \mathbb{R}$ .

Enfin, en combinant les équations (2.14), (2.16), (2.19) et (2.20), il devient évident que

$$\left| \mathbb{I}_0(x) - \tilde{\mathbb{I}}_0(x) \right| \leq O(2^{-2J}) + \frac{1}{2^{J+1}} O(2^{-J}) + \frac{\tau}{\sigma 2^{2J+2}} O(1),$$

et donc que

$$\left| \mathbb{I}_0(x) - \tilde{\mathbb{I}}_0(x) \right| = O(2^{-2J}),$$

uniformément en  $x$ . □

Remarquons que le raisonnement fait au théorème précédent ne peut être utilisé pour étudier le comportement de  $\tilde{\mathbb{I}}_l(x)$ , pour  $l \geq 1$ . En effet, les coefficients  $b_{l,jk}$  deviennent alors trop complexes pour qu'une telle façon de procéder amène un résultat simple. Cependant, comme il a déjà été dit, l'approximation de  $\tilde{\mathbb{I}}_0(x)$  tiendra un rôle principal au chapitre 4. Le gain de précision établi par ce dernier résultat sera alors extrêmement précieux.

Revenons maintenant à la construction des approximations de l'estimateur de Bayes et de sa perte espérée *a posteriori*.

### 2.3. Approximation de $\hat{\theta}_B(x)$ et de $\rho(\pi, \hat{\theta}_B|x)$

Nous nous intéressons ici à construire des approximations de l'estimateur de Bayes, soit  $\hat{\theta}_B(x)$ , et de sa perte espérée *a posteriori*,  $\rho(\pi, \hat{\theta}_B|x)$ . Pour ce faire, nous disposons déjà d'approximations de  $\mathbb{I}_l(x)$ . Il nous suffit donc d'utiliser les équations (2.4) et (2.5) qui relient l'estimateur de Bayes et sa perte espérée *a posteriori* aux  $\mathbb{I}_l(x)$ , ainsi que les résultats obtenus à la section précédente.

Posons d'abord

$$R_l(x) = \frac{\mathbb{I}_l(x)}{\mathbb{I}_0(x)},$$

pour  $l \geq 1$ . Les quantités d'intérêt s'écrivent donc comme

$$\widehat{\theta}_B(x) = \mu + \tau R_1(x),$$

et

$$\rho(\pi, \widehat{\theta}_B|x) = \tau^2 \left[ R_2(x) - (R_1(x))^2 \right].$$

En employant les équations précédentes et le résultat 2.1, il est tout à fait naturel d'approximer  $\widehat{\theta}_B(x)$  par

$$\widetilde{\theta}_B(x) = \mu + \tau \widetilde{R}_1(x),$$

et  $\rho(\pi, \widehat{\theta}_B|x)$  par

$$\widetilde{\rho}(\pi, \widetilde{\theta}_B|x) = \tau^2 \left[ \widetilde{R}_2(x) - (\widetilde{R}_1(x))^2 \right],$$

où

$$\widetilde{R}_l(x) = \frac{\widetilde{\mathbb{I}}_l(x)}{\widetilde{\mathbb{I}}_0(x)},$$

pour  $l \geq 1$ . Il nous faut donc maintenant étudier le comportement de  $\widetilde{R}_l(x)$ , en particulier pour  $l = 1, 2$ . Notre connaissance du comportement de  $\widetilde{\mathbb{I}}_l(x)$  (pour  $l \in \mathbb{N}$ ) est ici, bien sûr, primordiale puisqu'elle permet d'obtenir l'erreur d'approximation liée à  $\widetilde{R}_l(x)$  sans trop d'efforts.

**Théorème 2.3.** *Si  $f$ ,  $g$  et  $g_l$  respectent les conditions du théorème 2.1, alors*

$$\left| R_l(x) - \widetilde{R}_l(x) \right| = O(2^{-J}) \quad \forall x \in \mathbb{R},$$

et ce, quelque soit  $l \geq 1$ .

**Démonstration.** D'abord, notons que, d'après le théorème 2.1,

$$\begin{aligned} \widetilde{R}_l(x) &= \frac{\widetilde{\mathbb{I}}_l(x)}{\widetilde{\mathbb{I}}_0(x)} \\ &= \frac{\mathbb{I}_l(x) + O(2^{-J})}{\mathbb{I}_0(x) + O(2^{-J})}, \end{aligned} \tag{2.21}$$

puisque  $f$ ,  $g$  et  $g_l$  satisfont les hypothèses de ce théorème.

Maintenant, posons

$$p(y) = \frac{1}{y},$$



et remarquons que cette fonction est continue et continûment dérivable partout sauf en  $y = 0$ . Ainsi, elle peut être développée en série de Taylor de façon à obtenir, pour  $y \neq 0$ ,

$$p(y+h) = p(y) + hp'(\varepsilon),$$

pour  $\varepsilon \in (y, y+h)$  si  $h > 0$ , et pour  $\varepsilon \in (y+h, y)$  sinon. Ceci permet donc d'écrire

$$p(y + O(h)) = p(y) + O(h), \quad (2.22)$$

pour  $y \neq 0$ , puisque  $p'(\varepsilon)$  est une constante indépendante de  $h$ . Par contre, comme  $p'(\varepsilon)$  n'est pas indépendant du choix de  $y$ , il est important de noter que le résultat précédent n'est pas uniforme en  $y$ .

Remarquons ensuite que, d'après l'équation (2.3),

$$\mathbb{I}_0(x) = \sigma m(x) > 0,$$

et donc, le développement donné par l'équation (2.22) est valable pour  $y = \mathbb{I}_0(x)$ . Ainsi, nous avons

$$\frac{1}{\mathbb{I}_0(x) + O(2^{-J})} = \frac{1}{\mathbb{I}_0(x)} + O(2^{-J}),$$

ce résultat permettant de réduire l'équation (2.21).

En effet, en utilisant ce résultat dans l'équation (2.21), nous trouvons

$$\begin{aligned} \tilde{R}_l(x) &= (\mathbb{I}_l(x) + O(2^{-J})) \left( \frac{1}{\mathbb{I}_0(x)} + O(2^{-J}) \right) \\ &= \frac{\mathbb{I}_l(x)}{\mathbb{I}_0(x)} + \left( \mathbb{I}_l(x) + \frac{1}{\mathbb{I}_0(x)} \right) O(2^{-J}) + O(2^{-2J}) \\ &= R_l(x) + O(2^{-J}), \end{aligned}$$

pour  $l \geq 1$ , puisque  $\mathbb{I}_l(x)$  et  $\mathbb{I}_0(x)$  sont des constantes ne dépendant que de  $x$ . De là, nous avons finalement, quelque soit  $l \geq 1$ ,

$$\left| R_l(x) - \tilde{R}_l(x) \right| = O(2^{-J}),$$

pour  $x \in \mathbb{R}$ . Notons que ce résultat est valable seulement de façon ponctuelle. La borne obtenue n'est donc pas uniforme en  $x$ .  $\square$

Maintenant que le comportement de  $\tilde{R}_l(x)$  est connu, il nous est possible d'étudier la qualité des approximations données par  $\tilde{\theta}_B(x)$  et  $\tilde{\rho}(\pi, \tilde{\theta}_B|x)$ . En fait, le théorème précédent permet de conclure de façon assez directe que l'ordre d'approximation associé à  $\tilde{R}_l(x)$  est préservé lors de l'approximation de l'estimateur

de Bayes et de sa perte espérée *a posteriori*. Le prochain corollaire présente ce résultat.

**Corollaire 2.1.** *Si  $f, g, g_1$  et  $g_2$  respectent les conditions du théorème 2.1, alors*

$$\left| \widehat{\theta}_B(x) - \widetilde{\theta}_B(x) \right| = O(2^{-J}) \quad \forall x \in \mathbb{R},$$

et

$$\left| \rho(\pi, \widehat{\theta}_B|x) - \widetilde{\rho}(\pi, \widehat{\theta}_B|x) \right| = O(2^{-J}) \quad \forall x \in \mathbb{R}.$$

**Démonstration.** D'abord, le théorème 2.3 amène, pour  $l = 1$  et  $2$ ,

$$\widetilde{R}_l(x) = R_l(x) + O(2^{-J}),$$

puisque  $f, g, g_1$  et  $g_2$  satisfont aux conditions du théorème 2.1. Le premier résultat est ainsi trivial car

$$\widetilde{\theta}_B(x) = \mu + \tau \widetilde{R}_1(x),$$

et donc,

$$\begin{aligned} \widehat{\theta}_B(x) &= \mu + \tau (R_1(x) + O(2^{-J})) \\ &= \mu + \tau R_1(x) + O(2^{-J}) \\ &= \widetilde{\theta}_B(x) + O(2^{-J}). \end{aligned}$$

Il devient alors évident que

$$\left| \widehat{\theta}_B(x) - \widetilde{\theta}_B(x) \right| = O(2^{-J}),$$

pour  $x \in \mathbb{R}$ .

Pour le second résultat, il nous suffit de remarquer que

$$\begin{aligned} \widetilde{\rho}(\pi, \widehat{\theta}_B|x) &= \tau^2 \left[ \widetilde{R}_2(x) - (\widetilde{R}_1(x))^2 \right] \\ &= \tau^2 \left[ (R_2(x) + O(2^{-J})) - (R_1(x) + O(2^{-J}))^2 \right] \\ &= \tau^2 \left[ R_2(x) - (R_1(x))^2 + (2R_1(x) + 1)O(2^{-J}) + O(2^{-2J}) \right] \\ &= \tau^2 \left[ R_2(x) - (R_1(x))^2 \right] + O(2^{-J}), \end{aligned}$$

puisque

$$R_1(x) = \frac{\mathbb{I}_1(x)}{\mathbb{I}_0(x)}$$

est une constante dépendant uniquement de  $x$ . Ainsi,

$$\tilde{\rho}(\pi, \hat{\theta}_B|x) = \rho(\pi, \hat{\theta}_B|x) + O(2^{-J}),$$

et nous avons bien

$$\left| \rho(\pi, \hat{\theta}_B|x) - \tilde{\rho}(\pi, \hat{\theta}_B|x) \right| = O(2^{-J}),$$

pour  $x \in \mathbb{R}$ . □

À la lumière du résultat précédent, nous pouvons maintenant affirmer que les approximations  $\tilde{\theta}_B(x)$  et  $\tilde{\rho}(\pi, \hat{\theta}_B|x)$  ont le comportement souhaité. Malheureusement, ces approximations ne constituent pas des alternatives efficaces au calcul de  $\hat{\theta}_B(x)$  et de  $\rho(\pi, \hat{\theta}_B|x)$ . En effet, les coefficients  $b_{1,jk}$  et  $b_{2,jk}$ , donnés par la proposition 2.1, sont en général difficiles à calculer de façon explicite. Il nous apparaît donc important d'obtenir des approximations de l'estimateur de Bayes et de sa perte espérée *a posteriori* n'étant pas limitées par cette barrière potentielle.

## 2.4. Alternatives à $\tilde{\theta}_B(x)$ et $\tilde{\rho}(\pi, \hat{\theta}_B|x)$

Nous venons tout juste de faire remarquer au lecteur que la complexité des coefficients  $b_{1,jk}$  et  $b_{2,jk}$  peut limiter l'emploi des approximations présentées au corollaire 2.1. Cependant, il est très simple de borner ces coefficients. Ces bornes sont à l'origine de l'alternative au calcul de  $\tilde{\theta}_B(x)$  et de  $\tilde{\rho}(\pi, \hat{\theta}_B|x)$  que nous proposons plus loin. Intéressons nous d'abord au cas où  $l = 1$ . Le lemme suivant permet de borner les coefficients  $b_{1,jk}$ .

**Lemme 2.3.** *Les coefficients  $b_{1,jk}$  peuvent être bornés par*

$$\frac{k}{2^j} b_{0,jk} \leq b_{1,jk} \leq \frac{k+1}{2^j} b_{0,jk} \quad \forall j, k \in \mathbb{Z}.$$

**Démonstration.** D'abord, la proposition 2.1 nous donne l'expression exacte des coefficients  $b_{1,jk}$ . Nous avons, pour  $j, k \in \mathbb{Z}$ ,

$$b_{1,jk} = \frac{k}{2^j} b_{0,jk} + e_{1,jk},$$

où

$$e_{1,jk} = \frac{1}{2^{j/2}} G\left(\frac{k+1}{2^j}\right) - 2^{j/2} \int_{A_{jk}} G(s) ds. \quad (2.23)$$

Dans un premier temps, trouvons des bornes pour  $e_{1,jk}$ . Pour cela, notons que

$$\frac{1}{2^j} \inf_{s \in A_{jk}} G(s) \leq \int_{A_{jk}} G(s) ds \leq \frac{1}{2^j} \sup_{s \in A_{jk}} G(s).$$

Notons également que  $G$  est une fonction de répartition : elle est donc non décroissante. Ceci nous permet d'écrire l'inégalité précédente sous une forme plus simple, soit

$$\frac{1}{2^j} G\left(\frac{k}{2^j}\right) \leq \int_{A_{jk}} G(s) ds \leq \frac{1}{2^j} G\left(\frac{k+1}{2^j}\right).$$

En jumelant cette dernière inégalité à l'équation (2.23), nous obtenons finalement

$$0 \leq e_{1,jk} \leq \frac{1}{2^j} b_{0,jk},$$

où les coefficients  $b_{0,jk}$  sont donnés par l'équation (2.10). Ces bornes peuvent maintenant être appliquées à la définition de  $b_{1,jk}$  pour arriver au résultat recherché. En effet, nous obtenons alors

$$\frac{k}{2^j} b_{0,jk} \leq b_{1,jk} \leq \frac{k+1}{2^j} b_{0,jk},$$

pour  $j, k \in \mathbb{Z}$ . □

Les bornes construites au lemme précédent peuvent à leur tour être employées afin de trouver des bornes pour  $\tilde{R}_1(x)$ . Ces bornes permettent d'arriver à une alternative au calcul de  $\tilde{R}_1(x)$ .

**Théorème 2.4.** *Le rapport  $\tilde{R}_1(x)$  peut être borné par*

$$\tilde{R}_{1,\min}(x) \leq \tilde{R}_1(x) \leq \tilde{R}_{1,\max}(x),$$

où

$$\tilde{R}_{1,\min}(x) = \frac{1}{2^J} \frac{\sum_{k \in \mathbb{Z}} k a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}},$$

et

$$\tilde{R}_{1,\max}(x) = \frac{1}{2^J} + \frac{1}{2^J} \frac{\sum_{k \in \mathbb{Z}} k a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}} = \tilde{R}_{1,\min}(x) + \frac{1}{2^J}.$$

**Démonstration.** D'abord, par définition, nous savons que

$$\tilde{R}_1(x) = \frac{\tilde{\mathbb{I}}_1(x)}{\tilde{\mathbb{I}}_0(x)},$$

ce qui amène, d'après le résultat 2.1,

$$\tilde{R}_1(x) = \frac{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{1,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}}.$$

Le lemme 2.3 permet directement de conclure puisque

$$\begin{aligned} \tilde{R}_1(x) &\geq \frac{\sum_{k \in \mathbb{Z}} a_{Jk}(x) \left(\frac{k}{2^J}\right) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}} \\ &= \frac{1}{2^J} \frac{\sum_{k \in \mathbb{Z}} k a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}}, \end{aligned}$$

et

$$\begin{aligned} \tilde{R}_1(x) &\leq \frac{\sum_{k \in \mathbb{Z}} a_{Jk}(x) \left(\frac{k+1}{2^J}\right) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}} \\ &= \frac{1}{2^J} + \frac{1}{2^J} \frac{\sum_{k \in \mathbb{Z}} k a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}}, \end{aligned}$$

pour  $x \in \mathbb{R}$ . □

Le résultat précédent nous amène à imaginer l'alternative suivante au calcul de  $\tilde{R}_1(x)$ . Nous proposons d'approximer  $\tilde{R}_1(x)$  par

$$\begin{aligned} \tilde{R}_1^*(x) &= \frac{1}{2} \left( \tilde{R}_{1,\min}(x) + \tilde{R}_{1,\max}(x) \right) \\ &= \frac{1}{2^{J+1}} + \tilde{R}_{1,\min}(x) \\ &= \frac{1}{2^{J+1}} + \frac{1}{2^J} \frac{\sum_{k \in \mathbb{Z}} k a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}}, \end{aligned} \tag{2.24}$$

pour qu'ainsi,

$$\left| \tilde{R}_1(x) - \tilde{R}_1^*(x) \right| \leq \frac{1}{2} \left( \tilde{R}_{1,\max}(x) - \tilde{R}_{1,\min}(x) \right) = \frac{1}{2^{J+1}}, \tag{2.25}$$

et

$$\left| \tilde{R}_1(x) - \tilde{R}_1^*(x) \right| = O(2^{-J}), \tag{2.26}$$

pour  $x \in \mathbb{R}$ . Grâce à ce résultat, nous sommes maintenant en mesure de fournir une approximation adéquate et surtout efficace (puisque facile à calculer) de  $R_1(x)$ . Le théorème suivant confirme que  $\tilde{R}_1^*(x)$  est une alternative adéquate au calcul de  $\tilde{R}_1(x)$  puisque l'ordre de l'approximation est conservé en passant de  $\tilde{R}_1(x)$  à  $\tilde{R}_1^*(x)$ .

**Théorème 2.5.** *L'erreur d'approximation liée à  $\tilde{R}_1^*(x)$  est bornée par*

$$\left| R_1(x) - \tilde{R}_1^*(x) \right| \leq \left| R_1(x) - \tilde{R}_1(x) \right| + \frac{1}{2^{J+1}}.$$

*Ainsi, si  $f$ ,  $g$  et  $g_1$  satisfont les conditions du théorème 2.1, alors*

$$\left| R_1(x) - \tilde{R}_1^*(x) \right| = O(2^{-J}) \quad \forall x \in \mathbb{R}.$$

**Démonstration.** Il suffit d'abord d'employer l'inégalité du triangle, ce qui nous permet d'écrire

$$\left| R_1(x) - \tilde{R}_1^*(x) \right| \leq \left| R_1(x) - \tilde{R}_1(x) \right| + \left| \tilde{R}_1(x) - \tilde{R}_1^*(x) \right|. \quad (2.27)$$

En bornant le dernier terme de cette inégalité à l'aide de l'équation (2.25), nous obtenons la borne sur l'erreur d'approximation, soit

$$\left| R_1(x) - \tilde{R}_1^*(x) \right| \leq \left| R_1(x) - \tilde{R}_1(x) \right| + \frac{1}{2^{J+1}}.$$

Pour l'ordre de l'approximation, l'inégalité (2.27) amène aussi directement au résultat. En effet, d'après le théorème 2.3, comme  $f$ ,  $g$  et  $g_1$  satisfont les hypothèses du théorème 2.1, alors

$$\left| R_1(x) - \tilde{R}_1(x) \right| = O(2^{-J}).$$

De plus, il est évident que

$$\left| \tilde{R}_1(x) - \tilde{R}_1^*(x) \right| = O(2^{-J}).$$

Ainsi, en jumelant ces deux résultats à l'équation (2.27), nous avons bien

$$\left| R_1(x) - \tilde{R}_1^*(x) \right| = O(2^{-J}),$$

pour  $x \in \mathbb{R}$ . □

Passons maintenant au cas où  $l = 2$ . Les coefficients  $b_{2,jk}$  sont, en général, eux aussi difficiles à obtenir sous une forme explicite. Mais, comme pour le cas où  $l = 1$ , il est très simple de les borner. Cet exercice nous permettra d'accéder à une alternative efficace au calcul direct de  $\tilde{R}_2(x)$ .

**Lemme 2.4.** *Les coefficients  $b_{2,jk}$  peuvent être bornés, pour  $k \geq 0$ , par*

$$\frac{k^2}{2^{2j}} b_{0,jk} \leq b_{2,jk} \leq \frac{(k+1)^2}{2^{2j}} b_{0,jk} \quad \forall j \in \mathbb{Z},$$

et dans le cas où  $k < 0$ , par

$$\frac{(k+1)^2}{2^{2j}} b_{0,jk} \leq b_{2,jk} \leq \frac{k^2}{2^{2j}} b_{0,jk} \quad \forall j \in \mathbb{Z}.$$

**Démonstration.** D'après la proposition 2.1, les coefficients  $b_{2,jk}$  sont donnés par

$$b_{2,jk} = \frac{k^2}{2^{2j}} b_{0,jk} + e_{2,jk}, \quad (2.28)$$

pour  $j, k \in \mathbb{Z}$ , où

$$e_{2,jk} = \frac{2k+1}{2^{3j/2}} G\left(\frac{k+1}{2^j}\right) - 2^{j/2+1} \int_{A_{jk}} s G(s) ds. \quad (2.29)$$

Commençons par examiner le cas où  $k \geq 0$  (et  $A_{jk} \subset \mathbb{R}^+$ ). En procédant comme au lemme 2.3, nous avons

$$\frac{2k+1}{2^{2j+1}} \inf_{s \in A_{jk}} G(s) \leq \int_{A_{jk}} s G(s) ds \leq \frac{2k+1}{2^{2j+1}} \sup_{s \in A_{jk}} G(s),$$

et donc,

$$\frac{2k+1}{2^{2j+1}} G\left(\frac{k}{2^j}\right) \leq \int_{A_{jk}} s G(s) ds \leq \frac{2k+1}{2^{2j+1}} G\left(\frac{k+1}{2^j}\right).$$

En jumelant cette inégalité à la définition de  $e_{2,jk}$  donnée par l'équation (2.29), nous obtenons

$$0 \leq e_{2,jk} \leq \frac{2k+1}{2^{2j}} b_{0,jk},$$

où les coefficients  $b_{0,jk}$  sont définis par l'équation (2.10). En combinant ces bornes sur  $e_{2,jk}$  avec l'équation (2.28), nous avons finalement

$$\frac{k^2}{2^{2j}} b_{0,jk} \leq b_{2,jk} \leq \frac{(k+1)^2}{2^{2j}} b_{0,jk},$$

pour  $j \in \mathbb{Z}$ .

Le cas où  $k < 0$  (et  $A_{jk} \subset \mathbb{R}^-$ ) est résolu d'une manière similaire. Il suffit de remarquer que, cette fois,

$$\frac{2k+1}{2^{2j+1}} \sup_{s \in A_{jk}} G(s) \leq \int_{A_{jk}} s G(s) ds \leq \frac{2k+1}{2^{2j+1}} \inf_{s \in A_{jk}} G(s),$$

la fonction  $p(s) = s$  étant négative sur  $A_{jk}$ . Après avoir effectué les mêmes étapes que précédemment nous obtenons donc plutôt

$$\frac{2k+1}{2^{2j}} b_{0,jk} \leq e_{2,jk} \leq 0,$$

et, finalement,

$$\frac{(k+1)^2}{2^{2j}} b_{0,jk} \leq b_{2,jk} \leq \frac{k^2}{2^{2j}} b_{0,jk},$$

pour  $j \in \mathbb{Z}$ . □

Il nous est maintenant possible de borner  $\tilde{R}_2(x)$ . Ces bornes nous mèneront à une alternative efficace au calcul de  $\tilde{R}_2(x)$  et, éventuellement, à une approximation de  $R_2(x)$ .

**Théorème 2.6.** *Le rapport  $\tilde{R}_2(x)$  peut être borné par*

$$\tilde{R}_{2,\min}(x) \leq \tilde{R}_2(x) \leq \tilde{R}_{2,\max}(x),$$

où

$$\tilde{R}_{2,\min}(x) = \frac{1}{2^{2J}} \frac{\sum_{k \in \mathbb{Z}} k^2 a_{Jk}(x) b_{0,Jk} + \sum_{k < 0} (2k+1) a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}},$$

et

$$\tilde{R}_{2,\max}(x) = \frac{1}{2^{2J}} \frac{\sum_{k \in \mathbb{Z}} k^2 a_{Jk}(x) b_{0,Jk} + \sum_{k \geq 0} (2k+1) a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}}.$$

**Démonstration.** Il suffit ici de répéter le raisonnement fait au théorème 2.4, mais en employant plutôt les bornes fournies par le lemme 2.4. D'abord, du résultat 2.1, nous tirons

$$\begin{aligned} \tilde{R}_2(x) &= \frac{\tilde{\mathbb{I}}_2(x)}{\tilde{\mathbb{I}}_0(x)} \\ &= \frac{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{2,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}}. \end{aligned}$$

Ainsi,

$$\begin{aligned} \tilde{R}_2(x) &\geq \frac{\sum_{k \geq 0} a_{Jk}(x) \frac{k^2}{2^{2j}} b_{0,Jk} + \sum_{k < 0} a_{Jk}(x) \frac{(k+1)^2}{2^{2j}} b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}} \\ &= \frac{1}{2^{2J}} \frac{\sum_{k \in \mathbb{Z}} k^2 a_{Jk}(x) b_{0,Jk} + \sum_{k < 0} (2k+1) a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}}, \end{aligned}$$



et

$$\begin{aligned}\tilde{R}_2(x) &\leq \frac{\sum_{k \geq 0} a_{Jk}(x) \frac{(k+1)^2}{2^{2J}} b_{0,Jk} + \sum_{k < 0} a_{Jk}(x) \frac{k^2}{2^{2J}} b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}} \\ &= \frac{1}{2^{2J}} \frac{\sum_{k \in \mathbb{Z}} k^2 a_{Jk}(x) b_{0,Jk} + \sum_{k \geq 0} (2k+1) a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}}.\end{aligned}$$

□

En prenant la même approche que celle employée avec  $\tilde{R}_1(x)$ , nous proposons d'approximer  $\tilde{R}_2(x)$  par

$$\tilde{R}_2^*(x) = \frac{1}{2} \left( \tilde{R}_{2,\min}(x) + \tilde{R}_{2,\max}(x) \right),$$

et donc, par

$$\begin{aligned}\tilde{R}_2^*(x) &= \frac{1}{2^{2J+1}} \frac{2 \sum_{k \in \mathbb{Z}} k^2 a_{Jk}(x) b_{0,Jk} + \sum_{k \in \mathbb{Z}} (2k+1) a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}} \\ &= \frac{1}{2^{2J}} \frac{\sum_{k \in \mathbb{Z}} k^2 a_{Jk}(x) b_{0,Jk} + \sum_{k \in \mathbb{Z}} k a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}} + \frac{1}{2^{2J+1}} \\ &= \frac{1}{2^{2J}} \frac{\sum_{k \in \mathbb{Z}} k^2 a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}} + \frac{1}{2^J} \tilde{R}_1^*(x).\end{aligned}\tag{2.30}$$

Le prochain résultat traite du comportement de  $\tilde{R}_2^*(x)$ . Plus précisément, il quantifie l'erreur d'approximation associée à  $\tilde{R}_2^*(x)$  et permet de constater que le passage de  $\tilde{R}_2(x)$  à  $\tilde{R}_2^*(x)$  n'amène aucune détérioration de l'ordre de l'approximation.

**Théorème 2.7.** *L'erreur associée à l'approximation  $\tilde{R}_2^*(x)$  de  $R_2(x)$  est bornée supérieurement par*

$$\begin{aligned}\left| R_2(x) - \tilde{R}_2^*(x) \right| &\leq \left| R_2(x) - \tilde{R}_2(x) \right| + \frac{1}{2^{2J}} \frac{\sum_{k \in \mathbb{Z}} |k| a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}} \\ &\quad + \frac{1}{2^{2J+1}} \frac{\sum_{k \in \mathbb{Z}} \text{sign}(k) a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}},\end{aligned}$$

où

$$\text{sign}(k) = \begin{cases} |k|/k & \text{si } k \neq 0, \\ 0 & \text{sinon.} \end{cases}$$

De là, si  $f$ ,  $g$  et  $g_2$  respectent les hypothèses du théorème 2.1, alors

$$\left| R_2(x) - \tilde{R}_2^*(x) \right| = O(2^{-J}) \quad \forall x \in \mathbb{R}.$$

**Démonstration.** Nous débutons en faisant comme au théorème 2.5. Ainsi, nous écrivons d'abord

$$\left| R_2(x) - \tilde{R}_2^*(x) \right| \leq \left| R_2(x) - \tilde{R}_2(x) \right| + \left| \tilde{R}_2(x) - \tilde{R}_2^*(x) \right|, \quad (2.31)$$

et remarquons que d'après le théorème 2.3,

$$\left| R_2(x) - \tilde{R}_2(x) \right| = O(2^{-J}), \quad (2.32)$$

puisque  $f$ ,  $g$  et  $g_2$  satisfont les conditions du théorème 2.1. Il ne nous reste plus qu'à examiner le dernier terme de l'équation (2.31) pour connaître le comportement de  $\tilde{R}_2^*(x)$ .

Pour ce qui est de la borne supérieure sur l'erreur d'approximation, notons que

$$\begin{aligned} \left| \tilde{R}_2(x) - \tilde{R}_2^*(x) \right| &\leq \frac{1}{2} \left( \tilde{R}_{2, \max}(x) - \tilde{R}_{2, \min}(x) \right) \\ &= \frac{1}{2^{2J+1}} \frac{\sum_{k \geq 0} (2k+1) a_{Jk}(x) b_{0, Jk} - \sum_{k < 0} (2k+1) a_{Jk}(x) b_{0, Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0, Jk}} \\ &= \frac{1}{2^{2J}} \frac{\sum_{k \in \mathbb{Z}} |k| a_{Jk}(x) b_{0, Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0, Jk}} + \frac{1}{2^{2J+1}} \frac{\sum_{k \in \mathbb{Z}} \text{sign}(k) a_{Jk}(x) b_{0, Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0, Jk}}. \end{aligned}$$

Ainsi, en combinant cette inégalité à l'équation (2.31), nous obtenons la borne cherchée.

Pour obtenir le deuxième résultat, soit l'ordre de l'approximation, nous devons maintenant étudier le comportement des deux derniers termes de l'équation précédente. Tout d'abord, notons que (voir section 2.1)

$$\begin{aligned} \mathbb{E}(|\theta - \mu| | x) &= \frac{1}{\mathbb{I}_0(x)} \int_{\mathbb{R}} |\theta - \mu| f\left(\frac{x - \theta}{\sigma}\right) g\left(\frac{\theta - \mu}{\tau}\right) d\theta \\ &= \frac{\tau}{\mathbb{I}_0(x)} \int_{\mathbb{R}} |s| f_x(s) g(s) ds, \end{aligned}$$

et qu'une approximation *ad hoc* de cette quantité, obtenue à l'aide de la base de Haar, est donnée par

$$\tilde{\mathbb{E}}^*(x) = \frac{\tau}{2^J} \frac{\sum_{k \in \mathbb{Z}} |k| a_{Jk}(x) b_{0, Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0, Jk}}.$$

De là, nous avons

$$\tilde{\mathbb{E}}^*(x) = \mathbb{E}(|\theta - \mu| | x) + g_E(h),$$

où  $h = 2^{-J}$  et  $g_E(h)$  correspond à l'ordre (inconnu) de cette approximation. Maintenant, le comportement exact de  $g_E$  n'est pas important car il est suffisant d'avoir

$$\lim_{h \rightarrow 0} g_E(h) = 0.$$

En effet, nous avons ainsi

$$\begin{aligned} \frac{1}{2^{2J}} \frac{\sum_{k \in \mathbb{Z}} |k| a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}} &= \frac{1}{2^J} \frac{1}{\tau} \tilde{\mathbb{E}}^*(x) \\ &= \frac{1}{2^J} \frac{1}{\tau} [\mathbb{E}(|\theta - \mu| | x) + g_E(2^{-J})] \\ &= O(2^{-J}), \end{aligned} \tag{2.33}$$

puisque  $\mathbb{E}(|\theta - \mu| | x)$  est une constante bornée dépendant uniquement de  $x$ .

Pour le dernier terme à étudier, remarquons d'abord que

$$\begin{aligned} \frac{1}{2^{2J+1}} \frac{\sum_{k \in \mathbb{Z}} \text{sign}(k) a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}} &= \frac{1}{2^{2J+1}} \frac{2 \sum_{k \geq 0} a_{Jk}(x) b_{0,Jk} - \sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}} \\ &= \frac{1}{2^{2J}} \frac{\sum_{k \geq 0} a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}} - \frac{1}{2^{2J+1}} \\ &= \frac{1}{2^{2J}} \tilde{\mathbb{P}}^*(x) - \frac{1}{2^{2J+1}}, \end{aligned}$$

où

$$\tilde{\mathbb{P}}^*(x) = \frac{\sum_{k \geq 0} a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}},$$

et que  $\tilde{\mathbb{P}}^*(x)$  correspond à l'approximation, obtenue à l'aide de la base de Haar, de la probabilité *a posteriori* que  $\theta$  soit supérieur ou égal à  $\mu$ . En effet, cette probabilité est donnée (voir section 2.1) par

$$\begin{aligned} \mathbb{P}(\theta \geq \mu | x) &= \frac{1}{\mathbb{I}_0(x)} \int_{\mu}^{\infty} f\left(\frac{x - \theta}{\sigma}\right) g\left(\frac{\theta - \mu}{\tau}\right) d\theta \\ &= \frac{1}{\mathbb{I}_0(x)} \int_0^{\infty} f_x(s) g(s) ds, \end{aligned}$$

et devrait donc être approximée, en employant le raisonnement utilisé au résultat 2.1, par  $\tilde{\mathbb{P}}^*(x)$ . Ceci nous permet d'écrire, comme précédemment,

$$\tilde{\mathbb{P}}^*(x) = \mathbb{P}(\theta \geq \mu | x) + g_P(h),$$

où  $h = 2^{-J}$  et  $g_P(h)$  satisfait

$$\lim_{h \rightarrow 0} g_P(h) = 0.$$

Cette fois, c'est le comportement exact de  $g_P$  qui est sans importance. En effet, nous avons ici

$$\begin{aligned} \frac{1}{2^{2J+1}} \frac{\sum_{k \in \mathbb{Z}} \text{sign}(k) a_{Jk}(x) b_{0,Jk}}{\sum_{k \in \mathbb{Z}} a_{Jk}(x) b_{0,Jk}} &= \frac{1}{2^{2J}} [\mathbb{P}(\theta \geq \mu | x) + g_P(2^{-J})] - \frac{1}{2^{2J+1}} \\ &= O(2^{-2J}), \end{aligned} \quad (2.34)$$

puisque  $\mathbb{P}(\theta \geq \mu | x)$  est une constante bornée dont la valeur ne dépend que de  $x$ .

Finalement, les équations (2.33) et (2.34) permettent de conclure que

$$\left| \tilde{R}_2(x) - \tilde{R}_2^*(x) \right| = O(2^{-J}), \quad (2.35)$$

et, en jumelant ce dernier résultat aux équations (2.31) et (2.32), d'obtenir

$$\left| R_2(x) - \tilde{R}_2^*(x) \right| = O(2^{-J}),$$

pour  $x \in \mathbb{R}$ . □

Enfin, nous avons maintenant établi que  $\tilde{R}_1^*(x)$  et  $\tilde{R}_2^*(x)$  forment des approximations adéquates et efficaces de  $R_1(x)$  et  $R_2(x)$ . Il ne nous reste plus qu'à les employer correctement pour approximer l'estimateur de Bayes et sa perte espérée *a posteriori*. Pour ce faire, il nous suffit de définir

$$\tilde{\theta}_B^*(x) = \mu + \tau \tilde{R}_1^*(x),$$

et

$$\tilde{\rho}^*(\pi, \hat{\theta}_B | x) = \tau^2 \left[ \tilde{R}_2^*(x) - (\tilde{R}_1^*(x))^2 \right].$$

Il est très facile de démontrer que ces approximations exhibent le comportement désiré. Le prochain corollaire, qui est le dernier résultat introduit dans la présente section, confirme ce fait.

**Corollaire 2.2.** *Si  $f$ ,  $g$ ,  $g_1$  et  $g_2$  respectent les conditions du théorème 2.1, alors*

$$\left| \hat{\theta}_B(x) - \tilde{\theta}_B^*(x) \right| = O(2^{-J}) \quad \forall x \in \mathbb{R},$$

et

$$\left| \rho(\pi, \hat{\theta}_B | x) - \tilde{\rho}^*(\pi, \hat{\theta}_B | x) \right| = O(2^{-J}) \quad \forall x \in \mathbb{R}.$$

**Démonstration.** Il suffit de reprendre la réflexion faite au corollaire 2.1, mais en faisant évidemment les quelques ajustements nécessaires. D'abord, d'après les théorèmes 2.5 et 2.7, comme  $f, g, g_1$  et  $g_2$  satisfont les hypothèses du théorème 2.1,

$$\tilde{R}_l^*(x) = R_l(x) + O(2^{-J}),$$

pour  $l = 1, 2$ . Ainsi, nous avons, comme au corollaire 2.1,

$$\tilde{\theta}_B^*(x) = \mu + \tau \tilde{R}_1^*(x) = \hat{\theta}_B(x) + O(2^{-J}),$$

et

$$\tilde{\rho}^*(\pi, \hat{\theta}_B|x) = \tau^2 \left[ \tilde{R}_2^*(x) - (\tilde{R}_1^*(x))^2 \right] = \rho(\pi, \hat{\theta}_B|x) + O(2^{-J}),$$

ce qui permet de conclure que

$$\left| \hat{\theta}_B(x) - \tilde{\theta}_B^*(x) \right| = O(2^{-J}),$$

pour  $x \in \mathbb{R}$ , et

$$\left| \rho(\pi, \hat{\theta}_B|x) - \tilde{\rho}^*(\pi, \hat{\theta}_B|x) \right| = O(2^{-J}),$$

pour  $x \in \mathbb{R}$ . □

Ainsi, la difficulté associée au calcul des coefficients  $b_{1,jk}$  et  $b_{2,jk}$  peut effectivement être évitée. En effet, le corollaire précédent confirme que  $\tilde{\theta}_B^*(x)$  et  $\tilde{\rho}^*(\pi, \hat{\theta}_B|x)$  forment des approximations satisfaisantes de l'estimateur de Bayes et de sa perte espérée *a posteriori*, et ce, sans même faire intervenir le calcul de ces coefficients.

Nous soulignons tout de même au lecteur un problème technique qui se présente lors de l'évaluation des approximations obtenues. Dans les faits, les rapports  $\tilde{R}_1^*(x)$  et  $\tilde{R}_2^*(x)$  comportent tous deux des sommes infinies qui doivent, en pratique, être tronquées. Ceci soulève évidemment une nouvelle question : comment tronquer ces sommes sans détériorer la qualité de l'approximation ? Nous étudions cette question à la section 2.6.

Auparavant, nous nous tournons vers une autre alternative au calcul de  $\tilde{\theta}_B(x)$  et de  $\tilde{\rho}(\pi, \hat{\theta}_B|x)$ . Cette seconde approche nous permettra de définir une forme d'approximation simplifiée de  $\hat{\theta}_B(x)$  et de  $\rho(\pi, \hat{\theta}_B|x)$ .

## 2.5. Alternatives à $\tilde{\theta}_B(x)$ et $\tilde{\rho}(\pi, \hat{\theta}_B|x)$ : formes simples

Notre but est ici de proposer une forme encore plus simple pour l'approximation de l'estimateur de Bayes et de sa perte espérée *a posteriori*. Pour cela, nous

posons

$$S_l(x) = \sum_{k \in \mathbb{Z}} k^l a_{Jk}(x) b_{0,Jk},$$

pour  $l \in \mathbb{N}$ . Remarquons qu'avec cette notation,

$$\tilde{\mathbb{I}}_0(x) = S_0(x),$$

et que les approximations de  $R_1(x)$  et de  $R_2(x)$  développées précédemment peuvent s'écrire, d'après les équations (2.24) et (2.30), comme

$$\tilde{R}_1^*(x) = \frac{1}{2^J} \frac{S_1(x)}{S_0(x)} + \frac{1}{2^{J+1}}, \quad (2.36)$$

et

$$\tilde{R}_2^*(x) = \frac{1}{2^{2J}} \frac{S_2(x)}{S_0(x)} + \frac{1}{2^J} \tilde{R}_1^*(x). \quad (2.37)$$

Le résultat suivant est à la base de la forme simplifiée d'approximation de  $R_1(x)$  et de  $R_2(x)$ .

**Théorème 2.8.** *Sous les conditions du théorème 2.1,*

$$\left| \mathbb{I}_l(x) - \frac{1}{2^{lJ}} S_l(x) \right| = O(2^{-J}) \quad \forall x \in \mathbb{R},$$

pour  $l = 0, 1, 2$ .

**Démonstration.** D'abord, le cas où  $l = 0$  est trivial. En effet, d'après le théorème 2.1, nous avons directement

$$|\mathbb{I}_0(x) - S_0(x)| = |\mathbb{I}_0(x) - \tilde{\mathbb{I}}_0(x)| = O(2^{-J}).$$

Pour les deux autres cas, posons

$$\tilde{\mathbb{I}}_1^*(x) = \frac{1}{2^J} S_1(x) + \frac{1}{2^{J+1}} S_0(x),$$

et

$$\tilde{\mathbb{I}}_2^*(x) = \frac{1}{2^{2J}} S_2(x) + \frac{1}{2^J} \tilde{\mathbb{I}}_1^*(x),$$

et notons qu'alors

$$\tilde{R}_l^*(x) = \frac{\tilde{\mathbb{I}}_l^*(x)}{\tilde{\mathbb{I}}_0(x)},$$

pour  $l = 1, 2$ . Maintenant, l'inégalité du triangle permet d'écrire, toujours pour  $l = 1$  et  $2$ ,

$$\begin{aligned} \left| \mathbb{I}_l(x) - \frac{1}{2^{lJ}} S_l(x) \right| &\leq \left| \mathbb{I}_l(x) - \tilde{\mathbb{I}}_l(x) \right| + \left| \tilde{\mathbb{I}}_l(x) - \tilde{\mathbb{I}}_l^*(x) \right| \\ &\quad + \left| \tilde{\mathbb{I}}_l^*(x) - \frac{1}{2^{lJ}} S_l(x) \right|. \end{aligned} \quad (2.38)$$

Or, le théorème 2.1 amène

$$\left| \mathbb{I}_l(x) - \tilde{\mathbb{I}}_l(x) \right| = O(2^{-J}), \quad (2.39)$$

et, comme

$$\begin{aligned} \left| \tilde{\mathbb{I}}_l(x) - \tilde{\mathbb{I}}_l^*(x) \right| &= \tilde{\mathbb{I}}_0(x) \left| \frac{\tilde{\mathbb{I}}_l(x)}{\tilde{\mathbb{I}}_0(x)} - \frac{\tilde{\mathbb{I}}_l^*(x)}{\tilde{\mathbb{I}}_0(x)} \right| \\ &= \tilde{\mathbb{I}}_0(x) \left| \tilde{R}_l(x) - \tilde{R}_l^*(x) \right|, \end{aligned}$$

il amène aussi, en le jumelant aux équations (2.26) et (2.35),

$$\begin{aligned} \left| \tilde{\mathbb{I}}_l(x) - \tilde{\mathbb{I}}_l^*(x) \right| &= (\mathbb{I}_0(x) + O(2^{-J})) O(2^{-J}) \\ &= O(2^{-J}), \end{aligned} \quad (2.40)$$

pour  $l = 1, 2$ . Il nous suffit donc de déterminer le comportement du dernier terme de l'équation (2.38) pour arriver au résultat voulu.

Tout d'abord, dans le cas où  $l = 1$ ,

$$\begin{aligned} \left| \tilde{\mathbb{I}}_1^*(x) - \frac{1}{2^J} S_1(x) \right| &= \frac{1}{2^{J+1}} S_0(x) \\ &= \frac{1}{2^{J+1}} \tilde{\mathbb{I}}_0(x), \end{aligned}$$

et donc, en employant le même argument que précédemment,

$$\begin{aligned} \left| \tilde{\mathbb{I}}_1^*(x) - \frac{1}{2^J} S_1(x) \right| &= \frac{1}{2^{J+1}} (\mathbb{I}_0(x) + O(2^{-J})) \\ &= O(2^{-J}). \end{aligned}$$

Pour  $l = 2$ , nous avons plutôt

$$\begin{aligned} \left| \tilde{\mathbb{I}}_2^*(x) - \frac{1}{2^{2J}} S_2(x) \right| &= \frac{1}{2^J} \left| \tilde{\mathbb{I}}_1^*(x) \right| \\ &= \frac{1}{2^J} \tilde{\mathbb{I}}_0(x) \left| \frac{\tilde{\mathbb{I}}_1^*(x)}{\tilde{\mathbb{I}}_0(x)} \right| \\ &= \frac{1}{2^J} \tilde{\mathbb{I}}_0(x) \left| \tilde{R}_1^*(x) \right|, \end{aligned}$$

et, d'après les théorèmes 2.1 et 2.5,

$$\begin{aligned} \left| \tilde{\mathbb{I}}_2^*(x) - \frac{1}{2^{2J}} S_2(x) \right| &= \frac{1}{2^J} (\mathbb{I}_0(x) + O(2^{-J})) (|R_1(x)| + O(2^{-J})) \\ &= \frac{1}{2^J} [\mathbb{I}_0(x) |R_1(x)| + (\mathbb{I}_0(x) + |R_1(x)|) O(2^{-J}) + O(2^{-2J})] \\ &= \frac{1}{2^J} |\mathbb{I}_1(x)| + O(2^{-J}) \\ &= O(2^{-J}). \end{aligned}$$

Ainsi, nous obtenons

$$\left| \tilde{\mathbb{I}}_l^*(x) - \frac{1}{2^{lJ}} S_l(x) \right| = O(2^{-J}), \quad (2.41)$$

pour  $l = 1, 2$ .

Finalement, en substituant les équations (2.39) à (2.41) dans l'équation (2.38), nous avons bien, pour  $l = 1$  et  $2$ ,

$$\left| \mathbb{I}_l(x) - \frac{1}{2^{lJ}} S_l(x) \right| = O(2^{-J}),$$

pour  $x \in \mathbb{R}$ . □

Le résultat précédent nous amène naturellement à approximer plus simplement  $R_l(x)$  par

$$\tilde{R}_l^S(x) = \frac{1}{2^{lJ}} \frac{S_l(x)}{S_0(x)}, \quad (2.42)$$

pour  $l = 1, 2$ . Notons que les équations (2.36) et (2.37) laissaient déjà entrevoir cette possibilité. Notons aussi que la proposition 2.1 aurait pu nous lancer dès le départ dans cette direction. Le prochain théorème confirme que cette façon de procéder est acceptable.



**Théorème 2.9.** *Sous les hypothèses du théorème 2.1,*

$$\left| R_l(x) - \tilde{R}_l^S(x) \right| = O(2^{-J}) \quad \forall x \in \mathbb{R},$$

pour  $l = 1, 2$ .

**Démonstration.** Il nous suffit ici de procéder comme au théorème 2.3. En effet, le théorème 2.8 permet d'abord d'écrire

$$\frac{1}{2^{lJ}} S_l(x) = \mathbb{I}_l(x) + O(2^{-J}),$$

pour  $l = 0, 1, 2$ . Ainsi, en faisant comme au théorème 2.3, nous avons, pour  $l = 1$  et 2,

$$\begin{aligned} \tilde{R}_l^S(x) &= \frac{1}{2^{lJ}} \frac{S_l(x)}{S_0(x)} \\ &= \frac{\mathbb{I}_l(x) + O(2^{-J})}{\mathbb{I}_0(x) + O(2^{-J})} \\ &= (\mathbb{I}_l(x) + O(2^{-J})) \left( \frac{1}{\mathbb{I}_0(x)} + O(2^{-J}) \right) \\ &= R_l(x) + O(2^{-J}). \end{aligned}$$

Il devient donc évident que

$$\left| R_l(x) - \tilde{R}_l^S(x) \right| = O(2^{-J}),$$

pour  $x \in \mathbb{R}$  et  $l = 1, 2$ . □

Ce dernier résultat prouve que  $\tilde{R}_l^S(x)$  est une seconde approximation adéquate de  $R_l(x)$ , pour  $l = 1, 2$ . Il est donc normal de définir la forme simple des approximations de  $\hat{\theta}_B(x)$  et de  $\rho(\pi, \hat{\theta}_B|x)$  comme

$$\tilde{\theta}_B^S(x) = \mu + \tau \tilde{R}_1^S(x),$$

et

$$\tilde{\rho}^S(\pi, \hat{\theta}_B|x) = \tau^2 \left[ \tilde{R}_2^S(x) - (\tilde{R}_1^S(x))^2 \right].$$

Le corollaire qui suit confirme que ces formes simples d'approximation sont du même ordre que  $\tilde{\theta}_B^*(x)$  et  $\tilde{\rho}^*(\pi, \hat{\theta}_B|x)$ .

**Corollaire 2.3.** *Toujours sous les hypothèses du théorème 2.1,*

$$\left| \hat{\theta}_B(x) - \tilde{\theta}_B^S(x) \right| = O(2^{-J}) \quad \forall x \in \mathbb{R},$$

et

$$\left| \rho(\pi, \hat{\theta}_B|x) - \tilde{\rho}^S(\pi, \hat{\theta}_B|x) \right| = O(2^{-J}) \quad \forall x \in \mathbb{R}.$$

**Démonstration.** Pour une seconde fois, nous procédons exactement comme au corollaire 2.1. Cette fois, le théorème 2.9 permet d'abord d'écrire

$$\tilde{R}_l^S(x) = R_l(x) + O(2^{-J}),$$

pour  $l = 1, 2$ . Ainsi, nous avons

$$\tilde{\theta}_B^S(x) = \mu + \tau \tilde{R}_1^S(x) = \hat{\theta}_B(x) + O(2^{-J}),$$

et

$$\tilde{\rho}^S(\pi, \hat{\theta}_B|x) = \tau^2 \left[ \tilde{R}_2^S(x) - (\tilde{R}_1^S(x))^2 \right] = \rho(\pi, \hat{\theta}_B|x) + O(2^{-J}),$$

ce qui permet d'affirmer que, comme au corollaire 2.1,

$$\left| \hat{\theta}_B(x) - \tilde{\theta}_B^S(x) \right| = O(2^{-J}),$$

et

$$\left| \rho(\pi, \hat{\theta}_B|x) - \tilde{\rho}^S(\pi, \hat{\theta}_B|x) \right| = O(2^{-J}),$$

pour  $x \in \mathbb{R}$ . □

Maintenant, il est intéressant de constater qu'il existe des relations fort simples reliant les deux formes d'approximation de l'estimateur de Bayes et de sa perte espérée *a posteriori*. En effet, les équations (2.36), (2.37) et (2.42) permettent de relier les rapports  $\tilde{R}_l^*(x)$  et  $\tilde{R}_l^S(x)$  (pour  $l = 1, 2$ ) et donc, par le fait même, les deux types d'approximation.

**Proposition 2.2.** *Les deux formes d'approximation de  $\hat{\theta}_B(x)$  et de  $\rho(\pi, \hat{\theta}_B|x)$  peuvent être reliées à l'aide de*

$$\tilde{\theta}_B^*(x) = \tilde{\theta}_B^S(x) + \frac{\tau}{2^{J+1}},$$

et

$$\tilde{\rho}^*(\pi, \hat{\theta}_B|x) = \tilde{\rho}^S(\pi, \hat{\theta}_B|x) + \frac{\tau^2}{2^{2J+2}}.$$

**Démonstration.** Dans un premier temps, les équations (2.36) et (2.42) amènent

$$\begin{aligned}\tilde{R}_1^*(x) &= \frac{1}{2^J} \frac{S_1(x)}{S_0(x)} + \frac{1}{2^{J+1}} \\ &= \tilde{R}_1^S(x) + \frac{1}{2^{J+1}}.\end{aligned}\tag{2.43}$$

Ainsi, nous avons

$$\begin{aligned}\tilde{\theta}_B^*(x) &= \mu + \tau \tilde{R}_1^*(x) \\ &= \mu + \tau \left( \tilde{R}_1^S(x) + \frac{1}{2^{J+1}} \right) \\ &= \mu + \tau \tilde{R}_1^S(x) + \frac{\tau}{2^{J+1}},\end{aligned}$$

et les deux approximations de  $\hat{\theta}_B(x)$  peuvent être reliées à l'aide de

$$\tilde{\theta}_B^*(x) = \tilde{\theta}_B^S(x) + \frac{\tau}{2^{J+1}}.$$

Pour la perte espérée *a posteriori* de l'estimateur de Bayes, les équations (2.37) et (2.42) permettent d'abord d'écrire

$$\begin{aligned}\tilde{R}_2^*(x) &= \frac{1}{2^{2J}} \frac{S_2(x)}{S_0(x)} + \frac{1}{2^J} \tilde{R}_1^*(x) \\ &= \tilde{R}_2^S(x) + \frac{1}{2^J} \tilde{R}_1^*(x),\end{aligned}$$

et donc,

$$\begin{aligned}\tilde{\rho}^*(\pi, \hat{\theta}_B|x) &= \tau^2 \left[ \tilde{R}_2^*(x) - (\tilde{R}_1^*(x))^2 \right] \\ &= \tau^2 \left[ \tilde{R}_2^S(x) + \frac{1}{2^J} \tilde{R}_1^*(x) - (\tilde{R}_1^*(x))^2 \right].\end{aligned}$$

Maintenant, d'après l'équation (2.43),

$$\begin{aligned}\frac{1}{2^J} \tilde{R}_1^*(x) - (\tilde{R}_1^*(x))^2 &= \tilde{R}_1^*(x) \left( \frac{1}{2^J} - \tilde{R}_1^*(x) \right) \\ &= \left( \tilde{R}_1^S(x) + \frac{1}{2^{J+1}} \right) \left( \frac{1}{2^{J+1}} - \tilde{R}_1^S(x) \right) \\ &= \frac{1}{2^{2J+2}} - (\tilde{R}_1^S(x))^2,\end{aligned}$$

et, nous avons donc

$$\tilde{\rho}^*(\pi, \hat{\theta}_B|x) = \tau^2 \left[ \tilde{R}_2^S(x) - (\tilde{R}_1^S(x))^2 + \frac{1}{2^{2J+2}} \right].$$

Ainsi, nous obtenons finalement

$$\tilde{\rho}^*(\pi, \hat{\theta}_B|x) = \tilde{\rho}^S(\pi, \hat{\theta}_B|x) + \frac{\tau^2}{2^{2J+2}},$$

et les deux approximations de  $\rho(\pi, \hat{\theta}_B|x)$  sont effectivement liées par une relation simple.  $\square$

Notons que la proposition précédente permet d'obtenir la façon la plus efficace d'évaluer  $\tilde{\theta}_B^*(x)$  et  $\tilde{\rho}^*(\pi, \hat{\theta}_B|x)$ . En effet, nous voyons ici que le calcul direct de  $\tilde{R}_1^*(x)$  et  $\tilde{R}_2^*(x)$  peut être évité. Il suffit plutôt de calculer les formes simples d'approximation  $\tilde{\theta}_B^S(x)$  et  $\tilde{\rho}^S(\pi, \hat{\theta}_B|x)$  et de leur ajouter le bon terme correctif pour retrouver les formes plus complexes.

Nous discuterons plus en détail de l'emploi des deux formes d'approximation de l'estimateur de Bayes et de sa perte espérée *a posteriori* à la section 2.7. Nous porterons aussi, à ce moment, une attention particulière au choix de  $J$ .

Auparavant, nous revenons à un problème mentionné à la section précédente. Nous sommes effectivement maintenant confrontés à la dernière limite que comporte notre technique d'approximation. En effet, les différents rapports introduits jusqu'ici sont composés des sommes infinies  $S_l(x)$  ( $l = 0, 1, 2$ ) qui doivent, en pratique, être tronquées. À la prochaine section, nous étudions ce problème et proposons une façon de faire qui n'altère pas la qualité de l'approximation construite.

## 2.6. Considérations pratiques

Les approximations obtenues jusqu'ici sont toutes basées sur les sommes

$$S_l(x) = \sum_{k \in \mathbb{Z}} k^l a_{Jk}(x) b_{0,Jk},$$

pour  $l = 0, 1, 2$ . Il est clair que ces sommes doivent être tronquées lors de leur évaluation. Notre but ici est d'arriver à tronquer ces sommes, sans pour autant détériorer la qualité des approximations résultantes. Pour cela, nous définissons d'abord  $\mathcal{N}$ , l'ensemble des termes à conserver, comme

$$\mathcal{N} = \{k \in \mathbb{Z} \mid -N \leq k \leq N\}.$$

Les sommes tronquées résultantes sont donc définies par

$$S_l^T(x) = \sum_{k \in \mathcal{N}} k^l a_{Jk}(x) b_{0,Jk},$$

pour  $l = 0, 1, 2$ . Par le fait même, nous pouvons aussi définir les parties tronquées de chacune des sommes comme

$$\begin{aligned} T_l(x) &= S_l(x) - S_l^T(x) \\ &= \sum_{k \notin \mathcal{N}} k^l a_{Jk}(x) b_{0,Jk}, \end{aligned} \quad (2.44)$$

pour  $l = 0, 1, 2$ . Le problème rencontré ici peut maintenant être reformulé plus précisément : comment choisir  $N$  pour que la qualité des approximations demeure inchangée, ou encore, que les parties tronquées  $T_l(x)$  soient négligeables?

Dans le but d'arriver à une conclusion intéressante, nous faisons ici une hypothèse supplémentaire. Cette nouvelle hypothèse peut, à première vue, sembler injustifiée et/ou trop restrictive, mais elle permet simplement d'arriver à comprendre le comportement des parties tronquées. En fait, grâce à cette hypothèse, il est possible d'identifier un choix adéquat de  $N$ , c'est-à-dire un choix permettant de considérer que les sommes tronquées sont du même ordre que les sommes complètes. Il nous sera par la suite facile de confirmer que le fait de tronquer les sommes  $S_l(x)$  ne détériore pas la qualité des approximations construites précédemment.

**Théorème 2.10.** *Supposons que les hypothèses du théorème 2.1 soient satisfaites. Si  $g \in C^1(\mathbb{R})$ , si  $g'(l, \cdot) \in \mathcal{L}^1(\mathbb{R})$  où*

$$g'(l, s) = s^l g'(s),$$

*s'il existe  $\varepsilon > 0$ ,  $c > 0$  et  $N_x > 0$  tels que*

$$f_x(s)g(s) \leq \frac{c}{|s|^{3+\varepsilon}} \quad \text{lorsque } |s| \geq N_x,$$

*et si  $N \propto 2^{J(1+1/\varepsilon)}$ , alors*

$$\frac{1}{2^{lJ}} T_l(x) = O(2^{-J}) \quad \forall x \in \mathbb{R},$$

*pour  $l = 0, 1, 2$ .*

**Démonstration.** Nous procédons d'abord à la manière du théorème 2.2. Ainsi, d'après le lemme 2.2, nous avons

$$\begin{aligned} T_l(x) &= \sum_{k \notin \mathcal{N}} k^l a_{Jk}(x) b_{0,Jk} \\ &= \frac{1}{2^J} \sum_{k \notin \mathcal{N}} k^l \left[ f_x \left( \frac{k}{2^J} \right) - \frac{\tau}{\sigma 2^{J+1}} f'(x, \varepsilon_{1Jk}) \right] \left[ g \left( \frac{k}{2^J} \right) + \frac{1}{2^{J+1}} g'(\varepsilon_{2Jk}) \right], \end{aligned}$$

où  $\varepsilon_{1Jk}, \varepsilon_{2Jk} \in A_{Jk}$ , pour  $k \in \mathbb{Z}$ . Ceci nous amène à écrire

$$\frac{1}{2^{lJ}} |T_l(x)| \leq T_{R_1} + \frac{1}{2^{J+1}} T_{R_2} + \frac{\tau}{\sigma 2^{J+1}} T_{R_3} + \frac{\tau}{\sigma 2^{2J+2}} T_{R_4}, \quad (2.45)$$

où

$$\begin{aligned} T_{R_1} &= \frac{1}{2^J} \sum_{k \notin \mathcal{N}} \left( \frac{|k|}{2^J} \right)^l f_x \left( \frac{k}{2^J} \right) g \left( \frac{k}{2^J} \right), \\ T_{R_2} &= \frac{1}{2^J} \sum_{k \notin \mathcal{N}} \left( \frac{|k|}{2^J} \right)^l f_x \left( \frac{k}{2^J} \right) |g'(\varepsilon_{2Jk})|, \\ T_{R_3} &= \frac{1}{2^J} \sum_{k \notin \mathcal{N}} \left( \frac{|k|}{2^J} \right)^l |f'(x, \varepsilon_{1Jk})| g \left( \frac{k}{2^J} \right), \end{aligned}$$

et

$$T_{R_4} = \frac{1}{2^J} \sum_{k \notin \mathcal{N}} \left( \frac{|k|}{2^J} \right)^l |f'(x, \varepsilon_{1Jk}) g'(\varepsilon_{2Jk})|.$$

Étudions donc maintenant chacune des sommes précédentes.

Nous commençons par remarquer que

$$\begin{aligned} T_{R_1} &= -\frac{1}{2^{J+1}} \left[ \left( \frac{N}{2^J} \right)^l f_x \left( -\frac{N}{2^J} \right) g \left( -\frac{N}{2^J} \right) + \left( \frac{N}{2^J} \right)^l f_x \left( \frac{N}{2^J} \right) g \left( \frac{N}{2^J} \right) \right] \\ &\quad + \frac{1}{2^J} \left[ \frac{1}{2} \left( \frac{N}{2^J} \right)^l f_x \left( -\frac{N}{2^J} \right) g \left( -\frac{N}{2^J} \right) + \sum_{k=-\infty}^{-N-1} \left( \frac{|k|}{2^J} \right)^l f_x \left( \frac{k}{2^J} \right) g \left( \frac{k}{2^J} \right) \right] \\ &\quad + \frac{1}{2^J} \left[ \frac{1}{2} \left( \frac{N}{2^J} \right)^l f_x \left( \frac{N}{2^J} \right) g \left( \frac{N}{2^J} \right) + \sum_{k=N+1}^{\infty} \left( \frac{|k|}{2^J} \right)^l f_x \left( \frac{k}{2^J} \right) g \left( \frac{k}{2^J} \right) \right], \end{aligned}$$

et que, toujours à la manière du théorème 2.2,

$$\int_{\mathbb{R}} \left| \frac{\partial}{\partial t} (f_x(t)g(t)) \Big|_{t=s} \right| ds \leq M < +\infty,$$

où

$$M = \|f\|_{\infty} \|g'\|_1 + \frac{\tau}{\sigma} \|f'\|_{\infty} \|g\|_1.$$

Donc, nous pouvons ici appliquer le lemme C.3 (voir annexe C) et obtenir

$$\begin{aligned} T_{R_1} = & -\frac{1}{2^{J+1}} \left[ \left(\frac{N}{2^J}\right)^l f_x\left(-\frac{N}{2^J}\right) g\left(-\frac{N}{2^J}\right) + \left(\frac{N}{2^J}\right)^l f_x\left(\frac{N}{2^J}\right) g\left(\frac{N}{2^J}\right) \right] \\ & + c_l + O(2^{-J}), \end{aligned} \quad (2.46)$$

où la constante  $c_l$  est donnée par

$$c_l = \int_{-\infty}^{-N/2^J} |s|^l f_x(s)g(s) ds + \int_{N/2^J}^{\infty} |s|^l f_x(s)g(s) ds.$$

Supposons maintenant que  $N$  est assez grand et satisfait

$$\frac{N}{2^J} > N_x.$$

Par hypothèse,  $c_l$  respecte ainsi

$$\begin{aligned} c_l & \leq \int_{-\infty}^{-N/2^J} |s|^l \frac{c}{|s|^{3+\varepsilon}} ds + \int_{N/2^J}^{\infty} |s|^l \frac{c}{|s|^{3+\varepsilon}} ds \\ & = 2c \int_{N/2^J}^{\infty} \frac{1}{s^{3-l+\varepsilon}} ds, \end{aligned}$$

la dernière égalité découlant de la symétrie de la fonction intégrée. Ensuite, comme

$$\int_{N/2^J}^{\infty} \frac{1}{s^{3-l+\varepsilon}} ds = \frac{1}{2-l+\varepsilon} \left(\frac{2^J}{N}\right)^{2-l+\varepsilon},$$

alors la valeur de  $c_l$  peut être bornée par

$$c_l \leq \frac{2c}{2-l+\varepsilon} \left(\frac{2^J}{N}\right)^{2-l+\varepsilon}, \quad (2.47)$$

pour  $l = 0, 1, 2$ .

D'autre part, le premier terme de l'équation (2.46) peut aussi être borné d'une façon intéressante puisque

$$\begin{aligned} & \left(\frac{N}{2^J}\right)^l f_x\left(-\frac{N}{2^J}\right) g\left(-\frac{N}{2^J}\right) + \left(\frac{N}{2^J}\right)^l f_x\left(\frac{N}{2^J}\right) g\left(\frac{N}{2^J}\right) \\ & \leq \left(\frac{N}{2^J}\right)^l \frac{2c}{(N/2^J)^{3+\varepsilon}} \\ & = 2c \left(\frac{2^J}{N}\right)^{3-l+\varepsilon}, \end{aligned} \quad (2.48)$$

pour  $l = 0, 1, 2$  lorsque  $N > 2^J N_x$ .

Le comportement de  $T_{R_1}$  devient enfin clair. En effet, si  $N \propto 2^{J(1+1/\varepsilon)}$ , alors les équations (2.47) et (2.48) permettent d'écrire, à partir de l'équation (2.46),

$$\begin{aligned} |T_{R_1}| & \leq \frac{1}{2^{J+1}} O(2^{-J(1+\frac{3-l}{\varepsilon})}) + O(2^{-J(1+\frac{2-l}{\varepsilon})}) + O(2^{-J}) \\ & = O(2^{-J}), \end{aligned}$$

pour  $l = 0, 1, 2$ .

Passons maintenant à l'étude du comportement de  $T_{R_2}$ . Pour cela, notons que

$$T_{R_2} \leq \|f\|_\infty \frac{1}{2^J} \sum_{k \notin \mathcal{N}} \left(\frac{|k|}{2^J}\right)^l |g'(\varepsilon_{2^J k})|,$$

et que la somme apparaissant dans l'inégalité précédente s'apparente très fortement à une somme de Riemann. De ce point de vue, il est clair que cette somme converge lorsque  $g'(l, \cdot) \in \mathcal{L}^1(\mathbb{R})$ , et donc qu'elle correspond à une quantité bornée quelque soit la valeur de  $J$ . Ceci nous amène à écrire

$$T_{R_2} = O(1).$$

Le même raisonnement nous permet aussi d'affirmer que

$$T_{R_3} = O(1),$$

et que

$$T_{R_4} = O(1).$$

En effet, comme

$$T_{R_3} \leq \|f'\|_\infty \frac{1}{2^J} \sum_{k \notin \mathcal{N}} \left(\frac{|k|}{2^J}\right)^l g\left(\frac{k}{2^J}\right),$$



alors  $T_{R_3}$  est bornée lorsque  $g_l \in \mathcal{L}^1(\mathbb{R})$ . Aussi, comme

$$T_{R_4} \leq \|f'\|_\infty \frac{1}{2^J} \sum_{k \notin \mathcal{N}} \left(\frac{|k|}{2^J}\right)^l |g'(\varepsilon_{2^J k})|,$$

alors  $T_{R_4}$  est bornée, comme  $T_{R_2}$ , si  $g'(l, \cdot) \in \mathcal{L}^1(\mathbb{R})$ .

Revenons finalement à l'équation (2.45). Nous avons ainsi

$$\frac{1}{2^{lJ}} |T_l(x)| \leq O(2^{-J}) + \frac{1}{2^{J+1}} O(1) + \frac{1}{2^{J+1}} O(1) + \frac{\tau}{\sigma 2^{2J+2}} O(1),$$

et donc,

$$\frac{1}{2^{lJ}} T_l(x) = O(2^{-J}),$$

pour  $x \in \mathbb{R}$  et  $l = 0, 1, 2$ . □

Notons que le résultat précédent est un résultat ponctuel : la borne obtenue n'est, en effet, pas uniforme en  $x$ .

Notons aussi que comme la qualité des formes simples d'approximation n'est pas affectée par le fait de tronquer les sommes  $S_l(x)$ , alors, d'après la proposition 2.2, la qualité des formes plus complexes  $\tilde{\theta}_B^*(x)$  et  $\tilde{\rho}^*(\pi, \hat{\theta}_B|x)$  ne l'est pas non plus. Ceci s'explique par le fait que les deux formes d'approximation de l'estimateur de Bayes et de sa perte espérée *a posteriori* sont reliées par des facteurs constants par rapport aux sommes tronquées considérées (mais qui dépendent du niveau  $J$ ). Il nous suffit donc d'étudier l'effet de l'emploi des sommes tronquées pour l'évaluation des approximations simples  $\tilde{\theta}_B^S(x)$  et  $\tilde{\rho}^S(\pi, \hat{\theta}_B|x)$ . Pour cela, notons qu'il suffit que

$$\left| \mathbb{I}_l(x) - \frac{1}{2^{lJ}} S_l^T(x) \right| = O(2^{-J}),$$

pour  $l = 0, 1, 2$  et  $x \in \mathbb{R}$ , pour que les sommes  $S_l(x)$  puissent être tronquées sans affecter les approximations résultantes  $\tilde{\theta}_B^S(x)$  et  $\tilde{\rho}^S(\pi, \hat{\theta}_B|x)$ . En effet, si l'équation précédente est vérifiée, alors les sommes tronquées  $S_l^T(x)$  peuvent être substituées aux sommes complètes  $S_l(x)$  sans modification du comportement des approximations résultantes.

Le prochain corollaire confirme que si les sommes  $S_l(x)$  sont tronquées de façon adéquate, la qualité de l'approximation de  $\mathbb{I}_l(x)$  demeure intacte.

**Corollaire 2.4.** *Sous les conditions des théorèmes 2.1 et 2.10,*

$$\left| \mathbb{I}_l(x) - \frac{1}{2^{lJ}} S_l^T(x) \right| = O(2^{-J}) \quad \forall x \in \mathbb{R},$$

pour  $l = 0, 1, 2$ .

**Démonstration.** Remarquons tout d'abord que

$$\begin{aligned} \left| \mathbb{I}_l(x) - \frac{1}{2^{lJ}} S_l^T(x) \right| &\leq \left| \mathbb{I}_l(x) - \frac{1}{2^{lJ}} S_l(x) \right| + \left| \frac{1}{2^{lJ}} S_l(x) - \frac{1}{2^{lJ}} S_l^T(x) \right| \\ &= \left| \mathbb{I}_l(x) - \frac{1}{2^{lJ}} S_l(x) \right| + \frac{1}{2^{lJ}} |T_l(x)|. \end{aligned} \quad (2.49)$$

Maintenant, d'après le théorème 2.8,

$$\left| \mathbb{I}_l(x) - \frac{1}{2^{lJ}} S_l(x) \right| = O(2^{-J}),$$

et d'après le théorème 2.10,

$$\frac{1}{2^{lJ}} T_l(x) = O(2^{-J}).$$

Il découle de ces deux résultats et de l'équation (2.49) que

$$\left| \mathbb{I}_l(x) - \frac{1}{2^{lJ}} S_l^T(x) \right| = O(2^{-J}) \quad \forall x \in \mathbb{R},$$

pour  $l = 0, 1, 2$ . □

Ainsi, il est effectivement possible de tronquer les sommes  $S_l(x)$  sans détériorer, dans un premier temps, la qualité des approximations de  $\mathbb{I}_l(x)$ , puis la qualité des diverses approximations résultantes de l'estimateur de Bayes et de sa perte espérée *a posteriori*. La façon de tronquer adéquatement dépend cependant de la nature des densités utilisées dans le modèle bayésien de départ. Nous discuterons plus en détail de notre façon de procéder à la prochaine section lors de l'étude de quelques exemples.

Par ailleurs, avant de passer à cette dernière étape, nous faisons remarquer au lecteur que les résultats obtenus dans la présente section peuvent être raffinés dans le cas où  $l = 0$ . Comme c'était le cas pour le théorème 2.2, ce raffinement prendra tout son sens au chapitre 4.

**Théorème 2.11.** *Si les hypothèses du théorème 2.2 sont satisfaites, s'il existe  $\varepsilon > 0$ ,  $c > 0$  et  $N_g > 0$  tels que*

$$g(s) \leq \frac{c}{|s|^{1+\varepsilon}} \quad \text{lorsque } |s| \geq N_g,$$

*et si  $N \propto 2^{J(1+2/\varepsilon)}$ , alors*

$$T_0(x) = O(2^{-2J}) \quad \forall x \in \mathbb{R}.$$

*De plus, cette borne est uniforme en  $x$ .*

**Démonstration.** Notons d'abord que les coefficients  $a_{Jk}(x)$  peuvent être bornés selon

$$\begin{aligned} a_{Jk}(x) &= \langle f_x, \phi_{Jk} \rangle \\ &= 2^{J/2} \int_{A_{Jk}} f_x(s) ds \\ &\leq 2^{J/2} \|f\|_\infty \int_{A_{Jk}} ds \\ &= \frac{1}{2^{J/2}} \|f\|_\infty, \end{aligned}$$

pour tout  $k \in \mathbb{Z}$  et  $x \in \mathbb{R}$ . Ainsi, la partie tronquée de  $S_0(x)$ , donnée par l'équation (2.44), peut être à son tour bornée par

$$T_0(x) = \sum_{k \notin \mathcal{N}} a_{Jk}(x) b_{0,Jk} \leq \frac{1}{2^{J/2}} \|f\|_\infty \sum_{k \notin \mathcal{N}} b_{0,Jk}. \quad (2.50)$$

Maintenant, pour traiter la somme précédente, l'argument du lemme 2.2 peut être repris afin d'obtenir

$$b_{0,Jk} = \frac{1}{2^{J/2}} \left[ g\left(\frac{k}{2^J}\right) + \frac{1}{2^{J+1}} g'\left(\frac{k}{2^J}\right) + \frac{1}{3 \times 2^{2J+1}} g''(\varepsilon_{Jk}) \right],$$

où  $\varepsilon_{Jk} \in A_{Jk}$ . En effet, comme ici  $g \in C^2(\mathbb{R})$ , alors  $G \in C^3(\mathbb{R})$  et le développement en série utilisé au lemme 2.2 peut être poussé plus loin : nous passons maintenant à un développement en série de Taylor d'ordre 3. Ceci permet d'exprimer la borne donnée par l'équation (2.50) comme

$$T_0(x) \leq \|f\|_\infty \left[ T_{R_1} + \frac{1}{2^{J+1}} T_{R_2} + \frac{1}{3 \times 2^{2J+1}} T_{R_3} \right] \quad (2.51)$$

où

$$T_{R_1} = \frac{1}{2^J} \sum_{k \notin \mathcal{N}} g\left(\frac{k}{2^J}\right)$$

$$T_{R_2} = \frac{1}{2^J} \sum_{k \notin \mathcal{N}} g'\left(\frac{k}{2^J}\right),$$

et

$$T_{R_3} = \frac{1}{2^J} \sum_{k \notin \mathcal{N}} g''(\varepsilon_{Jk}),$$

chacune de ces quantités étant une somme de Riemann. Notons que ces sommes de Riemann sont toutes trois indépendantes de  $x$ . L'uniformité en  $x$  de la borne sur  $T_0(x)$  est donc essentiellement déjà établie.

Procédons à une étude terme à terme de la borne donnée par l'équation (2.51). Comme au théorème 2.10, nous écrivons d'abord

$$T_{R_1} = -\frac{1}{2^{J+1}} \left[ g\left(-\frac{N}{2^J}\right) + g\left(\frac{N}{2^J}\right) \right] + \frac{1}{2^J} \left[ \frac{1}{2} g\left(-\frac{N}{2^J}\right) + \sum_{k=-\infty}^{-N-1} g\left(\frac{k}{2^J}\right) \right]$$

$$+ \frac{1}{2^J} \left[ \frac{1}{2} g\left(\frac{N}{2^J}\right) + \sum_{k=N+1}^{\infty} g\left(\frac{k}{2^J}\right) \right].$$

Nous appliquons cette fois le lemme C.4 (voir annexe C) pour obtenir

$$T_{R_1} = -\frac{1}{2^{J+1}} \left[ g\left(-\frac{N}{2^J}\right) + g\left(\frac{N}{2^J}\right) \right] + c_1 + O(2^{-2J}), \quad (2.52)$$

où

$$c_1 = \int_{-\infty}^{-N/2^J} g(s) ds + \int_{N/2^J}^{\infty} g(s) ds.$$

Maintenant, si nous supposons que

$$\frac{N}{2^J} > N_g,$$

alors, la constante  $c_1$  peut être bornée par

$$\begin{aligned} c_1 &\leq \int_{-\infty}^{-N/2^J} \frac{c}{|s|^{1+\varepsilon}} ds + \int_{N/2^J}^{\infty} \frac{c}{|s|^{1+\varepsilon}} ds \\ &= 2c \int_{N/2^J}^{\infty} \frac{1}{|s|^{1+\varepsilon}} ds \\ &= \frac{2c}{\varepsilon} \left( \frac{2^J}{N} \right)^\varepsilon, \end{aligned}$$

et, le premier terme de l'équation (2.52) par

$$g\left(-\frac{N}{2^J}\right) + g\left(\frac{N}{2^J}\right) \leq 2c \left(\frac{2^J}{N}\right)^{1+\varepsilon}. \quad (2.53)$$

Ainsi, si  $N \propto 2^{J(1+2/\varepsilon)}$ , le comportement de  $T_{R_1}$  est donné par

$$\begin{aligned} |T_{R_1}| &\leq \frac{1}{2^{J+1}} O(2^{-2J(1+1/\varepsilon)}) + O(2^{-2J}) + O(2^{-2J}) \\ &= O(2^{-2J}), \end{aligned}$$

indépendamment de  $x \in \mathbb{R}$ .

Examinons ensuite  $T_{R_2}$ . D'une manière similaire à ce qui a été fait plus haut, nous exprimons  $T_{R_2}$  comme

$$\begin{aligned} T_{R_2} &= -\frac{1}{2^{J+1}} \left[ g'\left(-\frac{N}{2^J}\right) + g'\left(\frac{N}{2^J}\right) \right] + \frac{1}{2^J} \left[ \frac{1}{2} g'\left(-\frac{N}{2^J}\right) + \sum_{k=-\infty}^{-N-1} g'\left(\frac{k}{2^J}\right) \right] \\ &\quad + \frac{1}{2^J} \left[ \frac{1}{2} g'\left(\frac{N}{2^J}\right) + \sum_{k=N+1}^{\infty} g'\left(\frac{k}{2^J}\right) \right]. \end{aligned}$$

En appliquant le lemme C.3 (voir annexe C), nous avons ainsi

$$T_{R_2} = -\frac{1}{2^{J+1}} \left[ g'\left(-\frac{N}{2^J}\right) + g'\left(\frac{N}{2^J}\right) \right] + c_2 + O(2^{-J}), \quad (2.54)$$

où

$$c_2 = \int_{-\infty}^{-N/2^J} g'(s) ds + \int_{N/2^J}^{\infty} g'(s) ds.$$

Remarquons que  $c_2$  peut être calculée directement. En effet, nous avons

$$\begin{aligned} c_2 &= g(s) \Big|_{s=-\infty}^{s=-N/2^J} + g(s) \Big|_{s=N/2^J}^{s=+\infty} \\ &= g\left(-\frac{N}{2^J}\right) - g\left(\frac{N}{2^J}\right). \end{aligned}$$

De là, si  $N > 2^J N_g$ , alors nous avons aussi

$$\begin{aligned} |c_2| &\leq g\left(-\frac{N}{2^J}\right) + g\left(\frac{N}{2^J}\right) \\ &\leq 2c\left(\frac{2^J}{N}\right)^{1+\varepsilon}, \end{aligned}$$

la dernière inégalité ayant déjà été obtenue à l'équation (2.53).

Finalement, comme  $g'$  est bornée (puisque continue et absolument intégrable sur  $\mathbb{R}$ ), alors grâce à l'équation (2.54),  $T_{R_2}$  peut être bornée, indépendamment de  $x$ , par

$$|T_{R_2}| \leq \frac{1}{2^J} \|g'\|_\infty + |c_2| + O(2^{-J}),$$

et donc, si  $N \propto 2^{J(1+2/\varepsilon)}$ , par

$$\begin{aligned} |T_{R_2}| &\leq O(2^{-J}) + O(2^{-2J(1+1/\varepsilon)}) + O(2^{-J}) \\ &= O(2^{-J}). \end{aligned}$$

Le cas de  $T_{R_3}$  peut être traité de façon beaucoup plus simple. En effet, il nous suffit de montrer que  $T_{R_3}$  est une quantité bornée pour arriver plus tard au résultat souhaité. Pour cela, nous remarquons que  $T_{R_3}$  est une somme de Riemann convergente lorsque  $g'' \in \mathcal{L}^1(\mathbb{R})$ . Dans ce cas, quelque soit la valeur de  $J$ , elle est bornée et il est possible d'écrire

$$T_{R_3} = O(1),$$

encore une fois, indépendamment de  $x \in \mathbb{R}$ .

Un retour à l'équation (2.51) amène finalement

$$|T_0(x)| \leq \|f\|_\infty \left[ O(2^{-2J}) + \frac{1}{2^{J+1}} O(2^{-J}) + \frac{1}{3 \times 2^{2J+1}} O(1) \right],$$

d'où

$$T_0(x) = O(2^{-2J}),$$

et ce, uniformément en  $x$ . □

Après avoir caractérisé le comportement de  $T_0(x)$ , il est maintenant possible d'énoncer une version plus forte du corollaire 2.4. Notons que cette nouvelle variante n'est également valable que lorsque  $l = 0$ .

**Corollaire 2.5.** *Sous les conditions du théorème 2.11,*

$$|\mathbb{I}_0(x) - S_0^T(x)| = O(2^{-2J}) \quad \forall x \in \mathbb{R}.$$

*Cette borne est uniforme en  $x$ .*

Ce résultat est donné sans preuve puisqu'il suffit, pour le démontrer, de reprendre les arguments du corollaire 2.4. Tout comme les théorèmes 2.2 et 2.11, il prendra toute son importance au chapitre 4.

Faisons maintenant un bref retour sur les diverses hypothèses faites jusqu'ici sur  $f$  et  $g$ . D'abord, pour approximer  $\hat{\theta}_B(x)$  et  $\rho(\pi, \hat{\theta}_B|x)$  à l'aide des deux formes d'approximation données aux corollaires 2.2 et 2.3, la densité  $f$  doit satisfaire

- $f \in \mathcal{L}^2(\mathbb{R}) \cap C^1(\mathbb{R})$ ,
- $f' \in \mathcal{L}^\infty$ ,

et la densité  $g$  doit satisfaire

- $g \in \mathcal{L}^2(\mathbb{R}) \cap C^1(\mathbb{R})$ ,
- $g_l \in \mathcal{L}^1(\mathbb{R})$  pour  $l = 0, 1, 2$ ,
- $g'(l, \cdot) \in \mathcal{L}^1(\mathbb{R})$  pour  $l = 0, 1, 2$ .

De plus, nos résultats demandent essentiellement aux queues du produit des densités  $f$  et  $g$  de se comporter comme  $|s|^{-(3+\varepsilon)}$ , pour  $\varepsilon > 0$ . Notons que cette dernière condition est relativement peu restrictive puisqu'elle est satisfaite dans la très grande majorité des cas, dont entre autres, le cas extrême où les deux densités du modèle sont des densités de Cauchy. Cette condition est aussi satisfaite lorsque l'une des densités a un support borné ou contient un terme à décroissance exponentielle. Notons que dans ce dernier cas,  $N$  peut être choisi selon

$$N \propto 2^J,$$

et ce, sans affecter l'ordre des approximations construites. Ceci est possible puisque la décroissance exponentielle est plus rapide que toute décroissance polynomiale. En fait, en employant les mêmes arguments que O'Hagan (1990), nous pouvons considérer l'hypothèse précédente satisfaite avec  $\varepsilon = \infty$ , ce qui amène bien

$$N \propto 2^{J(1+1/\varepsilon)} \xrightarrow{\varepsilon \rightarrow \infty} 2^J.$$

Dans un tel cas, nous proposons donc de prendre

$$N = C2^J,$$

où  $C$  est choisi de façon à s'assurer que l'approximation obtenue soit adéquate.

Pour ce faire, notons d'abord que les coefficients  $b_{0,Jk}$ , contrairement aux  $a_{Jk}(x)$ , sont indépendants de la valeur observée  $x$ . Ainsi, pour une valeur de  $J$  fixée, il est possible de trouver des valeurs de  $k$  au-delà desquelles les coefficients  $b_{0,Jk}$  sont négligeables, et ce sans égard à la valeur de  $x$ . En effet, nous utilisons l'équation (2.11) afin d'obtenir

$$\begin{aligned} \sum_{k \in \mathbb{Z}} b_{0,Jk} &= 2^{J/2} \sum_{k \in \mathbb{Z}} \mathbb{P}_g(A_{Jk}) \\ &= 2^{J/2} \mathbb{P}_g\left(\bigcup_{k \in \mathbb{Z}} A_{Jk}\right) \\ &= 2^{J/2}, \end{aligned}$$

puisque  $\bigcup_{k \in \mathbb{Z}} A_{Jk} = \mathbb{R}$ . Par le même raisonnement, nous avons également

$$\begin{aligned} \sum_{k \in \mathcal{N}} b_{0,Jk} &= 2^{J/2} \mathbb{P}_g\left(\bigcup_{k \in \mathcal{N}} A_{Jk}\right) \\ &= 2^{J/2} \mathbb{P}_g\left(-\frac{N}{2^J} \leq \theta < \frac{N+1}{2^J}\right) \\ &= 2^{J/2} \mathbb{P}_g\left(-C \leq \theta < C + \frac{1}{2^J}\right) \\ &\geq 2^{J/2} \mathbb{P}_g(-C < \theta < C) \\ &= 2^{J/2} \mathbb{P}_g(|\theta| < C). \end{aligned}$$

De là, un choix de  $C$  tel que

$$\mathbb{P}_g(|\theta| < C) \simeq 1, \tag{2.55}$$

nous assure que tous les coefficients  $b_{0,Jk}$  non négligeables sont inclus dans les sommes  $S_l^T(x)$ , quelque soit la valeur de  $x \in \mathbb{R}$ . Cette façon de procéder permet également de voir que pour les coefficients  $a_{Jk}(x)$ , nous avons plutôt

$$\sum_{k \in \mathbb{Z}} a_{Jk}(x) = 2^{J/2} \frac{\sigma}{\tau},$$



et que,

$$\sum_{k \in \mathcal{N}} a_{Jk}(x) = 2^{J/2} \frac{\sigma}{\tau} \mathbb{P}_f \left( \bigcup_{k \in \mathcal{N}} \left\{ \theta \in \mathbb{R} \mid \theta \in A_{x, Jk} \right\} \right),$$

où les intervalles  $A_{x, Jk}$  sont définis par

$$A_{x, Jk} = \left[ \frac{x - \mu}{\sigma} - \frac{\tau(k+1)}{\sigma 2^J}, \frac{x - \mu}{\sigma} - \frac{\tau k}{\sigma 2^J} \right),$$

pour  $k \in \mathbb{Z}$  et  $x \in \mathbb{R}$ . Ainsi, nous avons

$$\begin{aligned} \sum_{k \in \mathcal{N}} a_{Jk}(x) &= 2^{J/2} \frac{\sigma}{\tau} \mathbb{P}_f \left( \frac{x - \mu}{\sigma} - \frac{\tau(N+1)}{\sigma 2^J} \leq \theta < \frac{x - \mu}{\sigma} + \frac{\tau N}{\sigma 2^J} \right) \\ &= 2^{J/2} \frac{\sigma}{\tau} \mathbb{P}_f \left( \frac{x - \mu}{\sigma} - \frac{\tau}{\sigma} \left( C + \frac{1}{2^J} \right) \leq \theta < \frac{x - \mu}{\sigma} + \frac{\tau}{\sigma} C \right) \\ &\geq 2^{J/2} \frac{\sigma}{\tau} \mathbb{P}_f \left( \frac{x - \mu}{\sigma} - \frac{\tau}{\sigma} C \leq \theta < \frac{x - \mu}{\sigma} + \frac{\tau}{\sigma} C \right) \\ &= 2^{J/2} \frac{\sigma}{\tau} \mathbb{P}_f \left( \left| \theta - \frac{x - \mu}{\sigma} \right| < \frac{\tau}{\sigma} C \right), \end{aligned}$$

et, pour que tous les coefficients  $a_{Jk(x)}$  non négligeables soient aussi inclus dans les sommes  $S_l^T(x)$ , il faut que  $C$  respecte également

$$\mathbb{P}_f \left( \left| \theta - \frac{x - \mu}{\sigma} \right| < \frac{\tau}{\sigma} C \right) \simeq 1. \quad (2.56)$$

Cette réflexion nous amène donc à conclure que si  $C$  respecte les équations (2.55) et (2.56), alors nous pouvons nous attendre à ce que les diverses approximations obtenues soient bonnes. Remarquons cependant que cette dernière équation permet de constater qu'une valeur adéquate de  $C$  peut certainement être influencée par la valeur observée  $x$ . Ceci dépendra ultimement de la nature des densités  $f$  et  $g$  composant le modèle bayésien.

Maintenant, notons qu'il est possible d'employer la technique développée (moyennant quelques ajustements) dès que l'une des densités du modèle satisfait les conditions imposées ici à  $f$  et l'autre, les conditions imposées à  $g$ . En effet, les rôles de  $f$  et  $g$  peuvent être inversés dans notre façon de construire les approximations des intégrales  $\mathbb{I}_l(x)$  puisque les équations (2.1) et (2.2) peuvent

aussi s'écrire comme

$$\int_{\mathbb{R}} \theta^n f\left(\frac{x-\theta}{\sigma}\right) g\left(\frac{\theta-\mu}{\tau}\right) d\theta = \sum_{l=0}^n \binom{n}{l} (-1)^l x^{n-l} \sigma^{l+1} \mathbb{I}_l(x),$$

pour  $n \geq 0$ , où l'intégrale  $\mathbb{I}_l(x)$  est donnée, cette fois, par

$$\mathbb{I}_l(x) = \int_{\mathbb{R}} s^l f(s) g\left(\frac{\sigma}{\tau} \left[\frac{x-\mu}{\sigma} - s\right]\right) ds. \quad (2.57)$$

Ce résultat est obtenu en posant

$$\theta = x - \sigma s,$$

dans la première intégrale. En utilisant cette approche, les équations (2.3) à (2.5) se retrouvent modifiées de la façon suivante. Tout d'abord, la densité marginale  $m$  est maintenant donnée par

$$m(x) = \frac{1}{\tau} \mathbb{I}_0(x). \quad (2.58)$$

De là, l'estimateur de Bayes devient

$$\hat{\theta}_B(x) = x - \sigma \frac{\mathbb{I}_1(x)}{\mathbb{I}_0(x)}, \quad (2.59)$$

et la perte espérée *a posteriori* de l'estimateur de Bayes devient

$$\rho(\pi, \hat{\theta}_B|x) = \sigma^2 \left[ \frac{\mathbb{I}_2(x)}{\mathbb{I}_0(x)} - \left( \frac{\mathbb{I}_1(x)}{\mathbb{I}_0(x)} \right)^2 \right]. \quad (2.60)$$

Sous cette nouvelle forme,  $\hat{\theta}_B(x)$  et  $\rho(\pi, \hat{\theta}_B|x)$  peuvent être approximés en faisant comme précédemment, sauf que l'équation (2.57) laisse voir que les rôles de  $f$  et  $g$  sont effectivement inversés. Ainsi, les coefficients  $a_{J_k}(x)$  sont maintenant calculés à l'aide de  $g$ , et les coefficients  $b_{0,J_k}$  à l'aide de  $f$ . À la prochaine section, nous verrons un exemple simple où une telle réflexion est nécessaire.

Notons finalement qu'il y a une deuxième situation où il peut être préférable d'inverser les rôles de  $f$  et  $g$  pour le calcul des intégrales  $\mathbb{I}_l(x)$ . En effet, si les deux densités du modèle considéré satisfont toutes les conditions imposées ici à  $f$  et  $g$ , et donc que les deux approches peuvent être employées, alors l'approche calculant les coefficients  $b_{0,J_k}$  à l'aide de la densité la plus concentrée permettra de minimiser le nombre de termes à conserver dans les sommes tronquées  $S_l^T(x)$ . Ceci permet, en quelque sorte, d'optimiser le rendement des approximations utilisées puisque des gains substantiels peuvent ainsi être réalisés au niveau du temps de calcul. Nous reviendrons sur ce point à la prochaine section.

## 2.7. Quelques exemples

Nous proposons maintenant une brève étude des diverses approximations de l'estimateur de Bayes et de sa perte espérée *a posteriori* construites aux sections 2.4 et 2.5. Lors de nos investigations, nous discutons de l'emploi des deux formes d'approximation de  $\hat{\theta}_B(x)$  et de  $\rho(\pi, \hat{\theta}_B|x)$ . Nous portons également une attention particulière aux choix de  $J$ , le niveau de l'approximation initiale de  $f_x$  obtenue à l'aide de la base de Haar, et de  $N$ , le paramètre contrôlant le nombre de termes à conserver dans le développement des sommes tronquées  $S_l^T(x)$ .

L'étude proposée est faite à l'aide de deux exemples. Le premier permet d'évaluer le comportement des diverses approximations de  $\hat{\theta}_B(x)$  et de  $\rho(\pi, \hat{\theta}_B|x)$  dans le cas d'un modèle simple où l'estimateur de Bayes et sa perte espérée *a posteriori* peuvent être obtenus de façon explicite. Ceci est fait en comparant les diverses approximations construites aux valeurs calculées de  $\hat{\theta}_B(x)$  et de  $\rho(\pi, \hat{\theta}_B|x)$ .

Le deuxième illustre l'utilité de la technique en obtenant des approximations de  $\hat{\theta}_B(x)$  et de  $\rho(\pi, \hat{\theta}_B|x)$  dans le cas d'un modèle où les quantités d'intérêt ne peuvent être calculées directement. De plus, cet exemple illustre également la procédure à suivre lorsque les rôles de  $f$  et  $g$  doivent être inversés pour satisfaire les hypothèses nous assurant de la précision des diverses approximations. Là également, nous examinons le comportement des approximations construites.

### 2.7.1. Le modèle normale/double exponentielle

Nous supposons ici que

$$X|\theta \sim N(\theta, 1),$$

et que

$$\theta \sim DE(0, 1).$$

Nous avons ainsi

$$L(\theta|x) = f(x - \theta) = \phi(x - \theta),$$

pour  $x \in \mathbb{R}$ , où  $\phi$  correspond à la fonction de densité normale centrée et réduite et

$$\pi(\theta) = g(\theta) = \frac{1}{2}e^{-|\theta|},$$

pour  $\theta \in \mathbb{R}$ . Dans ce contexte, il est possible de montrer que la densité marginale de l'observation  $X$  est donnée par

$$m(x) = \frac{1}{2} e^{(\frac{1}{2}-x)} [\Phi(x-1) + e^{2x} [1 - \Phi(x+1)]], \quad (2.61)$$

pour  $x \in \mathbb{R}$ , où  $\Phi$  correspond à la fonction de répartition associée à  $\phi$ . De plus, sous la perte quadratique, il est aussi possible de montrer que l'estimateur de Bayes de  $\theta$  est donné par

$$\hat{\theta}_B(x) = x + 1 + \frac{\phi(x-1) - e^{2x}\phi(x+1) - 2\Phi(x-1)}{\Phi(x-1) + e^{2x}[1 - \Phi(x+1)]},$$

et que sa perte espérée *a posteriori* est donnée par

$$\rho(\pi, \hat{\theta}_B|x) = 2 + \frac{(x-1)e^{2x}\phi(x+1) - (x+1)\phi(x-1)}{\Phi(x-1) + e^{2x}[1 - \Phi(x+1)]} - \left(x - \hat{\theta}_B(x)\right)^2.$$

Pour plus de détails, voir Leblanc (1995).

Notons qu'ici, toutes les hypothèses faites sur  $f$  et  $g$  à la section précédente sont satisfaites. Donc, les corollaires 2.2, 2.3 et 2.4 et la proposition 2.2 peuvent être employés directement afin de définir les formes tronquées des approximations construites précédemment qui sont utiles ici. Tout d'abord, les formes tronquées de  $\tilde{\theta}_B^S(x)$  et  $\tilde{\rho}^S(\pi, \hat{\theta}_B|x)$  sont données par

$$\tilde{\theta}_1(x) = \frac{1}{2^J} \frac{S_1^T(x)}{S_0^T(x)},$$

et

$$\tilde{\rho}_1(\pi, \hat{\theta}_B|x) = \frac{1}{2^{2J}} \left[ \frac{S_2^T(x)}{S_0^T(x)} - \left( \frac{S_1^T(x)}{S_0^T(x)} \right)^2 \right],$$

et les formes tronquées de  $\tilde{\theta}_B^*(x)$  et  $\tilde{\rho}^*(\pi, \hat{\theta}_B|x)$  par

$$\tilde{\theta}_2(x) = \tilde{\theta}_1(x) + \frac{1}{2^{J+1}},$$

et

$$\tilde{\rho}_2(\pi, \hat{\theta}_B|x) = \tilde{\rho}_1(\pi, \hat{\theta}_B|x) + \frac{1}{2^{2J+2}}.$$

Notons aussi que, comme  $\sigma = 1$ , l'équation (2.3) et le théorème 2.8 permettent de considérer  $S_0^T(x)$  comme approximation de  $m(x)$ , la densité marginale évaluée

au point  $x$ . Ceci prendra une importance particulière au chapitre 4. Rappelons nous également que les sommes tronquées  $S_l^T(x)$  peuvent s'exprimer comme

$$S_l^T(x) = \sum_{k=-N}^N k^l a_{Jk}(x) b_{0,Jk}, \quad (2.62)$$

où les coefficients  $a_{Jk}(x)$  et  $b_{0,Jk}$  sont donnés par les équations (2.9) et (2.10). Les choix de  $N$  et du niveau d'approximation  $J$  déterminent clairement la qualité des approximations résultantes.

Notons finalement que la décroissance des deux densités du modèle bayésien est ici exponentielle. Ainsi, comme nous l'avons mentionné auparavant, nous proposons de choisir  $N$  selon

$$N \propto 2^J,$$

et donc de prendre

$$N = C2^J.$$

Le comportement des approximations obtenues pour diverses valeurs de  $J$  et de  $C$  est donc maintenant le point à l'étude.

Pour avoir une meilleure idée des valeurs intéressantes de  $C$ , nous renvoyons le lecteur aux deux critères identifiés à la section précédente. Les équations (2.55) et (2.56) nous amènent ici en effet à exiger de  $C$  que

$$\mathbb{P}_g(|\theta| < C) \simeq 1, \quad (2.63)$$

et que, pour la valeur observée  $x$ ,

$$\mathbb{P}_f(|\theta - x| < C) \simeq 1.$$

Maintenant, supposons que nous souhaitions plutôt approximer  $\hat{\theta}_B(x)$  et  $\rho(\pi, \hat{\theta}_B|x)$  pour des valeurs de  $x \in [x_{\min}, x_{\max}]$  avec, pour simplifier notre réflexion,  $x_{\min} \leq 0$  et  $x_{\max} \geq 0$ . Nous désirons donc avoir ici, d'une part,

$$\mathbb{P}_f(|\theta - x_{\min}| < C) \simeq 1, \quad (2.64)$$

et d'autre part,

$$\mathbb{P}_f(|\theta - x_{\max}| < C) \simeq 1. \quad (2.65)$$

De là, si  $C$  satisfait

$$C \geq \max\{6, |-3 + x_{\min}|, 3 + x_{\max}\}, \quad (2.66)$$

alors les équations (2.63) à (2.65) sont vérifiées. En effet, nous avons alors

$$\mathbb{P}_g(|\theta| < C) \geq \mathbb{P}_g(|\theta| < 6) \simeq 0,9975,$$

puis

$$\begin{aligned}
 \mathbb{P}_f(|\theta - x_{\min}| < C) &\geq \mathbb{P}_f(|\theta - x_{\min}| < |-3 + x_{\min}|) \\
 &= \mathbb{P}_f(|\theta - x_{\min}| < 3 - x_{\min}) \\
 &= \mathbb{P}_f(-3 + 2x_{\min} < \theta < 3) \\
 &\geq \mathbb{P}_f(-3 < \theta < 3) \\
 &\simeq 0,9973,
 \end{aligned}$$

et finalement,

$$\begin{aligned}
 \mathbb{P}_f(|\theta - x_{\max}| < C) &\geq \mathbb{P}_f(|\theta - x_{\max}| < 3 + x_{\max}) \\
 &= \mathbb{P}_f(-3 < \theta < 3 + 2x_{\max}) \\
 &\geq \mathbb{P}_f(-3 < \theta < 3) \\
 &\simeq 0,9973.
 \end{aligned}$$

Comme nous le verrons plus loin, le critère donné par l'équation (2.66) est légèrement conservateur, mais en général assez juste.

Avant de débiter notre étude plus approfondie du comportement des approximations de  $\hat{\theta}_B(x)$  et de  $\rho(\pi, \hat{\theta}_B|x)$ , nous examinons rapidement la densité marginale  $m$ , donnée par l'équation (2.61), et son approximation  $S_0^T(x)$ , représentées graphiquement par la figure 2.1. De cette figure, il apparaît clairement que les valeurs de  $x$  les plus importantes, pour l'approximation de l'estimateur de Bayes et de sa perte espérée *a posteriori*, sont incluses dans l'intervalle  $[-6, 6]$ . Ceci est confirmé par le fait que

$$\mathbb{P}_m(|X| < 6) \simeq 0,9959,$$

et donc que la très grande majorité de la masse associée à  $m$  est comprise dans cet intervalle. Autrement dit, une observation à l'extérieur de cet intervalle peut pratiquement être considérée comme aberrante, ou du moins comme suffisamment anormale pour que l'approximation de  $\hat{\theta}_B(x)$  et de  $\rho(\pi, \hat{\theta}_B|x)$  ne soit pas, à moins d'avis contraire, d'un très grand intérêt (comme dans un cas, par exemple, où les queues de  $m$  sont d'un intérêt particulier).

Notons l'étonnante précision de  $S_0^T(x)$ . L'approximation de  $m$  est ici calculée au niveau  $J = 1$  et avec  $C = 6$ , et il est en effet extrêmement étonnant que ces choix soient, à première vue, suffisants. Tout d'abord, d'après l'équation (2.66), la valeur minimale de  $C$  devrait ici être de 9. Ensuite, le corollaire 2.4 permet de conclure que les approximations sont bonnes si  $J$  est assez grand. Clairement, il est normal de s'attendre à ce que  $J = 1$  ne soit pas considéré comme assez

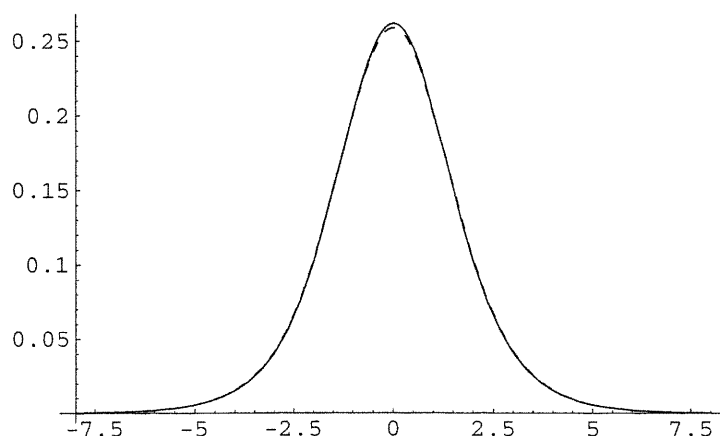


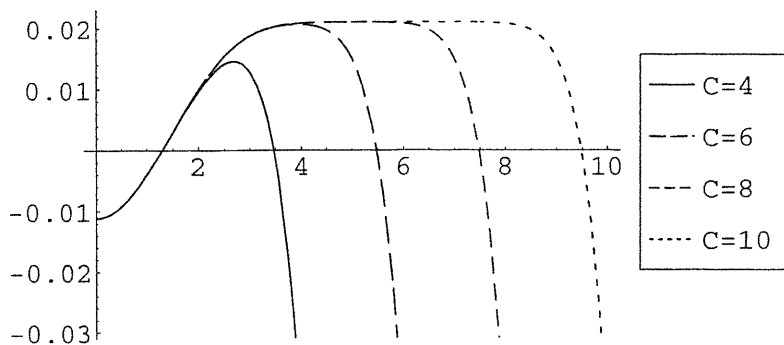
FIG. 2.1.  $m(x)$  (ligne pleine) et  $S_0^T(x)$  (tirets courts) au niveau  $J = 1$  pour  $C = 6$  et  $x \in [-8, 8]$ .

grand. C'est pourtant le cas ici. De plus, l'approximation initiale  $\tilde{f}_x$  de  $f_x$  sur laquelle repose toute notre procédure, donnée par l'équation (2.8), est une fonction constante par morceaux sur des intervalles de longueur  $2^{-J}$ . Il est surprenant de constater qu'ici, en approximant  $f_x$  par un histogramme dont les bandes ont une largeur égale à  $\frac{1}{2}$ , l'approximation obtenue est d'une telle qualité.

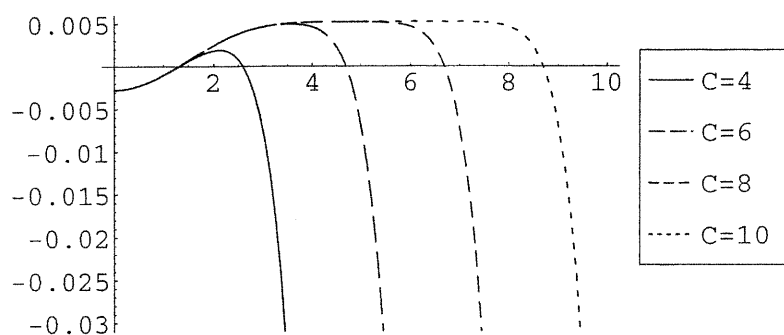
Remarquons finalement que  $S_0^T$  est une fonction très régulière. Même si cela peut paraître surprenant à cause de l'emploi de fonctions continues par morceaux pour la construction de  $\tilde{f}_x$  (et indirectement, de  $S_0^T$ ), l'approximation de  $m$  obtenue ici semble en effet très régulière. Une réflexion toute simple permet de comprendre pourquoi ceci est effectivement le cas. Notons que  $S_0^T(x)$  ne dépend de  $x$  qu'à travers les coefficients  $a_{Jk}(x)$  qui, à leur tour, ne dépendent de  $x$  qu'à travers  $F$ . Comme ici  $F$  correspond à la fonction de répartition normale centrée et réduite, alors  $F \in C^\infty(\mathbb{R})$  et c'est aussi le cas pour  $S_0^T$ . Le même argument s'applique aux approximations  $\tilde{\theta}_1(x)$  et  $\tilde{\theta}_2(x)$  de l'estimateur de Bayes, ainsi qu'aux approximations  $\tilde{\rho}_1(\pi, \hat{\theta}_B|x)$  et  $\tilde{\rho}_2(\pi, \hat{\theta}_B|x)$  de la perte espérée *a posteriori* de l'estimateur bayésien, qui sont donc aussi, comme nous le verrons plus loin, des fonctions (de  $x$ ) très régulières.

La figure 2.2(a) permet de confirmer la qualité de l'approximation de  $m(x)$ , de niveau  $J = 1$ , construite ici. Nous y présentons l'erreur relative de  $S_0^T(x)$  associée à l'approximation de  $m(x)$ , et donnée par

$$\text{Erreur relative de } S_0^T(x) = \frac{S_0^T(x) - m(x)}{m(x)},$$



(a)

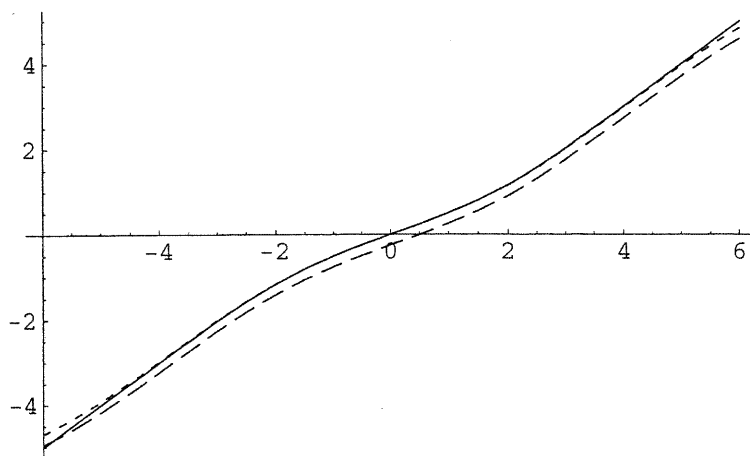


(b)

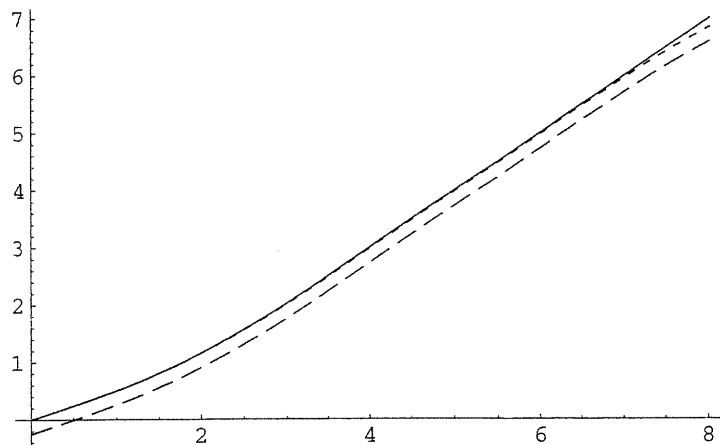
FIG. 2.2. Erreur relative de  $S_0^T(x)$  pour  $x \in [0, 10]$ , pour plusieurs valeurs de  $C$  et aux niveaux (a)  $J = 1$ , (b)  $J = 2$ .

au niveau  $J = 1$ , pour plusieurs valeurs de  $C$  et pour  $x \in [0, 10]$ . Ce graphique permet de constater que pour l'intervalle  $[0, 5]$  (ou  $[-5, 5]$ , par symétrie), l'approximation  $S_0^T(x)$  est effectivement excellente lorsque  $C = 6$ . Par contre, ce nouveau graphique met en lumière une détérioration rapide de l'approximation à l'extérieur de cet intervalle. Ainsi, dans l'éventualité où une qualité suffisante est nécessaire pour tout l'intervalle  $[-6, 6]$  ciblé plus haut, le passage à  $C = 8$  s'avère souhaitable. Dans ce cas, le phénomène de détérioration rapide survient en effet dans un voisinage de  $x = 7$ . La figure 2.2(b) présente, à des fins de comparaison, l'erreur relative de  $S_0^T(x)$  au niveau  $J = 2$ . Notons l'amélioration évidente apportée par l'augmentation du niveau, bien que pour  $C = 6$ , le comportement





(a)

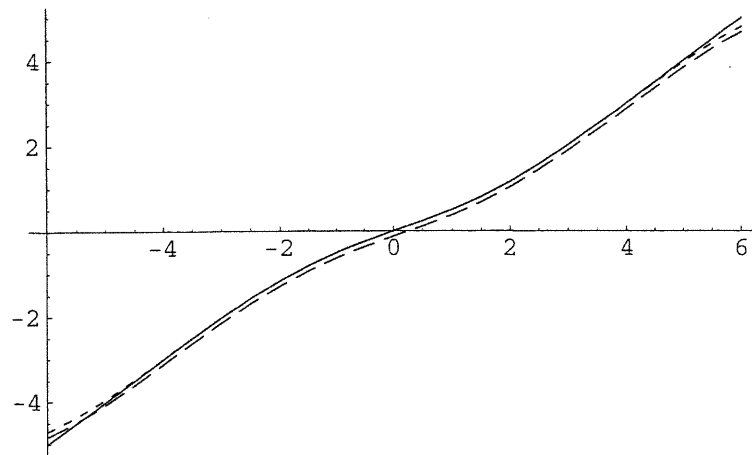


(b)

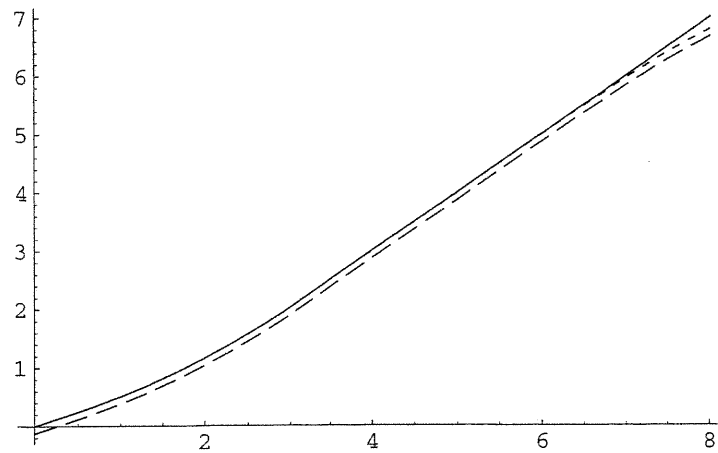
FIG. 2.3.  $\hat{\theta}_B(x)$  (ligne pleine),  $\tilde{\theta}_1(x)$  (tirets longs) et  $\tilde{\theta}_2(x)$  (tirets courts) au niveau  $J = 1$  pour (a)  $C = 6$  et  $x \in [-6, 6]$ , (b)  $C = 8$  et  $x \in [0, 8]$ .

de l'approximation demeure problématique à l'extérieur de l'intervalle  $[-5, 5]$ . Encore une fois, il peut donc s'avérer préférable d'augmenter la valeur de  $C$ .

Étudions maintenant le comportement des deux types d'approximation de l'estimateur de Bayes. Les figures 2.3 et 2.4 permettent de voir l'impact du terme correctif ajouté à  $\tilde{\theta}_1(x)$  pour obtenir la forme corrigée  $\tilde{\theta}_2(x)$ . En effet, même pour les petites valeurs que sont  $J = 1$  et 2, la forme corrigée semble extrêmement



(a)



(b)

FIG. 2.4.  $\hat{\theta}_B(x)$  (ligne pleine),  $\tilde{\theta}_1(x)$  (tirets longs) et  $\tilde{\theta}_2(x)$  (tirets courts) au niveau  $J = 2$  pour (a)  $C = 6$  et  $x \in [-6, 6]$ , (b)  $C = 8$  et  $x \in [0, 8]$ .

précise, alors que la forme simple reste relativement inefficace, même en augmentant la valeur de  $C$ . Ainsi, la correction utilisée nous permet d'obtenir une bonne approximation de l'estimateur de Bayes même si le niveau choisi est étonnamment petit.

Par ailleurs, il est clair qu'un mauvais choix de  $C$  est désastreux pour la performance des deux types d'approximation. En effet, la figure 2.5 permet de constater que si  $C$  est choisi trop petit (ici  $C = 4$ ), alors même l'ajout du terme

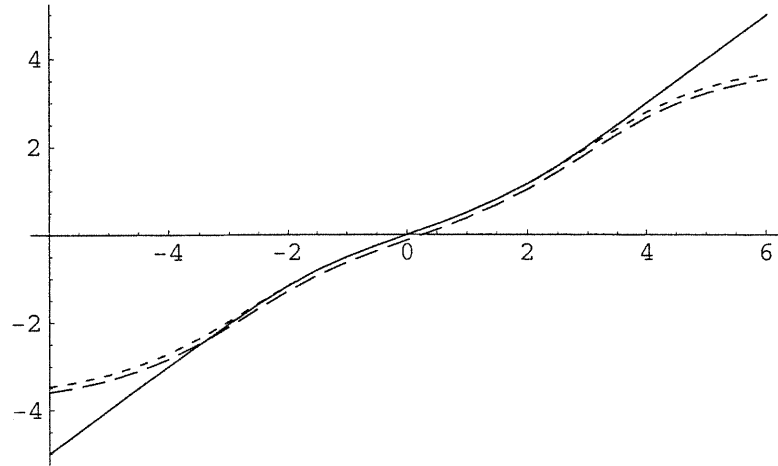


FIG. 2.5.  $\hat{\theta}_B(x)$  (ligne pleine),  $\tilde{\theta}_1(x)$  (tirets longs) et  $\tilde{\theta}_2(x)$  (tirets courts) au niveau  $J = 2$  pour  $C = 4$  et  $x \in [-6, 6]$ .

correctif ne corrige pas la situation. Dans un tel cas, la performance de  $\tilde{\theta}_1(x)$  et de  $\tilde{\theta}_2(x)$  est mauvaise dès que la valeur de  $x$  s'éloigne de la moyenne *a priori* qui est ici  $\mu = 0$ . Ceci est confirmé par la figure 2.6 où nous présentons l'erreur relative de  $\tilde{\theta}_2(x)$ , donnée par

$$\text{Erreur relative de } \tilde{\theta}_2(x) = \frac{\tilde{\theta}_2(x) - \hat{\theta}_B(x)}{\hat{\theta}_B(x)},$$

lorsque  $\hat{\theta}_B(x) \neq 0$ . En effet, nous voyons grâce à cette figure que même si le niveau est choisi aussi petit que  $J = 1$  ou  $2$ , l'approximation résultante est satisfaisante pourvu que  $C$  soit choisi assez grand. En particulier, la figure 2.6 permet de voir que lorsque  $C = 8$ ,  $\tilde{\theta}_2(x)$  a un bon comportement sur tout l'intervalle des valeurs les plus plausibles de  $x$ , soit l'intervalle  $[0, 6]$  (et par symétrie, sur  $[-6, 6]$ ), et ce pour les deux niveaux considérés ici. Les approximations de l'estimateur de Bayes semblent donc beaucoup plus sensibles au choix de  $C$  qu'à celui de  $J$ .

Cette dernière remarque est fort intéressante puisqu'il est ainsi possible d'obtenir une approximation satisfaisante, même pour les valeurs de  $x$  plus grandes, en augmentant  $C$  suffisamment, mais sans augmenter le niveau de l'approximation. L'intérêt de ceci vient du fait que les sommes tronquées  $S_i^T(x)$  contiennent chacune  $2N + 1 = C2^{J+1} + 1$  termes. Ainsi, une augmentation de  $C$  amène une augmentation linéaire du nombre de termes à considérer, alors qu'une augmentation du niveau  $J$  se traduit par une augmentation exponentielle du nombre de termes de ces sommes. Ceci permet donc, en cas de besoin, d'espérer améliorer

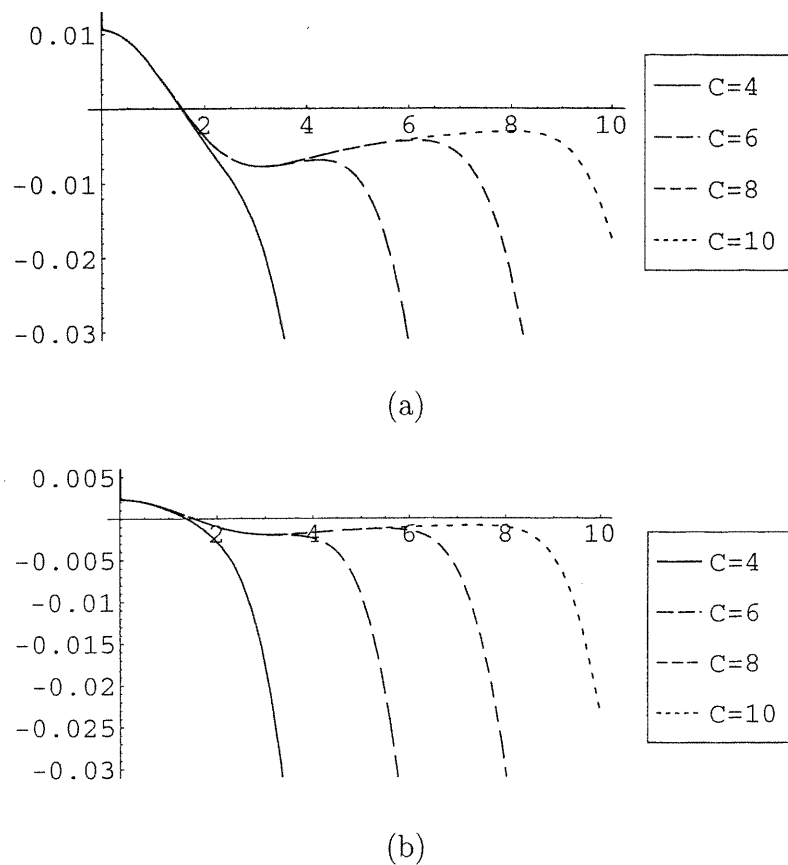


FIG. 2.6. Erreur relative de  $\tilde{\theta}_2(x)$  pour  $x \in [0, 10]$ , pour plusieurs valeurs de  $C$  et aux niveaux (a)  $J = 1$ , (b)  $J = 2$ .

la précision de l'approximation de l'estimateur bayésien sans pour autant voir croître le temps de calcul nécessaire de façon trop considérable.

Comparons maintenant les deux formes d'approximation de  $\rho(\pi, \hat{\theta}_B|x)$ . Une première constatation nous vient vite à l'esprit en examinant la figure 2.7(a). Les approximations  $\tilde{\rho}_1(\pi, \hat{\theta}_B|x)$  et  $\tilde{\rho}_2(\pi, \hat{\theta}_B|x)$  semblent beaucoup plus sensibles aux choix de  $J$  et de  $C$  que les approximations de l'estimateur de Bayes. En effet, cette figure illustre bien le fait que dans le cas où  $J = 1$ , les approximations calculées ne sont pas tout à fait adéquates, et ce même en passant de  $C = 6$  à  $C = 8$ . Un autre point nettement illustré par ces deux figures est le fait que le facteur correctif ajouté à  $\tilde{\rho}_1(\pi, \hat{\theta}_B|x)$  pour obtenir  $\tilde{\rho}_2(\pi, \hat{\theta}_B|x)$  n'a pas ici le rôle bénéfique

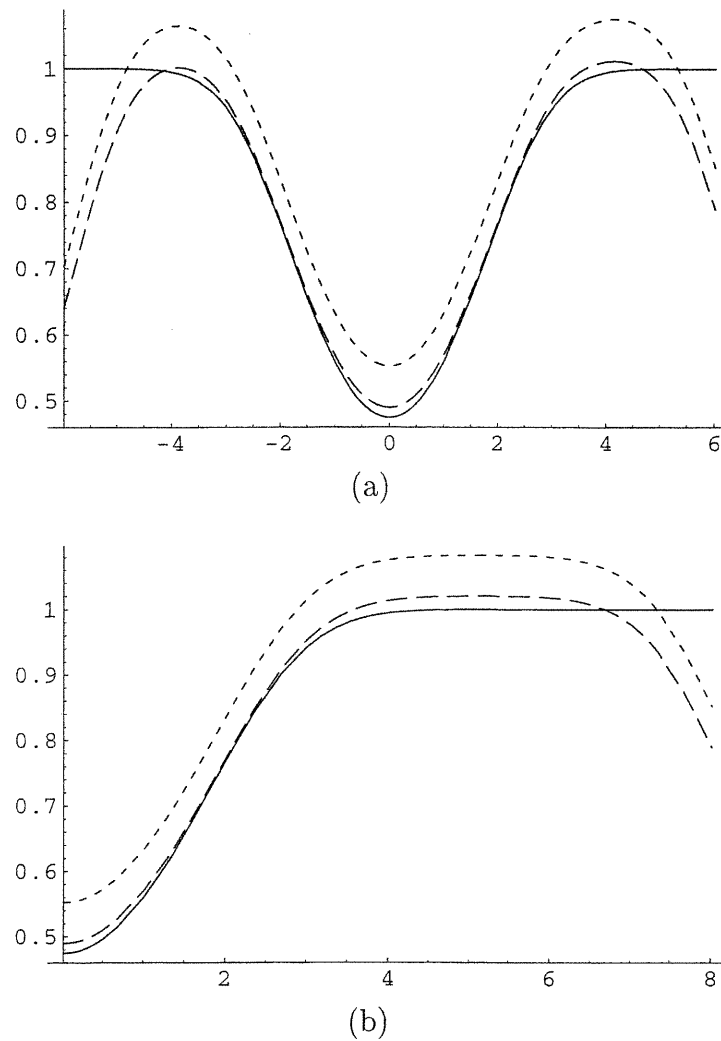


FIG. 2.7.  $\rho(\pi, \hat{\theta}_B|x)$  (ligne pleine),  $\tilde{\rho}_1(\pi, \hat{\theta}_B|x)$  (tirets longs) et  $\tilde{\rho}_2(\pi, \hat{\theta}_B|x)$  (tirets courts) au niveau  $J = 1$  pour (a)  $C = 6$  et  $x \in [-6, 6]$ , (b)  $C = 8$  et  $x \in [0, 8]$ .

observé plus tôt lors de l'approximation de l'estimateur de Bayes. En effet, l'ajout du terme correctif amène ici une détérioration de l'approximation. Ce phénomène étonnant est aussi observé aux figures 2.8 (lorsque  $J = 2$ ) et 2.9 (lorsque  $J = 3$ ). La qualité de la forme simple d'approximation semble donc ici supérieure à celle de la forme corrigée. L'emploi de la forme d'approximation corrigée semble donc inutile, et même à éviter dans le cas des plus petites valeurs de  $J$ . En effet, dans ce cas  $\tilde{\rho}_1(\pi, \hat{\theta}_B|x)$  permet quand même de bien saisir l'allure de  $\rho(\pi, \hat{\theta}_B|x)$  (si  $C$

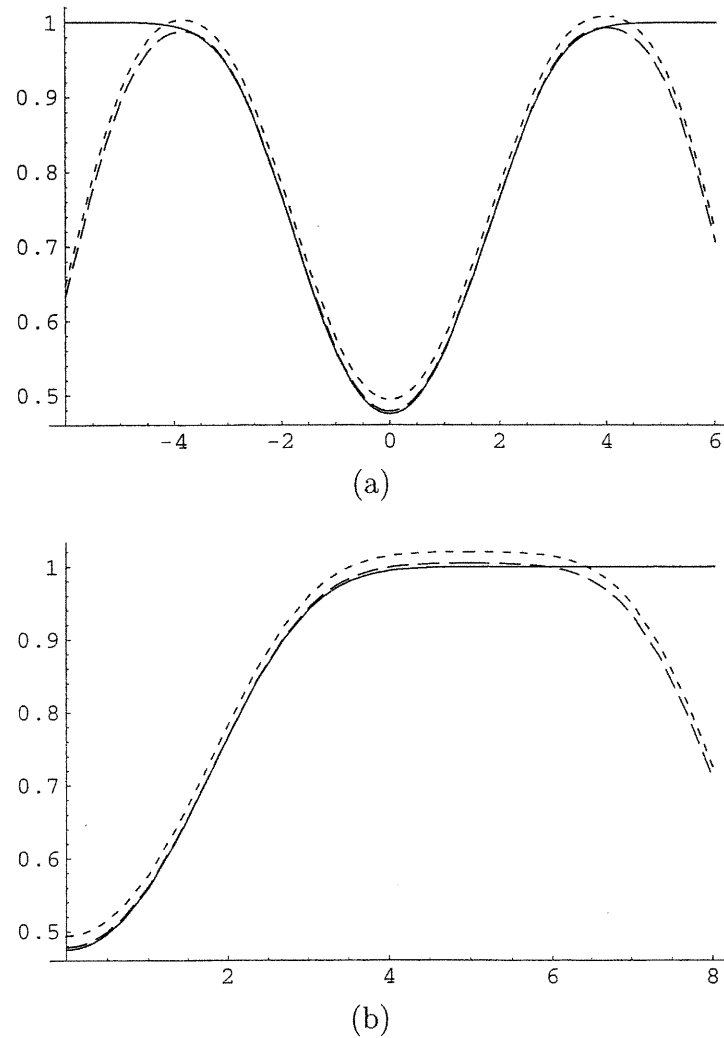


FIG. 2.8.  $\rho(\pi, \hat{\theta}_B|x)$  (ligne pleine),  $\tilde{\rho}_1(\pi, \hat{\theta}_B|x)$  (tirets longs) et  $\tilde{\rho}_2(\pi, \hat{\theta}_B|x)$  (tirets courts) au niveau  $J = 2$  pour (a)  $C = 6$  et  $x \in [-6, 6]$ , (b)  $C = 8$  et  $x \in [0, 8]$ .

est suffisamment grand) et peut très bien servir d'approximation grossière, alors que  $\tilde{\rho}_2(\pi, \hat{\theta}_B|x)$  a une tendance flagrante à la surestimation de  $\rho(\pi, \hat{\theta}_B|x)$ .

La figure 2.8 permet également de constater que même dans le cas où le niveau est augmenté à  $J = 2$ , le comportement de  $\tilde{\rho}_1(\pi, \hat{\theta}_B|x)$  n'est pas satisfaisant pour  $C = 6$ . Il nous faut passer à  $C = 8$  pour obtenir une approximation intéressante de la perte espérée *a posteriori* sur tout l'intervalle des valeurs plausibles pour  $x$

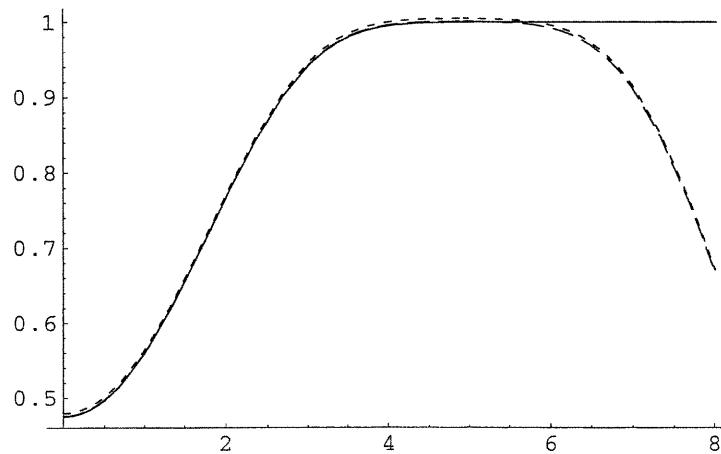


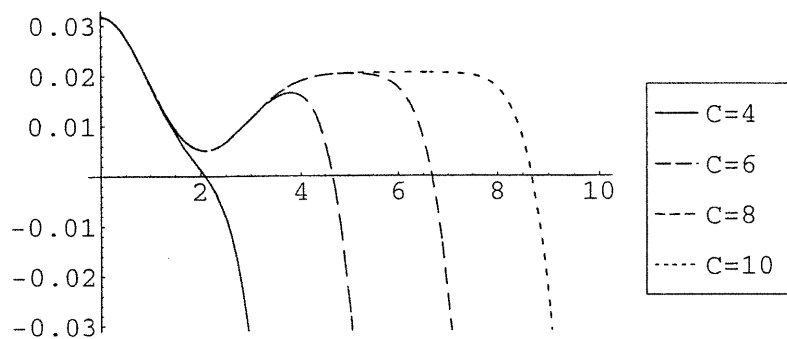
FIG. 2.9.  $\rho(\pi, \hat{\theta}_B|x)$  (ligne pleine),  $\tilde{\rho}_1(\pi, \hat{\theta}_B|x)$  (tirets longs) et  $\tilde{\rho}_2(\pi, \hat{\theta}_B|x)$  (tirets courts) au niveau  $J = 3$  pour  $C = 8$  et  $x \in [0, 8]$ .

donné par  $x \in [-6, 6]$ . En gardant  $C = 8$ , et en passant au niveau  $J = 3$ , l'approximation de la perte espérée *a posteriori*  $\rho(\pi, \hat{\theta}_B|x)$  prend une allure vraiment intéressante, ce qui est confirmé par la figure 2.9. En effet, cette figure permet de voir que pour cette combinaison de  $J$  et  $C$ , les approximations  $\tilde{\rho}_1(\pi, \hat{\theta}_B|x)$  et  $\tilde{\rho}_2(\pi, \hat{\theta}_B|x)$  offrent une excellente performance sur  $[0, 6]$  (et  $[-6, 6]$ , par symétrie).

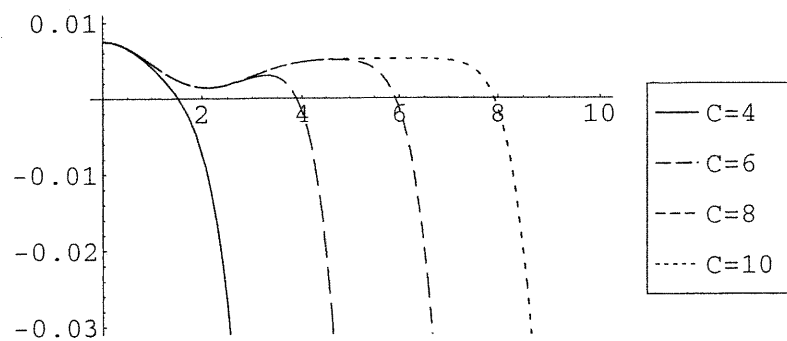
La figure 2.10 illustre bien le fait que, encore une fois, les approximations construites semblent beaucoup plus sensibles au choix de  $C$  qu'au choix de  $J$ . Nous y présentons l'erreur relative associée à  $\tilde{\rho}_1(\pi, \hat{\theta}_B|x)$ , donnée par

$$\text{Erreur relative de } \tilde{\rho}_1(\pi, \hat{\theta}_B|x) = \frac{\tilde{\rho}_1(\pi, \hat{\theta}_B|x) - \rho(\pi, \hat{\theta}_B|x)}{\rho(\pi, \hat{\theta}_B|x)},$$

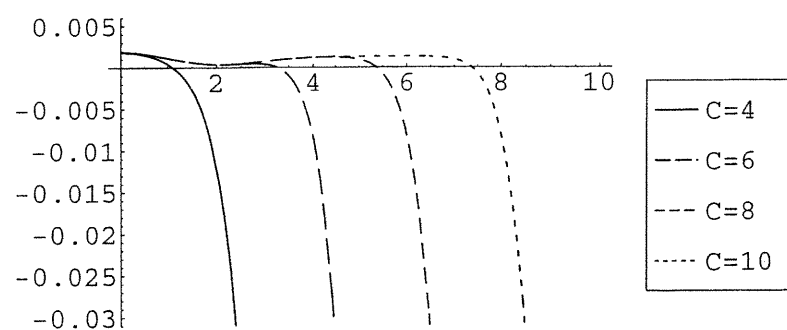
pour différentes combinaisons de  $J$  et  $C$ . Il est ainsi possible de constater que même si, pour l'approximation de l'estimateur de Bayes,  $J = 1$  et  $C = 6$  constituaient des choix minimaux, et  $J = 2$  et  $C = 8$  des choix amplement suffisants, ces mêmes choix de  $J = 2$  et  $C = 8$  sont maintenant minimaux, et il peut être préférable de passer, dans certains cas, à  $J = 3$  et/ou  $C = 10$ . En effet, la détérioration rapide de la qualité de  $\tilde{\rho}_1(\pi, \hat{\theta}_B|x)$  observée à la figure 2.10 pour  $x > 6$  permet de voir que  $C = 8$  n'est pas toujours un choix suffisant, et ce pour les trois niveaux considérés ici. La figure 10(c) permet de voir, en particulier, que ce phénomène de détérioration se produit pour  $x > 8$  dans le cas où  $C = 10$  et  $J = 3$ .



(a)



(b)



(c)

FIG. 2.10. Erreur relative de  $\tilde{\rho}_1(\pi, \hat{\theta}_B|x)$  pour  $x \in [0, 10]$ , pour plusieurs valeurs de  $C$  et aux niveaux (a)  $J = 1$ , (b)  $J = 2$ , (c)  $J = 3$ .



### 2.7.2. Le modèle normale/Cauchy

Nous supposons cette fois que

$$X|\theta \sim N(\theta, 1),$$

et que

$$\theta \sim \text{Cauchy}(0, 1).$$

Ainsi, la vraisemblance demeure inchangée et est toujours donnée par

$$L(\theta|x) = f(x - \theta) = \phi(x - \theta),$$

où  $\phi$  correspond à la densité normale centrée et réduite. Par ailleurs, la densité *a priori* du paramètre de position  $\theta$  est maintenant donnée par

$$g(\theta) = \frac{1}{\pi(1 + \theta^2)},$$

pour  $\theta \in \mathbb{R}$ . Notons qu'ici, contrairement à l'exemple précédent, la densité marginale  $m$ , l'estimateur de Bayes  $\hat{\theta}_B(x)$  et sa perte espérée *a posteriori* ne peuvent être obtenus sous forme explicite. De plus, remarquons que la densité  $g$  du modèle considéré ne satisfait pas les hypothèses permettant d'appliquer directement la méthodologie développée.

En effet,  $g_l \notin \mathcal{L}^1(\mathbb{R})$  pour  $l = 1, 2$ . Il nous faut donc utiliser l'approche alternative mentionnée à la fin de la section précédente, soit inverser les rôles de  $f$  et  $g$  dans l'approximation des intégrales  $\mathbb{I}_l(x)$ . Le lecteur peut vérifier que sous cette forme inversée,  $f$  et  $g$  respectent toutes les conditions nécessaires. Nous utilisons ici les approximations suivantes, définies dans la même foulée qu'à l'exemple précédent, mais en s'appuyant sur les équations (2.57) à (2.60) plutôt que sur les équations (2.2) à (2.5). Les formes tronquées de  $\hat{\theta}_B^S(x)$  et  $\tilde{\rho}^S(\pi, \hat{\theta}_B|x)$  sont ainsi données par

$$\tilde{\theta}_1(x) = x - \frac{1}{2^J} \frac{S_1^T(x)}{S_0^T(x)}, \quad (2.67)$$

et

$$\tilde{\rho}_1(\pi, \hat{\theta}_B|x) = \frac{1}{2^{2J}} \left[ \frac{S_2^T(x)}{S_0^T(x)} - \left( \frac{S_1^T(x)}{S_0^T(x)} \right)^2 \right], \quad (2.68)$$

et les formes tronquées de  $\tilde{\theta}_B^*(x)$  et  $\tilde{\rho}^*(\pi, \hat{\theta}_B|x)$  par

$$\tilde{\theta}_2(x) = \tilde{\theta}_1(x) + \frac{1}{2^{J+1}}, \quad (2.69)$$

et

$$\tilde{\rho}_2(\pi, \hat{\theta}_B | x) = \tilde{\rho}_1(\pi, \hat{\theta}_B | x) + \frac{1}{2^{2J+2}}. \quad (2.70)$$

De plus, comme  $\sigma = 1$ ,  $S_0^T(x)$  constitue ici aussi une approximation de  $m(x)$ . La différence notable avec les expressions employées à l'exemple précédent, à l'exception de la définition de  $\tilde{\theta}_1(x)$ , vient de la façon de calculer ici les coefficients  $a_{Jk}(x)$  et  $b_{0,Jk}$  composants les sommes tronquées  $S_l^T(x)$ , qui, quant à elles, sont toujours données par l'équation (2.62). En effet, comme nous l'avons déjà dit, cette forme d'approximation amène les rôles de  $f$  et  $g$  à être inversés lors du calcul des coefficients  $a_{Jk}(x)$  et  $b_{0,Jk}$ . Ces coefficients sont donc ici plutôt donnés par

$$b_{0,Jk} = 2^{J/2} \left[ F\left(\frac{k+1}{2^J}\right) - F\left(\frac{k}{2^J}\right) \right],$$

et

$$a_{Jk}(x) = 2^{J/2} \left[ G\left(x - \frac{k}{2^J}\right) - G\left(x - \frac{k+1}{2^J}\right) \right].$$

Pour  $N$ , la décroissance exponentielle de  $f$  nous amène à prendre, comme à l'exemple précédent,

$$N = C2^J.$$

Finalement, les choix de  $C$  et du niveau d'approximation  $J$  déterminent toujours la qualité des diverses approximations résultantes.

Afin d'identifier les valeurs intéressantes de  $C$ , étudions rapidement les équations (2.55) et (2.56), qui nous amènent cette fois à exiger de  $C$  que

$$\mathbb{P}_f(|\theta| < C) \simeq 1, \quad (2.71)$$

et que, pour la valeur observée  $x$ ,

$$\mathbb{P}_g(|\theta - x| < C) \simeq 1. \quad (2.72)$$

Notons d'abord que si  $C \geq 3$ , alors

$$\mathbb{P}_f(|\theta| < C) \geq \mathbb{P}_f(|\theta| < 3) \simeq 0,9973,$$

et donc la condition associée à l'équation (2.71) est satisfaite. Par ailleurs, l'application à la lettre de la deuxième condition n'a pas ici de sens. En effet, comme  $g$  correspond à la densité de Cauchy, l'équation (2.72) demande donc que  $C$  prenne une valeur très grande, ce qui est en pratique inutile. À titre d'exemple, pour l'approximation des quantités d'intérêts, lorsque  $x = 0$ , nous avons

$$\mathbb{P}_g(|\theta| < 64) \simeq 0,990054,$$

ce qui suggère qu'une valeur de  $C$  adéquate serait certainement supérieure à 64, peut-être autour de 75 ou de 80. Une telle valeur de  $C$  rend extrêmement lourd le calcul des diverses approximations définies plus haut. Heureusement, comme nous le verrons plus tard dans cette même section, des valeurs de  $C$  aussi grandes ne sont ici pas nécessaires.

Maintenant, revenons sur le fait que pour le modèle considéré, la densité marginale  $m$ , l'estimateur de Bayes ainsi que la perte espérée *a posteriori* de l'estimateur de Bayes ne peuvent être obtenus sous une forme explicite. Une difficulté nouvelle se présente donc ici. En effet, il nous est dorénavant impossible de comparer les approximations utilisées aux valeurs calculées des quantités d'intérêt. Afin de s'assurer du bon comportement des approximations définies plus haut, nous utilisons une deuxième méthode pour approximer la densité marginale, l'estimateur de Bayes et sa perte espérée *a posteriori*. Cette méthode est simple, quoique numériquement très exigeante, et s'apparente à une simulation Monte-Carlo. Tout d'abord, en revenant aux équations (2.2) à (2.5), nous avons ici

$$m(x) = \mathbb{I}_0(x),$$

ainsi que

$$\hat{\theta}_B(x) = \frac{\mathbb{I}_1(x)}{\mathbb{I}_0(x)},$$

et finalement,

$$\rho(\pi, \hat{\theta}_B|x) = \frac{\mathbb{I}_2(x)}{\mathbb{I}_0(x)} - \left( \frac{\mathbb{I}_1(x)}{\mathbb{I}_0(x)} \right)^2.$$

Dans le présent contexte, les intégrales  $\mathbb{I}_l(x)$  (pour  $l = 0, 1, 2$ ) peuvent être vues comme

$$\mathbb{I}_l(x) = \int_{\mathbb{R}} g_l(y) \phi(x-y) dy = \mathbb{E}[g_l(Y)], \quad (2.73)$$

où la variable aléatoire  $Y$  satisfait

$$Y \sim N(x, 1).$$

L'équation (2.73) rend naturelle l'approximation des intégrales  $\mathbb{I}_l(x)$  par

$$\tilde{\mathbb{E}}_l(x) = \frac{1}{n} \sum_{i=1}^n g_l(y_i) = \frac{1}{n\pi} \sum_{i=1}^n \frac{y_i^l}{1+y_i^2},$$

où les points  $y_1, y_2, \dots, y_n$  satisfont

$$y_i = x + \Phi^{-1} \left( \frac{i}{n+1} \right),$$

c'est-à-dire où  $y_i$  est le quantile d'ordre  $i/(n+1)$  de la densité  $N(x,1)$ . Cette façon de faire est en effet naturelle puisque  $y_1, y_2, \dots, y_n$  correspondent en quelque sorte à un pseudo-échantillon parfaitement représentatif de la densité normale d'espérance  $x$  et de variance 1. Cette méthode est donc effectivement très similaire à une simulation Monte-Carlo, mais peut être aussi vue comme l'application d'une règle d'intégration numérique classique. Davis et Rabinowitz (1984, section 2.6) présentent cette règle sous le nom de règle du point milieu généralisée ("generalized mid-point rule"). Les approximations de la densité marginale, de l'estimateur de Bayes et de sa perte espérée *a posteriori* découlant de cette réflexion sont données par

$$m_{gm}(x) = \tilde{\mathbb{E}}_0(x),$$

et

$$\hat{\theta}_{gm}(x) = \frac{\tilde{\mathbb{E}}_1(x)}{\tilde{\mathbb{E}}_0(x)},$$

ainsi que, finalement,

$$\rho_{gm}(x) = \frac{\tilde{\mathbb{E}}_2(x)}{\tilde{\mathbb{E}}_0(x)} - \left( \frac{\tilde{\mathbb{E}}_1(x)}{\tilde{\mathbb{E}}_0(x)} \right)^2.$$

Ces trois nouvelles approximations sont ici utilisées comme base de comparaison afin de valider les approximations construites à l'aide de la base de Haar. Dans chaque cas, nous avons utilisé  $n = 2000$  afin de nous assurer d'une très grande précision.

Nous débutons en étudiant le comportement de  $S_0^T$ . Pour cela, examinons rapidement la figure 2.11 qui présente  $S_0^T(x)$  et  $m_{gm}(x)$  pour  $x \in [-10, 10]$ . Tout d'abord, en examinant cette figure, il semble que  $S_0^T$  est, comme à l'exemple précédent, une fonction extrêmement régulière. Ceci est confirmé par le fait que, encore une fois,  $S_0^T \in C^\infty(\mathbb{R})$ . De plus, il est aussi clair que  $S_0^T$  et  $m_{gm}$  sont des approximations très similaires de  $m$ , et ce même si  $S_0^T$  a été calculée au niveau  $J = 1$  et avec  $C = 4$ . L'emploi de cette dernière valeur peut sembler particulièrement étonnante, surtout en considérant qu'à l'exemple précédent une valeur de  $C = 8$  était habituellement nécessaire, et ce dans un contexte où les deux densités du modèle bayésien exhibaient une décroissance exponentielle! Ceci illustre très bien l'affirmation faite à la toute fin de la section précédente. En effet, contrairement à l'exemple précédent, nous avons ici calculé les coefficients  $b_{0,Jk}$  à l'aide de la densité la plus concentrée, soit  $f$ . Dans de telles conditions, le choix

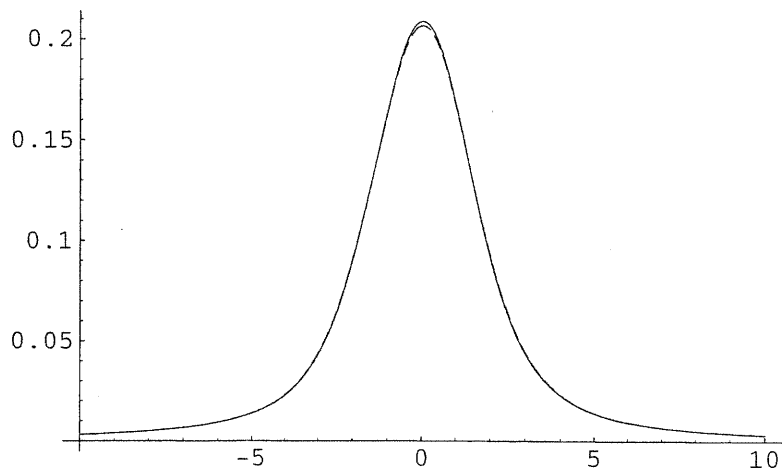


FIG. 2.11.  $m_{gm}(x)$  (ligne pleine) et  $S_0^T(x)$  (tirets) au niveau  $J = 1$ , pour  $C = 4$  et  $x \in [-10, 10]$ .

de  $C$  est principalement déterminé par

$$\mathbb{P}_f(|\theta| < C) \simeq 1,$$

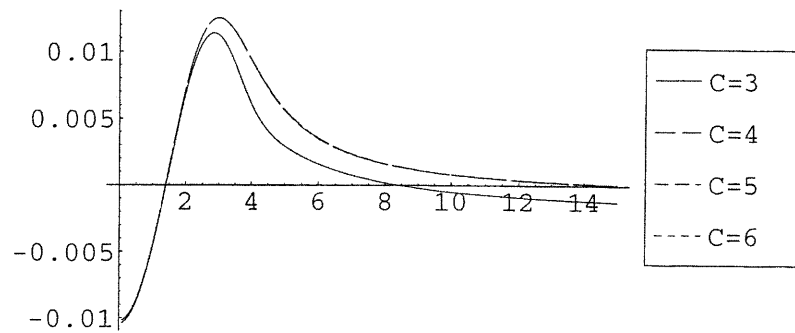
alors qu'à l'inverse, lorsque les coefficients  $a_{Jk}(x)$  étaient obtenus à partir de la densité la plus concentrée, ce choix reposait principalement sur

$$\mathbb{P}_f(|\theta - x| < C) \simeq 1,$$

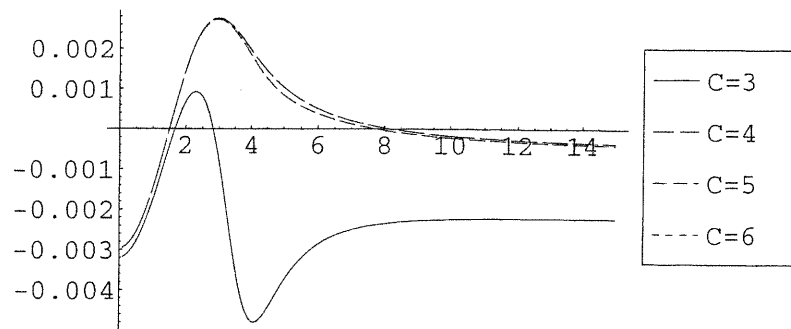
pour toutes les valeurs intéressantes de  $x \in [x_{\min}, x_{\max}]$ . Il est donc effectivement préférable de calculer les coefficients  $b_{0,Jk}$  à l'aide de la densité la plus concentrée puisqu'alors les coefficients de  $S_0^T(x)$  les plus importants sont moins nombreux et sont toujours les mêmes, quelque soit la valeur observée  $x$ . Ceci vient du fait que la densité la plus concentrée (ici la densité normale centrée et réduite) peut être considérée comme négligeable à l'extérieur d'un intervalle beaucoup plus court que les densités *a priori* considérées jusqu'ici (soient la densité de Cauchy, et à l'exemple précédent, la densité double exponentielle). Notons que cette réflexion vaut pour toutes les sommes tronquées  $S_l^T(x)$ , et donc aussi pour les diverses approximations de  $\hat{\theta}_B(x)$  et de  $\rho(\pi, \hat{\theta}_B|x)$ .

Dans le but de confirmer la qualité de  $S_0^T$  comme approximation de  $m$ , nous nous intéressons ensuite à l'erreur relative approximative de  $S_0^T(x)$ , associée à l'approximation de  $m(x)$ , que nous définissons comme

$$\text{Erreur relative approximative de } S_0^T(x) = \frac{S_0^T(x) - m_{gm}(x)}{m_{gm}(x)}.$$



(a)



(b)

FIG. 2.12. Erreur relative approximative de  $S_0^T(x)$  pour  $x \in [0, 15]$ , pour plusieurs valeurs de  $C$  et aux niveaux (a)  $J = 1$ , (b)  $J = 2$ .

À la figure 2.12, nous présentons cette erreur relative approximative aux niveaux  $J = 1$  et  $2$ , pour plusieurs valeurs de  $C$  et pour  $x \in [0, 15]$ . Les deux graphes présentés permettent de constater qu'encore une fois le comportement de l'approximation  $S_0^T(x)$  est satisfaisant. Par ailleurs, il met aussi en lumière un aspect intéressant lié au calcul des coefficients  $b_{0,Jk}$  à partir de la densité la plus concentrée. En effet, il est très intéressant de voir que le phénomène de détérioration rapide observé précédemment ne se produit plus ici. Ainsi, l'approximation obtenue demeure adéquate même pour des valeurs de  $x$  possiblement très éloignées de la moyenne *a priori*  $\mu = 0$ . De plus, ceci se traduit aussi visuellement par le fait que les courbes deviennent confondues, au niveau  $J = 1$  pour

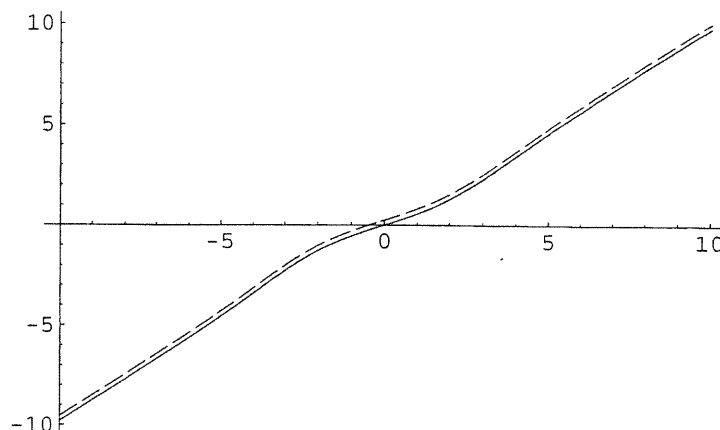


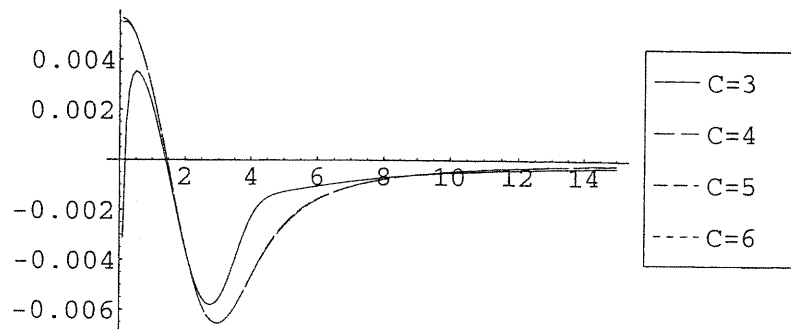
FIG. 2.13.  $\hat{\theta}_{gm}(x)$  (ligne pleine),  $\tilde{\theta}_1(x)$  (tirets longs) et  $\tilde{\theta}_2(x)$  (tirets courts) au niveau  $J = 1$  pour  $C = 4$  et  $x \in [-10, 10]$ .

des valeurs de  $C \geq 4$ , et au niveau  $J = 2$  pour des valeurs de  $C \geq 5$ , laissant entendre que l'augmentation de  $C$  au-delà d'une certaine valeur est inutile puisque la qualité de l'approximation résultante n'est pas vraiment affectée. Dans le cas du présent modèle, ces caractéristiques sont tout particulièrement intéressantes puisque

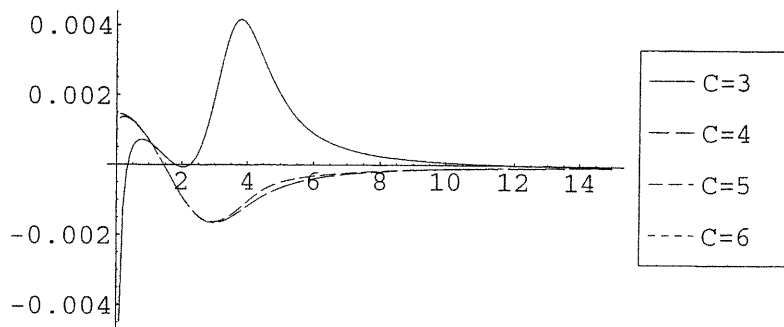
$$\mathbb{P}_m(|X| < 64) \simeq 0,990051,$$

et donc, la densité marginale  $m$  est une densité aux queues extrêmement relevées (en fait similaires aux queues de la densité de Cauchy), et il est tout à fait probable d'être confronté à une valeur de  $x$  éloignée de la moyenne *a priori*. Dans un tel contexte, il était normal de s'attendre à être contraint d'utiliser une très grande valeur de  $C$ , rendant par le fait même le calcul de l'approximation de  $m$  numériquement très exigeant. Heureusement, ce n'est pas le cas ici puisque le calcul des coefficients  $b_{0,Jk}$  est fait à l'aide de la densité la plus concentrée. À titre comparatif, nous présentons à la figure 2.12(b) cette même erreur relative approximative, mais cette fois pour l'approximation de niveau  $J = 2$ . Notons finalement que, des deux graphes de la figure 2.12, il se dégage que  $C = 3$  ne semble pas tout à fait suffisant, mais que  $C = 4$  fait très bien l'affaire et ce, autant au niveau  $J = 1$  qu'au niveau  $J = 2$ .

Nous nous portons maintenant vers l'étude du comportement des approximations de  $\hat{\theta}_B(x)$  données aux équations (2.67) et (2.69). La figure 2.13 permet de constater que  $\tilde{\theta}_2(x)$  semble avoir un comportement adéquat au niveau  $J = 1$



(a)



(b)

FIG. 2.14. Erreur relative approximative de  $\tilde{\theta}_2(x)$  pour plusieurs valeurs de  $C$ , pour  $x \in [0, 15]$  et aux niveaux (a)  $J = 1$ , (b)  $J = 2$ .

lorsque  $C = 4$ . En effet,  $\tilde{\theta}_2(x)$  et  $\hat{\theta}_{gm}(x)$  ont un comportement tout à fait similaire sur l'intervalle considéré, leur courbes étant même parfaitement confondues. Ainsi, comme à l'exemple précédent, l'approximation corrigée  $\tilde{\theta}_2(x)$  semble beaucoup plus performante que la forme simple  $\tilde{\theta}_1(x)$ , du moins lorsque la valeur de  $J$  est choisie aussi petite que  $J = 1$  ou  $2$ . Notons aussi que comme dans le cas de l'approximation de la densité marginale, l'approximation de l'estimateur de Bayes ne semble pas non plus se détériorer de façon dramatique lorsque la valeur observée  $x$  s'éloigne de la moyenne *a priori*  $\mu = 0$ . Pour confirmer ceci, nous présentons à la figure 2.14 l'erreur relative approximative de  $\tilde{\theta}_2(x)$  et liée à



l'approximation de  $\widehat{\theta}_B(x)$ , définie comme

$$\text{Erreur relative approximative de } \widetilde{\theta}_2(x) = \frac{\widetilde{\theta}_2(x) - \widehat{\theta}_{gm}(x)}{\widehat{\theta}_{gm}(x)},$$

lorsque  $\widehat{\theta}_{gm}(x) \neq 0$ . Plus particulièrement, la figure 2.14(a) présente l'erreur relative approximative de  $\widetilde{\theta}_2(x)$  au niveau  $J = 1$ , pour des valeurs bien choisies de  $C$  et pour  $x \in [0, 15]$ . La figure 2.14(b) fait de même, mais au niveau  $J = 2$ . En observant attentivement ces deux graphes, il devient effectivement apparent que le phénomène de détérioration rapide de la qualité de  $\widetilde{\theta}_2(x)$  observé à l'exemple 2.7.1 ne se produit pas non plus. Aussi, encore une fois, l'augmentation de  $C$ , au-delà de  $C = 4$  au niveau  $J = 1$  et de  $C = 5$  au niveau  $J = 2$ , n'améliore plus de façon concrète la performance des approximations construites et, surtout, ces approximations demeurent valides même si  $x$  s'éloigne sensiblement de la moyenne *a priori*. Tout ceci est dû, ici également, au fait que les coefficients  $b_{0,Jk}$  dépendent maintenant de la densité la plus concentrée.

Finalement, nous examinons le comportement des deux formes d'approximation de la perte espérée *a posteriori* données par les équations (2.68) et (2.70). Pour cela, considérons d'abord les figures 2.15 et 2.16. La première constatation qui s'impose en étudiant ces figures est que, comme à l'exemple précédent, c'est la forme corrigée d'approximation qui performe plutôt mal, alors que  $\widetilde{\rho}_1(\pi, \widehat{\theta}_B|x)$  s'en tire assez bien, même au niveau  $J = 1$ . De plus, une fois encore, le phénomène de détérioration rapide présent à l'exemple précédent n'est pas observé ici. Ceci est confirmé par la figure 2.17, où nous présentons l'erreur relative approximative de  $\widetilde{\rho}_1(\pi, \widehat{\theta}_B|x)$  à divers niveaux, pour quelques valeurs de  $C$  et  $x \in [-15, 15]$ . Nous définissons cette erreur relative comme prédéamment, soit

$$\text{Erreur relative approximative de } \widetilde{\rho}_1(\pi, \widehat{\theta}_B|x) = \frac{\widetilde{\rho}_1(\pi, \widehat{\theta}_B|x) - \rho_{gm}(x)}{\rho_{gm}(x)}.$$

Notons que la figure 2.17 permet de voir que, comme c'est le cas aux figures 2.12 et 2.14 avec les approximations de la densité marginale et de l'estimateur de Bayes, l'augmentation de  $C$  au-delà d'une certaine valeur ne change plus l'approximation résultante, ou du moins pas de façon appréciable. Dans le cas présent, ceci survient à partir de  $C = 5$  à chacun des trois niveaux considérés.

Ensuite, les figures 2.15(a) et 2.17(a) montrent clairement que  $C = 3$  n'est pas un choix adéquat au niveau  $J = 1$ . Notons l'asymétrie, particulièrement évidente à la figure 2.17(a), des deux formes d'approximation pour ces valeurs de  $C$  et de  $J$ . En effet, bien que toutes les approximations construites soient légèrement asymétriques, cette tendance est ici beaucoup plus prononcée. Ceci s'explique par

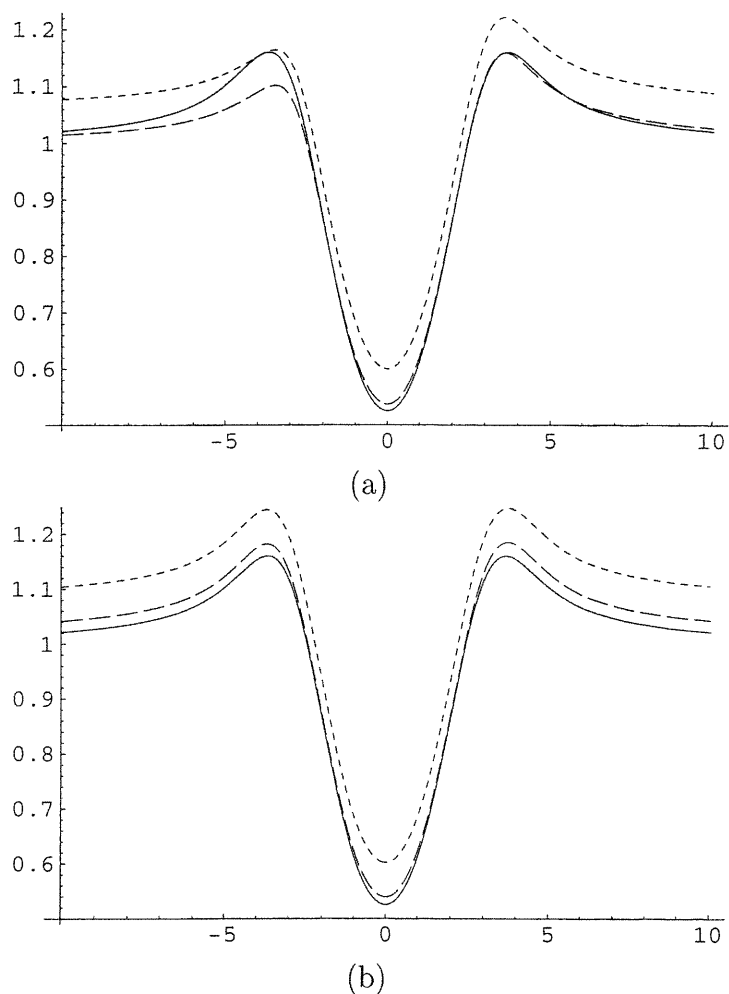
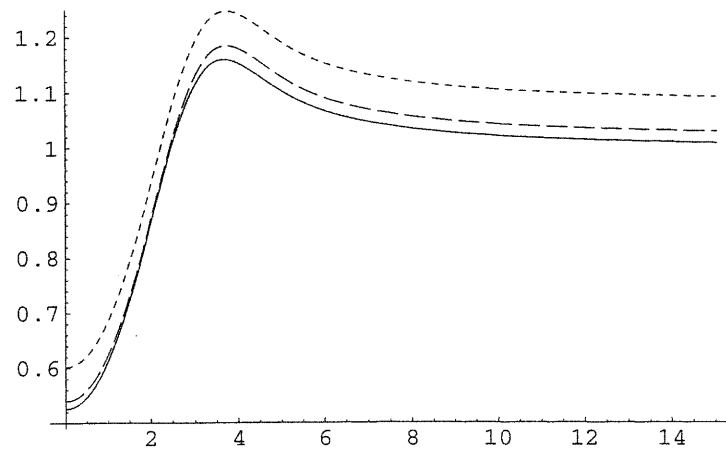
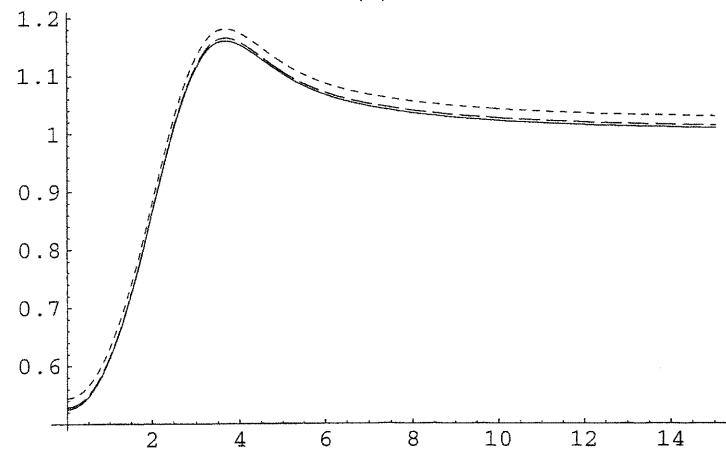


FIG. 2.15.  $\rho_{gm}(x)$  (ligne pleine),  $\tilde{\rho}_1(\pi, \hat{\theta}_B|x)$  (tirets longs) et  $\tilde{\rho}_2(\pi, \hat{\theta}_B|x)$  (tirets courts) au niveau  $J = 1$ , pour  $x \in [-10, 10]$ , et pour (a)  $C = 3$ , (b)  $C = 4$ .

le fait que les sommes tronquées  $S_l^T(x)$  comportent  $2N + 1$  termes dont  $N$  sont à indice  $k$  négatif ( $-N \leq k \leq -1$ ), et  $N + 1$  à indice  $k$  non négatif ( $0 \leq k \leq N$ ). Ainsi les diverses sommes tronquées comptent un terme non négatif de plus, donné par  $b_{0,JN} \times a_{JN}(x)$ . Mais, ce terme est non négligeable uniquement lorsque  $b_{0,JN}$  est non négligeable et  $x \simeq N/2^J$ . Ainsi, lorsque  $C$  est choisi trop petit, comme c'est le cas ici, ce terme supplémentaire n'apporte aucune contribution à  $S_l^T(x)$  lorsque  $x$  est négatif, alors que lorsque  $x$  est positif, il est assez important pour améliorer considérablement les approximations résultantes. Par contre, notons que lorsque  $C$  est choisi assez grand,  $b_{0,JN}$  devient négligeable et le terme positif



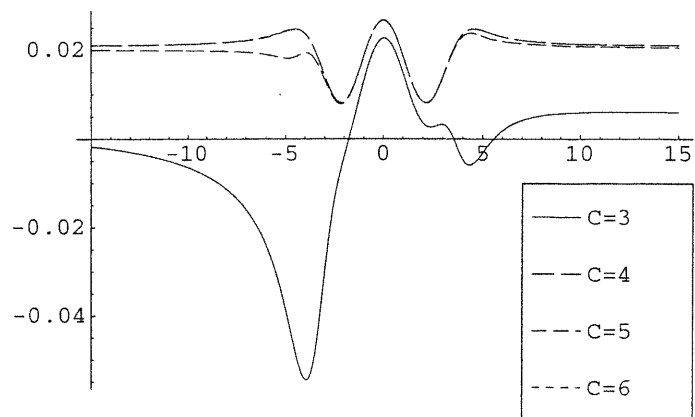
(a)



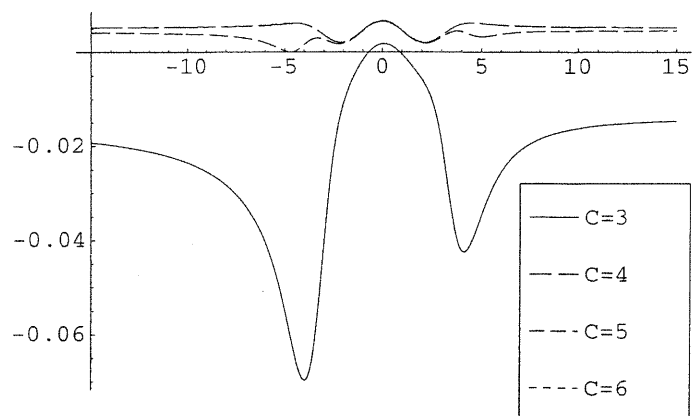
(b)

FIG. 2.16.  $\rho_{gm}(x)$  (ligne pleine),  $\tilde{\rho}_1(\pi, \hat{\theta}_B|x)$  (tirets longs) et  $\tilde{\rho}_2(\pi, \hat{\theta}_B|x)$  (tirets courts) pour  $C = 4$ , pour  $x \in [0, 15]$ , et aux niveaux (a)  $J = 1$ , (b)  $J = 2$ .

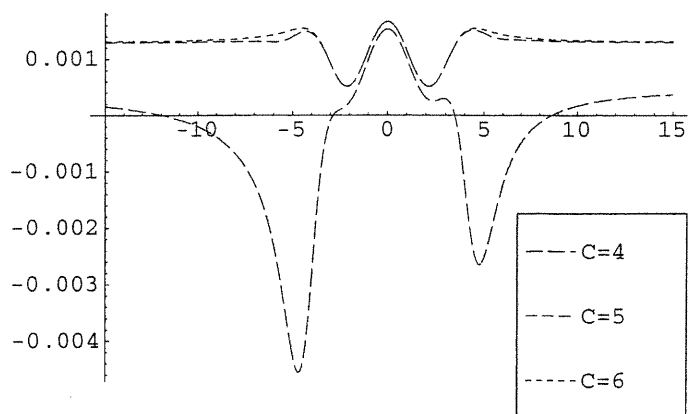
supplémentaire perd son influence, rendant par le fait même les diverses approximations presque symétriques. La figure 2.17(b) nous confirme que  $C = 3$  n'est vraiment pas un choix satisfaisant puisque cette valeur amène aussi de mauvais résultats au niveau  $J = 2$ . Dans ce cas, le mauvais comportement de  $\tilde{\rho}_1(\pi, \hat{\theta}_B|x)$  pour une telle valeur de  $C$  est encore plus apparent. Les figures 2.15(b), 2.16 et 2.17 illustrent bien le fait que  $C = 4$  est par contre un choix intéressant, et ce tant au niveau  $J = 1$  qu'aux niveaux  $J = 2$  et 3.



(a)



(b)



(c)

FIG. 2.17. Erreur relative de  $\tilde{\rho}_1(\pi, \hat{\theta}_B|x)$  pour  $x \in [-15, 15]$ , pour plusieurs valeurs de  $C$  et aux niveaux (a)  $J = 1$ , (b)  $J = 2$ , (c)  $J = 3$ .

### 2.7.3. Discussion

Nous faisons ici un bref retour sur quelques points reliés à l'utilisation de la technique développée. Tout d'abord, à la lumière des résultats obtenus lors des deux exemples présentés précédemment, certaines conclusions s'imposent. Re-voyons rapidement quelques-unes des caractéristiques principales des différentes approximations construites à travers les sections 2.1 à 2.6. Commençons d'abord par quelques commentaires plus généraux.

- La méthode proposée est très facile à utiliser. En effet, les formules employées pour le calcul des approximations obtenues sont relativement simples et faciles à manipuler. De très courts programmes écrit à l'aide du logiciel Mathematica ont permis de produire tous les résultats présentés aux exemples 2.7.1 et 2.7.2.
- Contrairement aux approximations obtenues à l'aide des techniques plus classiques de calcul bayésien, la méthode d'approximation développée ici permet d'approximer les quantités d'intérêt par des formes fonctionnelles. Ces formes fonctionnelles sont très régulières puisque, sous les conditions de régularité données aux théorèmes 2.1 et 2.10, elles sont deux fois continûment dérivables. La possibilité d'approximer des fonctions régulières par d'autres fonctions régulières nous semble logiquement fort intéressante.
- Les exemples 2.7.1 et 2.7.2 amènent tous les deux à la même conclusion au niveau de la qualité des différentes formes d'approximation. En effet, il semble que pour l'approximation de l'estimateur de Bayes, la forme corrigée d'approximation devrait être préférée à la forme simple alors que l'inverse semble vrai dans le cas de l'approximation de la perte espérée *a posteriori* de l'estimateur bayésien.

Maintenant, les choix de  $J$  et de  $C$ , où  $N = C2^J$  est le paramètre contrôlant le nombre de termes à inclure dans les sommes tronquées  $S_l^T(x)$ , déterminent la qualité des approximations obtenues. Voici quelques points fortement reliés au choix de ces paramètres lors de l'utilisation de la méthode d'approximation proposée.

- La précision des approximations construites est étonnante. En fait, l'efficacité de la méthode est telle qu'un niveau de  $J = 1, 2$  ou, au plus, 3 semble habituellement suffisant. Comme nos résultats théoriques sont de nature asymptotique, il est en effet étonnant, et surtout fort intéressant, de constater que  $J = 1$  puisse être vue comme une valeur suffisamment grande.

- Lorsque la précision des approximations construites fait défaut, il est habituellement possible de corriger la situation en augmentant la valeur de  $C$ . Ainsi, il est alors possible d'améliorer le comportement des approximations considérées d'une façon non négligeable avec une augmentation linéaire du nombre de termes de ces sommes. Ceci est fort utile puisqu'une augmentation du niveau va de pair avec une croissance exponentielle du nombre de termes à considérer pour le calcul des sommes  $S_l^T(x)$ , ce qui résulte en une augmentation considérable du temps nécessaire au calcul des diverses approximations.
- Si les deux densités composant le modèle bayésien satisfont toutes les conditions de régularité imposées par les théorèmes 2.1 et 2.10, il est possible d'obtenir des approximations adéquates en retenant peu de termes pour les sommes tronquées  $S_l^T(x)$ . En effet, il suffit pour cela d'obtenir les approximations de façon à calculer les coefficients  $b_{0,jk}$  à partir de la densité la plus concentrée. Ainsi, comme à l'exemple 2.7.2, lorsque l'une des densités est particulièrement concentrée, les sommes tronquées peuvent contenir peu de termes et mener tout de même à des approximations satisfaisantes, et ce même si la deuxième densité est aussi peu concentrée que la densité de Cauchy.
- La détermination d'une valeur adéquate de  $C$  est parfois complexe. En effet, les critères de sélection de  $C$  identifiés à la section 2.6 et donnés par les équations (2.55) et (2.56) mènent parfois à une valeur inutilement élevée pour  $C$ . Même si ceci n'a pas de répercussion négative sur la qualité des approximations résultantes, une brève étude du comportement de ces approximations peut alors s'avérer nécessaire dans un contexte où l'expérimentateur souhaite construire des approximations à la fois précises et plus efficaces au niveau du temps de calcul (par exemple, si la technique d'approximation est employée dans le cadre de simulations servant à solutionner un problème complexe).

Enfin, il est certain que l'élargissement du cadre d'utilisation de la méthode proposée est un point à considérer. En effet, il est inhabituel d'être confronté à un problème d'estimation bayésienne d'un paramètre de position à partir d'un échantillon composé d'une seule observation. Au chapitre suivant nous examinons les difficultés liées à l'estimation bayésienne lorsque l'échantillon est composé de plusieurs observations. Pour ce faire, nous indiquons comment la méthode se généralise à divers modèles bayésiens et surtout, quelles modifications apporter à la procédure dans ces cas plus généraux.

## Chapitre 3

---

### Généralisations à plusieurs observations

Nous cherchons ici à généraliser la technique introduite au chapitre précédent au cas où plusieurs observations sont considérées. Nous étudions ici deux généralisations particulières du modèle à une seule observation.

La première est triviale. En effet, si chaque observation possède son propre paramètre de position, et sous les hypothèses d'indépendance appropriées, alors le problème se réduit à l'estimation de plusieurs paramètres de position où, pour chacun d'entre eux, une seule observation est disponible. Pour illustrer ceci, considérons le modèle donné par

$$X_i|\theta_i \sim \frac{1}{\sigma} f\left(\frac{x_i - \theta_i}{\sigma}\right),$$

et

$$\theta_i \sim \frac{1}{\tau} g\left(\frac{\theta_i - \mu}{\tau}\right),$$

pour  $i = 1, 2, \dots, p$ . Nous supposons ici que  $X_1, X_2, \dots, X_p$  sont conditionnellement indépendantes, et que  $\theta_1, \theta_2, \dots, \theta_p$  sont aussi indépendants. Évidemment, nous supposons également que  $\mu \in \mathbb{R}$ ,  $\sigma > 0$  et  $\tau > 0$  sont des quantités connues, et que  $f \in \mathcal{L}^2(\mathbb{R})$  et  $g \in \mathcal{L}^2(\mathbb{R})$  sont entièrement spécifiées.

Dans ce contexte, il est facile de démontrer que les paramètres de position sont indépendants *a posteriori* et que

$$\hat{\theta}_i(x_i) = \mu + \tau R_1(x_i)$$

est l'estimateur bayésien de  $\theta_i$ , et que

$$\rho_i(\pi, \hat{\theta}_i|x_i) = \tau^2 \left[ R_2(x_i) - \left( R_1(x_i) \right)^2 \right]$$

est la perte espérée *a posteriori* de cet estimateur, pour  $i = 1, 2, \dots, p$ . Ainsi, le problème de l'estimation de plusieurs paramètres de position se réduit ici effectivement à plusieurs estimations d'un seul paramètre de position en employant à chaque fois une seule observation. Il suffit donc, dans ce cas, d'utiliser directement la technique développée au chapitre précédent.

Une deuxième généralisation du problème traité au chapitre 2 est intéressante. Cette fois, considérons le cas où plusieurs observations indépendantes dépendent toutes d'un même paramètre de position, soit, pour  $\mathbf{X} = (X_1, X_2, \dots, X_p)^t$  où  $t$  dénote la transposition,

$$\mathbf{X}|\theta \sim L(\theta|\mathbf{x}) = \frac{1}{\sigma^p} \prod_{i=1}^p f\left(\frac{x_i - \theta}{\sigma}\right),$$

et

$$\theta \sim \pi(\theta) = \frac{1}{\tau} g\left(\frac{\theta - \mu}{\tau}\right),$$

où, encore une fois, nous supposons que  $\mu, \sigma, \tau, f$  et  $g$  sont connues et respectent les mêmes hypothèses que précédemment.

Dans ces conditions, la technique doit généralement être modifiée de façon à obtenir un résultat satisfaisant, soit la construction d'une approximation du même ordre que celle obtenue dans le cas du modèle à une seule observation. Nous nous attaquons à ce problème à travers les trois premières sections de ce chapitre.

À la section 3.4, nous portons notre attention sur le cas particulier de la loi *a priori* uniforme sur  $\mathbb{R}$ , c'est-à-dire

$$\pi(\theta) \equiv 1.$$

Cette loi impropre à caractère non informatif ne satisfait pas les hypothèses imposées précédemment sur la densité *a priori* et doit donc être traitée à part. La popularité et la simplicité de cette loi *a priori* sont des aspects qui nous poussent à étudier ce cas particulier de façon plus approfondie.

### 3.1. Estimation de $\theta$ : cas de plusieurs observations

Nous nous intéressons ici au cas de plusieurs observations indépendantes, mais dépendant toutes d'un même paramètre de position. Le modèle considéré



est donc, comme nous l'avons mentionné précédemment,

$$\mathbf{X}|\theta \sim L(\theta|\mathbf{x}) = \frac{1}{\sigma^p} \prod_{i=1}^p f\left(\frac{x_i - \theta}{\sigma}\right), \quad (3.1)$$

où  $\sigma > 0$  et  $f \in \mathcal{L}^2(\mathbb{R})$  sont connues, et

$$\theta \sim \pi(\theta) = \frac{1}{\tau} g\left(\frac{\theta - \mu}{\tau}\right), \quad (3.2)$$

où  $\mu \in \mathbb{R}$ ,  $\tau > 0$  et  $g \in \mathcal{L}^2(\mathbb{R})$  sont aussi connues. Dans ce contexte, nous pouvons écrire la densité marginale de  $\mathbf{X}$  comme

$$\begin{aligned} m(\mathbf{x}) &= \int_{\mathbb{R}} L(\theta|\mathbf{x})\pi(\theta)d\theta \\ &= \frac{1}{\tau\sigma^p} \int_{\mathbb{R}} \left( \prod_{i=1}^p f\left(\frac{x_i - \theta}{\sigma}\right) \right) g\left(\frac{\theta - \mu}{\tau}\right) d\theta, \end{aligned}$$

et la densité *a posteriori* de  $\theta$  comme

$$\pi(\theta|\mathbf{x}) = \frac{\left( \prod_{i=1}^p f\left(\frac{x_i - \theta}{\sigma}\right) \right) g\left(\frac{\theta - \mu}{\tau}\right)}{\int_{\mathbb{R}} \left( \prod_{i=1}^p f\left(\frac{x_i - \theta}{\sigma}\right) \right) g\left(\frac{\theta - \mu}{\tau}\right) d\theta}.$$

Notons que si nous supposons l'existence d'une statistique exhaustive pour  $\theta$ , alors il est possible de factoriser la vraisemblance selon

$$L(\theta|\mathbf{x}) = h_1(s(\mathbf{x})|\theta)h_2(\mathbf{x}),$$

où  $s(\mathbf{x})$  est la statistique exhaustive pour  $\theta$ , et  $h_1$  est la densité conditionnelle (univariée) de cette statistique. La densité *a posteriori* de  $\theta$  peut alors s'écrire comme

$$\pi(\theta|\mathbf{x}) = \frac{h_1(s(\mathbf{x})|\theta)g\left(\frac{\theta - \mu}{\tau}\right)}{\int_{\mathbb{R}} h_1(s(\mathbf{x})|\theta)g\left(\frac{\theta - \mu}{\tau}\right) d\theta},$$

et il devient évident que le problème à traiter est similaire à celui étudié au chapitre précédent, sauf qu'ici l'unique observation est  $s(\mathbf{x})$ . Remarquons que si  $\theta$  n'est pas le paramètre de position associé à  $s(\mathbf{x})$ , il est possible (et même fort probable) qu'il soit nécessaire d'apporter des modifications mineures à la technique présentée auparavant, ces modifications ne touchant uniquement que la façon d'obtenir les coefficients  $a_{Jk}(s(\mathbf{x}))$  du développement en série de  $\mathbb{I}_l(s(\mathbf{x}))$ .

Maintenant, il existe plusieurs exemples simples où le modèle donné par les équations (3.1) et (3.2) est valable, mais où  $\theta$  n'admet pas de statistique exhaustive. Dans ce cas plus général, les moments *a posteriori* de  $\theta$  peuvent être obtenus, pour  $n > 0$ , comme étant

$$\mathbb{E}[\theta^n | \mathbf{x}] = \frac{\int_{\mathbb{R}} \theta^n \left( \prod_{i=1}^p f\left(\frac{x_i - \theta}{\sigma}\right) \right) g\left(\frac{\theta - \mu}{\tau}\right) d\theta}{\int_{\mathbb{R}} \left( \prod_{i=1}^p f\left(\frac{x_i - \theta}{\sigma}\right) \right) g\left(\frac{\theta - \mu}{\tau}\right) d\theta},$$

avec la propriété que

$$\int_{\mathbb{R}} \theta^n \left( \prod_{i=1}^p f\left(\frac{x_i - \theta}{\sigma}\right) \right) g\left(\frac{\theta - \mu}{\tau}\right) d\theta = \sum_{l=0}^n \binom{n}{l} \mu^{n-l} \tau^{l+1} \mathbb{I}_l(\mathbf{x}), \quad (3.3)$$

pour  $n \geq 0$ , et où, cette fois,

$$\mathbb{I}_l(\mathbf{x}) = \int_{\mathbb{R}} g_l(s) \left( \prod_{i=1}^p f_{x_i}(s) \right) ds. \quad (3.4)$$

Notons que les fonctions  $g_l$  et  $f_{x_i}$  sont toujours définies par

$$g_l(s) = s^l g(s),$$

pour  $l \geq 0$ , et

$$f_{x_i}(s) = f\left(\frac{\tau}{\sigma} \left[ \frac{x_i - \mu}{\tau} - s \right]\right),$$

pour  $i = 1, 2, \dots, p$ .

En procédant exactement comme dans le cas d'une seule observation, il est maintenant facile d'obtenir l'estimateur de Bayes de  $\theta$  et sa perte espérée *a posteriori*. En effet, nous avons ici, comme auparavant,

$$\hat{\theta}_B(\mathbf{x}) = \mu + \tau \frac{\mathbb{I}_1(\mathbf{x})}{\mathbb{I}_0(\mathbf{x})},$$

et

$$\rho(\pi, \hat{\theta}_B | \mathbf{x}) = \tau^2 \left[ \frac{\mathbb{I}_2(\mathbf{x})}{\mathbb{I}_0(\mathbf{x})} - \left( \frac{\mathbb{I}_1(\mathbf{x})}{\mathbb{I}_0(\mathbf{x})} \right)^2 \right].$$

Ainsi, les changements à l'estimation de  $\theta$  amenés par la généralisation du modèle se traduisent uniquement par une modification de l'expression des intégrales  $\mathbb{I}_l$  qui dépendent maintenant de  $p$  observations plutôt que d'une seule.

Cette modification des intégrales  $\mathbb{I}_l$  nous conduit cependant à une difficulté nouvelle quant à leur approximation. En effet, dans le cas du modèle à une seule observation  $x$ , l'approximation de  $\mathbb{I}_l(x)$  a été construite à l'aide d'une approximation initiale de  $f$ , la vraisemblance obtenue à partir de cette unique observation. Maintenant, dans le cas qui nous intéresse ici, la vraisemblance est en fait le produit de  $p$  densités. La technique développée doit donc être adaptée de façon à tenir compte de ce fait.

### 3.2. Approximation de $\mathbb{I}_l(\mathbf{x})$

Tout comme au chapitre précédent, nous construisons ici, pour  $l = 0, 1$  et  $2$ , une approximation initiale de l'intégrale  $\mathbb{I}_l(\mathbf{x})$ , donnée par l'équation (3.4), à l'aide de la base de Haar. Pour cela, nous procédons comme au résultat 2.1, mais cette fois en approximant chacune des fonctions  $f_{x_i}$  de l'expression de  $\mathbb{I}_l(\mathbf{x})$ , donnée par l'équation (3.4), avec la base de Haar. L'approximation résultante est présentée par le théorème qui suit.

**Résultat 3.1.** Si  $f \in \mathcal{L}^2(\mathbb{R})$ , si

$$a_{jk}(x_i) = \langle f_{x_i}, \phi_{jk} \rangle \quad \forall j, k \in \mathbb{Z},$$

et si

$$b_{l,jk} = \langle g_l, \phi_{jk} \rangle \quad \forall j, k \in \mathbb{Z},$$

alors l'approximation de  $\mathbb{I}_l(\mathbf{x})$ , obtenue à partir des approximations de Haar de niveau  $J$  des fonctions  $f_{x_i}$  (pour  $i = 1, 2, \dots, p$ ), est donnée par

$$\tilde{\mathbb{I}}_l(\mathbf{x}) = 2^{(p-1)J/2} \sum_{k \in \mathbb{Z}} \alpha_{Jk}(\mathbf{x}) b_{l,Jk},$$

où

$$\alpha_{Jk}(\mathbf{x}) = \prod_{i=1}^p a_{Jk}(x_i).$$

**Démonstration.** D'abord, en faisant comme au résultat 2.1, l'équation (2.7) permet d'écrire

$$\tilde{f}_{x_i}(s) = \sum_{k \in \mathbb{Z}} a_{Jk}(x_i) \phi_{Jk}(s),$$

pour  $i = 1, 2, \dots, p$ . Notons que le niveau d'approximation  $J$  est choisi identique pour toutes les fonctions  $f_{x_i}$  ( $i = 1, 2, \dots, p$ ). Il ne nous suffit plus qu'à remplacer

chaque  $f_{x_i}$  par son approximation dans l'expression de  $\mathbb{I}_l(\mathbf{x})$ . Nous avons ainsi

$$\tilde{\mathbb{I}}_l(\mathbf{x}) = \int_{\mathbb{R}} g_l(s) \left( \prod_{i=1}^p \tilde{f}_{x_i}(s) \right) ds,$$

comme approximation de  $\mathbb{I}_l(\mathbf{x})$ .

Par ailleurs, notons que

$$\begin{aligned} \prod_{i=1}^p \tilde{f}_{x_i}(s) &= \prod_{i=1}^p \left( \sum_{k_i \in \mathbb{Z}} a_{Jk_i}(x_i) \phi_{Jk_i}(s) \right) \\ &= \sum_{k_1 \in \mathbb{Z}} \sum_{k_2 \in \mathbb{Z}} \cdots \sum_{k_p \in \mathbb{Z}} \left( \prod_{i=1}^p a_{Jk_i}(x_i) \right) \left( \prod_{i=1}^p \phi_{Jk_i}(s) \right). \end{aligned} \quad (3.5)$$

Notons aussi que d'après l'équation (1.15), si  $k_1 = k_2 = \dots = k_p = k$ , alors

$$\begin{aligned} \prod_{i=1}^p \phi_{Jk_i}(s) &= (\phi_{Jk}(s))^p \\ &= \begin{cases} 2^{pJ/2} & \text{si } s \in A_{Jk} \\ 0 & \text{sinon} \end{cases} \\ &= 2^{(p-1)J/2} \phi_{Jk}(s), \end{aligned}$$

et sinon (c'est-à-dire s'il existe  $i$  et  $j$  tels que  $k_i \neq k_j$ ), alors

$$\prod_{i=1}^p \phi_{Jk_i}(s) = 0,$$

quelque soit  $s \in \mathbb{R}$ .

Ceci permet de simplifier l'équation (3.5) pour obtenir

$$\begin{aligned} \prod_{i=1}^p \tilde{f}_{x_i}(s) &= 2^{(p-1)J/2} \sum_{k \in \mathbb{Z}} \left( \prod_{i=1}^p a_{Jk}(x_i) \right) \phi_{Jk}(s) \\ &= 2^{(p-1)J/2} \sum_{k \in \mathbb{Z}} \alpha_{Jk}(\mathbf{x}) \phi_{Jk}(s). \end{aligned}$$

L'approximation de  $\mathbb{I}_l(\mathbf{x})$  qui en découle peut donc finalement s'exprimer comme

$$\begin{aligned}\tilde{\mathbb{I}}_l(\mathbf{x}) &= \int_{\mathbb{R}} g_l(s) \left( 2^{(p-1)J/2} \sum_{k \in \mathbb{Z}} \alpha_{Jk}(\mathbf{x}) \phi_{Jk}(s) \right) ds \\ &= 2^{(p-1)J/2} \sum_{k \in \mathbb{Z}} \alpha_{Jk}(\mathbf{x}) \int_{\mathbb{R}} g_l(s) \phi_{Jk}(s) ds \\ &= 2^{(p-1)J/2} \sum_{k \in \mathbb{Z}} \alpha_{Jk}(\mathbf{x}) b_{l,Jk},\end{aligned}$$

ce qui est étonnamment similaire à l'approximation obtenue dans le cas du modèle à une seule observation.  $\square$

Notons que les coefficients  $a_{Jk}(x_i)$  et  $b_{0,Jk}$  sont toujours donnés par les équations (2.9) et (2.10), et que les coefficients  $b_{1,Jk}$  et  $b_{2,Jk}$  sont également inchangés, et toujours donnés par la proposition 2.1. Il nous faut cependant nous assurer que  $\tilde{\mathbb{I}}_l(\mathbf{x})$  constitue toujours une approximation adéquate de  $\mathbb{I}_l(\mathbf{x})$ , même si la façon de procéder a été modifiée.

Le résultat qui suit est une adaptation du théorème 2.1 au cas qui nous intéresse ici, soit le modèle à plusieurs observations se partageant un même paramètre de position.

**Théorème 3.1.** *Sous les conditions du théorème 2.1, l'erreur d'approximation associée à  $\tilde{\mathbb{I}}_l(\mathbf{x})$  respecte, pour  $l \in \mathbb{N}$ ,*

$$\left| \mathbb{I}_l(\mathbf{x}) - \tilde{\mathbb{I}}_l(\mathbf{x}) \right| = O(2^{-J}) \quad \forall \mathbf{x} \in \mathbb{R}^p.$$

*De plus, cette borne est uniforme.*

**Démonstration.** D'abord, d'après le résultat 3.1 et l'équation (3.4), nous avons

$$\begin{aligned}\left| \mathbb{I}_l(\mathbf{x}) - \tilde{\mathbb{I}}_l(\mathbf{x}) \right| &= \left| \int_{\mathbb{R}} g_l(s) \left( \prod_{i=1}^p f_{x_i}(s) \right) ds - \int_{\mathbb{R}} g_l(s) \left( \prod_{i=1}^p \tilde{f}_{x_i}(s) \right) ds \right| \\ &= \left| \int_{\mathbb{R}} g_l(s) \left( \prod_{i=1}^p f_{x_i}(s) - \prod_{i=1}^p \tilde{f}_{x_i}(s) \right) ds \right| \\ &\leq \int_{\mathbb{R}} |g_l(s)| \left| \prod_{i=1}^p f_{x_i}(s) - \prod_{i=1}^p \tilde{f}_{x_i}(s) \right| ds.\end{aligned}\tag{3.6}$$

Or, nous pouvons écrire

$$\prod_{i=1}^p \tilde{f}_{x_i}(s) = \prod_{i=1}^p (f_{x_i}(s) + e_{x_i}(s)),$$

où

$$e_{x_i}(s) = \tilde{f}_{x_i}(s) - f_{x_i}(s),$$

pour  $i = 1, 2, \dots, p$ . Notons que  $e_{x_i}(s)$  correspond simplement à l'erreur d'approximation ponctuelle associée à  $\tilde{f}_{x_i}$  déjà caractérisée au lemme 2.1. Maintenant, en supposant que

$$f_{x_i}(s) \neq 0,$$

pour  $i = 1, 2, \dots, p$ , alors il est possible de pousser ce développement un peu plus loin afin d'obtenir

$$\begin{aligned} \prod_{i=1}^p \tilde{f}_{x_i}(s) &= \prod_{i=1}^p f_{x_i}(s) \left( 1 + \frac{e_{x_i}(s)}{f_{x_i}(s)} \right) \\ &= \left( \prod_{i=1}^p f_{x_i}(s) \right) \left( 1 + \sum_{i=1}^p \frac{e_{x_i}(s)}{f_{x_i}(s)} + \sum_{1 \leq i < j \leq p} \frac{e_{x_i}(s)e_{x_j}(s)}{f_{x_i}(s)f_{x_j}(s)} \right. \\ &\quad \left. + \dots + \prod_{i=1}^p \frac{e_{x_i}(s)}{f_{x_i}(s)} \right). \end{aligned}$$

De là, la différence de l'équation (3.6) peut être bornée selon

$$\begin{aligned} \left| \prod_{i=1}^p f_{x_i}(s) - \prod_{i=1}^p \tilde{f}_{x_i}(s) \right| &= \left( \prod_{i=1}^p f_{x_i}(s) \right) \left| \sum_{i=1}^p \frac{e_{x_i}(s)}{f_{x_i}(s)} + \sum_{1 \leq i < j \leq p} \frac{e_{x_i}(s)e_{x_j}(s)}{f_{x_i}(s)f_{x_j}(s)} \right. \\ &\quad \left. + \dots + \prod_{i=1}^p \frac{e_{x_i}(s)}{f_{x_i}(s)} \right| \\ &\leq \left( \prod_{i=1}^p f_{x_i}(s) \right) \left( \sum_{i=1}^p \frac{|e_{x_i}(s)|}{f_{x_i}(s)} + \sum_{1 \leq i < j \leq p} \frac{|e_{x_i}(s)||e_{x_j}(s)|}{f_{x_i}(s)f_{x_j}(s)} \right. \\ &\quad \left. + \dots + \prod_{i=1}^p \frac{|e_{x_i}(s)|}{f_{x_i}(s)} \right), \end{aligned}$$

et en remplaçant ce résultat dans l'équation (3.6), nous pouvons borner l'erreur d'approximation par

$$\begin{aligned}
\left| \mathbb{I}_l(\mathbf{x}) - \tilde{\mathbb{I}}_l(\mathbf{x}) \right| &\leq \sum_{i=1}^p \int_{\mathbb{R}} |g_l(s)| \frac{|e_{x_i}(s)|}{f_{x_i}(s)} \left( \prod_{k=1}^p f_{x_k}(s) \right) ds \\
&\quad + \sum_{1 \leq i < j \leq p} \int_{\mathbb{R}} |g_l(s)| \frac{|e_{x_i}(s)| |e_{x_j}(s)|}{f_{x_i}(s) f_{x_j}(s)} \left( \prod_{k=1}^p f_{x_k}(s) \right) ds \\
&\quad + \dots + \int_{\mathbb{R}} |g_l(s)| \prod_{i=1}^p |e_{x_i}(s)| ds. \tag{3.7}
\end{aligned}$$

Maintenant, comme  $f$  est continue et  $f \in \mathcal{L}^1(\mathbb{R})$  (puisque c'est une densité), alors  $f$  est bornée. De plus, par hypothèse,  $f'$  est aussi bornée. Nous définissons donc, pour alléger la notation,

$$c_f = \sup_{s \in \mathbb{R}} f(s) = \|f\|_{\infty},$$

et

$$d_f = \sup_{s \in \mathbb{R}} |f'(s)| = \|f'\|_{\infty},$$

puisque ces quantités existent. De plus, d'après le lemme 2.1, il nous est aussi possible de borner  $e_{x_i}(s)$  par

$$|e_{x_i}(s)| \leq \frac{\tau d_f}{\sigma 2^J},$$

pour  $i = 1, 2, \dots, p$  et  $s \in \mathbb{R}$ .

Revenons à l'équation (3.7). La réflexion précédente nous permet d'obtenir une nouvelle borne pour l'erreur d'approximation associée à  $\tilde{\mathbb{I}}_l(\mathbf{x})$ , soit

$$\begin{aligned}
\left| \mathbb{I}_l(\mathbf{x}) - \tilde{\mathbb{I}}_l(\mathbf{x}) \right| &\leq \sum_{i=1}^p c_f^{p-1} \frac{\tau d_f}{\sigma 2^J} \int_{\mathbb{R}} |g_l(s)| ds \\
&\quad + \sum_{1 \leq i < j \leq p} c_f^{p-2} \left( \frac{\tau d_f}{\sigma 2^J} \right)^2 \int_{\mathbb{R}} |g_l(s)| ds \\
&\quad + \dots + \left( \frac{\tau d_f}{\sigma 2^J} \right)^p \int_{\mathbb{R}} |g_l(s)| ds.
\end{aligned}$$

Ceci peut se réduire à

$$\begin{aligned} \left| \mathbb{I}_l(\mathbf{x}) - \tilde{\mathbb{I}}_l(\mathbf{x}) \right| &\leq \|g_l\|_1 \left[ \binom{p}{1} c_f^{p-1} \frac{\tau d_f}{\sigma 2^J} + \binom{p}{2} c_f^{p-2} \left( \frac{\tau d_f}{\sigma 2^J} \right)^2 \right. \\ &\quad \left. + \dots + \binom{p}{p} \left( \frac{\tau d_f}{\sigma 2^J} \right)^p \right] \\ &\leq \|g_l\|_1 c_f^{p-1} \frac{\tau d_f}{\sigma 2^J} \left[ p + \sum_{i=2}^p \binom{p}{i} \left( \frac{\tau d_f}{\sigma c_f 2^J} \right)^{i-1} \right]. \end{aligned} \quad (3.8)$$

Posons maintenant

$$w_i = \binom{p}{i} \left( \frac{\tau d_f}{\sigma c_f 2^J} \right)^{i-1},$$

pour  $i = 2, 3, \dots, p$ . Afin de simplifier encore davantage la borne donnée par l'équation (3.8), nous caractérisons ici le comportement de  $w_2, w_3, \dots, w_p$ . Tout d'abord, notons que

$$\frac{w_{i+1}}{w_i} = \left( \frac{p-i}{i+1} \right) \frac{\tau d_f}{\sigma c_f 2^J},$$

pour  $i = 2, 3, \dots, p-1$ . Notons également que le rapport précédent décroît avec l'augmentation de  $i$ . Donc, si

$$\frac{w_3}{w_2} \leq 1,$$

alors la série  $\{w_i\}_{i \in \{2, 3, \dots, p\}}$  sera strictement décroissante avec  $i$ . Étudions rapidement cette dernière inégalité. En fait, nous avons ici

$$\begin{aligned} \frac{(p-2) \tau d_f}{3 \sigma c_f 2^J} \leq 1 &\iff 2^J \geq \frac{(p-2) \tau d_f}{3 \sigma c_f} \\ &\iff J \geq \log_2(p-2) + \log_2 \left( \frac{\tau d_f}{\sigma 3 c_f} \right). \end{aligned} \quad (3.9)$$

Ainsi, si  $J$  est assez grand (c'est-à-dire s'il respecte la condition précédente), alors la suite  $\{w_i\}_{i \in \{2, 3, \dots, p\}}$  est effectivement strictement décroissante. Ce résultat est ici central puisqu'il permet d'écrire

$$\sum_{i=2}^p w_i \leq (p-1)w_2 = \frac{(p-1)(p+1)p \tau d_f}{2 \sigma c_f 2^J},$$

lorsque  $J$  satisfait l'inégalité (3.9).



Finalement, nous pouvons arriver à la conclusion recherchée puisqu'en utilisant le résultat précédent dans l'équation (3.8), nous avons

$$\begin{aligned} \left| \mathbb{I}_l(\mathbf{x}) - \tilde{\mathbb{I}}_l(\mathbf{x}) \right| &\leq p \|g_l\|_1 c_f^{p-1} \frac{\tau d_f}{\sigma 2^J} \left[ 1 + \frac{(p-1)(p+1) \tau d_f}{2 \sigma c_f 2^J} \right] \\ &= O(2^{-J}) (1 + O(2^{-J})), \end{aligned}$$

lorsque  $J$  est assez grand, puisque les diverses constantes apparaissant ici sont finies et indépendantes à la fois de  $\mathbf{x}$  et de  $J$ . Ceci permet de conclure que, pour  $l \in \mathbb{N}$ ,

$$\left| \mathbb{I}_l(\mathbf{x}) - \tilde{\mathbb{I}}_l(\mathbf{x}) \right| = O(2^{-J}),$$

indépendamment de  $\mathbf{x} \in \mathbb{R}^p$ . □

Il est maintenant clair que la généralisation du modèle n'a pas de conséquence majeure au niveau de l'approximation des intégrales  $\mathbb{I}_l(\mathbf{x})$ . Bien sûr, l'apparition d'un produit de densités dans l'expression de  $\mathbb{I}_l(\mathbf{x})$  rend l'approximation plus complexe, mais, et c'est ce qui est ici primordial, ceci n'amène pas de détérioration quant à la qualité de l'approximation résultante.

En effet, l'ordre de l'approximation est préservé, ce qui nous permet d'employer directement plusieurs résultats obtenus au chapitre 2 sans modification, ces résultats ne dépendant pas de la forme exacte de  $\mathbb{I}_l(\mathbf{x})$ , mais dépendant plutôt de la qualité de l'approximation construite. Il est donc vraiment simple d'utiliser l'approximation de  $\mathbb{I}_l(\mathbf{x})$  obtenue au théorème précédent afin de créer des approximations de l'estimateur de Bayes de  $\theta$  et de sa perte espérée *a posteriori* dans le contexte du modèle à plusieurs observations.

### 3.3. Approximation de $\hat{\theta}_B(\mathbf{x})$ et de $\rho(\pi, \hat{\theta}_B | \mathbf{x})$

Nous nous tournons maintenant vers la construction d'approximations de l'estimateur bayésien  $\hat{\theta}_B(\mathbf{x})$  et de sa perte espérée *a posteriori*. Pour cela, nous procédons comme à la section 2.3 et nous posons d'abord

$$\tilde{R}_l(\mathbf{x}) = \frac{\tilde{\mathbb{I}}_l(\mathbf{x})}{\tilde{\mathbb{I}}_0(\mathbf{x})},$$

pour  $l \geq 1$ . Nous définissons alors les approximations souhaitées comme

$$\tilde{\theta}_B(\mathbf{x}) = \mu + \tau \tilde{R}_1(\mathbf{x}),$$

et

$$\tilde{\rho}(\pi, \hat{\theta}_B | \mathbf{x}) = \tau^2 \left[ \tilde{R}_2(\mathbf{x}) - \left( \tilde{R}_1(\mathbf{x}) \right)^2 \right].$$

Il est intéressant de constater que le théorème 2.3 et le corollaire 2.1 peuvent ici s'appliquer directement (en employant le théorème 3.1 plutôt que le théorème 2.1), pour amener le prochain corollaire, qui est donné sans preuve.

**Corollaire 3.1.** *Sous les conditions du théorème 2.1, nous avons*

$$\left| \hat{\theta}_B(\mathbf{x}) - \tilde{\theta}_B(\mathbf{x}) \right| = O(2^{-J}) \quad \forall \mathbf{x} \in \mathbb{R}^p,$$

et

$$\left| \rho(\pi, \hat{\theta}_B | \mathbf{x}) - \tilde{\rho}(\pi, \hat{\theta}_B | \mathbf{x}) \right| = O(2^{-J}) \quad \forall \mathbf{x} \in \mathbb{R}^p.$$

Notons que tout comme avec le corollaire 2.1, le corollaire précédent est un résultat théorique intéressant, mais difficile à employer en pratique à cause de la complexité des coefficients  $b_{1,jk}$  et  $b_{2,jk}$ , et du fait que les sommes apparaissant dans les approximations de  $\mathbb{I}_0(\mathbf{x})$ ,  $\mathbb{I}_1(\mathbf{x})$  et  $\mathbb{I}_2(\mathbf{x})$  sont infinies. Il est donc encore une fois nécessaire de trouver des alternatives au calcul de  $\tilde{\theta}_B(\mathbf{x})$  et de  $\tilde{\rho}(\pi, \hat{\theta}_B | \mathbf{x})$ .

La première des limites mentionnées plus haut est contournée en adoptant la même approche qu'à la section 2.4. En effet, les lemmes 2.3 et 2.4, bornant les coefficients  $b_{1,jk}$  et  $b_{2,jk}$ , peuvent encore ici être employés de façon à obtenir des bornes pour  $\tilde{R}_1(\mathbf{x})$  et  $\tilde{R}_2(\mathbf{x})$ . Ces bornes sont similaires à celles données par les théorèmes 2.4 et 2.6. Il suffit en fait de remplacer ici  $a_{Jk}(x)$  par

$$\alpha_{Jk}(\mathbf{x}) = \prod_{i=1}^p a_{Jk}(x_i) \quad (3.10)$$

dans les expressions des diverses bornes pour trouver les résultats cherchés. Les bornes ainsi obtenues nous amènent à définir

$$S_l(\mathbf{x}) = \frac{1}{2^l} \sum_{k \in \mathbb{Z}} k^l \alpha_{Jk}(\mathbf{x}) b_{0,Jk}, \quad (3.11)$$

pour  $l = 0, 1, 2$ , puis

$$\tilde{R}_1^*(\mathbf{x}) = \frac{S_1(\mathbf{x})}{S_0(\mathbf{x})} + \frac{1}{2^{J+1}}, \quad (3.12)$$

et

$$\tilde{R}_2^*(\mathbf{x}) = \frac{S_2(\mathbf{x})}{S_0(\mathbf{x})} + \frac{1}{2^J} \tilde{R}_1^*(\mathbf{x}). \quad (3.13)$$

Avec ces définitions, il est maintenant possible de poser

$$\tilde{\theta}_B^*(\mathbf{x}) = \mu + \tau \tilde{R}_1^*(\mathbf{x}),$$

et

$$\tilde{\rho}^*(\pi, \hat{\theta}_B | \mathbf{x}) = \tau^2 \left[ \tilde{R}_2^*(\mathbf{x}) - \left( \tilde{R}_1^*(\mathbf{x}) \right)^2 \right],$$

et d'obtenir l'analogie, dans le cas du modèle à plusieurs observations, du corollaire 2.2. Ce résultat est donné sans preuve puisque pour l'obtenir, il suffit de reprendre les arguments invoqués aux théorèmes 2.5 et 2.7 ainsi qu'au corollaire 2.2 (en appliquant évidemment le théorème 3.1 plutôt que le théorème 2.1).

**Corollaire 3.2.** *Sous les conditions du théorème 2.1, nous avons*

$$\left| \hat{\theta}_B(\mathbf{x}) - \tilde{\theta}_B^*(\mathbf{x}) \right| = O(2^{-J}) \quad \forall \mathbf{x} \in \mathbb{R}^p,$$

et

$$\left| \rho(\pi, \hat{\theta}_B | \mathbf{x}) - \tilde{\rho}^*(\pi, \hat{\theta}_B | \mathbf{x}) \right| = O(2^{-J}) \quad \forall \mathbf{x} \in \mathbb{R}^p.$$

Ainsi, le fait de passer au modèle contenant plusieurs observations n'a pas non plus d'incidence majeure pour ce qui est de la première forme alternative d'approximation de l'estimateur de Bayes et de sa perte espérée *a posteriori*. En effet, la seule modification dans la façon de procéder est celle qui a déjà été mentionnée plus haut : il suffit de remplacer les coefficients  $a_{Jk}(x)$  des approximations à une seule observation  $x$ , par les coefficients  $\alpha_{Jk}(\mathbf{x})$ , donnés par l'équation (3.10), dans le cas de plusieurs.

Remarquons que ce changement mineur est aussi l'unique modification qui doit être faite pour les formes simples d'approximation introduites à la section 2.5. Effectivement, si nous posons

$$\tilde{R}_l^S(\mathbf{x}) = \frac{1}{2^{lJ}} \frac{S_l(\mathbf{x})}{S_0(\mathbf{x})}, \quad (3.14)$$

pour  $l = 0, 1$  et  $2$ , puis

$$\tilde{\theta}_B^S(\mathbf{x}) = \mu + \tau \tilde{R}_1^S(\mathbf{x}),$$

et

$$\tilde{\rho}^S(\pi, \hat{\theta}_B | \mathbf{x}) = \tau^2 \left[ \tilde{R}_2^S(\mathbf{x}) - \left( \tilde{R}_1^S(\mathbf{x}) \right)^2 \right],$$

alors les théorèmes 2.8 et 2.9 et le corollaire 2.3 peuvent être repris pour mener au résultat suivant.

**Corollaire 3.3.** *Sous les conditions du théorème 2.1, nous avons*

$$\left| \widehat{\theta}_B(\mathbf{x}) - \widetilde{\theta}_B^S(\mathbf{x}) \right| = O(2^{-J}) \quad \forall \mathbf{x} \in \mathbb{R}^p,$$

et

$$\left| \rho(\pi, \widehat{\theta}_B | \mathbf{x}) - \widetilde{\rho}^S(\pi, \widehat{\theta}_B | \mathbf{x}) \right| = O(2^{-J}) \quad \forall \mathbf{x} \in \mathbb{R}^p.$$

Le corollaire 3.3 est présenté sans démonstration puisqu'il découle directement des résultats mentionnés plus haut.

Notons que les formes simples d'approximation  $\widetilde{\theta}_B^S(\mathbf{x})$  et  $\widetilde{\rho}^S(\pi, \widehat{\theta}_B | \mathbf{x})$  et les formes complexes  $\widehat{\theta}_B^*(\mathbf{x})$  et  $\widehat{\rho}^*(\pi, \widehat{\theta}_B | \mathbf{x})$  respectent toujours la proposition 2.2, même après les changements dus à la présence de plusieurs observations.

Maintenant, le dernier point à considérer ici est le fait que les sommes  $S_i(\mathbf{x})$  employées pour les diverses approximations sont, tout comme auparavant, des sommes infinies. Ce problème, étudié à la section 2.6 dans le cas du modèle à une seule observation, doit lui aussi être reconsidéré.

Pour ce faire, remarquons d'abord que les conclusions du théorème 2.10 demeurent valides si, cette fois, nous supposons plutôt qu'il existe  $\varepsilon > 0$ ,  $c > 0$  et  $N_{\mathbf{x}} > 0$  tels que

$$\left( \prod_{i=1}^p f_{x_i}(s) \right) g(s) \leq \frac{c}{|s|^{3+\varepsilon}},$$

dès que  $|s| \geq N_{\mathbf{x}}$ . Notons que cette nouvelle hypothèse permet de diminuer la contribution minimale de la densité de chaque observation et de la loi *a priori*, même si, globalement, le comportement du produit de la vraisemblance basée sur les  $p$  observations et de la densité *a priori* doit demeurer inchangé. Notons également que ceci laisse entrevoir dès maintenant la possibilité d'employer une densité *a priori* impropre, pourvu que  $f$  décroisse assez rapidement.

Finalement, comme le théorème 2.10 est toujours vérifié (en faisant la modification donnée plus haut), alors les sommes tronquées, données par

$$S_i^T(\mathbf{x}) = \sum_{k \in \mathcal{N}} k^l \alpha_{Jk}(\mathbf{x}) b_{0,Jk}, \quad (3.15)$$

pour le cas de plusieurs observations, peuvent être substituées aux sommes complètes  $S_i(\mathbf{x})$ , comme précédemment. Encore une fois, ces substitutions peuvent être effectuées sans affecter la qualité des diverses approximations de l'estimateur bayésien de  $\theta$  et de sa perte espérée *a posteriori*.

Ici se termine notre revue des modifications nécessaires à l'utilisation de la technique développée au chapitre 2 dans le cas du modèle à plusieurs observations, et où la loi *a priori* de  $\theta$  est une densité intégrable. Maintenant, que faire si la loi *a priori* est impropre? La technique peut-elle être employée sans modification? Qu'en est-il si des changements sont apportés? Nous répondons à ces questions à la section suivante en étudiant le cas de la loi *a priori* uniforme sur  $\mathbb{R}$ .

### 3.4. Cas de la loi *a priori* uniforme sur $\mathbb{R}$

Nous nous intéressons encore une fois au modèle comportant plusieurs observations indépendantes partageant un même paramètre de position. Ainsi, la structure de la vraisemblance demeure inchangée et elle est toujours donnée par l'équation (3.1). Par contre, nous tournons maintenant vers la loi *a priori* impropre uniforme sur  $\mathbb{R}$ , soit

$$\theta \sim \pi(\theta) \equiv 1,$$

pour étudier les répercussions de l'utilisation de lois *a priori* impropres sur la méthode développée aux sections précédentes.

D'abord, notons que dans ces conditions, la densité marginale des observations est donnée par

$$\begin{aligned} m(\mathbf{x}) &= \int_{\mathbb{R}} L(\theta|\mathbf{x})\pi(\theta) d\theta \\ &= \frac{1}{\sigma^p} \int_{\mathbb{R}} \left( \prod_{i=1}^p f\left(\frac{x_i - \theta}{\sigma}\right) \right) d\theta, \end{aligned}$$

et la densité *a posteriori* de  $\theta$ , par

$$\pi(\theta|\mathbf{x}) = \frac{\prod_{i=1}^p f\left(\frac{x_i - \theta}{\sigma}\right)}{\int_{\mathbb{R}} \left( \prod_{i=1}^p f\left(\frac{x_i - \theta}{\sigma}\right) \right) d\theta}.$$

De là, il est possible de calculer les moments *a posteriori* de  $\theta$ , qui sont donnés, pour  $n > 0$ , par

$$\mathbb{E}[\theta^n | \mathbf{x}] = \frac{\int_{\mathbb{R}} \theta^n \left( \prod_{i=1}^p f\left(\frac{x_i - \theta}{\sigma}\right) \right) d\theta}{\int_{\mathbb{R}} \left( \prod_{i=1}^p f\left(\frac{x_i - \theta}{\sigma}\right) \right) d\theta}.$$

Remarquons immédiatement qu'en employant l'approche de la section 3.1 telle quelle, les équations (3.3) et (3.4) demeurent valides, mais avec  $\mu = 0$ ,  $\tau = 1$  et

$$g(s) \equiv 1,$$

et donc,

$$g_l(s) = s^l,$$

pour  $l \geq 0$ .

Ainsi, l'utilisation de la loi *a priori* uniforme sur  $\mathbb{R}$  mène à un problème technique majeur quant aux résultats obtenus précédemment. En effet, au moment de borner l'erreur d'approximation associée à  $\tilde{\mathbb{I}}_l(\mathbf{x})$ , le théorème 3.1 est sans équivoque. Notre façon de faire exige que  $g_l \in \mathcal{L}^1(\mathbb{R})$  pour espérer borner cette erreur en procédant comme nous l'avons fait jusqu'ici. Il est clair que ce n'est pas le cas ici, et que de façon générale, cette condition ne sera pas non plus respectée dès que la loi *a priori* employée est impropre. Notons que ce problème ne survient pas au résultat 3.1, puisque les coefficients  $b_{l,jk}$  de  $g_l$  peuvent habituellement être calculés (ou, tout au moins, approximés) même si  $g_l \notin \mathcal{L}^1(\mathbb{R})$ . Cependant, le comportement de l'approximation résultante de  $\mathbb{I}_l(\mathbf{x})$  est alors indéterminé, ce qui rend son usage douteux.

Heureusement, une modification simple à la façon de procéder permet ici de contourner cette difficulté. Il suffit en fait d'utiliser la densité de l'une des observations pour jouer le rôle de  $g$  dans le développement de l'approximation de  $\mathbb{I}_l(\mathbf{x})$ . Pour cela, il faut d'abord adapter les équations (3.3) et (3.4) de façon à obtenir plutôt, pour  $n \geq 0$ ,

$$\int_{\mathbb{R}} \theta^n \left( \prod_{i=1}^p f\left(\frac{x_i - \theta}{\sigma}\right) \right) d\theta = \sum_{l=0}^n \binom{n}{l} (-1)^l x_p^{n-l} \sigma^{l+1} \mathbb{I}_l(\mathbf{x}), \quad (3.16)$$

où, ici, les intégrales  $\mathbb{I}_l(\mathbf{x})$  sont données par

$$\mathbb{I}_l(\mathbf{x}) = \int_{\mathbb{R}} f_l(s) \left( \prod_{i=1}^{p-1} f_{\Delta_i}(s) \right) ds, \quad (3.17)$$

avec, bien sûr,

$$f_l(s) = s^l f(s),$$

pour  $l \geq 0$ , et où

$$f_{\Delta}(s) = f(s + \Delta), \quad (3.18)$$

et

$$\Delta_i = \frac{x_i - x_p}{\sigma},$$

pour  $i = 1, 2, \dots, p-1$ . Notons que ce résultat est obtenu en procédant comme précédemment, mais en utilisant le changement de variable donné par

$$\theta = x_p - \sigma s.$$

Notons également que ce changement de variable peut aussi être fait selon

$$\theta = x_i - \sigma s,$$

quelque soit  $i = 1, 2, \dots, p-1$ . Il est ainsi possible de faire apparaître n'importe laquelle des  $p-1$  autres observations à la place de  $x_p$  dans le développement donné par l'équation (3.16). Cette nouvelle approche au problème est similaire à celle utilisée par O'Hagan (1994, section 7.60) en s'intéressant aux cotes de Bayes dites intrinsèques. Notons finalement que cette façon de faire est équivalente à employer la densité d'une observation comme loi *a priori*. En particulier, ici, les résultats précédents sont identiques à ceux obtenus par l'approche de la section 3.1 lorsque nous disposons de  $p-1$  observations et lorsque la densité *a priori* de  $\theta$  est donnée par

$$\theta \sim \pi(\theta) = \frac{1}{\sigma} f\left(\frac{\theta - x_p}{\sigma}\right),$$

où  $f \in \mathcal{L}^2(\mathbb{R})$  est symétrique.

Revenons maintenant à l'équation (3.16). Grâce à cette équation, nous pouvons écrire la marginale des observations comme

$$m(\mathbf{x}) = \frac{1}{\sigma^{p-1}} \mathbb{I}_0(\mathbf{x}),$$

et les principales quantités d'intérêt que sont l'estimateur bayésien de  $\theta$  et sa perte espérée *a posteriori* comme

$$\widehat{\theta}_B(\mathbf{x}) = x_p - \sigma \frac{\mathbb{I}_1(\mathbf{x})}{\mathbb{I}_0(\mathbf{x})},$$

et

$$\rho(\pi, \widehat{\theta}_B | \mathbf{x}) = \sigma^2 \left[ \frac{\mathbb{I}_2(\mathbf{x})}{\mathbb{I}_0(\mathbf{x})} - \left( \frac{\mathbb{I}_1(\mathbf{x})}{\mathbb{I}_0(\mathbf{x})} \right)^2 \right].$$

Pour obtenir des approximations utiles de ces quantités, il nous suffit de construire les nouvelles approximations des intégrales  $\mathbb{I}_l(\mathbf{x})$  ( $l = 0, 1, 2$ ) qui, dans le contexte présent, sont maintenant données par l'équation (3.17). Le théorème qui suit donne la procédure à suivre.

**Résultat 3.2.** *Si  $f \in \mathcal{L}^2(\mathbb{R})$ , si*

$$a_{jk}(\Delta) = \langle f_\Delta, \phi_{jk} \rangle \quad \forall j, k \in \mathbb{Z},$$

*et si*

$$b_{l,jk} = \langle f_l, \phi_{jk} \rangle \quad \forall j, k \in \mathbb{Z},$$

*alors l'intégrale  $\mathbb{I}_l(\mathbf{x})$  est ici approximée à l'aide de*

$$\widetilde{\mathbb{I}}_l(\mathbf{x}) = 2^{(p-2)J/2} \sum_{k \in \mathbb{Z}} \alpha_{Jk}(\mathbf{x}) b_{l,Jk},$$

*où les coefficients  $\alpha_{Jk}(\mathbf{x})$  sont maintenant donnés par*

$$\alpha_{Jk}(\mathbf{x}) = \prod_{i=1}^{p-1} a_{Jk}(\Delta_i) \quad \forall k \in \mathbb{Z}.$$

Ce résultat est donné sans preuve. Pour l'obtenir, il suffit de reprendre pas à pas la réflexion faite au résultat 3.1, mais en apportant les quelques ajustements mineurs qui s'imposent. Notons que les coefficients  $a_{jk}(\Delta)$  ne sont plus donnés par l'équation (2.9). En effet, la définition de  $f_\Delta$  donnée par l'équation (3.18) amène maintenant que

$$a_{Jk}(\Delta) = 2^{J/2} \left[ F \left( \frac{k+1}{2^J} + \Delta \right) - F \left( \frac{k}{2^J} + \Delta \right) \right],$$

pour  $k \in \mathbb{Z}$  et  $\Delta \in \mathbb{R}$ . Les coefficients  $b_{l,jk}$  sont quant à eux inchangés, sauf pour le fait qu'ils dépendent maintenant de la densité  $f$  et de la fonction de répartition  $F$ , plutôt que de  $g$  et de  $G$ . L'équation (2.10) et la proposition 2.1 sont donc toujours valables et donnent les valeurs de ces coefficients.



Le théorème suivant permet de conclure que, dans le présent contexte,  $\tilde{\mathbb{I}}_l(\mathbf{x})$  constitue toujours une approximation adéquate de  $\mathbb{I}_l(\mathbf{x})$ .

**Théorème 3.2.** *Si  $f$  respecte les conditions du théorème 2.1 et si  $f_l \in \mathcal{L}^1(\mathbb{R})$ , alors pour  $l \in \mathbb{N}$ ,*

$$\left| \mathbb{I}_l(\mathbf{x}) - \tilde{\mathbb{I}}_l(\mathbf{x}) \right| = O(2^{-J}) \quad \forall \mathbf{x} \in \mathbb{R}^p.$$

Ce résultat n'est pas non plus démontré puisqu'il suffit encore une fois de reprendre des arguments invoqués précédemment pour y arriver. Ici, de légères modifications des arguments présentés aux théorèmes 2.1 et 3.1 permettent d'obtenir le résultat assez facilement.

Les difficultés liées au fait que la loi *a priori* de  $\theta$  soit uniforme sur  $\mathbb{R}$ , et donc impropre, sont ainsi contournées en employant la vraisemblance liée à l'une des observations pour imiter le rôle d'une densité *a priori* intégrable. De plus, ceci est fait sans même affecter l'ordre d'approximation de  $\tilde{\mathbb{I}}_l(\mathbf{x})$ . En fait, de façon générale, une loi *a priori* impropre sur  $\theta$  pourra donc être employée sans problème si le nombre d'observations est supérieur à un, et si  $f$  satisfait les bonnes conditions. Une approximation adéquate  $\tilde{\mathbb{I}}_l(\mathbf{x})$  de  $\mathbb{I}_l(\mathbf{x})$  devrait alors pouvoir être construite sans trop de difficulté en utilisant l'approche présentée plus haut.

Maintenant, pour obtenir des approximations de l'estimateur de Bayes  $\hat{\theta}_B(\mathbf{x})$  et de sa perte espérée *a posteriori*  $\rho(\pi, \hat{\theta}_B | \mathbf{x})$ , il suffit de faire comme à la section 3.3. Pour cela, nous définissons d'abord, pour  $l \geq 1$ ,

$$\tilde{R}_l(\mathbf{x}) = \frac{\tilde{\mathbb{I}}_l(\mathbf{x})}{\tilde{\mathbb{I}}_0(\mathbf{x})}.$$

Les approximations initiales de  $\hat{\theta}_B(\mathbf{x})$  et de  $\rho(\pi, \hat{\theta}_B | \mathbf{x})$  peuvent alors être définies comme

$$\tilde{\theta}_B(\mathbf{x}) = x_p - \sigma \tilde{R}_1(\mathbf{x}),$$

et

$$\tilde{\rho}(\pi, \hat{\theta}_B | \mathbf{x}) = \sigma^2 \left[ \tilde{R}_2(\mathbf{x}) - \left( \tilde{R}_1(\mathbf{x}) \right)^2 \right].$$

Ces approximations initiales respectent le corollaire 3.1, mais, clairement, souffrent toujours des limites déjà mentionnées: elles sont difficiles d'utilisation à cause de la complexité des coefficients  $b_{l,jk}$  et elles sont composées de sommes infinies qui doivent, en pratique, être tronquées. Les alternatives au calcul de  $\tilde{\theta}_B(\mathbf{x})$  et de  $\tilde{\rho}(\pi, \hat{\theta}_B | \mathbf{x})$  sont les mêmes qu'auparavant.

Nous rappelons d'abord au lecteur qu'ici,

$$\alpha_{Jk}(\mathbf{x}) = \prod_{i=1}^{p-1} a_{Jk}(\Delta_i) \quad (3.19)$$

remplace l'expression des coefficients  $\alpha_{Jk}(\mathbf{x})$  donnée par l'équation (3.10), et que les coefficients  $b_{0,Jk}$  sont calculés avec  $F$  plutôt qu'avec  $G$ . Maintenant, en définissant  $S_l(\mathbf{x})$  comme à l'équation (3.11) et  $\tilde{R}_1^*(\mathbf{x})$  et  $\tilde{R}_2^*(\mathbf{x})$  comme aux équations (3.12) et (3.13), alors il est possible de définir

$$\tilde{\theta}_B^*(\mathbf{x}) = x_p - \sigma \tilde{R}_1^*(\mathbf{x}),$$

et

$$\tilde{\rho}^*(\pi, \hat{\theta}_B | \mathbf{x}) = \sigma^2 \left[ \tilde{R}_2^*(\mathbf{x}) - \left( \tilde{R}_1^*(\mathbf{x}) \right)^2 \right].$$

Ces nouvelles approximations plus utiles satisfont le corollaire 3.2.

En définissant  $\tilde{R}_l^S(\mathbf{x})$  comme à l'équation (3.14), il est aussi possible de donner les formes simplifiées d'approximation de  $\hat{\theta}_B(\mathbf{x})$  et de  $\rho(\pi, \hat{\theta}_B | \mathbf{x})$ . Nous avons ainsi

$$\tilde{\theta}_B^S(\mathbf{x}) = x_p - \sigma \tilde{R}_1^S(\mathbf{x}),$$

et

$$\tilde{\rho}^S(\pi, \hat{\theta}_B | \mathbf{x}) = \sigma^2 \left[ \tilde{R}_2^S(\mathbf{x}) - \left( \tilde{R}_1^S(\mathbf{x}) \right)^2 \right],$$

et le comportement de ces quantités est toujours donné par le corollaire 3.3. Notons que les deux dernières formes d'approximation ne peuvent plus être reliées à l'aide de la proposition 2.2. Cependant, il est facile de voir qu'elles sont toujours reliées par des facteurs du même ordre.

Finalement, nous avons déjà remarqué que les approximations développées ici sont toujours basées sur des sommes infinies. En effet,  $\tilde{\theta}_B^*(\mathbf{x})$ ,  $\tilde{\rho}^*(\pi, \hat{\theta}_B | \mathbf{x})$ ,  $\tilde{\theta}_B^S(\mathbf{x})$  et  $\tilde{\rho}^S(\pi, \hat{\theta}_B | \mathbf{x})$  sont construites à l'aide des sommes infinies  $S_l(\mathbf{x})$  données par l'équation (3.11). L'évaluation de ces approximations demande donc que les sommes  $S_l(\mathbf{x})$  soient tronquées. Dans le cas du modèle à plusieurs observations où la loi *a priori* de  $\theta$  est une densité intégrable, nous avons vu, à la section précédente, que le théorème 2.10 peut toujours être appliqué, pourvu qu'une modification mineure soit apportée à l'une des hypothèses. Il est possible, ici, de faire la même chose. En appliquant le même principe que précédemment, il suffit de contrôler le comportement du produit de la vraisemblance et de la densité *a priori* de manière adéquate pour que le théorème 2.10 demeure valable.

Dans le présent contexte, l'hypothèse à faire est l'existence de  $\varepsilon > 0$ ,  $c > 0$  et  $N_{\mathbf{x}} > 0$  telles que

$$\left( \prod_{i=1}^{p-1} f_{\Delta_i}(s) \right) f(s) \leq \frac{c}{|s|^{3+\varepsilon}},$$

ou encore, plus simplement, telles que

$$\prod_{i=1}^p f\left(\frac{x_i - s}{\sigma}\right) \leq \frac{c}{|s|^{3+\varepsilon}},$$

dès que  $|s| > N_{\mathbf{x}}$ . Cette hypothèse permet de voir clairement que, même si la densité *a priori* de  $\theta$  utilisée est la loi uniforme sur  $\mathbb{R}$ , l'approximation de l'estimateur bayésien et de sa perte espérée *a posteriori* est possible si  $f$  satisfait des conditions appropriées. Notons que la condition précédente, jumelée aux autres hypothèses du théorème 2.10, nous assure en même temps de l'existence des deux quantités d'intérêt. Ainsi, il est fort intéressant de constater que les mêmes hypothèses amènent à la fois l'existence de  $\hat{\theta}_B(\mathbf{x})$  et de  $\rho(\pi, \hat{\theta}_B|\mathbf{x})$ , et l'assurance de pouvoir construire des approximations calculables en pratique de ces quantités.

Avec la modification mentionnée plus haut, le théorème 2.10 peut aussi être appliqué dans le cas du modèle à plusieurs observations et où la densité *a priori* est la loi uniforme sur  $\mathbb{R}$ . Les sommes tronquées données par l'équation (3.15), mais où les coefficients  $\alpha_{Jk}(\mathbf{x})$  sont donnés par l'équation (3.19), peuvent donc être substituées aux sommes complètes, comme précédemment. De plus, ceci n'amène encore une fois aucune détérioration de la qualité des approximations résultantes de  $\hat{\theta}_B(\mathbf{x})$  et de  $\rho(\pi, \hat{\theta}_B|\mathbf{x})$ .

Pour finir, nous faisons remarquer au lecteur que le raisonnement effectué dans cette dernière section peut être repris, de façon générale, quelque soit la densité *a priori* impropre de  $\theta$ . Des changements mineurs peuvent être nécessaires, mais l'approche demeure essentiellement la même et les conclusions obtenues sont alors identiques.

Au prochain chapitre, nous portons notre attention sur une différente application des approximations  $\tilde{\mathbb{I}}_l(x)$  construites au chapitre 2. En effet, dans le cas du modèle à une seule observation  $X$ , nous avons déjà indiqué auparavant que la densité marginale de cette observation est donnée par

$$m(x) = \frac{1}{\sigma} \mathbb{I}_0(x),$$

où  $\mathbb{I}_0(x)$  est ici donnée par l'équation (2.2). Il apparaît naturel d'approximer  $m(x)$  à l'aide de

$$\tilde{m}(x) = \frac{1}{\sigma} \tilde{\mathbb{I}}_0(x),$$

avec l'avantage intéressant que le comportement de l'approximation  $\tilde{\mathbb{I}}_0(x)$  de  $\mathbb{I}_0(x)$  a déjà été étudié en détail à travers le chapitre 2. Cette remarque nous mènera à une façon simple d'estimer une densité *a priori* inconnue.

## Chapitre 4

---

### Estimation de la densité *a priori* d'un paramètre de position

Au début de cette thèse, nous avons soulevé deux questions fort importantes reliées au modèle simple donné par

$$X|\theta \sim \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right),$$

et par

$$\theta \sim \pi(\theta),$$

où  $\sigma > 0$  et  $f \in \mathcal{L}^2(\mathbb{R})$  sont connues. La première, associée à l'estimation de  $\theta$  lorsque la densité *a priori*  $\pi$  est aussi connue, a été étudiée en profondeur aux chapitres 2 et 3. La seconde, associée quant à elle à la détermination même de  $\pi$ , n'a pas encore été considérée.

En effet, nous avons jusqu'ici adopté le point de vue du statisticien intéressé par l'estimation bayésienne de  $\theta$  dans le cas où la densité *a priori* est disponible. Ceci suppose bien sûr que la densité  $\pi$  a été préalablement identifiée, et qu'au moment de se tourner vers l'estimation de  $\theta$ , cette densité est entièrement spécifiée. Cette étape, qui consiste à faire le choix de  $\pi$ , est selon l'avis de plusieurs absolument primordiale. À ce sujet, le lecteur peut consulter Berger (1985, chapitre 3).

Notre but n'est pas ici d'identifier la meilleure façon de procéder au choix de  $\pi$ . Cette question a déjà été étudiée par plusieurs et aucun consensus n'a encore été établi. Cependant, plusieurs méthodes permettent d'obtenir une solution intéressante à ce problème. Ces diverses méthodes constituent d'excellents outils d'exploration pour le statisticien confronté à une décision quant au choix de  $\pi$ .

Dans cet ordre d'idées, la méthode d'approximation développée aux chapitres précédents nous a suggéré la conception d'un outil supplémentaire qui peut

être ajouté à l'arsenal du statisticien aux prises avec cette difficulté. C'est à la présentation de ce nouvel outil que nous vouons ce quatrième et dernier chapitre.

#### 4.1. Introduction au problème

Le problème de l'identification de la densité *a priori* n'est pas simple. À ce sujet, Berger (1984) explique que les lois *a priori* peuvent rarement être déterminées de façon exacte. Ainsi, la loi *a priori*  $\pi$  choisie, dans un modèle bayésien, ne demeure qu'une approximation de la loi *a priori* "réelle" sous-jacente à la question étudiée (que nous notons pour l'instant  $\pi_r$ ). Dans ce contexte, il est impératif d'approcher le problème du choix de  $\pi$  comme la sélection d'un substitut adéquat à  $\pi_r$ , ou encore comme la construction d'une approximation adéquate de  $\pi_r$ .

Plusieurs possibilités s'offrent au statisticien confronté au choix d'une loi *a priori*. La plus simple d'entre elles est évidemment l'emploi de la loi *a priori* conjuguée, lorsqu'elle existe. Malheureusement, l'emploi d'une telle loi ne se fait pas sans risque. Outre un manque de flexibilité lié à l'obligation d'utiliser une forme fonctionnelle bien déterminée, les lois conjuguées amènent également souvent un grave manque de robustesse de l'inférence résultante. Berger (1984, 1985, section 4.7) discute de ces limites des lois *a priori* conjuguées. Une solution partielle à ces problèmes, proposée par Dalal et Hall (1983), consiste à employer un mélange fini de lois *a priori* conjuguées.

Une autre approche consiste à adopter une modélisation hiérarchique: la loi *a priori* peut elle-même dépendre de paramètres inconnus (appelés hyperparamètres) auxquels est associée une seconde loi *a priori*. Ce procédé peut être répété plusieurs fois pour ainsi obtenir des niveaux successifs. Ceci permet, entre autres, d'enrichir une classe de lois *a priori* conjuguées, tout en conservant la simplicité associée à une modélisation *a priori* conjuguée au premier niveau. Lindley et Smith (1972), Good (1980), Goel et DeGroot (1981) et Goel (1983) discutent de divers aspects de ce type de modélisation.

Les densités *a priori* non informatives sont aussi un choix à considérer. Cependant, il n'existe pas pour l'instant de consensus menant à une définition unique du concept de densité non informative. Plusieurs auteurs ont quand même proposé des techniques permettant de construire de telles densités pour un problème donné, les travaux les plus reconnus étant certainement ceux de Jeffreys (1961), de Bernardo (1979) et de Berger et Bernardo (1992). Même si les densités *a priori* ainsi obtenues sont souvent impropres, ces techniques sont habituellement

efficaces et permettent de construire des densités *a priori* suffisamment robustes lorsque très peu d'information *a priori* est disponible.

Plusieurs autres options s'offrent au statisticien. Entre autres, les lois *a priori* MDIP ("Maximal Data Information Prior") dues à Zellner (1977) et les lois *a priori* à entropie maximale dues à Jaynes (1968). Ces dernières connaissent une certaine popularité puisqu'elles constituent, en quelque sorte, une forme de compromis entre l'utilisation d'information *a priori* partielle, et le souhait d'avoir la loi *a priori* la plus non informative possible. Pour plus de détails sur ces deux techniques, ainsi que pour la présentation de plusieurs autres règles de sélection de la loi *a priori*, le lecteur peut consulter Berger (1985, chapitre 3), ainsi que la très vaste revue faite par Kass et Wasserman (1996).

Finalement, il existe un large éventail de techniques bayésiennes empiriques permettant de construire une loi *a priori* en employant des données antérieures et/ou actuelles à notre disposition. L'approche bayésienne empirique, introduite puis formalisée par Robbins (1955, 1964), diffère fondamentalement des précédentes. En effet, il n'est plus ici question d'espérer que la loi *a priori* choisie constitue une bonne approximation de la loi *a priori* "réelle"  $\pi_r$ . Le but visé est plutôt une estimation directe de  $\pi_r$ , l'approximation ainsi obtenue pouvant être employée comme densité *a priori*  $\pi$  dans le modèle bayésien, ou permettant de confirmer le bien-fondé d'un choix plus standard de  $\pi$  fait avec l'une des méthodes précédentes.

C'est dans cette optique que nous nous intéressons, au cours du présent chapitre, à présenter une technique d'estimation de  $\pi_r$  à partir des observations disponibles. Cette technique repose sur l'emploi de l'approximation de  $\mathbb{I}_0(x)$  construite aux théorèmes 2.2 et 2.11 à l'aide de la base d'ondelettes de Haar.

Pour cela, il nous faut d'abord considérer que plusieurs observations indépendantes sont disponibles, toutes issues du modèle présenté au début de ce chapitre. Nous nous intéressons donc en fait à un nouveau modèle, donné par

$$X_i | \theta_i \sim \frac{1}{\sigma} f\left(\frac{x_i - \theta_i}{\sigma}\right), \quad (4.1)$$

et par

$$\theta_i \sim \pi(\theta_i), \quad (4.2)$$

pour  $i = 1, 2, \dots, p$  et où  $\sigma > 0$  et  $f \in \mathcal{L}^2(\mathbb{R})$  sont connues. Nous supposons aussi l'indépendance conditionnelle des observations  $X_i$  et l'indépendance des paramètres de position  $\theta_i$ . Cette fois, contrairement aux chapitres précédents, la densité *a priori*  $\pi$  n'est pas supposée connue. Notre but est d'arriver à l'estimer.

Comme nous l'avons dit précédemment, cette façon de procéder n'est pas nouvelle. Elle est due à Robbins (1955, 1964) et a été largement approfondie depuis. En fait, depuis les travaux de Morris (1983), on divise même en deux sous-classes les diverses méthodologies bayésiennes empiriques. Il s'agit des approches bayésiennes empiriques paramétriques et non paramétriques.

Dans la première, il est usuel de supposer que la densité *a priori* inconnue appartient à une famille de densités dépendant de paramètres inconnus. L'estimation de  $\pi$  est alors équivalente à l'estimation de ces hyperparamètres inconnus. Des exemples d'application de cette approche sont présentés par Efron et Morris (1972), Morris (1983) et Berger et Berliner (1986).

Dans la seconde, aucune forme paramétrique n'est supposée pour  $\pi$ . La seule hypothèse faite est habituellement l'indépendance *a priori* des paramètres  $\theta_i$ , une hypothèse que nous faisons ici nous-même. Des exemples d'application de cette deuxième approche sont donnés par Laird et Louis (1991) et Healy et Kim (1996).

Finalement, notons que le problème de l'estimation de la loi *a priori*, en apparence un problème essentiellement bayésien, se retrouve sous une forme équivalente dans des domaines fort différents de la statistique. Nous pensons ici à l'estimation d'une densité mélangeante, étudiée, entre autres, par Laird (1978), Carroll et Hall (1988), Zhang (1990), Eggermont et LaRiccia (1995), Magder et Zeger (1996) et Goutis (1997). Nous pensons également à l'estimation de la densité d'observations contaminées par des erreurs additives et à la régression, dans le cas où les régresseurs sont observés avec une erreur. Divers exemples de traitement de ces problèmes sont étudiés par Bickel et Ritov (1987), Fan (1991a, 1991b), Fan et Truong (1993) et Roeder *et al.* (1996).

## 4.2. Expression alternative de la densité marginale

Comme nous l'avons mentionné à la section précédente, la méthode d'estimation de  $\pi$  que nous proposons repose sur l'emploi des approximations construites au chapitre 2. En fait, nous utilisons principalement les théorèmes 2.2 et 2.11 pour obtenir d'abord, à travers la présente section, une nouvelle approximation de  $m$ , la densité marginale commune à toutes les observations. Cette approximation nous permettra, à la prochaine section, de construire ensuite une approximation de  $\pi$ , la densité *a priori* inconnue.

Nous débutons notre réflexion en remarquant que le modèle d'intérêt, donné par les équations (4.1) et (4.2), amène que  $X_1, X_2, \dots, X_p$  partagent la même



densité marginale, soit

$$\begin{aligned} m(x) &= \int_{\mathbb{R}} \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right) \pi(\theta) d\theta \\ &= \frac{1}{\sigma} \mathbb{I}_0(x), \end{aligned} \quad (4.3)$$

où l'intégrale  $\mathbb{I}_0(x)$  est donnée par

$$\mathbb{I}_0(x) = \int_{\mathbb{R}} f\left(\frac{x-\theta}{\sigma}\right) \pi(\theta) d\theta.$$

Notons que ce résultat est identique à celui qui est obtenu aux équations (2.2) et (2.3), en posant  $\mu = 0$ ,  $\tau = 1$  et  $g \equiv \pi$ . Remarquons également que la densité marginale  $m$  donnée par l'équation (4.3) peut être approximée directement à l'aide de résultats déjà obtenus au chapitre 2. En effet, le corollaire 2.5 permet d'arriver au résultat suivant.

**Corollaire 4.1.** *Sous les conditions des théorèmes 2.2 et 2.11, nous avons*

$$m(x) = \frac{1}{\sigma} S_0^T(x) + O(2^{-2J}) \quad \forall x \in \mathbb{R},$$

où  $S_0^T(x)$  est donné par

$$S_0^T(x) = \sum_{k \in \mathcal{N}} a_{Jk}(x) b_{0,Jk},$$

et où les coefficients  $a_{Jk}(x)$  et  $b_{0,Jk}$  sont donnés par les équations (2.9) et (2.10), mais en posant  $\mu = 0$  et  $\tau = 1$ .

De plus, l'erreur d'approximation est bornée uniformément en  $x$ .

Notons que le gain de précision obtenu par l'emploi du théorème 2.2 et du corollaire 2.5 plutôt que du théorème 2.1 et du corollaire 2.4 est ici primordial. La raison exacte de ceci sera évoquée plus précisément à la section 4.4.

Notons également que les coefficients  $a_{Jk}(x)$  peuvent s'écrire ici sous la forme suivante,

$$\begin{aligned} a_{Jk}(x) &= \sigma 2^{J/2} \left[ F\left(\frac{1}{\sigma} \left(x - \frac{k}{2^J}\right)\right) - F\left(\frac{1}{\sigma} \left(x - \frac{k+1}{2^J}\right)\right) \right] \\ &= \sigma 2^{J/2} \left[ F\left(\frac{2^J x - k}{\sigma 2^J}\right) - F\left(\frac{2^J x - k - 1}{\sigma 2^J}\right) \right], \end{aligned} \quad (4.4)$$

pour  $k \in \mathbb{Z}$  et  $x \in \mathbb{R}$ . Remarquons aussi que ces coefficients sont toujours connus, ou tout au moins calculables, puisque  $f$  et  $\sigma$  sont encore considérées connues.

Par contre, ce n'est plus le cas pour les coefficients  $b_{0,Jk}$ . En effet, la densité *a priori*  $\pi$  est maintenant inconnue. Donc, même si les coefficients  $b_{0,Jk}$  sont toujours donnés par l'équation (2.10), où cette fois  $G$  correspond à la fonction de répartition associée à  $\pi$ , cette information supplémentaire s'avère en fait ici inutile puisque  $G$  est évidemment aussi inconnue. Cependant, il est très intéressant de remarquer que ces coefficients sont exactement les coefficients des fonctions  $\phi_{Jk}$  apparaissant dans l'approximation de  $\pi$  obtenue à l'aide de la base de Haar. En effet, d'après l'équation (2.7),  $\pi$  peut être approximée par

$$\tilde{\pi}_J(s) = \sum_{k \in \mathbb{Z}} c_{Jk} \phi_{Jk}(s),$$

où ici,

$$c_{Jk} = \langle \pi, \phi_{Jk} \rangle = b_{0,Jk}. \quad (4.5)$$

Ainsi, si les coefficients  $b_{0,Jk}$  peuvent être estimés, l'expression précédente nous permettra de définir une approximation de  $\pi$ .

Pour cela, nous nous attardons d'abord à construire  $\tilde{m}_J$ , l'approximation de  $m$  obtenue à l'aide de la base de Haar et résultant de l'équation (2.7). Cette approximation sera utilisée à la section 4.3 afin de déduire une méthode d'estimation des coefficients  $b_{0,Jk}$ , et ultimement de  $\pi$ .

Tout d'abord, il est clair que

$$\tilde{m}_J(x) = \sum_{k \in \mathbb{Z}} m_{Jk} \phi_{Jk}(x), \quad (4.6)$$

où les coefficients  $m_{Jk}$  sont donnés par

$$m_{Jk} = \langle m, \phi_{Jk} \rangle, \quad (4.7)$$

pour  $k \in \mathbb{Z}$ . Le prochain résultat donne une expression utile de ces coefficients. Cette expression nous permettra plus tard de relier les approximations  $\tilde{m}_J$  et  $\tilde{\pi}_J$ .

**Théorème 4.1.** *Sous les conditions des théorèmes 2.2 et 2.11, et si  $f \in C^3(\mathbb{R})$  et  $f^{(3)} \in \mathcal{L}^\infty(\mathbb{R})$ , alors*

$$m_{Jk} = \frac{1}{6\sigma 2^{J/2}} \sum_{l \in \mathcal{N}} b_{0,Jl} \left( a_{Jl} \left( \frac{k}{2^J} \right) + 4a_{Jl} \left( \frac{2k+1}{2^{J+1}} \right) + a_{Jl} \left( \frac{k+1}{2^J} \right) \right) + O(2^{-5J/2}) \quad \forall k \in \mathbb{Z},$$

et l'erreur d'approximation est ici bornée uniformément en  $k$ .

**Démonstration.** Notons d'abord que

$$\begin{aligned} m_{Jk} &= \langle m, \phi_{Jk} \rangle \\ &= \int_{\mathbb{R}} m(x) \phi_{Jk}(x) dx. \end{aligned}$$

Ainsi, d'après le corollaire 4.1, nous avons

$$\begin{aligned} m_{Jk} &= \int_{\mathbb{R}} \left( \frac{1}{\sigma} S_0^T(x) + O(2^{-2J}) \right) \phi_{Jk}(x) dx \\ &= \frac{1}{\sigma} \int_{\mathbb{R}} S_0^T(x) \phi_{Jk}(x) dx + O(2^{-2J}) \int_{\mathbb{R}} \phi_{Jk}(x) dx \\ &= \frac{1}{\sigma} \langle S_0^T, \phi_{Jk} \rangle + O(2^{-5J/2}), \end{aligned} \quad (4.8)$$

puisque le terme  $O(2^{-2J})$  est ici indépendant de  $x$  et de  $k$ , et puisque

$$\int_{\mathbb{R}} \phi_{Jk}(x) dx = 2^{J/2} \int_{A_{Jk}} dx = 2^{-J/2}.$$

Notons que ce résultat est uniforme en  $k$ .

Maintenant, intéressons nous au produit scalaire de l'équation (4.8). Ce produit scalaire peut facilement s'écrire sous une forme plus simple en tirant profit de la simplicité de la fonction  $\phi$ . En effet, nous avons ici

$$\begin{aligned} \langle S_0^T, \phi_{Jk} \rangle &= \int_{\mathbb{R}} S_0^T(x) \phi_{Jk}(x) dx \\ &= 2^{J/2} \int_{A_{Jk}} S_0^T(x) dx. \end{aligned} \quad (4.9)$$

Or, pour toute fonction  $p \in C^4(\mathbb{R})$ , la règle d'intégration numérique de Simpson (voir Davis et Rabinowitz, 1984, section 2.2) amène que

$$\int_a^{a+h} p(x) dx = \frac{h}{6} (p(a) + 4p(a+h/2) + p(a+h)) - \frac{h^5}{2880} p^{(4)}(\varepsilon),$$

où  $\varepsilon \in (a, a+h)$ . Autrement dit, la règle de Simpson nous permet d'écrire

$$\int_a^{a+h} p(x) dx = \frac{h}{6} (p(a) + 4p(a+h/2) + p(a+h)) + O(h^5),$$

dès que  $p \in C^4(\mathbb{R})$ . De plus, l'erreur d'approximation sera indépendante du choix de l'intervalle  $(a, a+h)$  si  $p^{(4)}$  est une fonction bornée sur  $\mathbb{R}$ .

Ce résultat peut être employé ici. En effet, nous rappelons que  $S_0^T$  ne dépend de  $x$  qu'à travers les coefficients  $a_{Jk}(x)$  qui, à leur tour, ne dépendent de  $x$  qu'à

travers  $F$ . Comme  $f \in C^3(\mathbb{R})$ , alors  $F \in C^4(\mathbb{R})$  et il en va de même pour  $S_0^T$ . La règle de Simpson peut donc effectivement être utilisée pour approximer l'intégrale apparaissant dans l'équation (4.9). De plus, d'après l'équation (4.4), la quatrième dérivée de  $S_0^T$  satisfait

$$\begin{aligned} \frac{\partial^4 S_0^T}{\partial x^4}(x) &= \sigma 2^{J/2} \sum_{k \in \mathcal{N}} b_{0,Jk} \frac{\partial^4}{\partial x^4} \left[ F \left( \frac{2^J x - k}{\sigma 2^J} \right) - F \left( \frac{2^J x - k - 1}{\sigma 2^J} \right) \right] \\ &= \frac{2^{J/2}}{\sigma^3} \sum_{k \in \mathcal{N}} b_{0,Jk} \left[ F^{(4)} \left( \frac{2^J x - k}{\sigma 2^J} \right) - F^{(4)} \left( \frac{2^J x - k - 1}{\sigma 2^J} \right) \right] \\ &= \frac{2^{J/2}}{\sigma^3} \sum_{k \in \mathcal{N}} b_{0,Jk} \left[ f^{(3)} \left( \frac{2^J x - k}{\sigma 2^J} \right) - f^{(3)} \left( \frac{2^J x - k - 1}{\sigma 2^J} \right) \right]. \end{aligned}$$

Elle peut donc être bornée par

$$\begin{aligned} \left| \frac{\partial^4 S_0^T}{\partial x^4}(x) \right| &\leq \frac{2^{J/2}}{\sigma^3} \sum_{k \in \mathcal{N}} b_{0,Jk} \left| f^{(3)} \left( \frac{2^J x - k}{\sigma 2^J} \right) - f^{(3)} \left( \frac{2^J x - k - 1}{\sigma 2^J} \right) \right| \\ &\leq \frac{2^{J/2+1}}{\sigma^3} \|f^{(3)}\|_\infty \sum_{k \in \mathcal{N}} b_{0,Jk} \\ &\leq \frac{2^{J/2+1}}{\sigma^3} \|f^{(3)}\|_\infty \sum_{k \in \mathbb{Z}} b_{0,Jk} \\ &= \frac{2^{J+1}}{\sigma^3} \|f^{(3)}\|_\infty, \end{aligned}$$

lorsque  $f^{(3)} \in \mathcal{L}^\infty$ , puisque l'équation (4.5) amène

$$\begin{aligned} \sum_{k \in \mathbb{Z}} b_{0,Jk} &= 2^{J/2} \sum_{k \in \mathbb{Z}} \int_{A_{Jk}} \pi(s) ds \\ &= 2^{J/2} \int_{\mathbb{R}} \pi(s) ds \\ &= 2^{J/2}. \end{aligned}$$

Grâce à ce résultat, nous écrivons ici

$$\begin{aligned}
\langle S_0^T, \phi_{Jk} \rangle &= 2^{J/2} \int_{\frac{k}{2^J}}^{\frac{k+1}{2^J}} S_0^T(x) dx \\
&= 2^{J/2} \left[ \frac{1}{6 \times 2^J} \left( S_0^T \left( \frac{k}{2^J} \right) + 4S_0^T \left( \frac{k+1/2}{2^J} \right) + S_0^T \left( \frac{k+1}{2^J} \right) \right) \right. \\
&\quad \left. + \frac{2^{J+1}}{\sigma^3} \|f^{(3)}\|_\infty O(2^{-5J}) \right] \\
&= \frac{1}{6 \times 2^{J/2}} \left( S_0^T \left( \frac{k}{2^J} \right) + 4S_0^T \left( \frac{2k+1}{2^{J+1}} \right) + S_0^T \left( \frac{k+1}{2^J} \right) \right) + O(2^{-7J/2}),
\end{aligned}$$

la borne sur l'erreur d'approximation étant ici indépendante de la valeur de  $k \in \mathbb{Z}$ . En substituant ce dernier résultat dans l'équation (4.8), nous obtenons finalement l'expression recherchée, puisqu'alors

$$\begin{aligned}
m_{Jk} &= \frac{1}{\sigma} \left[ \frac{1}{6 \times 2^{J/2}} \left( S_0^T \left( \frac{k}{2^J} \right) + 4S_0^T \left( \frac{2k+1}{2^{J+1}} \right) + S_0^T \left( \frac{k+1}{2^J} \right) \right) + O(2^{-7J/2}) \right] \\
&\quad + O(2^{-5J/2}) \\
&= \frac{1}{6\sigma 2^{J/2}} \left( S_0^T \left( \frac{k}{2^J} \right) + 4S_0^T \left( \frac{2k+1}{2^{J+1}} \right) + S_0^T \left( \frac{k+1}{2^J} \right) \right) + O(2^{-5J/2}),
\end{aligned}$$

et donc, en utilisant la définition de  $S_0^T(x)$  donnée au corollaire 4.1,

$$\begin{aligned}
m_{Jk} &= \frac{1}{6\sigma 2^{J/2}} \left( \sum_{l \in \mathcal{N}} a_{Jl} \left( \frac{k}{2^J} \right) b_{0,Jl} + 4 \sum_{l \in \mathcal{N}} a_{Jl} \left( \frac{2k+1}{2^{J+1}} \right) b_{0,Jl} \right. \\
&\quad \left. + \sum_{l \in \mathcal{N}} a_{Jl} \left( \frac{k+1}{2^J} \right) b_{0,Jl} \right) + O(2^{-5J/2}) \\
&= \frac{1}{6\sigma 2^{J/2}} \sum_{l \in \mathcal{N}} b_{0,Jl} \left( a_{Jl} \left( \frac{k}{2^J} \right) + 4a_{Jl} \left( \frac{2k+1}{2^{J+1}} \right) + a_{Jl} \left( \frac{k+1}{2^J} \right) \right) \\
&\quad + O(2^{-5J/2}),
\end{aligned}$$

de façon uniforme pour  $k \in \mathbb{Z}$ . □

Ce résultat permet de relier les coefficients  $m_{Jk}$  et  $b_{0,Jk}$ . Bien sûr, comme la densité *a priori*  $\pi$  est ici inconnue, la densité marginale des observations  $m$  l'est également. Il est donc impossible de calculer directement les coefficients  $m_{Jk}$  à l'aide de l'équation (4.7). Le résultat qui précède ne permet pas non plus d'y arriver puisque les coefficients  $b_{0,Jk}$  sont aussi inconnus. Par contre, notons que

les coefficients  $m_{Jk}$  sont en fait directement estimables à partir des observations  $X_1, X_2, \dots, X_p$ . Il apparaît donc normal d'espérer utiliser le théorème 4.1 pour en déduire une estimation des coefficients  $b_{0,Jk}$ . L'estimation résultante est alors basée sur les coefficients  $a_{Jl}(k/2^J)$  qui sont calculables à partir de  $f$ , et les coefficients  $m_{Jk}$  qui sont estimables. La réflexion qui mène à cette conclusion est présentée en détail à la prochaine section.

### 4.3. Estimation non paramétrique de $\pi$

Supposons un instant que les coefficients  $m_{Jk}$  soient calculables directement à l'aide de l'équation (4.7) (c'est-à-dire supposons que  $m$  est connue). L'énoncé du théorème 4.1 ferait alors appel à des quantités connues et calculables, sauf pour les coefficients  $b_{0,Jk}$  de la décomposition de Haar de la densité *a priori*  $\pi$ . Dans ce contexte, il est possible d'obtenir un système d'équations satisfait par les coefficients inconnus. Nous pouvons alors espérer résoudre ce système afin de retrouver les coefficients inconnus.

D'abord, remarquons que d'après l'équation (4.4),

$$\begin{aligned} a_{Jl} \left( \frac{k}{2^J} \right) &= \sigma 2^{J/2} \left[ F \left( \frac{k-l}{\sigma 2^J} \right) - F \left( \frac{k-l-1}{\sigma 2^J} \right) \right] \\ &= a_{J0} \left( \frac{k-l}{2^J} \right), \end{aligned} \quad (4.10)$$

pour  $k, l \in \mathbb{Z}$ . Remarquons aussi qu'en employant ce résultat et le théorème 4.1, nous pouvons écrire

$$\begin{aligned} m_{Jk} &= \frac{1}{6\sigma 2^{J/2}} \sum_{l \in \mathcal{N}} b_{0,Jl} \left( a_{J0} \left( \frac{k-l}{2^J} \right) + 4a_{J0} \left( \frac{2(k-l)+1}{2^{J+1}} \right) \right. \\ &\quad \left. + a_{J0} \left( \frac{k-l+1}{2^J} \right) \right) + O(2^{-5J/2}), \end{aligned} \quad (4.11)$$

pour  $k \in \mathbb{Z}$ .

Maintenant, définissons les vecteurs

$$\mathbf{b} = (b_{0,Jk})_{k \in \mathcal{N}} = \begin{pmatrix} b_{0,J(-N)} \\ b_{0,J(-N+1)} \\ \vdots \\ b_{0,JN} \end{pmatrix},$$

et

$$\mathbf{a}_k = \left( a_{J0} \left( \frac{k-l}{2^J} \right) \right)_{l \in \mathcal{N}} = \begin{pmatrix} a_{J0} \left( \frac{k+N}{2^J} \right) \\ a_{J0} \left( \frac{k+N-1}{2^J} \right) \\ \vdots \\ a_{J0} \left( \frac{k-N}{2^J} \right) \end{pmatrix},$$

pour  $k \in \mathbb{Z}$ . L'équation (4.11) peut être écrite sous forme vectorielle grâce à ces deux quantités. En effet, si  $\mathbf{b}^t$  dénote le vecteur  $\mathbf{b}$  après transposition, alors

$$\begin{aligned} \sum_{l \in \mathcal{N}} b_{0,l} \left( a_{J0} \left( \frac{k-l}{2^J} \right) + 4a_{J0} \left( \frac{2(k-l)+1}{2^{J+1}} \right) + a_{J0} \left( \frac{k-l+1}{2^J} \right) \right) \\ = \mathbf{b}^t (\mathbf{a}_k + 4\mathbf{a}_{k+1/2} + \mathbf{a}_{k+1}) \\ = (\mathbf{a}_k^t + 4\mathbf{a}_{k+1/2}^t + \mathbf{a}_{k+1}^t) \mathbf{b}, \end{aligned}$$

et nous avons l'égalité suivante,

$$m_{Jk} = \frac{1}{6\sigma 2^{J/2}} (\mathbf{a}_k^t + 4\mathbf{a}_{k+1/2}^t + \mathbf{a}_{k+1}^t) \mathbf{b} + O(2^{-5J/2}),$$

pour  $k \in \mathbb{Z}$ , et où le terme  $O(2^{-5J/2})$  est indépendant de  $k$ . Finalement, en définissant le vecteur

$$\mathbf{m} = (m_{Jk})_{k \in \mathcal{N}} = \begin{pmatrix} m_{J(-N)} \\ m_{J(-N+1)} \\ \vdots \\ m_{JN} \end{pmatrix}, \quad (4.12)$$

et la matrice

$$\mathbf{A} = (\mathbf{a}_k^t + 4\mathbf{a}_{k+1/2}^t + \mathbf{a}_{k+1}^t)_{k \in \mathcal{N}} = \begin{pmatrix} \mathbf{a}_{-N}^t + 4\mathbf{a}_{-N+1/2}^t + \mathbf{a}_{-N+1}^t \\ \mathbf{a}_{-N+1}^t + 4\mathbf{a}_{-N+3/2}^t + \mathbf{a}_{-N+2}^t \\ \vdots \\ \mathbf{a}_N^t + 4\mathbf{a}_{N+1/2}^t + \mathbf{a}_{N+1}^t \end{pmatrix}, \quad (4.13)$$

nous obtenons, sous forme matricielle, le système d'équations suivant,

$$\mathbf{m} = \frac{1}{6\sigma 2^{J/2}} \mathbf{A} \mathbf{b} + O(2^{-5J/2}) \mathbf{1}_{2N+1}, \quad (4.14)$$

où  $\mathbf{1}_{2N+1}$  correspond simplement à un vecteur de longueur  $2N+1$  dont tous les éléments sont égaux à 1.

Notons que ce système est bien défini car les vecteurs  $\mathbf{b}$  et  $\mathbf{m}$  sont de longueur  $2N+1$  et la matrice  $\mathbf{A}$  est carrée de dimension  $(2N+1) \times (2N+1)$ . C'est cette raison qui nous pousse à définir  $\mathbf{m}$  selon l'équation (4.12). En effet, comme le

vecteur  $\mathbf{b}$  est composé de  $2N + 1$  valeurs inconnues,  $\mathbf{m}$  doit aussi être de longueur  $2N + 1$  si nous espérons obtenir un système possédant une unique solution.

Étudions brièvement la matrice  $\mathbf{A}$ . D'après l'équation (4.13), la  $i^{\text{ème}}$  ligne de  $\mathbf{A}$ , que nous notons  $\mathbf{A}_{i\bullet}$ , est donnée par

$$\mathbf{A}_{i\bullet}^t = \mathbf{a}_{i-(N+1)} + 4\mathbf{a}_{i-(N+1/2)} + \mathbf{a}_{i-N},$$

pour  $i \in \{1, 2, \dots, 2N + 1\}$ . Ainsi, cette ligne peut être obtenue, sous une forme explicite, comme étant

$$\begin{aligned} \mathbf{A}_{i\bullet}^t &= \left( a_{J0} \left( \frac{i - (N + 1) - l}{2^J} \right) \right)_{l \in \mathcal{N}} + 4 \left( a_{J0} \left( \frac{i - (N + 1/2) - l}{2^J} \right) \right)_{l \in \mathcal{N}} \\ &\quad + \left( a_{J0} \left( \frac{i - N - l}{2^J} \right) \right)_{l \in \mathcal{N}} \\ &= \begin{pmatrix} a_{J0} \left( \frac{i-1}{2^J} \right) + 4a_{J0} \left( \frac{i-1/2}{2^J} \right) + a_{J0} \left( \frac{i}{2^J} \right) \\ a_{J0} \left( \frac{i-2}{2^J} \right) + 4a_{J0} \left( \frac{i-3/2}{2^J} \right) + a_{J0} \left( \frac{i-1}{2^J} \right) \\ \vdots \\ a_{J0} \left( \frac{i-(2N+1)}{2^J} \right) + 4a_{J0} \left( \frac{i-(2N+1/2)}{2^J} \right) + a_{J0} \left( \frac{i-2N}{2^J} \right) \end{pmatrix} \\ &= \left( a_{J0} \left( \frac{i-j}{2^J} \right) + 4a_{J0} \left( \frac{i-j+1/2}{2^J} \right) + a_{J0} \left( \frac{i-j+1}{2^J} \right) \right)_{j \in \{1, 2, \dots, 2N+1\}} \\ &= \left( a_{J0} \left( \frac{i-j}{2^J} \right) + 4a_{J0} \left( \frac{2(i-j)+1}{2^{J+1}} \right) + a_{J0} \left( \frac{i-j+1}{2^J} \right) \right)_{j \in \{1, 2, \dots, 2N+1\}}, \end{aligned}$$

pour  $i \in \{1, 2, \dots, 2N + 1\}$ .

En combinant ce résultat à l'équation (4.10) et en notant

$$\mathbf{A} = (\mathbf{A}_{ij})_{i,j \in \{1, 2, \dots, 2N+1\}},$$



nous pouvons écrire les éléments de  $\mathbf{A}$  comme

$$\begin{aligned}
\mathbf{A}_{ij} &= a_{J0} \left( \frac{i-j}{2^J} \right) + 4a_{J0} \left( \frac{2(i-j)+1}{2^{J+1}} \right) + a_{J0} \left( \frac{i-j+1}{2^J} \right) \\
&= \sigma 2^{J/2} \left[ F \left( \frac{i-j}{\sigma 2^J} \right) - F \left( \frac{i-j-1}{\sigma 2^J} \right) \right] \\
&\quad + 4\sigma 2^{J/2} \left[ F \left( \frac{2(i-j)+1}{\sigma 2^{J+1}} \right) - F \left( \frac{2(i-j)-1}{\sigma 2^{J+1}} \right) \right] \\
&\quad + \sigma 2^{J/2} \left[ F \left( \frac{i-j+1}{\sigma 2^J} \right) - F \left( \frac{i-j}{\sigma 2^J} \right) \right] \\
&= \sigma 2^{J/2} \left[ F \left( \frac{i-j+1}{\sigma 2^J} \right) - F \left( \frac{i-j-1}{\sigma 2^J} \right) \right. \\
&\quad \left. + 4 F \left( \frac{2(i-j)+1}{\sigma 2^{J+1}} \right) - 4 F \left( \frac{2(i-j)-1}{\sigma 2^{J+1}} \right) \right],
\end{aligned}$$

pour  $i, j \in \{1, 2, \dots, 2N+1\}$ . Ce dernier résultat nous amène à définir une nouvelle matrice  $\mathbf{B}$ , dont les éléments sont donnés par

$$\begin{aligned}
\mathbf{B}_{ij} &= F \left( \frac{i-j+1}{\sigma 2^J} \right) - F \left( \frac{i-j-1}{\sigma 2^J} \right) + 4 F \left( \frac{2(i-j)+1}{\sigma 2^{J+1}} \right) \\
&\quad - 4 F \left( \frac{2(i-j)-1}{\sigma 2^{J+1}} \right) \\
&= \frac{1}{\sigma 2^{J/2}} \mathbf{A}_{ij},
\end{aligned} \tag{4.15}$$

pour  $i, j \in \{1, 2, \dots, 2N+1\}$ . Ainsi, nous avons

$$\mathbf{B} = \frac{1}{\sigma 2^{J/2}} \mathbf{A}, \tag{4.16}$$

ce qui nous permet d'écrire l'équation (4.14) sous une forme plus simple, soit

$$\mathbf{m} = \frac{1}{6} \mathbf{B} \mathbf{b} + O(2^{-5J/2}) \mathbf{1}_{2N+1}. \tag{4.17}$$

Nous nous tournons maintenant vers l'étude des propriétés de la matrice  $\mathbf{B}$ .

Pour commencer, remarquons que l'équation (4.15) implique que les éléments de  $\mathbf{B}$  respectent

$$\mathbf{B}_{(i+1)(j+1)} = \mathbf{B}_{ij},$$

pour  $i, j \in \{1, 2, \dots, 2N\}$ . Remarquons également que si  $f$  est une densité symétrique, alors  $\mathbf{B}$  est une matrice symétrique. En effet, si  $f$  est symétrique,

alors  $F$ , la fonction de répartition associée à  $f$ , satisfait

$$F(x) = 1 - F(-x) \quad \forall x \in \mathbb{R}.$$

Ainsi, toujours d'après l'équation (4.15), nous avons

$$\begin{aligned} \mathbf{B}_{ij} &= 1 - F\left(\frac{-i+j-1}{\sigma 2^J}\right) - \left[1 - F\left(\frac{-i+j+1}{\sigma 2^J}\right)\right] \\ &\quad + 4 \left[1 - F\left(\frac{-2(i-j)-1}{\sigma 2^{J+1}}\right)\right] - 4 \left[1 - F\left(\frac{-2(i-j)+1}{\sigma 2^{J+1}}\right)\right] \\ &= F\left(\frac{j-i+1}{\sigma 2^J}\right) - F\left(\frac{j-i-1}{\sigma 2^J}\right) + 4 F\left(\frac{2(j-i)+1}{\sigma 2^{J+1}}\right) \\ &\quad - 4 F\left(\frac{2(j-i)-1}{\sigma 2^{J+1}}\right) \\ &= \mathbf{B}_{ji}, \end{aligned}$$

pour  $i, j \in \{1, 2, \dots, 2N+1\}$ . La matrice  $\mathbf{B}$  est donc effectivement symétrique. Notons que ces deux propriétés de  $\mathbf{B}$  sont partagées par la matrice  $\mathbf{A}$  puisque, d'après l'équation (4.16),  $\mathbf{A}$  et  $\mathbf{B}$  diffèrent uniquement d'un facteur multiplicatif constant.

Revenons maintenant à l'équation (4.17). En supposant toujours que les coefficients  $m_{jk}$  sont calculables directement à partir de l'équation (4.7), l'équation (4.17) nous suggère d'approximer les coefficients  $b_{0,jk}$  à l'aide de

$$\mathbf{b}^* = 6\mathbf{B}^{-1}\mathbf{m}, \quad (4.18)$$

lorsque  $\mathbf{B}$  est non singulière. Le théorème suivant confirme que cette manière de faire est acceptable. Le résultat est donné sans preuve.

**Théorème 4.2.** *Supposons d'abord que les hypothèses des théorèmes 2.2 et 2.11 soient satisfaites. Si  $\mathbf{B}$  est non singulière et s'il existe  $\delta < 5/2$  tel que*

$$\mathbf{B}^{-1}\mathbf{1}_{2N+1} = O(2^{\delta J})\mathbf{1}_{2N+1},$$

alors

$$\mathbf{b} = 6\mathbf{B}^{-1}\mathbf{m} + O(2^{-J(5/2-\delta)})\mathbf{1}_{2N+1}.$$

L'hypothèse supplémentaire faite au théorème précédent peut paraître, à première vue, déraisonnable, puisqu'elle est très difficile à vérifier dans un contexte aussi général. Par contre, notre expérience démontre qu'elle est habituellement satisfaite en pratique. Afin d'en convaincre le lecteur, nous avons procédé à trois études de cas. Pour chaque cas, notre façon de faire est la suivante. D'abord, pour

TAB. 4.1. Composition des modèles utilisés et choix de  $N$  associé à chaque modèle.

No du modèle	Vraisemblance	Densité <i>a priori</i>	$N$
1	$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$	$\pi(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta^2}$	$N = 6 \times 2^J$
2	$f(x) = \frac{x^2}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$	$\pi(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta^2}$	$N = 6 \times 2^J$
3	$f(x) = \frac{1}{\pi(1+x^2)}$	$\pi(\theta) = \frac{1}{\pi(1+\theta^2)}$	$N = 15 \times 2^{4J/3}$

TAB. 4.2. Coefficients de détermination  $R^2$  et estimations de  $\delta$  obtenus pour chaque modèle.

No du modèle	$R^2$	$\delta_{mc}$	Ordre approximatif de l'erreur
1	0,9980	0,2714	$y(J) = O(2^{-2,2286J})$
2	0,9980	0,2738	$y(J) = O(2^{-2,2262J})$
3	0,9706	1,3329	$y(J) = O(2^{-1,1671J})$

une vraisemblance et une densité *a priori* fixées, nous avons fait un choix de  $N$  satisfaisant les hypothèses des théorèmes 2.2 et 2.11. Ensuite, nous avons calculé, pour plusieurs valeurs de  $J$  différentes (en fait  $J = -1, 0, 1, 2$ ),

$$y(J) = \sup_{i \in \{1, 2, \dots, 2N+1\}} |b_i - b_i^*|,$$

soit l'erreur absolue maximale, composante par composante, faite en approximant  $\mathbf{b}$  par  $\mathbf{b}^*$ . Finalement, nous avons ajusté un modèle linéaire du type

$$\log(y(J)) = \alpha - (5/2 - \delta)J \log 2 + e(J), \quad (4.19)$$

où les variables  $e(J)$  sont des erreurs aléatoires d'espérance nulle et de variance constante, afin de nous assurer de la présence d'une relation du type

$$y(J) = O(2^{-J(5/2-\delta)}). \quad (4.20)$$

Le tableau 4.1 présente les trois modèles considérés, ainsi que les valeurs de  $N$  choisies pour la reconstruction de chaque densité *a priori*. Le tableau 4.2 donne, pour chacun des trois modèles, le coefficient de détermination  $R^2$ , l'estimation des moindres carrés de  $\delta$ , notée  $\delta_{mc}$ , ainsi que l'ordre approximatif de l'erreur d'approximation résultant de l'estimation de  $\delta$ .

De ces deux tableaux, il se dégage quelques points intéressants. Tout d'abord, les coefficients  $R^2$  associés à chaque modèle sont très élevés. Ceci semble confirmer l'existence d'une relation linéaire du même type que l'équation (4.19) puisque l'ajustement d'un tel modèle sur les erreurs d'approximation  $y(J)$  est très bon. Par

le fait même, il est tout à fait plausible que la relation donnée par l'équation (4.20) régisse le comportement des erreurs d'approximation  $y(J)$ . Notons également que les valeurs estimées de  $\delta$  permettent de penser que la perte de précision due à l'inversion est minimale lorsque les densités du modèle considéré possèdent un noyau exponentiel. Ceci n'est pas vérifié lorsque les densités considérées sont toutes deux associées à la loi de Cauchy. Dans ce cas extrême, la perte de précision due à l'inversion semble beaucoup plus importante, quoique encore raisonnable puisqu'elle n'empêche pas la reconstruction de  $\pi$ . Notons finalement que le même type de réflexion est présenté à la section 4.7 pour le cas de modèles où la densité *a priori* utilisée est constituée d'un mélange de densités normales. Comme nous le verrons à ce moment, les conclusions tirées dans ce cas sont similaires.

Revenons maintenant à la discussion entamée avant la présentation du théorème 4.2. En supposant que les coefficients  $m_{Jk}$  sont connus ou calculables, l'équation (4.18) donne donc effectivement une façon adéquate d'approximer les coefficients  $b_{0,Jk}$ . Cependant, comme nous l'avons déjà expliqué, la densité marginale  $m$  commune aux observations est en fait inconnue.

Heureusement, cette difficulté n'est pas insurmontable. En effet, nous avons déjà vu, à la section 1.5, que les coefficients du développement en série d'ondelettes d'une densité peuvent être estimés sans biais à l'aide d'observations obtenues à partir de cette densité. En appliquant ici ce raisonnement, il est possible d'estimer les coefficients  $m_{Jk}$  inconnus, en utilisant les observations disponibles, selon

$$\hat{m}_{Jk} = \frac{1}{p} \sum_{i=1}^p \phi_{Jk}(X_i), \quad (4.21)$$

pour  $k \in \mathcal{N}$ . De plus, les estimateurs ainsi définis sont tous sans biais (ce résultat sera vérifié à la proposition 4.1).

Ceci suggère d'estimer  $\mathbf{m}$  par  $\hat{\mathbf{m}}$ , où

$$\hat{\mathbf{m}} = (\hat{m}_{Jk})_{k \in \mathcal{N}}, \quad (4.22)$$

puisque  $\hat{\mathbf{m}}$  est alors sans biais pour estimer  $\mathbf{m}$ , puis d'utiliser ce nouvel estimateur pour remplacer  $\mathbf{m}$  dans l'équation (4.18). Nous obtenons alors l'estimateur suivant pour  $\mathbf{b}$ ,

$$\hat{\mathbf{b}} = 6\mathbf{B}^{-1}\hat{\mathbf{m}}, \quad (4.23)$$

où

$$\hat{\mathbf{b}} = (\hat{b}_{0,Jk})_{k \in \mathcal{N}},$$

lorsque  $\mathbf{B}$  est non singulière. Nous verrons plus loin que cet estimateur n'est pas sans biais pour estimer  $\mathbf{b}$ . L'équation (4.23) nous permet, à l'aide de  $\widehat{\mathbf{b}}$ , d'arriver enfin à définir l'estimateur de la densité *a priori*  $\pi$  inconnue.

**Définition 4.1.** *L'estimateur de niveau  $J$  de  $\pi$ , aussi appelé estimateur par déconvolution de  $\pi$  et noté  $\hat{\pi}_J$ , est défini par*

$$\hat{\pi}_J(s) = \sum_{k \in \mathcal{N}} \hat{b}_{0,Jk} \phi_{Jk}(s),$$

pour  $s \in \mathbb{R}$ .

Sachant que  $\widehat{\mathbf{b}}$  est lui-même un estimateur biaisé, il devient évident que  $\hat{\pi}_J(s)$  est un estimateur biaisé de  $\pi(s)$ . Par contre, même si cette propriété peut paraître indésirable, elle ne nous cause en fait pas réellement de problème de conscience. En effet, Rosenblatt (1956) a déjà démontré qu'aucun estimateur non paramétrique n'est sans biais pour toutes les densités. Il est donc clairement inutile d'espérer atteindre l'absence de biais dans le cas qui nous intéresse. Rosenblatt (1956) propose de s'assurer plutôt que les estimateurs non paramétriques de densités sont asymptotiquement sans biais et vérifient certains critères de convergence.

À la prochaine section, nous étudions le comportement de l'estimateur  $\widehat{\mathbf{b}}$  défini par l'équation (4.23). Nous pourrions par la suite nous attaquer à l'étude de la convergence de  $\hat{\pi}_J$ , l'estimateur de la densité *a priori*  $\pi$ .

#### 4.4. Propriétés de $\widehat{\mathbf{b}}$

Notre but ultime est d'arriver à établir la convergence de l'estimateur de  $\pi$  donné par la définition 4.1. Cependant, pour y parvenir, nous devons auparavant connaître le comportement de l'estimateur  $\widehat{\mathbf{b}}$  défini à l'équation (4.23). Dans la présente section, nous établissons divers résultats permettant de bien saisir le comportement de cet estimateur.

Tout d'abord, l'équation (4.23) établit clairement que le comportement de  $\widehat{\mathbf{b}}$  est en majeure partie déterminé par celui de  $\widehat{\mathbf{m}}$ , dont les éléments sont donnés par l'équation (4.21). Il nous apparaît donc essentiel d'étudier  $\widehat{\mathbf{m}}$ . Pour cela, nous commençons par établir la loi de probabilité exacte de cet estimateur. Aussi, nous utilisons la connaissance de la loi de  $\widehat{\mathbf{m}}$  pour en déduire son espérance et sa variance.

**Proposition 4.1.** *L'estimateur  $\widehat{\mathbf{m}}$  satisfait*

$$\frac{p}{2^{J/2}} \widehat{\mathbf{m}} \sim \mathcal{M} \left( p, 2N + 1, \frac{1}{2^{J/2}} \mathbf{m} \right),$$

*c'est-à-dire  $p2^{-J/2}\widehat{\mathbf{m}}$  est de loi multinomiale de paramètres  $p$ ,  $2N + 1$  et  $2^{-J/2}\mathbf{m}$ . De là, il est possible de conclure que*

$$\mathbb{E}(\widehat{\mathbf{m}}) = \mathbf{m},$$

*et que*

$$\text{Cov}(\widehat{\mathbf{m}}) = \frac{1}{p} \Sigma,$$

*où les éléments de la matrice  $\Sigma$  sont donnés par*

$$\Sigma_{ij} = \begin{cases} m_{J(i-N-1)} (2^{J/2} - m_{J(i-N-1)}) & \text{si } i = j, \\ -m_{J(i-N-1)} m_{J(j-N-1)} & \text{si } i \neq j, \end{cases}$$

*pour  $i, j \in \{1, 2, \dots, 2N + 1\}$ .*

**Démonstration.** D'abord, notons que

$$\phi_{Jk}(X_i) = 2^{J/2} \mathbb{I}_{A_{Jk}}(X_i),$$

où  $\mathbb{I}_{A_{Jk}}$  correspond à la fonction indicatrice de l'intervalle  $A_{Jk}$ , c'est-à-dire

$$\mathbb{I}_{A_{Jk}}(x) = \begin{cases} 1 & \text{si } x \in A_{Jk}, \\ 0 & \text{sinon.} \end{cases}$$

Ainsi, d'après l'équation (4.21), nous avons

$$\begin{aligned} \widehat{m}_{Jk} &= \frac{1}{p} \sum_{i=1}^p \phi_{Jk}(X_i) \\ &= \frac{2^{J/2}}{p} \sum_{i=1}^p \mathbb{I}_{A_{Jk}}(X_i), \end{aligned}$$

et donc si nous posons, pour  $k \in \mathcal{N}$ ,

$$Y_k = \frac{p}{2^{J/2}} \widehat{m}_{Jk} = \sum_{i=1}^p \mathbb{I}_{A_{Jk}}(X_i),$$

alors  $Y_k$  correspond au nombre d'observations (parmi  $X_1, X_2, \dots, X_p$ ) prenant une valeur dans l'intervalle  $A_{Jk}$ . Posons maintenant

$$\mathbf{Y} = (Y_k)_{k \in \mathcal{N}}.$$

Comme  $X_1, X_2, \dots, X_p$  sont indépendantes et identiquement distribuées, alors il est clair que  $\mathbf{Y}$  est de loi multinomiale. Plus précisément, la loi de probabilité de  $\mathbf{Y}$  est donnée par

$$\mathbf{Y} \sim \mathcal{M}(p, 2N + 1, \mathbf{P}), \quad (4.24)$$

où

$$\mathbf{P} = (\mathbb{P}_m(A_{Jk}))_{k \in \mathcal{N}},$$

et où  $\mathbb{P}_m(A_{Jk})$  représente la probabilité marginale qu'une observation prenne une valeur dans l'intervalle  $A_{Jk}$ . Or, notons que d'après l'équation (4.7),

$$\begin{aligned} m_{Jk} &= \langle m, \phi_{Jk} \rangle \\ &= \int_{\mathbb{R}} m(x) \phi_{Jk}(x) dx \\ &= 2^{J/2} \int_{A_{Jk}} m(x) dx \\ &= 2^{J/2} \mathbb{P}_m(A_{Jk}), \end{aligned}$$

pour  $k \in \mathcal{N}$ . Cette égalité est fort intéressante puisqu'elle amène que

$$\mathbb{P}_m(A_{Jk}) = \frac{1}{2^{J/2}} m_{Jk},$$

pour  $k \in \mathcal{N}$ , et donc que

$$\mathbf{P} = \frac{1}{2^{J/2}} \mathbf{m}.$$

Finalement, en utilisant le fait que

$$\mathbf{Y} = \frac{p}{2^{J/2}} \hat{\mathbf{m}},$$

l'équation (4.24) nous permet d'écrire que

$$\frac{p}{2^{J/2}} \hat{\mathbf{m}} \sim \mathcal{M}\left(p, 2N + 1, \frac{1}{2^{J/2}} \mathbf{m}\right),$$

ce qui établit la loi de  $\hat{\mathbf{m}}$ .

Ce résultat nous permet maintenant d'obtenir facilement l'espérance et la covariance du vecteur  $\hat{\mathbf{m}}$ . En effet, nous avons d'abord

$$\mathbb{E}\left(\frac{p}{2^{J/2}} \hat{\mathbf{m}}\right) = \frac{p}{2^{J/2}} \mathbf{m},$$

ce qui amène clairement que

$$\mathbb{E}(\hat{\mathbf{m}}) = \mathbf{m}.$$

Ceci confirme que  $\widehat{\mathbf{m}}$  est bien, comme nous l'avons annoncé précédemment, un estimateur sans biais de  $\mathbf{m}$ . Nous avons également

$$\begin{aligned}\text{Cov}\left(\frac{p}{2^{J/2}}\widehat{\mathbf{m}}\right) &= p\left(\text{Diag}\left(\frac{1}{2^{J/2}}\mathbf{m}\right) - \frac{1}{2^J}\mathbf{m}\mathbf{m}^t\right) \\ &= \frac{p}{2^J}\left(2^{J/2}\text{Diag}(\mathbf{m}) - \mathbf{m}\mathbf{m}^t\right) \\ &= \frac{p}{2^J}\Sigma,\end{aligned}$$

où la matrice  $\Sigma$  est donnée par

$$\Sigma = 2^{J/2}\text{Diag}(\mathbf{m}) - \mathbf{m}\mathbf{m}^t, \quad (4.25)$$

et où  $\text{Diag}(\mathbf{m})$  correspond à la matrice diagonale dont les éléments sont donnés par le vecteur  $\mathbf{m}$ . Ce résultat permet d'écrire

$$\text{Cov}(\widehat{\mathbf{m}}) = \frac{1}{p}\Sigma,$$

et il ne nous reste plus qu'à caractériser les éléments de

$$\Sigma = (\Sigma_{ij})_{i,j \in \{1, \dots, 2N+1\}},$$

pour avoir le résultat voulu. Pour cela, revenons à l'équation (4.25). Nous avons ainsi

$$\begin{aligned}\Sigma_{ii} &= 2^{J/2}m_{J(i-N-1)} - m_{J(i-N-1)}^2 \\ &= m_{J(i-N-1)}\left(2^{J/2} - m_{J(i-N-1)}\right),\end{aligned}$$

pour  $i \in \{1, 2, \dots, 2N+1\}$ . Aussi, si  $i \neq j$ , nous avons

$$\Sigma_{ij} = -m_{J(i-N-1)}m_{J(j-N-1)},$$

pour  $i, j \in \{1, 2, \dots, 2N+1\}$ . Ceci correspond à la forme recherchée pour la matrice  $\Sigma$ .  $\square$

Avant de nous tourner vers l'étude de  $\widehat{\mathbf{b}}$ , nous obtenons deux dernières caractérisations de  $\widehat{\mathbf{m}}$  : sa normalité asymptotique et sa convergence en probabilité, qui vont de paire.

**Proposition 4.2.** *L'estimateur  $\widehat{\mathbf{m}}$  est de loi asymptotique normale multivariée donnée par*

$$\widehat{\mathbf{m}} \sim AN_{2N+1}\left(\mathbf{m}, \frac{1}{p}\Sigma\right),$$



où  $\Sigma$  est la matrice définie à la proposition 4.1. Ceci entraîne que  $\hat{\mathbf{m}}$  converge en probabilité selon

$$\hat{\mathbf{m}} \xrightarrow{P} \mathbf{m}.$$

**Démonstration.** D'abord, posons

$$\mathbf{Y} = \frac{p}{2^{J/2}} \hat{\mathbf{m}}.$$

D'après la proposition 4.1, nous savons que

$$\mathbf{Y} \sim \mathcal{M} \left( p, 2N + 1, \frac{1}{2^{J/2}} \mathbf{m} \right).$$

La version multivariée du théorème de la limite centrale de Lindeberg-Lévy (voir Serfling, 1980, section 1.9) permet d'établir que

$$\sqrt{p} \left( \frac{\mathbf{Y}}{p} - \frac{1}{2^{J/2}} \mathbf{m} \right) \xrightarrow{L} N_{2N+1} \left( \mathbf{0}_{2N+1}, \frac{1}{2^J} \Sigma \right),$$

où  $\xrightarrow{L}$  dénote la convergence en loi,  $\mathbf{0}_{2N+1}$  correspond au vecteur nul de longueur  $2N + 1$ , et  $\Sigma$  est la matrice définie à la proposition 4.1. Mais, comme

$$\sqrt{p} \left( \frac{\mathbf{Y}}{p} - \frac{1}{2^{J/2}} \mathbf{m} \right) = \frac{\sqrt{p}}{2^{J/2}} (\hat{\mathbf{m}} - \mathbf{m}),$$

alors nous avons également

$$\sqrt{p} (\hat{\mathbf{m}} - \mathbf{m}) \xrightarrow{L} N_{2N+1} (\mathbf{0}_{2N+1}, \Sigma),$$

ou, écrit autrement,

$$\hat{\mathbf{m}} \sim AN_{2N+1} \left( \mathbf{m}, \frac{1}{p} \Sigma \right).$$

Finalement, pour obtenir la convergence en probabilité de  $\hat{\mathbf{m}}$ , il suffit de remarquer que

$$\text{Cov}(\hat{\mathbf{m}}) \xrightarrow{p \rightarrow \infty} \mathbf{0}_{(2N+1) \times (2N+1)},$$

où  $\mathbf{0}_{(2N+1) \times (2N+1)}$  correspond à la matrice nulle de taille  $(2N + 1) \times (2N + 1)$ . Ainsi (voir Serfling, 1980, section 1.7), nous avons

$$\hat{\mathbf{m}} \xrightarrow{P} \mathbf{m},$$

et la convergence en probabilité de  $\hat{\mathbf{m}}$  est bien établie.  $\square$

Concentrons nous maintenant sur le comportement de  $\hat{\mathbf{b}}$ . Comme nous l'avons déjà mentionné, l'équation (4.23) permet de voir que les propriétés de  $\hat{\mathbf{b}}$  peuvent

être déduites directement de celles de  $\widehat{\mathbf{m}}$ , puisque la matrice  $\mathbf{B}$  n'est pas aléatoire. Nous débutons notre réflexion en portant notre attention sur l'espérance et la covariance de l'estimateur  $\widehat{\mathbf{b}}$ .

**Théorème 4.3.** *L'estimateur  $\widehat{\mathbf{b}}$  satisfait*

$$\mathbb{E}(\widehat{\mathbf{b}}) = 6\mathbf{B}^{-1}\mathbf{m},$$

et

$$\text{Cov}(\widehat{\mathbf{b}}) = \frac{36}{p}\mathbf{B}^{-1}\Sigma\mathbf{B}^{-1t}.$$

**Démonstration.** La première propriété est facile à vérifier. En effet, nous avons

$$\begin{aligned}\mathbb{E}(\widehat{\mathbf{b}}) &= \mathbb{E}(6\mathbf{B}^{-1}\widehat{\mathbf{m}}) \\ &= 6\mathbf{B}^{-1}\mathbb{E}(\widehat{\mathbf{m}}) \\ &= 6\mathbf{B}^{-1}\mathbf{m},\end{aligned}$$

puisque, d'après la proposition 4.1, l'estimateur  $\widehat{\mathbf{m}}$  est sans biais pour estimer  $\mathbf{m}$ . Pour la matrice de covariance de  $\widehat{\mathbf{b}}$ , nous avons,

$$\begin{aligned}\text{Cov}(\widehat{\mathbf{b}}) &= \text{Cov}(6\mathbf{B}^{-1}\widehat{\mathbf{m}}) \\ &= 36\mathbf{B}^{-1}\text{Cov}(\widehat{\mathbf{m}})\mathbf{B}^{-1t} \\ &= \frac{36}{p}\mathbf{B}^{-1}\Sigma\mathbf{B}^{-1t},\end{aligned}$$

la dernière égalité découlant aussi de la proposition 4.1.  $\square$

Le théorème précédent permet de conclure que  $\widehat{\mathbf{b}}$  est bien un estimateur biaisé de  $\mathbf{b}$ . Nous caractérisons ici ce biais. Tout d'abord, nous avons déjà vu, au théorème 4.2, qu'en faisant les hypothèses nécessaires,

$$\mathbf{b} = 6\mathbf{B}^{-1}\mathbf{m} + O(2^{-J(5/2-\delta)})\mathbf{1}_{2N+1}.$$

L'espérance de  $\widehat{\mathbf{b}}$  peut donc être réécrite comme

$$\begin{aligned}\mathbb{E}(\widehat{\mathbf{b}}) &= 6\mathbf{B}^{-1}\mathbf{m} \\ &= \mathbf{b} + O(2^{-J(5/2-\delta)})\mathbf{1}_{2N+1},\end{aligned}$$

et il est tout à fait évident, sous cette forme, que  $\widehat{\mathbf{b}}$  est biaisé pour l'estimation de  $\mathbf{b}$ . Remarquons que le biais de  $\widehat{\mathbf{b}}$  est indépendant du nombre d'observations  $p$ . Remarquons aussi qu'il est possible de contrôler ce biais à travers le choix de  $J$ , le niveau de l'approximation initiale de  $m$  obtenue à l'aide de la base de Haar

(voir corollaire 4.1). Notons finalement que  $\widehat{\mathbf{b}}$  est en fait un estimateur sans biais de

$$\mathbf{b}^* = 6\mathbf{B}^{-1}\mathbf{m},$$

l'approximation initiale de  $\mathbf{b}$  que nous avons établie à la section précédente. Comme nous l'avons déjà fait remarquer au lecteur, cette approximation de  $\mathbf{b}$  n'est pas utilisable directement puisque le vecteur  $\mathbf{m}$  est ici inconnu.

Le prochain résultat donne les propriétés asymptotiques de l'estimateur  $\widehat{\mathbf{b}}$ .

**Théorème 4.4.** *L'estimateur  $\widehat{\mathbf{b}}$  est aussi de loi asymptotique normale multivariée. Plus précisément, nous avons*

$$\widehat{\mathbf{b}} \sim AN_{2N+1} \left( 6\mathbf{B}^{-1}\mathbf{m}, \frac{36}{p}\mathbf{B}^{-1}\Sigma\mathbf{B}^{-1t} \right).$$

De plus,  $\widehat{\mathbf{b}}$  converge en probabilité selon

$$\widehat{\mathbf{b}} \xrightarrow{P} 6\mathbf{B}^{-1}\mathbf{m}.$$

**Démonstration.** Il suffit ici de jumeler la proposition 4.2 à la définition de  $\widehat{\mathbf{b}}$  donnée par l'équation (4.23). Pour commencer, comme

$$\widehat{\mathbf{m}} \sim AN_{2N+1} \left( \mathbf{m}, \frac{1}{p}\Sigma \right),$$

alors nous avons (voir Serfling, 1980, section 1.7)

$$\widehat{\mathbf{b}} = 6\mathbf{B}^{-1}\widehat{\mathbf{m}} \sim AN_{2N+1} \left( 6\mathbf{B}^{-1}\mathbf{m}, \frac{36}{p}\mathbf{B}^{-1}\Sigma\mathbf{B}^{-1t} \right).$$

Puis, comme

$$\widehat{\mathbf{m}} \xrightarrow{P} \mathbf{m},$$

alors la convergence en probabilité de  $\widehat{\mathbf{b}}$  est donnée (voir Serfling, 1980, section 1.7) par

$$\widehat{\mathbf{b}} = 6\mathbf{B}^{-1}\widehat{\mathbf{m}} \xrightarrow{P} 6\mathbf{B}^{-1}\mathbf{m},$$

c'est-à-dire  $\widehat{\mathbf{b}}$  converge en probabilité vers l'approximation initiale  $\mathbf{b}^*$  de  $\mathbf{b}$ .  $\square$

Notons que les résultats du théorème précédent peuvent aussi s'écrire comme

$$\widehat{\mathbf{b}} \sim AN_{2N+1} \left( \mathbf{b}^*, \frac{36}{p}\mathbf{B}^{-1}\Sigma\mathbf{B}^{-1t} \right),$$

et

$$\widehat{\mathbf{b}} \xrightarrow{P} \mathbf{b}^*.$$

Notons aussi que ce dernier résultat peut s'écrire, si les conditions des théorèmes 2.2 et 2.11 sont satisfaites, sous la forme plus intéressante

$$\widehat{\mathbf{b}} \xrightarrow{P} \mathbf{b} + O(2^{-J(5/2-\delta)}) \mathbf{1}_{2N+1}.$$

Sous cette forme, l'effet du choix de  $J$  sur la convergence de  $\widehat{\mathbf{b}}$  devient évident.

Le comportement de  $\widehat{\mathbf{b}}$  étant enfin déterminé, nous passons à l'étude de l'estimateur  $\widehat{\pi}_J$  proposé à la section 4.3 et donné par la définition 4.1. Les résultats obtenus ici prendront alors toute leur importance.

#### 4.5. Propriétés de $\widehat{\pi}_J$

Avant de commencer l'étude de  $\widehat{\pi}_J$ , nous rappelons d'abord la définition de  $\tilde{\pi}_J$ , l'approximation de  $\pi$  obtenue à partir de l'équation (2.7), soit

$$\tilde{\pi}_J(s) = \sum_{k \in \mathbb{Z}} b_{0,Jk} \phi_{Jk}(s).$$

Nous définissons également la forme tronquée de  $\tilde{\pi}_J$  comme étant

$$\begin{aligned} \tilde{\pi}_J^*(s) &= \sum_{k \in \mathcal{N}} b_{0,Jk} \phi_{Jk}(s) \\ &= \sum_{k=-N}^N b_{0,Jk} \phi_{Jk}(s), \end{aligned}$$

et faisons remarquer au lecteur que d'après le théorème 2.11, nous avons en fait déjà établi que  $N = N(J)$ , puisque  $N$  doit être choisi selon

$$N \propto 2^{J(1+2/\varepsilon)},$$

où  $\varepsilon$  est déterminé par la nature des densités  $f$  et  $\pi$ . Maintenant, la convergence de  $\tilde{\pi}_J$  (lorsque  $J \rightarrow \infty$ ) a déjà été établie à la section 1.5. En effet, d'après les équations (1.38) et (1.39), l'approximation  $\tilde{\pi}_J$  est convergente ponctuellement et en norme quadratique puisqu'elle satisfait

$$\tilde{\pi}_J(s) \xrightarrow{J \rightarrow \infty} \pi(s) \quad \forall s \in \mathbb{R},$$

et

$$\|\tilde{\pi}_J - \pi\|_2 \xrightarrow{J \rightarrow \infty} 0.$$

Donc, comme

$$N = N(J) \xrightarrow{J \rightarrow \infty} \infty,$$

alors nous avons aussi clairement que

$$\tilde{\pi}_J^*(s) \xrightarrow{J \rightarrow \infty} \pi(s) \quad \forall s \in \mathbb{R},$$

et que

$$\|\tilde{\pi}_J^* - \pi\|_2 \xrightarrow{J \rightarrow \infty} 0.$$

Les propriétés de l'estimateur  $\hat{\pi}$  mettront en lumière l'importance de ces résultats.

Nous débutons avec les deux propriétés les plus simples de  $\hat{\pi}_J$ . En effet, l'espérance et la variance ponctuelles de  $\hat{\pi}_J$  sont faciles à obtenir et seront utiles pour la suite de notre réflexion.

**Proposition 4.3.** *L'estimateur  $\hat{\pi}_J$  satisfait*

$$\mathbb{E}(\hat{\pi}_J(s)) = \mathbf{b}^{*t} \Phi(s) \quad \forall s \in \mathbb{R},$$

et

$$\text{Var}(\hat{\pi}_J(s)) = \frac{36}{p} \Phi^t(s) \mathbf{B}^{-1} \Sigma \mathbf{B}^{-1t} \Phi(s) \quad \forall s \in \mathbb{R},$$

où le vecteur  $\Phi(s)$  est donné par

$$\Phi(s) = (\phi_{Jk}(s))_{k \in \mathcal{N}}.$$

**Démonstration.** Notons d'abord que

$$\begin{aligned} \hat{\pi}_J(s) &= \sum_{k \in \mathcal{N}} \hat{b}_{0,Jk} \phi_{Jk}(s) \\ &= \hat{\mathbf{b}}^t \Phi(s). \end{aligned}$$

De là, le théorème 4.3 amène que

$$\begin{aligned} \mathbb{E}(\hat{\pi}_J(s)) &= \mathbb{E}(\hat{\mathbf{b}}^t \Phi(s)) \\ &= \mathbb{E}(\hat{\mathbf{b}})^t \Phi(s) \\ &= \mathbf{b}^{*t} \Phi(s), \end{aligned}$$

pour  $s \in \mathbb{R}$ , et où

$$\mathbf{b}^* = 6\mathbf{B}^{-1}\mathbf{m}.$$

Pour trouver la variance cherchée, nous appliquons le même raisonnement. Ainsi, toujours grâce au théorème 4.3, nous trouvons

$$\begin{aligned}\mathrm{Var}(\hat{\pi}_J(s)) &= \mathrm{Var}(\widehat{\mathbf{b}}^t \Phi(s)) \\ &= \Phi(s)^t \mathrm{Cov}(\widehat{\mathbf{b}}) \Phi(s) \\ &= \frac{36}{p} \Phi(s)^t \mathbf{B}^{-1} \Sigma \mathbf{B}^{-1} \Phi(s),\end{aligned}$$

pour  $s \in \mathbb{R}$ . □

Remarquons immédiatement une première conséquence intéressante de ce résultat. Comme nous l'avons déjà vu, le théorème 4.2 permet d'écrire, en faisant les hypothèses nécessaires,

$$\mathbf{b}^* = \mathbf{b} + O(2^{-J(5/2-\delta)}) \mathbf{1}_{2N+1}.$$

Ainsi, comme

$$\mathbb{E}(\hat{\pi}_J(s)) = \mathbf{b}^{*t} \Phi(s), \quad (4.26)$$

alors nous avons également

$$\begin{aligned}\mathbb{E}(\hat{\pi}_J(s)) &= (\mathbf{b} + O(2^{-J(5/2-\delta)}) \mathbf{1}_{2N+1})^t \Phi(s) \\ &= \mathbf{b}^t \Phi(s) + O(2^{-J(5/2-\delta)}) \mathbf{1}_{2N+1}^t \Phi(s).\end{aligned}$$

Par ailleurs, notons que

$$\mathbf{b}^t \Phi(s) = \sum_{k \in \mathcal{N}} b_{0,Jk} \phi_{Jk}(s) = \tilde{\pi}_J^*(s),$$

et que

$$\begin{aligned}\mathbf{1}_{2N+1}^t \Phi(s) &= \sum_{k \in \mathcal{N}} \phi_{Jk}(s) \\ &\leq \sum_{k \in \mathbb{Z}} \phi_{Jk}(s) \\ &= 2^{J/2},\end{aligned}$$

pour  $s \in \mathbb{R}$ . L'espérance de  $\hat{\pi}_J$  peut donc s'écrire sous la forme plus significative

$$\mathbb{E}(\hat{\pi}_J(s)) = \tilde{\pi}_J^*(s) + O(2^{-J(2-\delta)}), \quad (4.27)$$

pour  $s \in \mathbb{R}$ .

Remarquons également que la proposition 4.3 implique que

$$\mathrm{Var}(\hat{\pi}_J(s)) = 0 \quad \forall s \notin \bigcup_{k \in \mathcal{N}} A_{Jk}.$$

Ceci est tout à fait normal, puisque la définition de  $\hat{\pi}_J$  (voir définition 4.1) amène que

$$\hat{\pi}_J(s) \equiv 0 \quad \forall s \notin \bigcup_{k \in \mathcal{N}} A_{Jk},$$

c'est-à-dire que  $\hat{\pi}_J$  est dégénéré en 0 pour ces valeurs de  $s$ . Dans le cas d'une analyse où les queues de la densité *a priori*  $\pi$  sont d'un intérêt particulier, la technique d'estimation présentée ici est donc certainement moins intéressante puisqu'elle ne permet pas de reconstruire les queues de  $\pi$  correctement. Si, dans un tel contexte, la technique est quand même employée, alors le choix de  $N$  devra être fait avec un soin extrême.

Une conséquence directe de la réflexion précédente et de la normalité asymptotique de  $\hat{\mathbf{b}}$  (voir théorème 4.4) est donnée par le résultat qui suit. Ce résultat est donné sans preuve.

**Corollaire 4.2.** *L'estimateur  $\hat{\pi}_J$  respecte*

$$\hat{\pi}_J(s) \sim AN \left( \mathbf{b}^{*t} \Phi(s), \frac{36}{p} \Phi(s)^t \mathbf{B}^{-1} \Sigma \mathbf{B}^{-1t} \Phi(s) \right) \quad \forall s \in \bigcup_{k \in \mathcal{N}} A_{Jk},$$

et

$$\mathbb{P}(\hat{\pi}_J(s) = 0) = 1 \quad \forall s \notin \bigcup_{k \in \mathcal{N}} A_{Jk}.$$

Nous établissons maintenant le plus simple des résultats ayant trait à la convergence de  $\hat{\pi}_J$ . En effet, la convergence en probabilité de  $\hat{\pi}_J$  s'obtient ici avec très peu de travail puisque la convergence en probabilité de  $\hat{\mathbf{b}}$  a déjà été démontrée.

**Théorème 4.5.** *Sous les conditions des théorèmes 2.2, 2.11 et 4.2, si  $\delta < 2$ , alors*

$$\hat{\pi}_J(s) \xrightarrow{P} \tilde{\pi}_J^*(s) + O(2^{-J(2-\delta)}) \quad \forall s \in \mathbb{R}.$$

**Démonstration.** D'abord, nous avons vu à la proposition précédente que l'estimateur  $\hat{\pi}_J$  peut s'écrire comme

$$\hat{\pi}_J(s) = \hat{\mathbf{b}}^t \Phi(s),$$

où

$$\Phi(s) = (\phi_{Jk}(s))_{k \in \mathcal{N}}.$$

Or, le théorème 4.4 amène que

$$\widehat{\mathbf{b}} \xrightarrow{P} 6\mathbf{B}^{-1}\mathbf{m} = \mathbf{b}^*,$$

et donc que

$$\hat{\pi}_J(s) = \widehat{\mathbf{b}}^t \Phi(s) \xrightarrow{P} \mathbf{b}^{*t} \Phi(s).$$

Mais, d'après les équations (4.26) et (4.27), nous avons

$$\mathbf{b}^{*t} \Phi(s) = \mathbb{E}(\hat{\pi}_J(s)) = \tilde{\pi}_J^*(s) + O(2^{-J(2-\delta)}).$$

Ceci confirme que

$$\hat{\pi}_J(s) \xrightarrow{P} \tilde{\pi}_J^*(s) + O(2^{-J(2-\delta)}),$$

quelque soit  $s \in \mathbb{R}$ . □

Un deuxième type de convergence est facile à obtenir : la convergence en moyenne quadratique ou convergence en EQM (écart quadratique moyen). Cette convergence s'obtient directement à partir des résultats déjà vus.

**Théorème 4.6.** *Sous les conditions des théorèmes 2.2, 2.11 et 4.2, si  $\delta < 2$ , alors*

$$\text{EQM}(\hat{\pi}_J(s)) \xrightarrow{p \rightarrow \infty} (\tilde{\pi}_J^*(s) - \pi(s))^2 + O(2^{-J(2-\delta)}) \quad \forall s \in \mathbb{R}.$$

**Démonstration.** Nous débutons notre réflexion en nous rappelant que

$$\begin{aligned} \text{EQM}(\hat{\pi}_J(s)) &= \mathbb{E} \left[ (\hat{\pi}_J(s) - \pi(s))^2 \right] \\ &= \text{Var}(\hat{\pi}_J(s)) + [\mathbb{E}(\hat{\pi}_J(s)) - \pi(s)]^2. \end{aligned}$$

Mais, l'espérance et la variance de  $\hat{\pi}_J$  ont déjà été calculées. En effet, nous avons d'après l'équation (4.27),

$$\mathbb{E}(\hat{\pi}_J(s)) = \tilde{\pi}_J^*(s) + O(2^{-J(2-\delta)}),$$

et d'après la proposition 4.3,

$$\text{Var}(\hat{\pi}_J(s)) = \frac{36}{p} \Phi(s)^t \mathbf{B}^{-1} \Sigma \mathbf{B}^{-1} \Phi(s).$$



Nous pouvons donc exprimer l'EQM de  $\hat{\pi}_J$  sous une forme explicite, soit

$$\begin{aligned} \text{EQM}(\hat{\pi}_J(s)) &= \frac{36}{p} \Phi(s)^t \mathbf{B}^{-1} \Sigma \mathbf{B}^{-1t} \Phi(s) + [\tilde{\pi}_J^*(s) - \pi(s) + O(2^{-J(2-\delta)})]^2 \\ &= \frac{36}{p} \Phi(s)^t \mathbf{B}^{-1} \Sigma \mathbf{B}^{-1t} \Phi(s) + (\tilde{\pi}_J^*(s) - \pi(s))^2 \\ &\quad + 2(\tilde{\pi}_J^*(s) - \pi(s))O(2^{-J(2-\delta)}) + O(2^{-2J(2-\delta)}). \end{aligned} \quad (4.28)$$

Par ailleurs,  $\pi$  est ici une densité continûment dérivable. Ainsi, il est possible de définir

$$c_\pi = \sup_{s \in \mathbb{R}} \pi(s),$$

puisque sous ces conditions,  $\pi$  est bornée. Notons que  $\tilde{\pi}_J^*(s)$  peut aussi être bornée de la même façon. En effet, d'après l'équation (4.5), nous avons

$$\begin{aligned} b_{0,Jk} &= \langle \pi, \phi_{Jk} \rangle \\ &= 2^{J/2} \int_{A_{Jk}} \pi(s) ds \\ &\leq \frac{1}{2^{J/2}} c_\pi, \end{aligned}$$

ce qui permet d'écrire

$$\begin{aligned} \tilde{\pi}_J^*(s) &= \sum_{k \in \mathcal{N}} b_{0,Jk} \phi_{Jk}(s) \\ &\leq \sum_{k \in \mathbb{Z}} b_{0,Jk} \phi_{Jk}(s) \\ &\leq \frac{1}{2^{J/2}} c_\pi \sum_{k \in \mathbb{Z}} \phi_{Jk}(s) = c_\pi, \end{aligned}$$

pour  $s \in \mathbb{R}$ .

Il est enfin possible de conclure puisque, grâce au résultat précédent, l'équation (4.28) se réduit à

$$\text{EQM}(\hat{\pi}_J(s)) = \frac{36}{p} \Phi(s)^t \mathbf{B}^{-1} \Sigma \mathbf{B}^{-1t} \Phi(s) + (\tilde{\pi}_J^*(s) - \pi(s))^2 + O(2^{-J(2-\delta)}),$$

d'où

$$\text{EQM}(\hat{\pi}_J(s)) \xrightarrow{p \rightarrow \infty} (\tilde{\pi}_J^*(s) - \pi(s))^2 + O(2^{-J(2-\delta)}),$$

pour  $s \in \mathbb{R}$ . □

Nous obtenons maintenant un résultat fort important pour la suite de nos travaux. Le lemme qui suit nous permettra en effet d'établir la convergence de la norme quadratique de  $\hat{\pi}_J$ .

**Lemme 4.1.** *L'estimateur  $\hat{\pi}_J$  respecte*

$$\|\hat{\pi}_J - \pi\|_2^2 = \|\hat{\pi}_J - \tilde{\pi}_J^*\|_2^2 + \|\tilde{\pi}_J^* - \pi\|_2^2.$$

*De plus, nous avons*

$$\|\hat{\pi}_J - \tilde{\pi}_J^*\|_2^2 = (\hat{\mathbf{b}} - \mathbf{b})^t (\hat{\mathbf{b}} - \mathbf{b}).$$

**Démonstration.** Commençons avec la deuxième partie du résultat. Puisque

$$\begin{aligned} \hat{\pi}_J(s) - \tilde{\pi}_J^*(s) &= \sum_{k \in \mathcal{N}} \hat{b}_{0,Jk} \phi_{Jk}(s) - \sum_{k \in \mathcal{N}} b_{0,Jk} \phi_{Jk}(s) \\ &= \sum_{k \in \mathcal{N}} (\hat{b}_{0,Jk} - b_{0,Jk}) \phi_{Jk}(s), \end{aligned} \quad (4.29)$$

alors la norme cherchée peut être exprimée comme

$$\begin{aligned} \|\hat{\pi}_J - \tilde{\pi}_J^*\|_2^2 &= \int_{\mathbb{R}} \left( \sum_{k \in \mathcal{N}} (\hat{b}_{0,Jk} - b_{0,Jk}) \phi_{Jk}(s) \right)^2 ds \\ &= \int_{\mathbb{R}} \left( \sum_{k \in \mathcal{N}} \sum_{l \in \mathcal{N}} (\hat{b}_{0,Jk} - b_{0,Jk}) (\hat{b}_{0,Jl} - b_{0,Jl}) \phi_{Jk}(s) \phi_{Jl}(s) \right) ds \\ &= \sum_{k \in \mathcal{N}} \sum_{l \in \mathcal{N}} (\hat{b}_{0,Jk} - b_{0,Jk}) (\hat{b}_{0,Jl} - b_{0,Jl}) \langle \phi_{Jk}, \phi_{Jl} \rangle. \end{aligned}$$

L'orthonormalité de la suite  $\{\phi_{Jk}\}_{k \in \mathbb{Z}}$  amène donc que

$$\begin{aligned} \|\hat{\pi}_J - \tilde{\pi}_J^*\|_2^2 &= \sum_{k \in \mathcal{N}} \sum_{l \in \mathcal{N}} (\hat{b}_{0,Jk} - b_{0,Jk}) (\hat{b}_{0,Jl} - b_{0,Jl}) \delta_{kl} \\ &= \sum_{k \in \mathcal{N}} (\hat{b}_{0,Jk} - b_{0,Jk})^2, \end{aligned}$$

et donc que

$$\|\hat{\pi}_J - \tilde{\pi}_J^*\|_2^2 = (\hat{\mathbf{b}} - \mathbf{b})^t (\hat{\mathbf{b}} - \mathbf{b}).$$

Pour obtenir la première partie du résultat, remarquons d'abord que

$$\|\hat{\pi}_J - \pi\|_2^2 = \|\hat{\pi}_J - \tilde{\pi}_J^*\|_2^2 + \|\tilde{\pi}_J^* - \pi\|_2^2 + 2 \langle \hat{\pi}_J - \tilde{\pi}_J^*, \tilde{\pi}_J^* - \pi \rangle. \quad (4.30)$$

Remarquons également que d'après les équations (1.34) et (4.5), il est possible d'écrire

$$\pi(s) = \sum_{k \in \mathbb{Z}} b_{0,Jk} \phi_{Jk}(s) + \sum_{j \geq J} \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk}(s),$$

où les fonctions  $\psi_{jk}$  sont les ondelettes associées à la base de Haar, et où

$$d_{jk} = \langle \pi, \psi_{jk} \rangle,$$

pour  $j, k \in \mathbb{Z}$ . Ainsi, nous pouvons exprimer la différence  $\tilde{\pi}_J^*(s) - \pi(s)$  comme

$$\begin{aligned} \tilde{\pi}_J^*(s) - \pi(s) &= \sum_{k \in \mathcal{N}} b_{0,Jk} \phi_{Jk}(s) - \left( \sum_{k \in \mathbb{Z}} b_{0,Jk} \phi_{Jk}(s) + \sum_{j \geq J} \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk}(s) \right) \\ &= - \left( \sum_{k \notin \mathcal{N}} b_{0,Jk} \phi_{Jk}(s) + \sum_{j \geq J} \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk}(s) \right), \end{aligned}$$

pour  $s \in \mathbb{R}$ .

En comparant l'égalité précédente à l'équation (4.29), il devient clair que

$$\langle \hat{\pi}_J - \tilde{\pi}_J^*, \tilde{\pi}_J^* - \pi \rangle = 0,$$

puisque les suites  $\{\phi_{Jk}\}_{k \in \mathbb{Z}}$  et  $\{\psi_{jk}\}_{j \geq J, k \in \mathbb{Z}}$  sont orthonormées et orthogonales entre elles. Enfin, cette conclusion nous permet de reformuler l'équation (4.30) sous la forme recherchée, soit

$$\|\hat{\pi}_J - \pi\|_2^2 = \|\hat{\pi}_J - \tilde{\pi}_J^*\|_2^2 + \|\tilde{\pi}_J^* - \pi\|_2^2.$$

□

Le lemme précédent peut être appliqué afin d'obtenir deux formes de convergence de la norme quadratique de  $\hat{\pi}_J$  : la convergence en probabilité et la convergence en moyenne. Le théorème suivant établit ces résultats.

**Théorème 4.7.** *Si les hypothèses des théorèmes 2.2, 2.11 et 4.2 sont satisfaites, et si  $\delta < 2 - 1/\varepsilon$  alors*

$$\|\hat{\pi}_J - \pi\|_2^2 \xrightarrow{P} \|\tilde{\pi}_J^* - \pi\|_2^2 + O(2^{-2J(2-1/\varepsilon-\delta)}),$$

et

$$\mathbb{E} [\|\hat{\pi}_J - \pi\|_2^2] \xrightarrow{p \rightarrow \infty} \|\tilde{\pi}_J^* - \pi\|_2^2 + O(2^{-2J(2-1/\varepsilon-\delta)}).$$

**Démonstration.** D'abord, le lemme 4.1 permet d'écrire que

$$\|\hat{\pi}_J - \pi\|_2^2 = \|\tilde{\pi}_J^* - \pi\|_2^2 + (\hat{\mathbf{b}} - \mathbf{b})^t (\hat{\mathbf{b}} - \mathbf{b}).$$

Commençons par étudier la convergence en probabilité. Pour cela, nous portons notre attention sur le dernier terme de l'égalité précédente. La convergence en probabilité de  $\widehat{\mathbf{b}}$  a déjà été obtenue au théorème 4.4, soit

$$\widehat{\mathbf{b}} \xrightarrow{P} \mathbf{b}^*.$$

Nous avons donc (voir Serfling, 1980, section 1.7) la propriété suivante,

$$(\widehat{\mathbf{b}} - \mathbf{b})^t (\widehat{\mathbf{b}} - \mathbf{b}) \xrightarrow{P} (\mathbf{b}^* - \mathbf{b})^t (\mathbf{b}^* - \mathbf{b}). \quad (4.31)$$

Par ailleurs, grâce au théorème 4.2, nous savons que

$$\mathbf{b} = \mathbf{b}^* + O(2^{-J(5/2-\delta)}) \mathbf{1}_{2N+1},$$

ou encore, que

$$\mathbf{b}^* - \mathbf{b} = O(2^{-J(5/2-\delta)}) \mathbf{1}_{2N+1}.$$

Ainsi, le deuxième terme de l'équation (4.31) est équivalent à

$$\begin{aligned} (\mathbf{b}^* - \mathbf{b})^t (\mathbf{b}^* - \mathbf{b}) &= O(2^{-2J(5/2-\delta)}) \mathbf{1}_{2N+1}^t \mathbf{1}_{2N+1} \\ &= O(2^{-2J(5/2-\delta)}) (2N+1), \end{aligned}$$

et comme  $N \propto 2^{J(1+2/\varepsilon)}$ , nous trouvons finalement

$$(\mathbf{b}^* - \mathbf{b})^t (\mathbf{b}^* - \mathbf{b}) = O(2^{-2J(2-1/\varepsilon-\delta)}). \quad (4.32)$$

La convergence en probabilité qui s'en suit est donnée par

$$(\widehat{\mathbf{b}} - \mathbf{b})^t (\widehat{\mathbf{b}} - \mathbf{b}) \xrightarrow{P} O(2^{-2J(2-1/\varepsilon-\delta)}),$$

et permet de conclure que

$$\|\widehat{\pi}_J - \pi\|_2^2 \xrightarrow{P} \|\widehat{\pi}_J^* - \pi\|_2^2 + O(2^{-2J(2-1/\varepsilon-\delta)}).$$

Pour la convergence en moyenne, notons d'abord que

$$\mathbb{E} [\|\widehat{\pi}_J - \pi\|_2^2] = \|\widehat{\pi}_J^* - \pi\|_2^2 + \mathbb{E} [(\widehat{\mathbf{b}} - \mathbf{b})^t (\widehat{\mathbf{b}} - \mathbf{b})], \quad (4.33)$$

et qu'en employant le fait que

$$\mathbb{E}(\widehat{\mathbf{b}}) = \mathbf{b}^*,$$

le dernier terme de l'égalité précédente peut être exprimé comme

$$\begin{aligned}
\mathbb{E} \left[ (\widehat{\mathbf{b}} - \mathbf{b})^t (\widehat{\mathbf{b}} - \mathbf{b}) \right] &= \mathbb{E} \left[ (\widehat{\mathbf{b}} - \mathbf{b}^*)^t (\widehat{\mathbf{b}} - \mathbf{b}^*) \right] + (\mathbf{b}^* - \mathbf{b})^t (\mathbf{b}^* - \mathbf{b}) \\
&= \mathbb{E} \left[ \text{tr} \left( (\widehat{\mathbf{b}} - \mathbf{b}^*)^t (\widehat{\mathbf{b}} - \mathbf{b}^*) \right) \right] + (\mathbf{b}^* - \mathbf{b})^t (\mathbf{b}^* - \mathbf{b}) \\
&= \text{tr} \left( \mathbb{E} \left[ (\widehat{\mathbf{b}} - \mathbf{b}^*) (\widehat{\mathbf{b}} - \mathbf{b}^*)^t \right] \right) + (\mathbf{b}^* - \mathbf{b})^t (\mathbf{b}^* - \mathbf{b}) \\
&= \text{tr} \left( \text{Cov}(\widehat{\mathbf{b}}) \right) + (\mathbf{b}^* - \mathbf{b})^t (\mathbf{b}^* - \mathbf{b}).
\end{aligned}$$

Maintenant, d'après l'équation (4.32) et le théorème 4.3, le résultat précédent est équivalent à

$$\mathbb{E} \left[ (\widehat{\mathbf{b}} - \mathbf{b})^t (\widehat{\mathbf{b}} - \mathbf{b}) \right] = \frac{36}{p} \text{tr} \left( \mathbf{B}^{-1} \Sigma \mathbf{B}^{-1t} \right) + O(2^{-2J(2-1/\varepsilon-\delta)}),$$

ce qui permet de récrire l'équation (4.33) comme

$$\mathbb{E} \left[ \|\widehat{\pi}_J - \pi\|_2^2 \right] = \|\widehat{\pi}_J^* - \pi\|_2^2 + \frac{36}{p} \text{tr} \left( \mathbf{B}^{-1} \Sigma \mathbf{B}^{-1t} \right) + O(2^{-2J(2-1/\varepsilon-\delta)}).$$

Cette dernière égalité permet d'obtenir

$$\mathbb{E} \left[ \|\widehat{\pi}_J - \pi\|_2^2 \right] \xrightarrow{p \rightarrow \infty} \|\widehat{\pi}_J^* - \pi\|_2^2 + O(2^{-2J(2-1/\varepsilon-\delta)}),$$

soit la convergence en moyenne recherchée.  $\square$

Nous faisons remarquer au lecteur qu'il est possible de montrer que

$$\begin{aligned}
\mathbb{E} \left[ \|\widehat{\pi}_J - \pi\|_2^2 \right] &= \left\| \mathbb{E} \left[ (\widehat{\pi}_J(s) - \pi(s))^2 \right] \right\|_2^2 \\
&= \left\| \text{EQM}(\widehat{\pi}_J(s)) \right\|_2^2,
\end{aligned}$$

et que plusieurs auteurs utilisent cette dernière quantité pour caractériser le comportement de l'estimateur d'une densité. Cette quantité est habituellement appelée *écart quadratique moyen intégré*, et notée  $\text{EQMI}(\widehat{\pi}_J)$ . Pour plus d'information à ce sujet, le lecteur peut se référer à Izenman (1991).

Finalement, pour bien illustrer la teneur des résultats obtenus ici, nous faisons un bref retour sur quelques points que nous jugeons essentiels. D'abord, l'estimateur  $\widehat{\pi}_J$  de la densité *a priori*  $\pi$  possède les propriétés suivantes, sous les hypothèses des théorèmes 2.2, 2.11 et 4.2, si  $\delta < 2 - 1/\varepsilon$  et pour  $J$  fixé :

(i) convergence ponctuelle en probabilité :

$$\widehat{\pi}_J(s) \xrightarrow{P} \widehat{\pi}_J^*(s) + O(2^{-J(2-\delta)}) \quad \forall s \in \mathbb{R}.$$

(ii) convergence ponctuelle en moyenne quadratique :

$$\text{EQM}(\hat{\pi}_J(s)) \xrightarrow{p \rightarrow \infty} (\tilde{\pi}_J^*(s) - \pi(s))^2 + O(2^{-J(2-\delta)}) \quad \forall s \in \mathbb{R}.$$

(iii) convergence en probabilité de la norme quadratique :

$$\|\hat{\pi}_J - \pi\|_2^2 \xrightarrow{P} \|\tilde{\pi}_J^* - \pi\|_2^2 + O(2^{-2J(2-1/\varepsilon-\delta)}).$$

(iv) convergence en moyenne de la norme quadratique (ou en moyenne quadratique intégrée):

$$\mathbb{E} [\|\hat{\pi}_J - \pi\|_2^2] \xrightarrow{p \rightarrow \infty} \|\tilde{\pi}_J^* - \pi\|_2^2 + O(2^{-2J(2-1/\varepsilon-\delta)}).$$

Maintenant, au début de la présente section, nous avons déjà discuté du fait que

$$\tilde{\pi}_J^*(s) \xrightarrow{J \rightarrow \infty} \pi(s) \quad \forall s \in \mathbb{R}, \quad (4.34)$$

et que

$$\|\tilde{\pi}_J^* - \pi\|_2^2 \xrightarrow{J \rightarrow \infty} 0.$$

Ces deux résultats prennent ici tout leur sens. En effet, grâce à eux, il devient évident que la performance de  $\hat{\pi}_J$  s'améliore au fur et à mesure que  $J$  augmente. Ultimement, ceci nous amène à conclure que, sous les hypothèses des théorèmes 2.2, 2.11 et 4.2 et si  $\delta < 2 - 1/\varepsilon$ , alors  $\hat{\pi}_J$  satisfait les critères de convergence classiques que sont les quatre critères mentionnés plus haut.

Notons également qu'en jumelant les équations (4.27) et (4.34) il devient clair que  $\hat{\pi}_J$  satisfait

$$\mathbb{E}(\hat{\pi}_J(s)) \xrightarrow{J \rightarrow \infty} \pi(s) \quad \forall s \in \mathbb{R},$$

et donc que  $\hat{\pi}_J$  est, en ce sens (c'est-à-dire lorsque  $J \rightarrow \infty$ ), asymptotiquement sans biais. L'estimateur  $\hat{\pi}_J$  satisfait donc aussi cette dernière propriété qui est habituellement recherchée en estimation de densité.

Le comportement de l'estimateur de la densité *a priori* que nous avons construit nous apparaît donc comme étant adéquat. En effet, les cinq critères mentionnés ici (quatre critères de convergence et absence de biais au niveau asymptotique), qui sont les plus largement employés, sont tous satisfaits par  $\hat{\pi}_J$ . Le lecteur intéressé par une discussion portant sur le bien-fondé et l'utilisation des divers critères d'évaluation d'un estimateur de densité peut consulter Izenman (1991) et les nombreuses références qui y sont fournies.

À travers la prochaine section, nous discutons des difficultés rencontrées et de quelques points dont il faut tenir compte, en pratique, lors de l'utilisation

de  $\hat{\pi}_J$ . Aussi, nous indiquons une manière de contourner les problèmes identifiés. Finalement, à la section 4.7, nous évaluerons le rendement réel de notre estimateur à l'aide d'un exemple simulé.

## 4.6. Considérations pratiques

Nous examinons ici quelques points à considérer lors de l'emploi de  $\hat{\pi}_J$ . Tout d'abord, les définitions de  $\hat{\pi}_J$  (voir définition 4.1) et des fonctions  $\phi_{Jk}$  de la base de Haar, données par l'équation (1.15), impliquent, comme nous l'avons dit plus tôt, que  $\hat{\pi}_J$  est une fonction constante par morceaux. Ainsi,  $\hat{\pi}_J$  est en quelque sorte une approximation par histogramme de la densité *a priori* inconnue  $\pi$ . De plus, chaque bande de cet histogramme est de largeur  $2^{-J}$ . Il est donc clair que  $J$  ne doit pas être choisi trop grand, puisqu'un histogramme ne permet de dégager que l'allure globale d'une densité.

Notons aussi que les définitions de  $\hat{\mathbf{m}}$ , donnée par l'équation (4.22), et de  $\hat{m}_J$ , donnée par l'équation (4.6), permettent de voir que

$$\hat{m}_J(x) = \sum_{k \in \mathcal{N}} \hat{m}_{Jk} \phi_{Jk}(x), \quad (4.35)$$

est une estimation par histogramme de la densité marginale des observations. Ici tout particulièrement, il est évident que si  $J$  est choisi trop grand, alors  $\hat{m}_J$  deviendra trop chaotique pour que les tendances des observations recueillies soient vraiment saisies par  $\hat{\mathbf{m}}$  et  $\hat{m}_J$ . À la limite,  $\hat{m}_J$  tendra vers une suite de points de masse situés à chacune des observations. Évidemment, il est utopique, dans de telles conditions, d'espérer obtenir une estimation intéressante de la densité *a priori* à l'aide de  $\hat{\pi}_J$ .

Un deuxième argument, celui-là d'ordre calculatoire, appuie le choix de petites valeurs de  $J$ . En effet, le vecteur  $\hat{\mathbf{b}}$  est de longueur  $2N+1$  avec  $N \propto 2^{J(1+2/\varepsilon)}$ . Ainsi, le nombre de paramètres à estimer augmente de façon exponentielle avec  $J$ . Dans un tel contexte, il est tout à fait normal de souhaiter conserver un nombre raisonnable de paramètres. De plus, le calcul de  $\hat{\mathbf{b}}$  à partir de l'équation (4.23) fait intervenir la matrice  $\mathbf{B}$ , de taille  $(2N+1) \times (2N+1)$ . Ici plus particulièrement, il est clair qu'une augmentation de  $J$ , même légère, peut se traduire par une augmentation astronomique de la taille de  $\mathbf{B}$ , ceci ayant comme conséquence de rendre le calcul de  $\hat{\mathbf{b}}$  extrêmement lourd. Il semble donc souhaitable, encore une fois, de se restreindre à des valeurs de  $J$  relativement petites, et ce malgré les arguments théoriques présentés à la section précédente militant en faveur d'un choix de  $J$  "assez grand".

Il nous faut aussi nous intéresser au choix de  $N$ , qui détermine ici le nombre de termes à inclure dans l'estimation de  $\pi$ . En effet, rappelons que  $\hat{\pi}_J$  est donné par

$$\hat{\pi}_J(s) = \sum_{k=-N}^N \hat{b}_{0,Jk} \phi_{Jk}(s),$$

$N$  correspondant à l'indice du dernier terme à inclure. Nous proposons de choisir  $N$  selon

$$N = C2^J.$$

Ce choix de  $N$ , basé sur les résultats obtenus à la section 2.6 et sur la discussion présentée à la section 2.7, représente un choix adéquat dans le cas d'un modèle où l'une des deux densités est à décroissance exponentielle. Cependant, seule  $f$  est ici connue. Bien sûr, dans le cas où celle-ci exhibe une décroissance exponentielle, ce choix de  $N$  est parfaitement justifié. Par simplicité, nous proposons de faire le même choix dans les autres cas.

Maintenant, nous proposons de choisir  $C$  de façon à s'assurer que toutes les observations soient situées à l'intérieur du support de l'estimateur  $\hat{m}_J$  (de la densité marginale  $m$ ) qui a été défini à l'équation (4.35). Autrement dit, le choix de  $C$  devrait être fait pour que

$$X_i \in \bigcup_{k=-N}^N A_{Jk} \quad \forall i \in \{1, 2, \dots, p\},$$

ou encore, pour que  $\hat{m}_J$  soit une fonction de densité.

Nous nous tournons maintenant vers les deux principales difficultés rencontrées lors de l'utilisation de  $\hat{\pi}_J$ . La première est commune à beaucoup d'autres techniques d'estimation de densité:  $\hat{\pi}_J$  n'est habituellement pas une estimation *bona fide* de  $\pi$ . Cette locution latine signifiant "de bonne foi" est employée dans la littérature pour signifier que l'estimation obtenue est bien elle-même une densité. Notons qu'ici, la définition de l'estimateur par déconvolution  $\hat{\pi}_J$  permet d'obtenir les conditions nécessaires et suffisantes pour que la fonction  $\hat{\pi}_J(s)$  soit vue comme une densité. En effet, en jumelant l'équation (1.15) à la définition de  $\hat{\pi}_J$ , nous avons d'abord

$$\hat{\pi}_J(s) \geq 0 \quad \forall s \in \mathbb{R} \quad \iff \quad \hat{b}_{0,Jk} \geq 0 \quad \forall k \in \mathcal{N}. \quad (4.36)$$



Ensuite, comme

$$\begin{aligned} \int_{\mathbb{R}} \hat{\pi}_J(s) ds &= \sum_{k \in \mathcal{N}} \hat{b}_{0,Jk} \int_{\mathbb{R}} \phi_{Jk}(s) ds \\ &= \frac{1}{2^{J/2}} \sum_{k \in \mathcal{N}} \hat{b}_{0,Jk}, \end{aligned}$$

alors une deuxième condition essentielle est donnée par

$$\int_{\mathbb{R}} \hat{\pi}_J(s) ds = 1 \quad \iff \quad \sum_{k \in \mathcal{N}} \hat{b}_{0,Jk} = 2^{J/2}. \quad (4.37)$$

Ainsi, les conditions nécessaires et suffisantes pour que  $\hat{\pi}_J$  soit une estimation *bona fide* de  $\pi$  sont toutes simples et peuvent être vérifiées facilement. Par ailleurs, rien n'indique qu'elles seront satisfaites pour un problème donné. En fait, nous pouvons même nous attendre à ce que, de façon générale, les équations (4.36) et (4.37) ne soient pas satisfaites puisque la construction de  $\hat{\pi}_J$  n'a pas tenu compte de ces deux critères.

La solution à ce problème est cependant assez simple. En effet, il nous semble naturel de définir la version tronquée et normalisée de  $\hat{\pi}_J$  selon

$$\hat{\pi}_J^+(s) = \sum_{k \in \mathcal{N}} \hat{\beta}_{Jk} \phi_{Jk}(s),$$

où

$$\hat{\beta}_{Jk} = \frac{2^{J/2}}{\sum_{l \in \mathcal{N}} [\hat{b}_{0,Jl}]^+} [\hat{b}_{0,Jk}]^+,$$

et où, finalement,

$$[\hat{b}_{0,Jk}]^+ = \begin{cases} \hat{b}_{0,Jk} & \text{si } \hat{b}_{0,Jk} \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

Il est facile de voir que, ainsi défini,  $\hat{\pi}_J^+$  est bien un estimateur *bona fide*.

La deuxième difficulté dont nous voulons discuter ici est peut-être déjà apparue au lecteur au moment de définir le vecteur  $\hat{\mathbf{b}}$ , à l'équation (4.23), ou à la lecture de l'énoncé du théorème 4.2. Dans ces deux cas, nous avons fait l'hypothèse que la matrice  $\mathbf{B}$  est non singulière. Même si, en pratique, cette condition est habituellement vérifiée, l'inversion de  $\mathbf{B}$  amène quand même une difficulté supplémentaire. En effet, la matrice  $\mathbf{B}$  est habituellement mal conditionnée, et

parfois d'une façon extrême. Ce problème survient lorsque  $\mathbf{B}$  est "presque" singulière, c'est-à-dire lorsqu'une très légère perturbation de ses éléments peut la rendre singulière.

Dans une telle situation, l'inverse de  $\mathbf{B}$  contient souvent plusieurs très grands éléments en valeur absolue. Le comportement de l'estimateur  $\hat{\mathbf{b}}$  peut alors devenir très instable. Ce phénomène est clairement illustré par le fait que la matrice de covariance de  $\hat{\mathbf{b}}$ , donnée par (voir théorème 4.3)

$$\text{Cov}(\hat{\mathbf{b}}) = \frac{36}{p} \mathbf{B}^{-1} \Sigma \mathbf{B}^{-1t},$$

peut alors aussi contenir de très grands éléments, amenant certains des  $\hat{b}_{jk}$  à avoir une trop grande variance. Pour plus de détails sur le mauvais conditionnement des matrices, ainsi que sur les répercussions du mauvais conditionnement des matrices sur la solution des systèmes linéaires et la régression linéaire multiple, le lecteur peut consulter Stewart (1973, 1987), Belsley *et al.* (1980, chapitre 3), Vinod et Ullah (1981, chapitre 5), et Lancaster et Tismenetsky (1985, chapitres 10 et 11).

Pour remédier à la situation, nous proposons d'utiliser la décomposition en valeurs singulières (SVD) de la matrice  $\mathbf{B}$  et le principe de la régression par composantes principales. Plus particulièrement, nous adoptons la même approche que Mandel (1982). Tout d'abord, la décomposition SVD de  $\mathbf{B}$  est donnée par

$$\mathbf{B} = \mathbf{U}^t \text{Diag}(\Lambda) \mathbf{V}, \quad (4.38)$$

où  $\mathbf{U}$  et  $\mathbf{V}$  sont des matrices orthogonales de dimension  $(2N + 1) \times (2N + 1)$ , et  $\text{Diag}(\Lambda)$  est la matrice diagonale dont les éléments de la diagonale sont

$$\Lambda = (\lambda_i)_{i \in \{1, 2, \dots, 2N+1\}},$$

et respectent

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{2N+1} \geq 0.$$

Les valeurs  $\lambda_1, \lambda_2, \dots, \lambda_{2N+1}$  sont appelées valeurs singulières de la matrice  $\mathbf{B}$  et correspondent en fait aux racines positives des valeurs propres ordonnées de la matrice  $\mathbf{B}^t \mathbf{B}$ . Notons que  $\mathbf{B}$  est non singulière si, et seulement si, toutes ses valeurs singulières sont non nulles. Ainsi, si  $\mathbf{B}$  est singulière, nous savons qu'il existe un certain  $n$  tel que  $1 \leq n \leq 2N$ , et pour lequel

$$\lambda_i > 0 \quad \forall i \in \{1, 2, \dots, n\},$$

et

$$\lambda_i = 0 \quad \forall i \in \{n + 1, n + 2, \dots, 2N + 1\}.$$

Nous excluons ici le cas où  $n = 0$ , soit le cas où

$$\lambda_i = 0 \quad \forall i \in \{1, 2, \dots, 2N + 1\},$$

puisqu'il correspond au cas où

$$\mathbf{B} = \mathbf{0}_{(2N+1) \times (2N+1)},$$

qui est clairement inintéressant.

Examinons maintenant certaines des implications de la décomposition donnée par l'équation (4.38). Pour cela, posons

$$\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{2N+1}),$$

et

$$\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{2N+1}),$$

où  $\mathbf{u}_i$  correspond à la  $i^{\text{ème}}$  colonne de la matrice  $\mathbf{U}$ , et  $\mathbf{v}_i$  à la  $i^{\text{ème}}$  colonne de  $\mathbf{V}$ . Notons que les vecteurs  $\mathbf{u}_i$  sont orthonormés et que c'est aussi le cas pour les vecteurs  $\mathbf{v}_i$ . Maintenant, en utilisant cette notation, nous pouvons écrire  $\mathbf{B}$  sous la forme alternative suivante,

$$\begin{aligned} \mathbf{B} &= \begin{pmatrix} \mathbf{u}_1^t \\ \mathbf{u}_2^t \\ \vdots \\ \mathbf{u}_{2N+1}^t \end{pmatrix} \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \dots & \\ 0 & & & \lambda_{2N+1} \end{pmatrix} (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{2N+1}) \\ &= \sum_{i=1}^{2N+1} \lambda_i \mathbf{u}_i \mathbf{v}_i^t. \end{aligned} \quad (4.39)$$

Cette dernière expression est toujours valable, même lorsque  $\mathbf{B}$  est singulière. De plus, elle illustre bien le rôle des valeurs singulières dans la décomposition SVD de  $\mathbf{B}$  et permet de voir que les petites valeurs singulières contribuent très peu à la structure générale de  $\mathbf{B}$ .

Supposons, pour le moment, que  $\mathbf{B}$  est non singulière, et voyons comment caractériser son inverse  $\mathbf{B}^{-1}$ . Remarquons, tout d'abord, que

$$\mathbf{B}^{-1} = \mathbf{V}^t \mathbf{Diag}^{-1}(\Lambda) \mathbf{U},$$

puisque  $\mathbf{U}$  et  $\mathbf{V}$  sont des matrices orthogonales. De plus, il est clair que

$$\begin{aligned} \text{Diag}^{-1}(\Lambda) &= \begin{pmatrix} 1/\lambda_1 & & & 0 \\ & 1/\lambda_2 & & \\ & & \ddots & \\ 0 & & & 1/\lambda_{2N+1} \end{pmatrix} \\ &= \text{Diag}(\Lambda^{-1}), \end{aligned}$$

où  $\Lambda^{-1}$  est le vecteur défini, par convention, comme

$$\Lambda^{-1} = \left( \frac{1}{\lambda_i} \right)_{i \in \{1, 2, \dots, 2N+1\}}.$$

Ainsi, une expression de la même forme que celle de l'équation (4.39) peut être obtenue pour  $\mathbf{B}^{-1}$ , soit

$$\begin{aligned} \mathbf{B}^{-1} &= \mathbf{V}^t \text{Diag}(\Lambda^{-1}) \mathbf{U} \\ &= \sum_{i=1}^{2N+1} \frac{1}{\lambda_i} \mathbf{v}_i \mathbf{u}_i^t. \end{aligned} \quad (4.40)$$

Cette façon d'écrire  $\mathbf{B}^{-1}$  est particulièrement instructive. En effet, elle permet d'établir que si certaines valeurs singulières de  $\mathbf{B}$  sont trop petites (par exemple si  $\mathbf{B}$  est "presque" singulière), alors il est fort possible que  $\mathbf{B}^{-1}$  contienne de très grands éléments. Dans une telle situation, il est normal de s'attendre à beaucoup d'instabilité lors de l'estimation de  $\mathbf{b}$ . De plus, il est fort troublant de constater que les valeurs singulières les moins influentes, au niveau de la structure de  $\mathbf{B}$ , sont possiblement les plus influentes quant à la structure de  $\mathbf{B}^{-1}$ . Tout ceci illustre bien le fait que la matrice  $\mathbf{B}$  peut être très mal conditionnée lorsqu'elle possède de très petites valeurs singulières.

Maintenant, supposons que

$$\lambda_i > 0 \quad \forall i \in \{1, 2, \dots, n\},$$

et que

$$\lambda \simeq 0 \quad \forall i \in \{n+1, n+2, \dots, 2N+1\}.$$

Il est clair, d'après l'équation (4.39), que nous avons alors

$$\mathbf{B} \simeq \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{v}_i^t.$$

Posons donc

$$\begin{aligned}\mathbf{B}_n &= \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{v}_i^t \\ &= \mathbf{U}^t \mathbf{D} \text{Diag}(\Lambda_n) \mathbf{V}\end{aligned}$$

où

$$\Lambda_n = (\lambda_1, \lambda_2, \dots, \lambda_n, 0, \dots, 0)^t.$$

Notons que si  $\lambda_{n+1}, \lambda_{n+2}, \dots, \lambda_{2N+1}$  sont très petits, alors  $\mathbf{B}_n$  devrait être une bonne approximation de  $\mathbf{B}$ . Notons également que cette façon de faire est équivalente à fixer artificiellement

$$\lambda_i = 0 \quad \forall i \in \{n+1, n+2, \dots, 2N+1\},$$

et par le fait même, d'amener artificiellement  $\mathbf{B}$  à être singulière. Remarquons finalement que si  $\mathbf{B}$  est singulière dès le départ, alors il existe un  $n < 2N+1$  pour lequel nous avons directement

$$\mathbf{B} = \mathbf{B}_n.$$

Revenons à l'équation (4.23), où nous définissons  $\widehat{\mathbf{b}}$ , et notons que cette définition est équivalente à dire que  $\widehat{\mathbf{b}}$  est la solution du système linéaire

$$\widehat{\mathbf{m}} = \frac{1}{6} \mathbf{B} \widehat{\mathbf{b}}.$$

Si, comme précédemment, nous avons

$$\lambda_i \simeq 0 \quad \forall i \in \{n+1, n+2, \dots, 2N+1\},$$

alors

$$\widehat{\mathbf{m}} = \frac{1}{6} \mathbf{B}_n \widehat{\mathbf{b}} \tag{4.41}$$

est une approximation du système que nous souhaitons résoudre. Par contre, comme  $\mathbf{B}_n$  est singulière, nous sommes ici confrontés à l'une des deux situations suivantes :

- le nouveau système n'a pas de solution;
- le nouveau système possède une infinité de solutions.

Dans un cas comme dans l'autre, l'approche à adopter est la même. Nous proposons d'approximer  $\widehat{\mathbf{b}}$  par

$$\widehat{\mathbf{b}}_n = 6 \mathbf{B}_n^- \widehat{\mathbf{m}}, \tag{4.42}$$

où  $\mathbf{B}_n^-$  correspond à l'inverse généralisée de Moore-Penrose de  $\mathbf{B}_n$ , soit

$$\mathbf{B}_n^- = \mathbf{V}^t \text{Diag}(\Lambda_n^-) \mathbf{U}, \quad (4.43)$$

où

$$\Lambda_n^- = \left( \frac{1}{\lambda_1}, \frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_n}, 0, \dots, 0 \right)^t.$$

Le lecteur souhaitant se familiariser avec l'inverse de Moore-Penrose et les inverses généralisées en général peut consulter Graybill (1969, chapitre 6), Searle (1982, chapitre 8) et Lancaster et Tismenetsky (1985, chapitre 12).

Notons que l'équation (4.43) nous permet d'écrire  $\mathbf{B}_n^-$  comme

$$\mathbf{B}_n^- = \sum_{i=1}^n \frac{1}{\lambda_i} \mathbf{v}_i \mathbf{u}_i^t. \quad (4.44)$$

Il est intéressant de comparer cette forme de  $\mathbf{B}_n^-$  à l'expression de  $\mathbf{B}^{-1}$  donnée par l'équation (4.40). En effet, si

$$\lambda_i \simeq 0 \quad \forall i \in \{n+1, n+2, \dots, 2N+1\},$$

alors les termes correspondant à ces valeurs singulières dans le développement de  $\mathbf{B}^{-1}$  vont amener l'instabilité de  $\widehat{\mathbf{b}}$  dont nous avons discuté précédemment. Par contre, ces mêmes termes ne sont pas inclus dans le développement de  $\mathbf{B}_n^-$  présenté plus haut. Ainsi, en passant de  $\widehat{\mathbf{b}}$  à  $\widehat{\mathbf{b}}_n$ , nous avons en quelque sorte stabilisé l'estimateur  $\widehat{\mathbf{b}}$ , ce qui est reflété par la matrice de covariance de  $\widehat{\mathbf{b}}_n$ , soit

$$\text{Cov}(\widehat{\mathbf{b}}_n) = \frac{36}{p} \mathbf{B}_n^- \Sigma (\mathbf{B}_n^-)^t.$$

En effet, cette matrice correspond essentiellement à la matrice de covariance de  $\widehat{\mathbf{b}}$ , mais débarrassée de l'influence des composantes problématiques de  $\mathbf{B}$ . Notons que  $\text{Cov}(\widehat{\mathbf{b}}_n)$  est une matrice singulière et donc, que dans ces conditions,  $\widehat{\mathbf{b}}_n$  ne possède pas de densité multivariée. Ceci vient du fait que  $2N+1-n$  relations linéaires lient les éléments de  $\widehat{\mathbf{b}}_n$ . En fait, il est possible d'écrire une densité conjointe pour  $n$  composantes bien choisies de  $\widehat{\mathbf{b}}_n$ , toutes les autres pouvant alors être déterminées à partir de la connaissance de ces  $n$  composantes principales.

Maintenant, les arguments présentés précédemment semblent indiquer que l'emploi de  $\widehat{\mathbf{b}}_n$ , quand la situation le demande, est justifié. Une question demeure pourtant : comment choisir  $n$ ? Autrement dit, combien de termes doit-on retrancher dans l'expression de  $\mathbf{B}$ , donnée par l'équation (4.39), pour que le comportement de  $\widehat{\mathbf{b}}_n$  soit satisfaisant?

Cette question, dont une étude approfondie serait certainement très complexe, obtient heureusement ici une réponse assez simple. La technique que nous proposons n'est cependant pas valable dans le contexte de la régression en composantes principales. Dans ce contexte plus général, plusieurs solutions ont déjà été proposées, dont celles de Hill et Judge (1990) et Foucart (1996).

L'emploi de ces différents critères n'est pas ici nécessaire pour obtenir une réponse satisfaisante. En effet, nous avons la possibilité de voir graphiquement l'effet des différents choix de  $n$  sur l'allure de  $\hat{\pi}_J$ , ce qui rend le problème beaucoup plus facile à traiter. Il suffit ainsi à l'utilisateur d'obtenir  $\hat{\pi}_J$  pour quelques valeurs intéressantes de  $n$  et de choisir l'estimation qui lui apparaît comme étant visuellement la plus satisfaisante. En ce sens, le choix de  $n$  peut être comparé aux choix des paramètres de lissage en régression et en estimation de densité par la méthode du noyau.

Pour avoir une idée des valeurs intéressantes de  $n$ , nous proposons le critère suivant. Comme Belsley *et al.* (1980, section 3.2), nous définissons l'indice de conditionnement de  $\mathbf{B}_n$  comme

$$\text{Cond}(\mathbf{B}_n) = \frac{\lambda_1}{\lambda_n}.$$

Ainsi,  $\text{Cond}(\mathbf{B}_n)$  est le rapport entre la plus grande valeur singulière de  $\mathbf{B}$ , et la plus petite considérée comme non nulle dans  $\mathbf{B}_n$ . Nous proposons d'étudier l'allure de  $\hat{\pi}_J$  pour les valeurs de  $n$  telles que

$$5 \leq \text{Cond}(\mathbf{B}_n) \leq 10.$$

Le choix de ces valeurs de base est fondé sur la discussion faite par Belsley *et al.* (1980, section 3.2). Nous sommes bien conscients du caractère relativement arbitraire de ce choix, c'est pourquoi nous avançons que le choix de  $n$  fait par l'utilisateur devrait être basé sur une étude graphique. Ces valeurs de l'indice de conditionnement ne sont réellement qu'un point de départ pour l'étude graphique du problème.

Notons que pour arriver à une exploration plus rapide, les équations (4.42) et (4.44) permettent d'obtenir une expression fort utile pour  $\hat{\mathbf{b}}_n$ . En effet, nous avons

$$\begin{aligned} \hat{\mathbf{b}}_n &= 6 \left( \sum_{i=1}^n \frac{1}{\lambda_i} \mathbf{u}_i \mathbf{v}_i^t \right) \hat{\mathbf{m}} \\ &= \sum_{i=1}^n \frac{6w_i}{\lambda_i} \mathbf{u}_i, \end{aligned}$$

où  $\{w_i\}_{i \in \{1, 2, \dots, 2N+1\}}$  est une suite de scalaires définis par

$$w_i = \mathbf{v}_i^t \widehat{\mathbf{m}}.$$

Cette forme alternative est intéressante pour deux raisons différentes. La première est qu'elle permet de calculer  $\widehat{\mathbf{b}}_n$  sans avoir à effectuer de multiplication matricielle. Ceci peut représenter un gain d'efficacité considérable, au niveau du temps de calcul, lorsque  $n$  est assez grand. Ensuite, et surtout, grâce à cette relation il est possible de relier deux approximations successives de  $\widehat{\mathbf{b}}$ . Nous avons ainsi

$$\begin{aligned} \widehat{\mathbf{b}}_{k+1} &= \sum_{i=1}^{k+1} \frac{6w_i}{\lambda_i} \mathbf{u}_i \\ &= \widehat{\mathbf{b}}_k + \frac{6w_{k+1}}{\lambda_{k+1}}, \end{aligned}$$

et le passage de  $\widehat{\mathbf{b}}_k$  à  $\widehat{\mathbf{b}}_{k+1}$  (ou l'inverse) est très simple, mais aussi très peu exigeant numériquement (bien sûr, une fois que la décomposition SVD de  $\mathbf{B}$  a été obtenue). Ainsi, il est facile d'obtenir graphiquement plusieurs estimations de  $\pi$  basées sur des approximations successives de  $\widehat{\mathbf{b}}$ , comme par exemple les différentes valeurs de  $\widehat{\mathbf{b}}_k$  telles que  $n_1 \leq k \leq n_2$ . Grâce au résultat précédent, nous pouvons donc comparer les estimations obtenues à partir d'un très large éventail de valeurs possiblement intéressantes pour  $n$ , le tout dans un temps de calcul raisonnable.

## 4.7. Exemple simulé

Nous passons maintenant à une étude du comportement de l'estimateur par déconvolution de la densité *a priori* faite à l'aide d'un exemple simulé. Cet exemple a tout particulièrement pour but d'illustrer les caractéristiques pratiques de l'estimateur  $\hat{\pi}_J$ .

### 4.7.1. Cas d'un mélange de deux normales

Nous supposons ici que  $p$  observations conditionnellement indépendantes sont distribuées selon

$$X_i | \theta_i \sim N(\theta_i, 1),$$

pour  $i = 1, 2, \dots, p$ . Nous supposons également que les paramètres de position  $\theta_1, \theta_2, \dots, \theta_p$  sont aussi indépendants et qu'ils partagent la même densité *a priori*.



Tour à tour, nous considérons les cas où la densité *a priori* des paramètres de position est donnée par

$$\pi_j(\theta) \equiv \frac{1}{3}N(-\mu_j, 1) + \frac{2}{3}N(\mu_j, 1),$$

pour  $j = 1, 2, 3$ . Nous considérons les valeurs  $\mu_1 = 3$ ,  $\mu_2 = 2$  et  $\mu_3 = 3/2$ . Notre but est ici de reconstruire les densités *a priori*  $\pi_1$ ,  $\pi_2$  et  $\pi_3$  en procédant de deux manières différentes. Dans un premier temps, pour illustrer la justesse du théorème 4.2, nous reconstruisons les densités *a priori* en utilisant le fait que les densités marginales peuvent être facilement calculées, et donc que le vecteur  $\mathbf{m}$  peut être obtenu, dans chaque cas, directement à l'aide de l'équation (4.7). Ensuite, pour illustrer le comportement de l'estimateur par déconvolution  $\hat{\pi}_J$ , nous reconstruisons toujours les mêmes densités *a priori*, mais cette fois à partir de données simulées. Cet exercice nous permettra de voir lesquelles des caractéristiques (particulièrement la position et la hauteur des modes, ainsi que l'écart entre les modes) des trois densités *a priori* considérées sont détectées par l'estimateur  $\hat{\pi}_j$ .

Notons que dans le présent contexte, les densités marginales des observations associées à chacune des densités *a priori* précédentes sont données par

$$m_j(x) \equiv \frac{1}{3}N(-\mu_j, 2) + \frac{2}{3}N(\mu_j, 2), \quad (4.45)$$

pour  $j = 1, 2, 3$ . De là, il est donc effectivement facile d'obtenir les vecteurs, notés  $\mathbf{m}_j$ , des coefficients de la décomposition de Haar de chacune des densités marginales  $m_j$ , et de générer des observations provenant de chaque modèle à partir de ces densités marginales.

Pour notre étude, nous considérons les valeurs de  $J = 0, 1, 2$ . Ces valeurs, quoique clairement très petites (en considérant la nature des résultats asymptotiques de la section 4.5), sont ici tout à fait appropriées, comme le démontrerons les différentes reconstructions obtenues. Ces choix ont été faits pour les deux parties de notre étude.

Pour  $N$ , nous utilisons dans chaque cas

$$N = C2^J,$$

avec ici  $C = 8$ . Ceci permet d'obtenir une reconstruction des diverses densités *a priori* pour  $\theta \in [-8, 8]$ . Ce choix a également été fait pour les deux parties de notre étude. Tout d'abord, en considérant les densités marginales données par l'équation (4.45), nous avons

$$\mathbb{P}_{m_j}(|X| < 8) > 0,9995,$$

TAB. 4.3. Coefficients de détermination  $R^2$  et estimations de  $\delta$  obtenus pour chaque modèle.

No du modèle	$R^2$	$\delta_{mc}$	Ordre approximatif de l'erreur
1	0,9977	0,2532	$y(J) = O(2^{-2,2468J})$
2	0,9981	0,2688	$y(J) = O(2^{-2,2312J})$
3	0,9996	0,0037	$y(J) = O(2^{-2,4973J})$

pour  $j = 1, 2, 3$ . Ainsi, il est clair que la partie significative du support de chacune des densités marginales  $m_j$  est incluse dans l'intervalle identifié plus haut. Ensuite, pour l'estimation des densités *a priori* à partir d'échantillons simulés, ce choix nous assure, comme nous le verrons plus tard, que toutes les observations simulées sont incluses dans le support de l'estimation par histogramme de  $m_j$  définie selon l'équation (4.35). Par convention, nous dénotons cet histogramme par  $\hat{m}_{J(j)}$ . De même, nous dénotons l'estimateur par déconvolution de  $\pi_j$  comme étant  $\hat{\pi}_{J(j)}$ .

Nous commençons notre étude du problème en considérant les reconstructions purement théoriques faites à partir des coefficients  $\mathbf{m}_j$  de chacune des densités marginales. L'étude de l'erreur d'approximation présentée aux tableaux 4.1 et 4.2 est reprise avec les trois modèles considérés ici. Le tableau 4.3 présente les résultats obtenus. Notons que les différents coefficients de détermination sont, encore une fois, très élevés. L'existence d'une relation du type

$$y(J) = O(2^{-(5/2-\delta)}),$$

avec  $\delta < 2$ , est donc plausible. La détérioration de l'erreur d'approximation due au processus d'inversion semble également être minimale pour les exemples considérés ici et même que, dans le cas du troisième modèle, la perte de précision est, étonnamment, négligeable.

Voyons maintenant graphiquement les reconstructions obtenues des diverses densités *a priori*. D'abord, à la figure 4.1, nous présentons les reconstructions de  $\pi_1$  obtenues à partir du vecteur  $\mathbf{m}_1$  dans les cas où  $J = 0, 1, 2$ . Il est clair que la technique de reconstruction est ici très efficace pour les trois valeurs de  $J$  considérées. Il en va de même pour les figures 4.2 et 4.3, présentant les reconstructions de  $\pi_2$  et de  $\pi_3$  obtenues à partir des vecteurs  $\mathbf{m}_2$  et  $\mathbf{m}_3$  (pour  $J = 0, 1, 2$ ). Il est intéressant de noter que pour obtenir ces différentes reconstructions, il n'a pas été nécessaire de tronquer et de renormaliser les densités obtenues. De plus il n'a pas non plus été nécessaire d'employer la technique basée sur la décomposition SVD de la matrice  $\mathbf{B}$ . Notons finalement que, pour une valeur de  $J$  fixée, la même matrice  $\mathbf{B}$  est utilisée pour la reconstruction des trois différentes densités *a priori* considérées ici. En effet, une fois les valeurs de  $C$  et  $J$  fixées, l'équation (4.15)

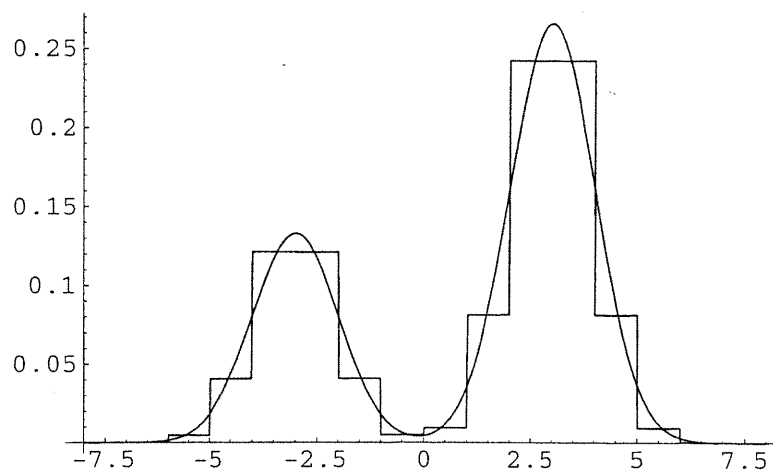
TAB. 4.4. Quelques statistiques sur les trois échantillons simulés.

	Échantillon 1	Échantillon 2	Échantillon 3
Maximum	7,07002	6,76291	6,02234
Minimum	-7,56927	-5,73968	-4,69093
Médiane	1,94649	1,12994	0,71870
Moyenne	0,93065	0,69368	0,49787
Écart type	3,10735	2,32663	2,04442

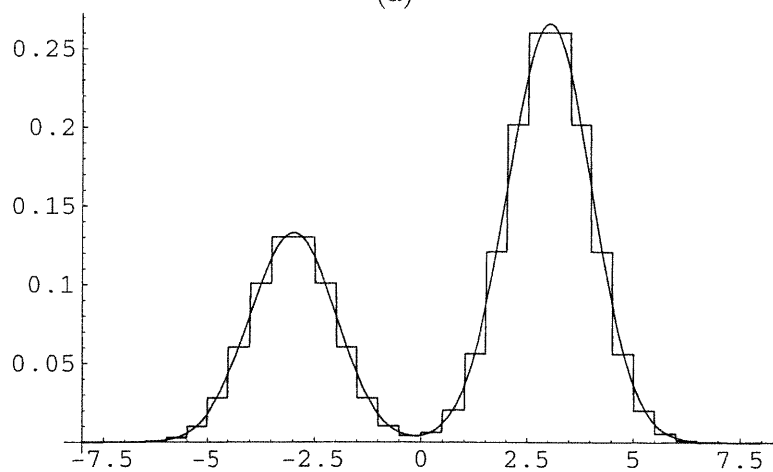
permet de voir que  $\mathbf{B}$  dépend de la vraisemblance associée à chacune des observations, mais pas des moyennes *a priori*  $\mu_j$ . Cette vraisemblance est ici la même pour chacun des modèles étudiés.

Maintenant, revenons à la figure 4.1. Un examen rapide de cette figure permet clairement d'établir que dans l'exemple considéré, le choix de  $J = 1$  est tout à fait satisfaisant. En effet, lorsque  $J = 0$ , la reconstruction obtenue nous semble un peu grossière, alors que le gain de précision réalisé en passant de  $J = 1$  à  $J = 2$  ne semble pas être assez impressionnant pour compenser l'augmentation du temps de calcul nécessaire. Ceci est bien sûr valable parce que  $\pi_1$  est une densité relativement lisse et qu'elle n'est pas trop concentrée. Les figures 4.2 et 4.3 permettent de constater que le choix de  $J = 1$  est tout aussi satisfaisant dans le cas des deux autres densités *a priori* considérées. Elles laissent également voir que dans ces deux cas, le gain de précision réalisé en passant de  $J = 1$  à  $J = 2$  ne compense pas non plus pour l'augmentation du temps de calcul nécessaire. Cette discussion laisse clairement entendre qu'un choix éclairé de  $J$  dépend de la densité à estimer et donc, dans le cas plus réaliste de l'estimation de la densité *a priori* à partir d'un échantillon, de l'échelle des données disponibles. Ainsi, par exemple, dans le cas de données restreintes à l'intervalle  $[0, 1]$ , le passage à  $J = 3$  ou  $J = 4$  peut s'avérer nécessaire, et dans le cas de données extrêmement dispersées, il peut être préférable de descendre vers  $J = -1$ ,  $-2$  ou même  $J = -3$ , ce qui correspond à une estimation par histogramme de la densité *a priori* dont la largeur des bandes est respectivement de 2, 4 et 8.

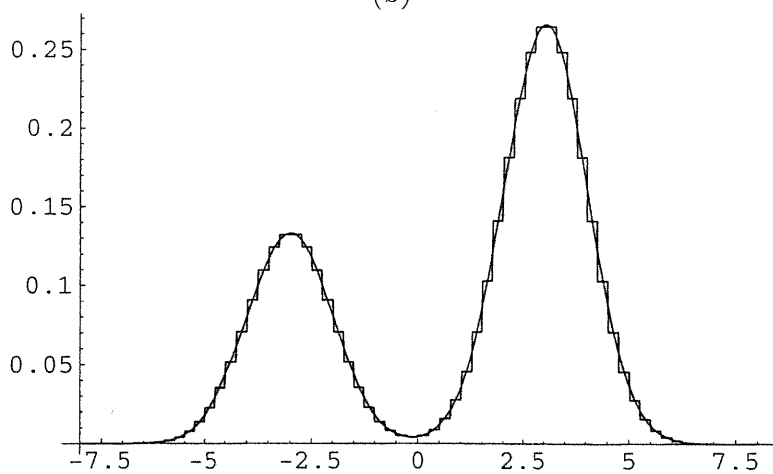
Passons maintenant à la deuxième partie de notre étude : l'estimation des densités  $\pi_1$ ,  $\pi_2$ , et  $\pi_3$  à partir d'échantillons simulés. Notons que nous utilisons ici la convention suivante : pour  $j = 1, 2$  et  $3$ , l'échantillon  $j$  est obtenu à partir de  $m_j$  et sera employé pour reconstruire  $\pi_j$ . Les échantillons obtenus pour cette dernière partie sont de taille  $p = 600$ . Le tableau 4.4 présente quelques statistiques reliées à chacun des trois échantillons simulés. Y sont présentés, entre autres, les éléments



(a)

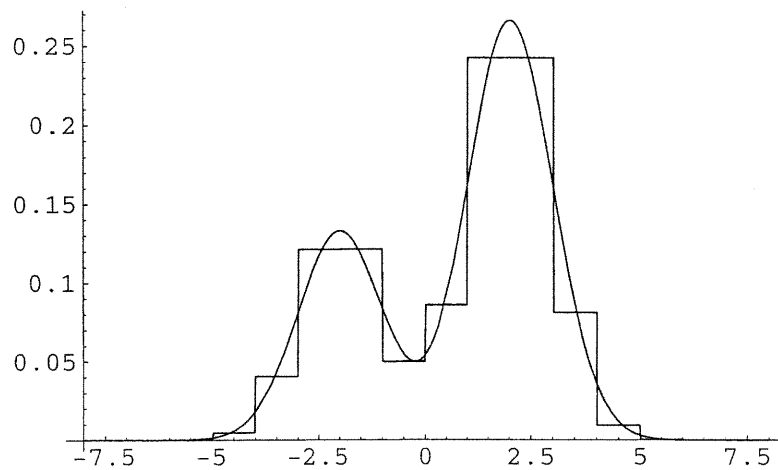


(b)

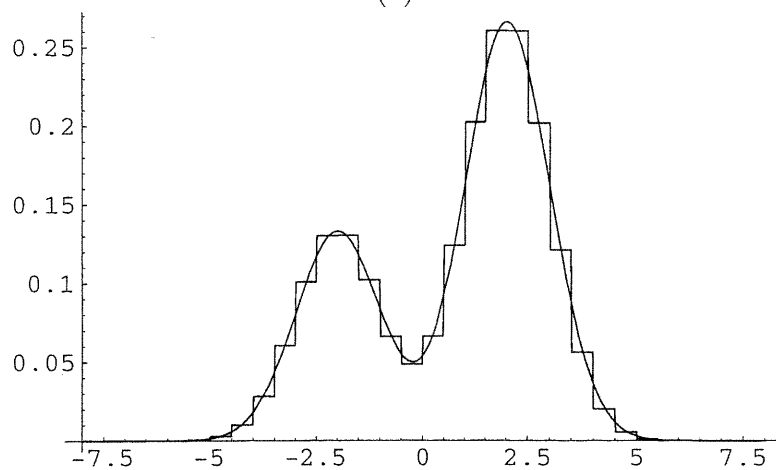


(c)

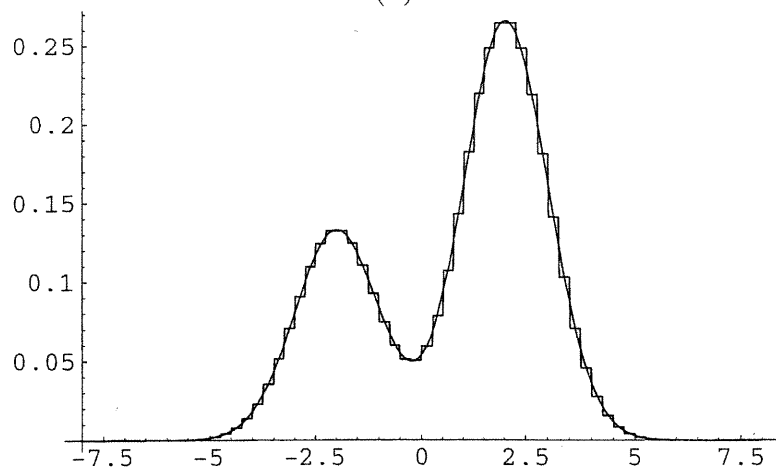
FIG. 4.1. Densité a priori réelle  $\pi_1$  et densité reconstruite à partir de  $m_1$  (a) au niveau  $J = 0$  (b) au niveau  $J = 1$  (c) au niveau  $J = 2$ .



(a)

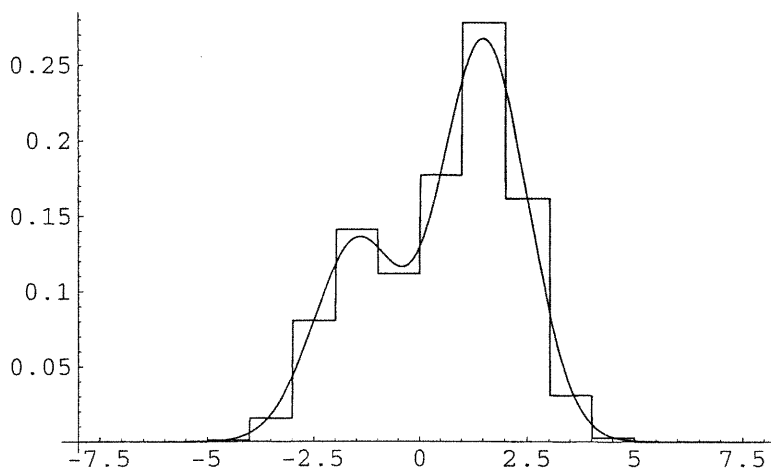


(b)

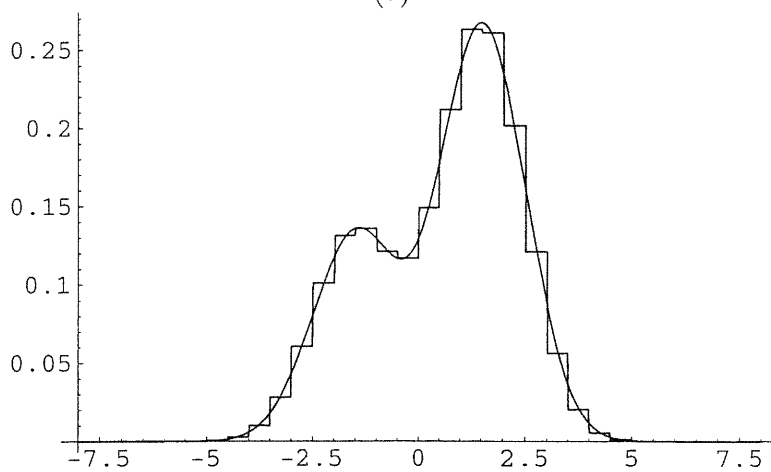


(c)

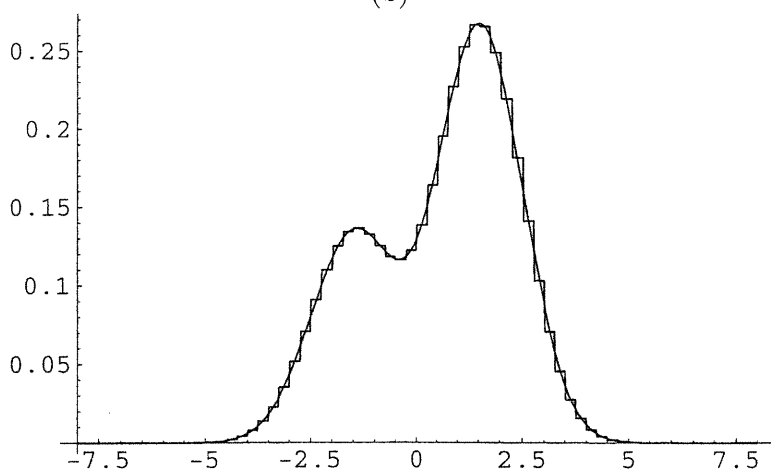
FIG. 4.2. Densité a priori réelle  $\pi_2$  et densité reconstruite à partir de  $m_2$  (a) au niveau  $J = 0$  (b) au niveau  $J = 1$  (c) au niveau  $J = 2$ .



(a)

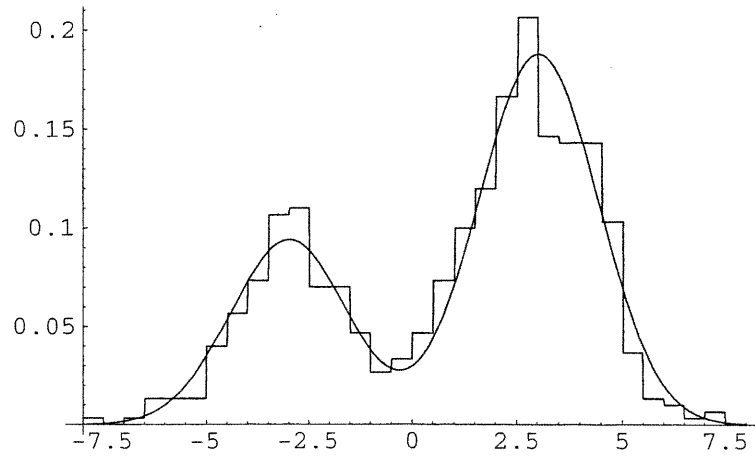


(b)

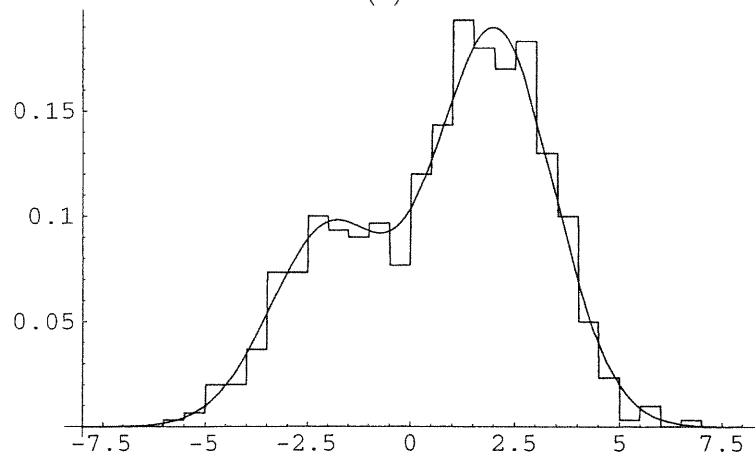


(c)

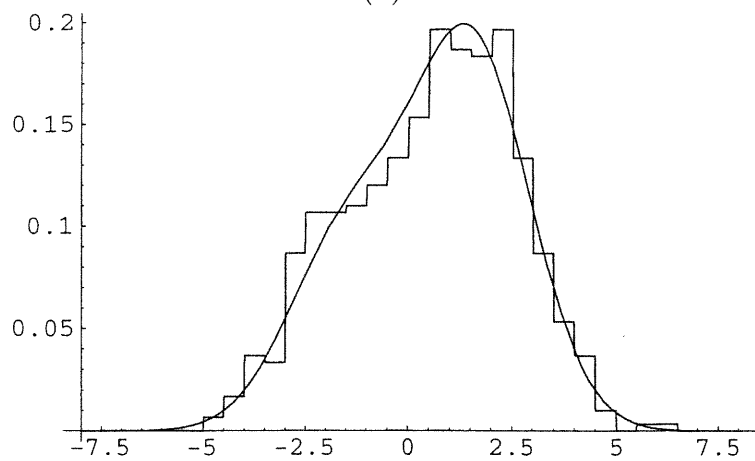
FIG. 4.3. Densité a priori réelle  $\pi_3$  et densité reconstruite à partir de  $m_3$  (a) au niveau  $J=0$  (b) au niveau  $J=1$  (c) au niveau  $J=2$ .



(a)



(b)



(c)

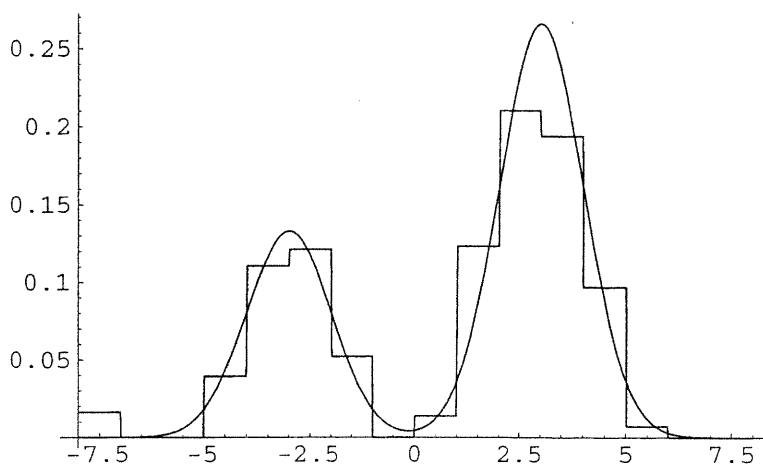
FIG. 4.4. Densité marginale réelle  $m_j$  et son estimation par histogramme  $\hat{m}_{1(j)}$  obtenue à partir du  $j^{\text{ième}}$  échantillon simulé (a)  $m_1$  et  $\hat{m}_{1(1)}$  (b)  $m_2$  et  $\hat{m}_{1(2)}$  (c)  $m_3$  et  $\hat{m}_{1(3)}$ .

maximaux et minimaux de chaque échantillon. Ceci permet au lecteur de constater que le choix de  $C = 8$  est bien ici un choix adéquat pour l'estimation des trois densités *a priori* à partir des observations simulées. Nous présentons également, à la figure 4.4, les histogrammes  $\widehat{m}_{1(j)}$  associés à chacun des échantillons ainsi que les densités marginales  $m_j$ . Ces trois graphiques permettent de bien visualiser la répartition des éléments des trois échantillons. Nous discuterons plus en détail de ces graphiques lors de l'étude de la reconstruction de chacune des densités *a priori*. Finalement, notons que toutes les estimations construites à partir des échantillons simulés correspondent à des formes tronquées et renormalisées des estimations obtenues en utilisant la technique présentée à la section 4.6 et basée sur la décomposition SVD de  $\mathbf{B}$ . Nous reviendrons sur l'emploi de cette technique après avoir présenté graphiquement les reconstructions obtenues avec chacun des échantillons simulés. Cette présentation est bien sûr accompagnée d'une brève discussion.

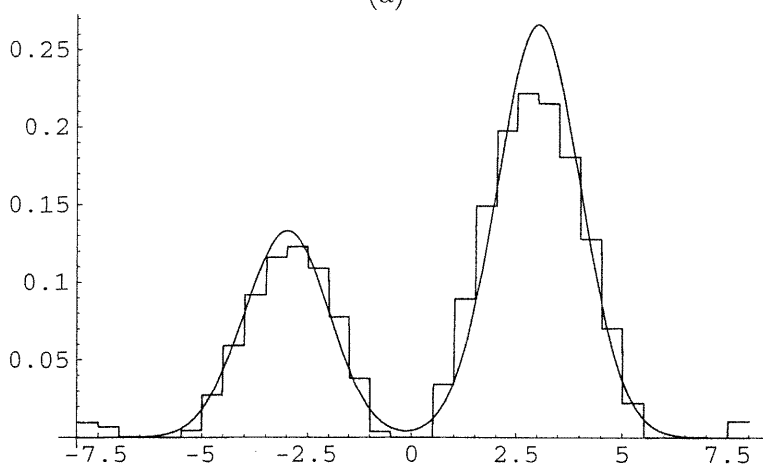
Nous débutons en présentant les estimations de  $\pi_1$  obtenues à partir de l'échantillon 1. La figure 4.5 donne les reconstructions de  $\pi_1$  faites pour les valeurs de  $J = 0, 1, 2$ . Tout d'abord, de ces trois graphiques, il se dégage clairement que l'estimateur par déconvolution saisit ici assez bien les caractéristiques principales de  $\pi_1$ , et ce pour les trois valeurs considérées de  $J$ . Comme précédemment avec les reconstructions théoriques de  $\pi_1$ , l'estimation au niveau  $J = 0$  semble un peu grossière, alors que celles de niveaux  $J = 1$  et  $J = 2$  semblent très satisfaisantes. Encore une fois, le passage de  $J = 1$  à  $J = 2$  ne semble pas ici nécessaire puisque  $\widehat{\pi}_{1(1)}$  et  $\widehat{\pi}_{2(1)}$  exhibent des comportements très similaires. Dans les trois cas, la position des deux modes de  $\pi_1$  est bien identifiée par  $\widehat{\pi}_{J(1)}$ . De plus, les modes sont bien séparés.

Par ailleurs, l'intensité du deuxième mode (situé en  $s = 3$ ) n'a pu être saisie par  $\widehat{\pi}_{J(1)}$  pour aucun des trois niveaux considérés. La raison de cette difficulté n'est pas tout à fait évidente, cependant la figure 4.4(a) nous semble amener une explication au moins partielle à ce phénomène. En effet, en examinant plus attentivement la répartition des observations simulées de l'échantillon 1, il est assez clair qu'une partie importante de la région du deuxième mode de  $m_1$  est sous-représentée. Or, comme l'estimateur par déconvolution  $\widehat{\pi}_J$  est construit en supposant la vraisemblance des observations connue, la sous-représentation d'une région dans l'échantillon utilisé se traduit inévitablement par une sous-représentation d'une région plus ou moins équivalente lors de l'estimation de la densité *a priori*. Ceci est ici vérifié puisque la région sous-représentée lors de l'estimation de  $\pi_1$  correspond à la région sous-représentée par l'échantillon simulé à partir de  $m_1$ .

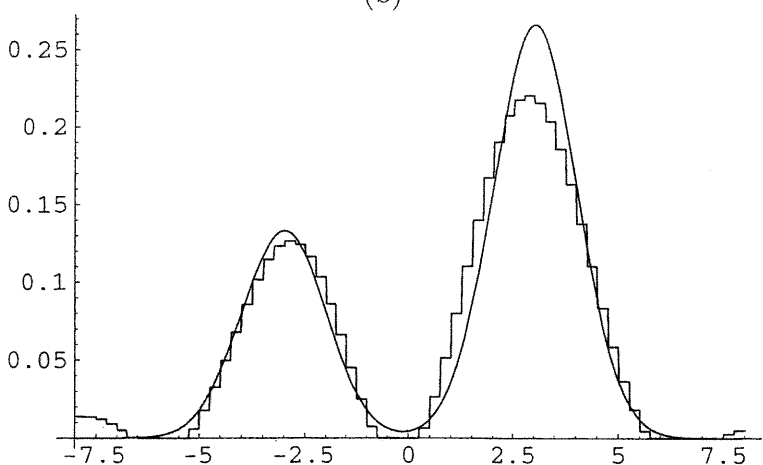




(a)



(b)



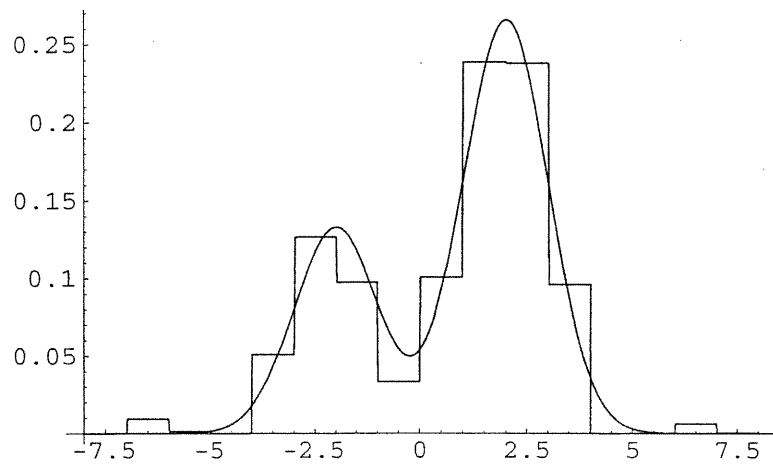
(c)

FIG. 4.5. Densité a priori réelle  $\pi_1$  et densité estimée à partir du premier échantillon simulé  $\hat{\pi}_{J(1)}$  (a) au niveau  $J = 0$  (b) au niveau  $J = 1$  (c) au niveau  $J = 2$ .

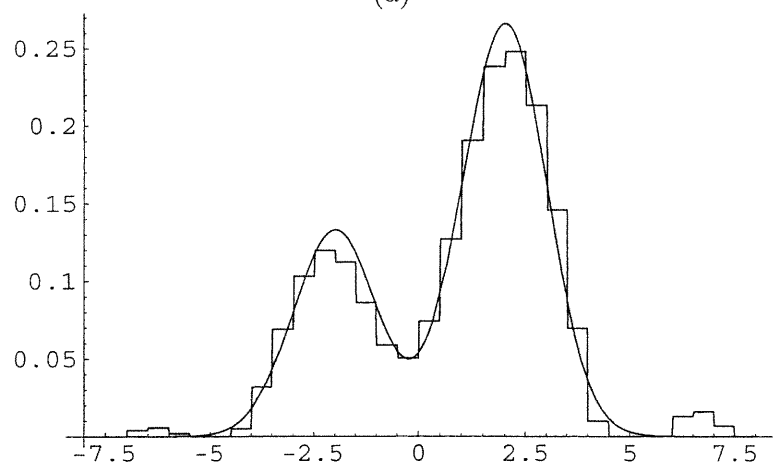
Passons maintenant à l'estimation de  $\pi_2$ . La figure 4.6 représente  $\hat{\pi}_{J(2)}$  pour les niveaux  $J = 0, 1, 2$ . Même si, comme précédemment, l'estimation obtenue au niveau  $J = 0$  semble un peu grossière, l'estimateur par déconvolution  $\hat{\pi}_{J(2)}$  réussit très bien ici à saisir toutes les caractéristiques principales de  $\pi_2$ . En effet, les modes de  $\pi_2$  sont parfaitement localisés par  $\hat{\pi}_{J(2)}$ , et ce pour les trois niveaux considérés. De plus, dans chaque cas, l'intensité des modes et leur séparation sont aussi parfaitement respectées.

Par contre, l'étude du comportement de  $\hat{\pi}_{J(2)}$  met en lumière un phénomène qui n'était pas clairement apparent lors de l'estimation de  $\pi_1$ . En effet, en étudiant plus attentivement les figures 4.6(b) et (c), il est possible de remarquer que les estimations obtenues aux niveaux  $J = 1$  et  $J = 2$  présentent, dans chaque cas, deux légères anomalies autour des valeurs  $s = -6$  et  $s = 6$ . En revenant à la figure 4.4(b), il est clair que la répartition de l'échantillon simulé ne permet pas d'envisager la présence de modes autour de  $x = 6$  et  $x = -6$ . Ceci est particulièrement frappant autour de  $x = -6$ , puisqu'aucune observation n'est présente dans cette région. Les anomalies présentes aux figures 4.6(b) et (c) ne peuvent donc pas être expliquées par les données utilisées. Ceci nous laisse croire que la technique d'estimation développée comporte un défaut inapparent jusqu'ici, celui de souffrir d'un problème apparenté à l'effet de bord présent en estimation non paramétrique à l'aide de la méthode du noyau (voir, par exemple, Hall et Wehrly, 1991). Ici, un léger phénomène d'oscillation semble en fait être présent dans les queues de la densité *a priori* estimée  $\hat{\pi}_{J(2)}$ . Notons que ce phénomène apparaît aussi, quoique de façon moins frappante, lors de l'estimation de  $\pi_1$ . En effet, l'unique anomalie du même type présente aux figures 4.5(a) à (c) pouvait aussi être vue comme le résultat d'observations extrêmes visibles à la figure 4.4(a). À la lumière de la discussion précédente, cette anomalie nous apparaît maintenant comme étant possiblement due à l'effet de bord oscillatoire mentionné plus haut.

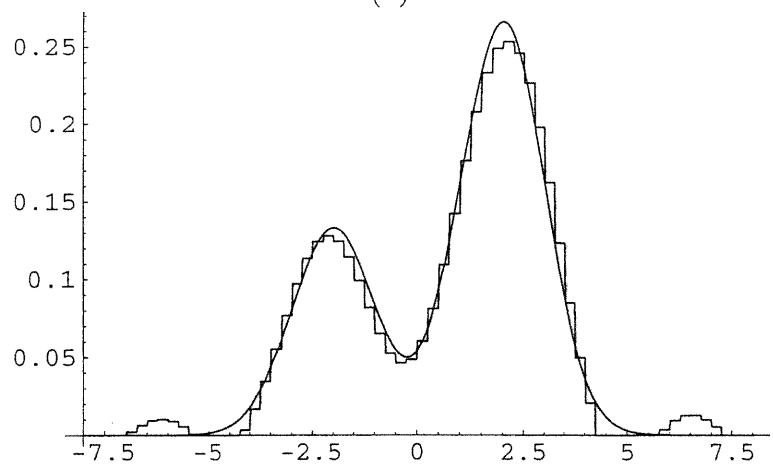
Nous arrivons finalement à l'estimation de  $\pi_3$ . Les diverses reconstructions de  $\pi_3$  obtenues à partir de l'échantillon 3 sont illustrées à la figure 4.7. Notons d'abord que l'estimation obtenue dans le cas de  $J = 0$  est encore une fois un peu grossière. Notons également que la position et l'intensité des modes de  $\pi_3$  ont été correctement caractérisées. Cette fois, c'est plutôt la séparation entre les modes qui a été mal saisie par l'estimateur par déconvolution. En effet, les modes de  $\hat{\pi}_{J(3)}$  sont beaucoup plus clairement séparés que ceux de  $\pi_3$ . Par ailleurs, comme pour le cas de l'estimation de  $\pi_1$ , un rapide examen de l'échantillon simulé utilisé permet d'expliquer le phénomène. En effet, un bref retour à la figure 4.4(c) permet de constater que l'échantillon 3 est distribué de façon sensiblement plus



(a)

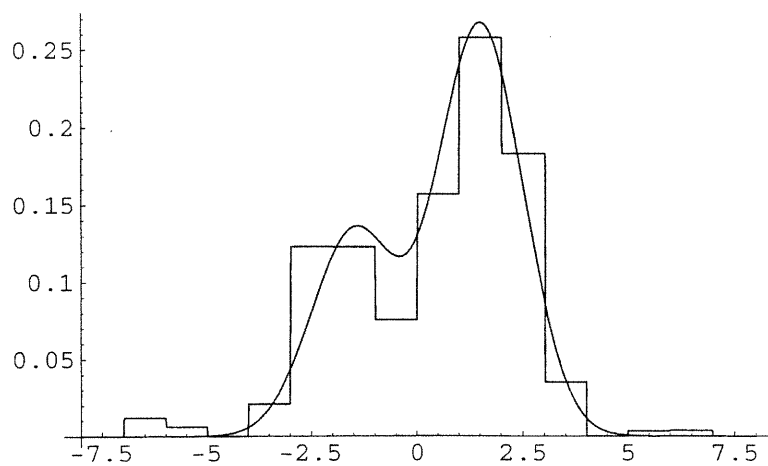


(b)

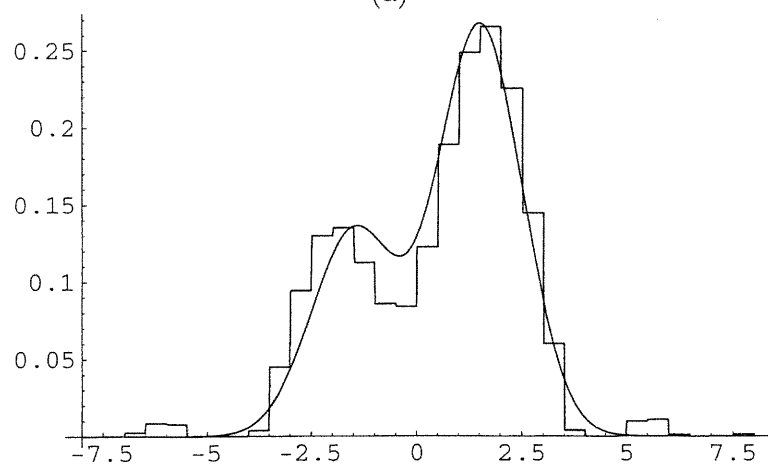


(c)

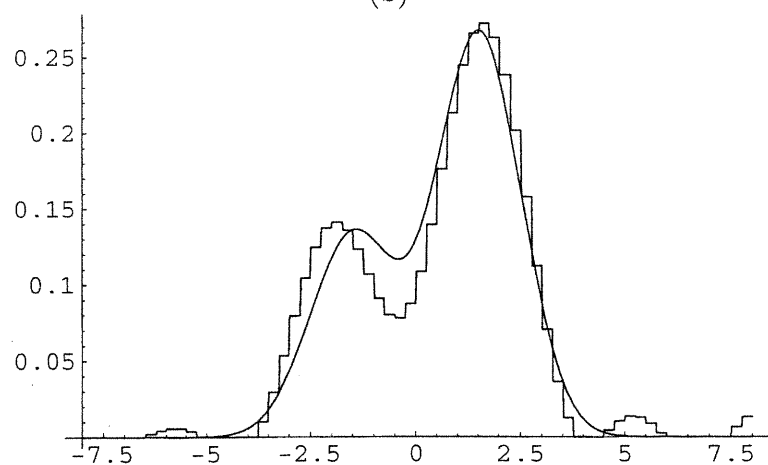
FIG. 4.6. Densité a priori réelle  $\pi_2$  et densité estimée à partir du deuxième échantillon simulé  $\hat{\pi}_{J(2)}$  (a) au niveau  $J = 0$  (b) au niveau  $J = 1$  (c) au niveau  $J = 2$ .



(a)



(b)



(c)

FIG. 4.7. Densité a priori réelle  $\pi_3$  et densité estimée à partir du troisième échantillon simulé  $\hat{\pi}_{J(3)}$  (a) au niveau  $J = 0$  (b) au niveau  $J = 1$  (c) au niveau  $J = 2$ .

asymétrique que la densité marginale  $m_3$ . Il est tout à fait logique de penser que cette asymétrie marquée de l'échantillon simulé augmente l'écart entre les modes de la densité *a priori* estimée. Notons finalement qu'un effet de bord oscillatoire, particulièrement frappant dans le cas où  $J = 2$ , est ici aussi présent.

Il ne reste plus qu'un point dont nous souhaitons discuter ici. Nous avons déjà mentionné que pour obtenir les diverses estimations à partir des échantillons simulés, nous avons été dans l'obligation d'utiliser la technique présentée à la section 4.6 et basée sur la décomposition SVD de  $\mathbf{B}$ . L'emploi de cette technique a été nécessaire puisque pour chaque valeur de  $J$  considérée ici, la matrice  $\mathbf{B}$  est assez sévèrement mal conditionnée. Notons que malgré ce fait, les reconstructions théoriques illustrées aux figures 4.1 à 4.3 ont pu être obtenues sans avoir recours à cette technique. Notons aussi que ceci est typique de la méthode d'estimation : en procédant à ce genre d'étude, il est habituellement possible de reconstruire la densité *a priori* de façon théorique sans utiliser la décomposition SVD, alors que, lors de l'estimation à partir d'observations simulées, il est nécessaire de le faire. C'est le cas ici, ainsi qu'avec les trois autres modèles présentés au tableau 4.1.

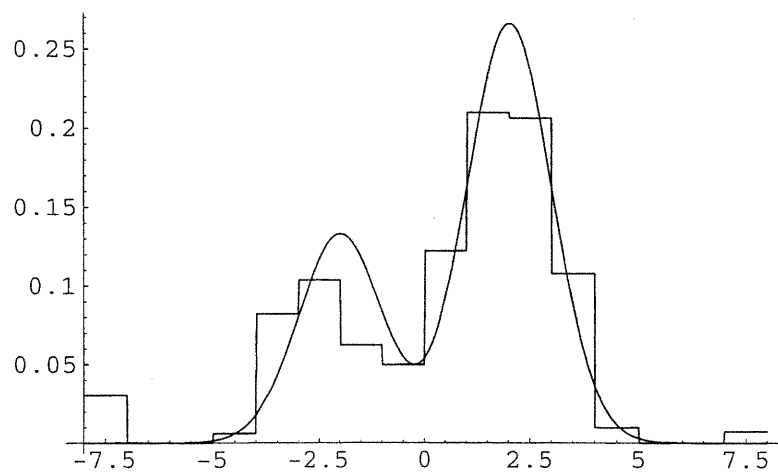
Nous rappelons au lecteur que le principe à employer (voir section 4.6), dans un tel cas, est l'approximation du vecteur des paramètres estimés de la densité *a priori*  $\hat{\mathbf{b}}$  selon

$$\hat{\mathbf{b}} \simeq \hat{\mathbf{b}}_n = 6\mathbf{B}_n^{-1}\hat{\mathbf{m}},$$

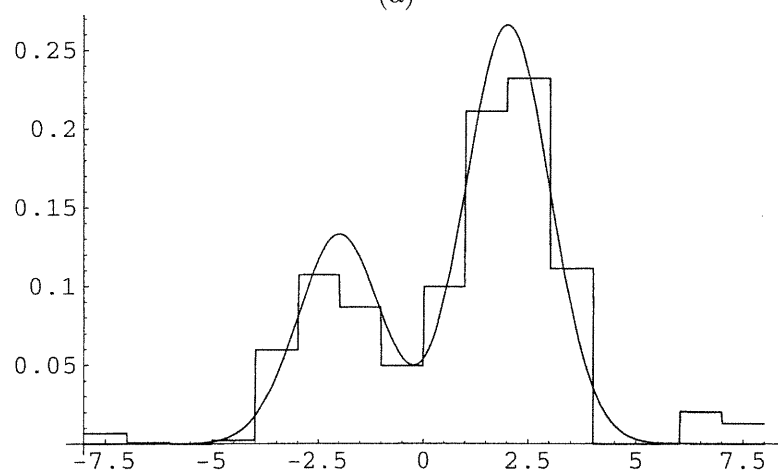
pour une certaine valeur de  $n$  telle que  $1 \leq n \leq 2N + 1$ . De plus,  $n$  devrait être choisi en examinant successivement les estimations obtenues en considérant d'abord les valeurs de  $n$  telles que

$$5 \leq \text{Cond}(\mathbf{B}_n) \leq 10,$$

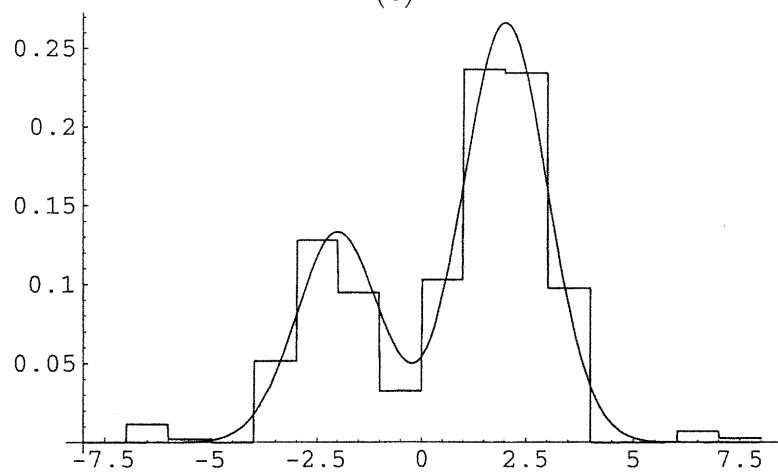
ainsi que, en cas de nécessité, certaines valeurs de  $n$  amenant  $\text{Cond}(\mathbf{B}_n)$  à être assez près de cet intervalle. Nous présentons au tableau 4.5 les indices de conditionnement  $\text{Cond}(\mathbf{B}_n)$  pour les trois niveaux considérés et pour les valeurs de  $n$  telles que  $1 \leq n \leq 17$ . Dans ce tableau, nous avons aussi indiqué les valeurs de  $n$  utilisées pour obtenir les différentes estimations présentées aux figures 4.5 à 4.7. Pour ce faire, nous avons en fait identifié, entre parenthèses et dans la colonne associée au niveau correspondant, la valeur de  $n$  sélectionnée pour chacun des trois modèles considérés. À titre d'exemple, les différentes estimations de  $\pi_2$  ont été obtenues en utilisant  $n = 11$  au niveau  $J = 0$ , et  $n = 9$  aux niveaux  $J = 1$  et  $2$ . En consultant ce tableau, il est assez clair que les bornes données précédemment sur l'indice de conditionnement ne sont pas toujours satisfaites. Cependant il est aussi évident qu'elles sont un bon point de départ pour l'exploration des diverses estimations obtenues en faisant varier la valeur de  $n$ .



(a)

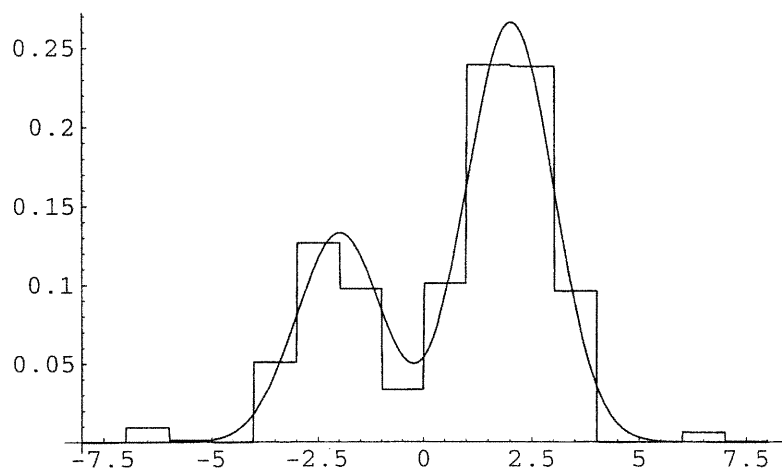


(b)

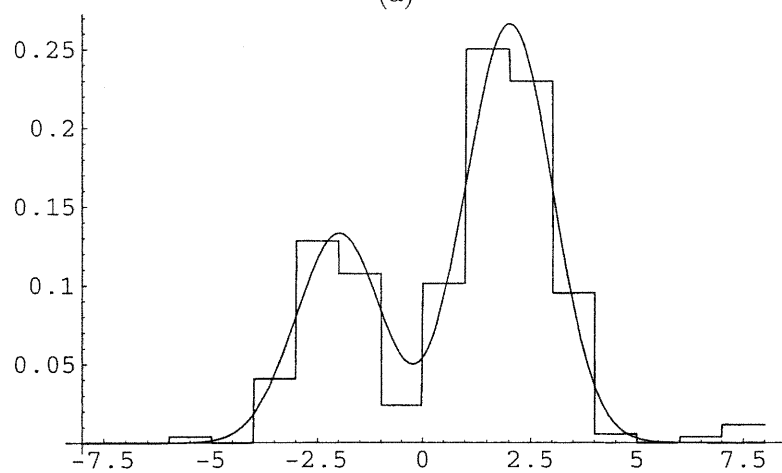


(c)

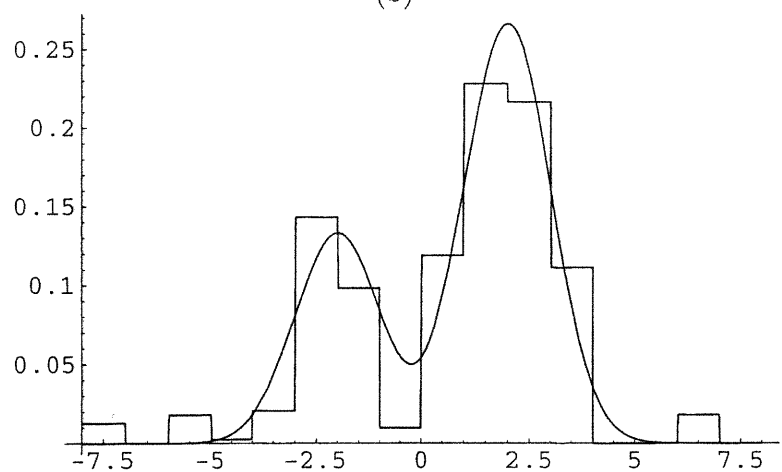
FIG. 4.8. Densité a priori réelle  $\pi_2$  et densité estimée à partir du deuxième échantillon simulé  $\hat{\pi}_{0(2)}$  pour plusieurs valeurs de  $n$  (a)  $n=8$  (b)  $n=9$  (c)  $n=10$ .



(a)



(b)



(c)

FIG. 4.9. Densité a priori réelle  $\pi_2$  et densité estimée à partir du deuxième échantillon simulé  $\hat{\pi}_{0(2)}$  pour plusieurs valeurs de  $n$  (a)  $n = 11$  (b)  $n = 12$  (c)  $n = 13$ .

TAB. 4.5. Indices de conditionnement  $\text{Cond}(\mathbf{B}_n)$  pour  $J = 0, 1, 2$ .

$n$	$\text{Cond}(\mathbf{B}_n) (J = 0)$	$\text{Cond}(\mathbf{B}_n) (J = 1)$	$\text{Cond}(\mathbf{B}_n) (J = 2)$
1	1,000	1,000	1,000
2	1,052	1,050	1,050
3	1,145	1,140	1,140
4	1,290	1,278	1,279
5	1,503	1,480	1,481
6	1,813	1,770	1,772
7	2,263	2,187	2,190
8	2,925	2,790 <sup>(1)</sup>	2,794 <sup>(1)</sup>
9	3,915	3,675 <sup>(2)</sup>	3,678 <sup>(2)</sup>
10	5,432 <sup>(1)</sup>	4,997	4,997
11	7,816 <sup>(2)</sup>	7,014 <sup>(3)</sup>	7,001 <sup>(3)</sup>
12	11,679 <sup>(3)</sup>	10,160	10,113
13	18,143	15,185	15,056
14	29,312	23,417	23,090
15	49,059	37,260	36,463
16	83,213	61,171	59,271
17	131,810	103,638	99,129

Afin d'illustrer comment nous avons ici utilisé la technique avec facilité, nous présentons maintenant, à l'aide des figures 4.8 et 4.9, la réflexion qui nous a mené au choix de  $n = 11$  lors de l'estimation de  $\pi_2$  au niveau  $J = 0$ . Tout d'abord, la figure 4.8(a) montre clairement que l'estimation obtenue n'est pas satisfaisante dans le cas où  $n = 8$ , les modes de  $\pi_2$  étant mal rendus. Ensuite, la figure 4.8(b) permet de constater que le choix de  $n = 9$  n'est pas non plus tout à fait satisfaisant puisque les deux modes de  $\hat{\pi}_{0(2)}$  ne semblent pas encore assez accentués. Finalement, la figure 4.9(c) montre aussi que l'estimation obtenue lorsque  $n = 13$  n'est pas non plus intéressante puisqu'elle semble légèrement trop chaotique. De plus, le premier mode de  $\pi_2$  est particulièrement mal rendu. Maintenant, les figures 4.8(c) et 4.9(a) et (b) montrent que les choix de  $n = 10, 11, 12$  sont en fait assez équivalents, et relativement satisfaisants. Les différences entre les estimations obtenues avec ces trois valeurs de  $n$  sont minimales et la préférence de l'une de ces valeurs plutôt que d'une autre est assez difficile à justifier. Ceci peut être



comparé au choix du paramètre de lissage qui doit être fait par l'utilisateur lors de l'emploi de la méthode du noyau en estimation non paramétrique. Nous avons ici choisi l'estimation obtenue avec  $n = 11$ , présentée par la figure 4.9(a), pour la simple raison que cette valeur de  $n$  nous semble être celle qui amène  $\hat{\pi}_{0(2)}$  à souffrir le moins de l'effet de bord oscillatoire dont nous avons parlé précédemment. Bien sûr, nous sommes conscients que ce choix est subjectif. Encore une fois, les trois valeurs identifiées plus haut nous semblent adéquates et amènent des estimations de  $\pi_2$  tout à fait satisfaisantes.

#### 4.7.2. Discussion

Faisons finalement un rappel des principales caractéristiques de l'estimateur non paramétrique de la densité *a priori*  $\hat{\pi}_J$  identifiées à l'aide de l'exemple précédent. Nous discutons par le fait même des modifications possibles à apporter afin de parfaire la technique d'estimation.

- La méthode proposée est facile à utiliser puisque toutes les manipulations nécessaires au calcul de  $\hat{\pi}_J$  ou d'une forme corrigée de  $\hat{\pi}_J$  sont faciles à effectuer. La grande simplicité de la méthode est principalement liée au fait que la densité *a priori* est estimée en résolvant un système d'équations linéaires. Un court programme utilisant Mathematica permet donc habituellement d'obtenir une estimation satisfaisante de la densité *a priori* de  $\theta$ .
- L'estimateur  $\hat{\pi}_J$  amène à une estimation par histogramme de la densité *a priori*  $\pi$ . À ce titre,  $\hat{\pi}_J$  est facile à manipuler et à interpréter, mais n'amène jamais une estimation continue de  $\pi$ . Une correction naturelle à apporter consiste à utiliser, dans le développement de l'approximation de la densité marginale des observations  $X_1, X_2, \dots, X_p$ , une base d'ondelettes continues comme les ondelettes de Daubechies d'ordre supérieur à 1, à la place de la base de Haar.
- L'estimation de  $\pi$  obtenue à partir de  $\hat{\pi}_J$  n'est habituellement pas *bona fide*. Il faut alors employer la forme tronquée et renormalisée de l'estimateur par déconvolution, notée  $\hat{\pi}_J^+$ . Même si ceci est très simple en pratique, les conséquences théoriques de cette modification ne sont pas claires.
- L'utilisation de la décomposition SVD pour le calcul d'une forme corrigée d'estimation de la densité *a priori* a été, à l'exemple de la section précédente, très efficace pour contourner le problème de mauvais conditionnement de la matrice  $\mathbf{B}$ . Par contre, cette approche n'est pas toujours utile et il peut arriver, dans les cas extrêmes, qu'il soit impossible d'obtenir une estimation

satisfaisante de  $\pi$ , et ce malgré l'utilisation de la décomposition SVD selon la méthode proposée à la section 4.6.

- L'effet de bord oscillatoire mis en lumière précédemment est, par moment, assez troublant puisqu'il amène parfois la formation de modes secondaires dans les queues de la densité estimée correspondant à des régions où aucune observation n'est retrouvée. De plus, cet effet de bord semble assez irrégulier puisqu'il est présent à divers degrés d'intensité dans les densités reconstruites à partir des échantillons simulés.

Il est certain que l'attrait principal de la technique proposée ici est son étonnante simplicité. Les techniques déjà existantes (voir section 4.1) ont ceci en commun qu'elles sont toutes beaucoup plus complexes et extrêmement exigeantes numériquement. De ce fait, la méthodologie développée ici présente un intérêt certain. Finalement, même si elle demande certainement quelques ajustements, la méthode nous semble très prometteuse, vu la popularité croissante des méthodes bayésiennes empiriques.

## Conclusion

---

Notre but premier était ici d'obtenir, en utilisant les bases d'ondelettes, une technique de calcul bayésien alternative aux méthodes classiques telles que l'approximation gaussienne de Naylor et Smith (1982), l'approximation laplacienne de Tierney et Kadane (1986), l'échantillonneur de Gibbs et les simulations de Monte-Carlo avec fonction d'importance ou avec chaînes de Markov. Par le fait même, nous souhaitons élargir l'utilisation des ondelettes à d'autres champs de la statistique.

Pour cela, nous nous sommes intéressés d'abord au cas des modèles bayésiens où le paramètre d'intérêt est le paramètre de position, noté  $\theta$ , de la vraisemblance associée à une seule observation  $X$ . Dans ce contexte bien particulier, nous avons utilisé la base d'ondelettes de Haar afin de construire des approximations de la densité marginale de  $X$ , de l'estimateur bayésien de  $\theta$  en considérant la fonction de perte quadratique, et de la perte espérée *a posteriori* de cet estimateur. Les résultats obtenus ont par la suite été généralisés au cas beaucoup plus intéressant d'un modèle basé sur plusieurs observations  $X_1, X_2, \dots, X_p$ . Ce cas est en effet plus intéressant puisqu'il amène souvent à l'emploi de l'une des techniques classiques mentionnées plus haut, tout particulièrement lorsqu'il n'existe pas de statistique exhaustive pour  $\theta$ .

Voyons maintenant les avantages et les inconvénients reliés à l'utilisation des approximations développées aux chapitres 2 et 3 dans le contexte de l'estimation bayésienne d'un paramètre de position. Tout d'abord, les approximations de la densité marginale de  $X$  (dans le cas du modèle à une seule observation) et de l'estimateur de Bayes et de sa perte espérée *a posteriori* (dans le cas des modèles à une ou plusieurs observations) ont été conçues pour les avantages techniques suivants.

- Les méthodes classiques donnent habituellement des approximations point à point des fonctions à approximer. Ainsi les approximations résultantes doivent être calculées une fois les valeurs  $X_1, X_2, \dots, X_p$  observées. La méthode d'approximation développée ici permet d'approximer les quantités d'intérêt par des formes fonctionnelles pouvant facilement être construites

avant la prise des données. De plus, ces formes fonctionnelles permettent, lorsqu'elles sont construites avec soin, l'étude de certaines caractéristiques théoriques liées au modèle considéré, telle la robustesse de l'estimateur bayésien.

- Les approximations gaussienne et laplacienne demandent une étude théorique plus approfondie du modèle particulier considéré, alors que l'application de la méthode d'approximation que nous avons proposée ne requiert que la connaissance des fonctions de répartition associées aux densités employées dans le modèle et des valeurs fixées des hyperparamètres. Pour le reste, la méthode d'approximation est automatique.
- Les méthodes itératives (comme l'échantillonneur de Gibbs et les simulations de Monte-Carlo), quoique généralement simples d'application, peuvent devenir assez difficiles à employer. En effet, dans certains cas, la densité *a posteriori* amenée par le modèle bayésien ne peut être écrite de façon explicite. Il est alors bien difficile (du point de vue numérique) de l'employer pour générer des variables aléatoires. Il arrive aussi que ces techniques aient des problèmes de convergence. Ce n'est pas le cas avec notre méthode puisqu'elle n'est pas itérative.
- La méthode proposée est très simple à utiliser. En effet, toutes les formules nécessaires au calcul des approximations obtenues sont relativement simples et faciles à manipuler.

De plus, grâce aux deux exemples étudiés à la section 2.7, nous avons pu établir, pour les approximations construites à l'aide de la base de Haar, les avantages pratiques suivants.

- La technique de calcul semble extrêmement efficace au niveau de la précision des approximations construites. En fait, l'efficacité de la méthode est telle qu'un niveau de  $J = 1, 2$  ou, au plus,  $3$  est habituellement suffisant, même si les différents résultats théoriques obtenus nous assurent de la qualité des approximations uniquement lorsque le niveau  $J$  est choisi assez grand.
- Il est parfois possible d'augmenter de manière considérable la précision des approximations utilisées en augmentant la valeur de  $C$ , plutôt que celle de  $J$ , où  $N = C2^J$  est le paramètre contrôlant le nombre de termes des sommes tronquées  $S_J^T(x)$ . Ainsi, une nette amélioration du comportement des approximations considérées peut être obtenue avec une augmentation linéaire du nombre de termes de ces sommes.
- Si les deux densités composant le modèle bayésien satisfont toutes les conditions de régularité données aux théorèmes 2.1 et 2.10, il est possible d'avoir des approximations adéquates en retenant relativement peu de termes pour

les sommes tronquées  $S_i^T(x)$ . En effet, il suffit pour cela d'obtenir les approximations de façon à calculer les coefficients  $b_{0,jk}$  à partir de la densité la plus concentrée.

La technique d'approximation comporte aussi certains désavantages. Les principaux, aussi mis en lumière par les exemples de la section 2.7, sont les suivants.

- Le choix de  $J$  est assez simple, mais la détermination d'une valeur adéquate de  $C$  peut devenir plus complexe. En effet, les critères de sélection de  $C$  identifiés à la section 2.6 et donnés par les équations (2.55) et (2.56) mènent souvent à une valeur très grande de  $C$ , parfois inutilement élevée.
- L'approximation de la perte espérée *a posteriori* de l'estimateur bayésien semble être plus difficile que celle de la densité marginale ou de l'estimateur de Bayes lui-même. Il nous faut donc considérer des plus grands niveaux (typiquement,  $J = 2$  ou  $3$ ) et des plus grandes valeurs de  $C$  lors de l'approximation de la perte espérée *a posteriori* de  $\widehat{\theta}_B(x)$ . Ceci peut augmenter considérablement la difficulté numérique du problème et se traduire, en pratique, par une forte croissance du temps de calcul requis.

Comme nous l'avons mentionné au début de cette thèse, nous pensons que les ondelettes offrent une multitude de possibilités dans le cadre du calcul bayésien. La technique d'approximation présentée aux chapitres 2 et 3 en est un exemple. Cependant, comme nous l'avons aussi mentionné, nous souhaitons ici explorer plusieurs avenues dans lesquelles les ondelettes peuvent s'engager afin d'offrir de nouvelles solutions à des problèmes bayésiens. Au chapitre 4, nous avons ainsi proposé une méthode d'estimation non paramétrique de la densité *a priori* d'un paramètre de position. Cette méthode repose essentiellement sur l'approximation de la densité marginale d'une observation  $X$  développée au chapitre 2. Elle permet de constater que les ondelettes s'appliquent effectivement à plusieurs types de problèmes bayésiens, mais surtout que la technique d'approximation proposée aux chapitres 2 et 3 a elle-même un potentiel intéressant d'application à d'autres problèmes.

L'estimateur par déconvolution de la densité *a priori* d'un paramètre de position, noté  $\hat{\pi}_J$ , est donné par la définition 4.1. Nous avons démontré, à la section 4.5, qu'en rajoutant une nouvelle hypothèse de régularité à celles faites précédemment, cet estimateur satisfait les critères de convergence habituels. Revoiyons maintenant les caractéristiques pratiques importantes de l'estimateur  $\hat{\pi}_J$  identifiées à la section 4.6 et lors de l'étude de l'exemple simulé présenté à la

section 4.7.1. Commençons d'abord par discuter des avantages principaux reliés à l'utilisation de l'estimateur de la densité *a priori* de  $\theta$ .

- Encore une fois, la méthode développée n'est pas une méthode itérative. Ainsi, aucune simulation longue et difficile n'est nécessaire.
- Comme précédemment, la méthode proposée est très simple à utiliser. En effet, toutes les manipulations nécessaires à l'obtention de  $\hat{\pi}_J$ , ou à une forme corrigée de  $\hat{\pi}_J$  (voir section 4.6), sont relativement simples.
- L'estimateur  $\hat{\pi}_J$  correspond à une estimation par histogramme de la densité *a priori* de  $\theta$ . Ainsi,  $\hat{\pi}_J$  est un objet mathématique en général facile à interpréter et à manipuler. Plusieurs applications de la technique d'estimation peuvent être envisagées en utilisant comme base  $\hat{\pi}_J$ . En effet, cet estimateur nous semble être tout indiqué pour des problèmes plus généraux tels que l'estimation bayésienne empirique (“Bayes empirical Bayes estimation”) et la construction d'intervalles de confiance bayésiens empiriques ou de cotes de Bayes empiriques.

Les principaux désavantages liés à l'utilisation de l'estimateur  $\hat{\pi}_J$  sont les suivants.

- L'estimateur  $\hat{\pi}_J$  n'amène habituellement pas une estimation *bona fide* de la densité *a priori*. Il faut alors utiliser la forme tronquée et renormalisée de  $\hat{\pi}_J$  définie à la section 4.6. Izenman (1991) discute plus en détail de cette difficulté rencontrée avec plusieurs estimateurs non paramétriques de densités.
- Comme nous l'avons dit plus haut, la méthodologie présentée amène à une estimation par histogramme de la densité *a priori*. Bien que ceci puisse être vu à plusieurs égards comme un avantage, il reste qu'à cause de ce fait, la densité estimée n'est jamais continue.
- Les difficultés liées au mauvais conditionnement de la matrice  $\mathbf{B}$  sont parfois très importantes. Ainsi, même en utilisant la décomposition SVD comme nous l'avons proposé à la section 4.6, il arrive qu'une reconstruction satisfaisante de la densité *a priori* ne puisse être obtenue.
- L'exemple simulé de la section 4.7.1 a mis en lumière un phénomène d'oscillation qui semble être présent à divers degrés d'intensité dans les densités estimées à l'aide de la technique proposée. L'effet de bord oscillatoire ainsi identifié est par moment particulièrement troublant puisqu'il peut amener l'apparition de modes secondaires dans les queues de la densité estimée associés à des régions où aucune observation n'est présente.

Finalement, le but premier de cette thèse a été atteint. En effet, notre but était d'explorer diverses possibilités d'application des ondelettes dans le domaine du

calcul bayésien. Pour cela, nous avons ici proposé une technique d'approximation numérique des moments *a posteriori* d'un paramètre de position  $\theta$ . À partir de cette méthode, nous avons montré comment construire des approximations de l'estimateur de Bayes de  $\theta$  et de sa perte espérée *a posteriori* et, dans le cas d'un modèle composé d'une seule observation, de la densité marginale de celle-ci. La méthode semble à la fois simple et précise, et elle nous paraît facilement applicable à d'autres problèmes bayésiens reliés aux moments *a posteriori* de  $\theta$  ou à la densité marginale des observations. Ceci sera certainement le sujet d'investigations futures.

Nous avons aussi présenté, dans un contexte bayésien empirique, une méthode d'estimation de la densité *a priori* de  $\theta$ . Cette méthode nous semble très prometteuse et recevra beaucoup d'attention de notre part dans l'avenir. Comme nous l'avons mentionné plus haut, elle nous semble particulièrement indiquée pour certains problèmes bayésiens empiriques. Ceci devrait nous amener à étudier plus à fond certains aspects théoriques reliés à la méthode proposée. Par exemple, l'utilisation de la forme tronquée et renormalisée de l'estimateur de la densité *a priori* est très simple en pratique, mais les conséquences théoriques de cette modification de  $\hat{\pi}_J$  ne sont pas claires. Les conséquences théoriques de l'utilisation de la décomposition SVD pour le calcul de  $\hat{\pi}_J$ , lorsque la matrice  $\mathbf{B}$  est mal conditionnée, ne sont pas non plus évidentes. Aussi, une correction naturelle à apporter à la discontinuité des densités estimées pourrait être de substituer à la base de Haar, dans le développement de l'approximation de la densité marginale des observations  $X_1, X_2, \dots, X_p$ , une base d'ondelettes continues comme les ondelettes de Daubechies d'ordre supérieur à un. Finalement, d'un point de vue pratique, il serait certainement intéressant d'amener un correctif à la méthode employée afin de minimiser l'effet de bord oscillatoire dont nous avons parlé plus haut. Tous ces détails feront l'objet de recherches futures.

## Annexe A

---

Cette annexe a pour but ultime de prouver deux propriétés extrêmement utiles découlant de l'analyse multirésolution et employées au chapitre 1. D'abord, si  $\phi$  est une fonction d'échelle, alors

$$\int_{\mathbb{R}} \phi(x) dx = \pm 1. \quad (\text{A.1})$$

Puis, si  $\psi$  est la mère des ondelettes qui lui est associée, alors

$$\int_{\mathbb{R}} \psi(x) dx = 0. \quad (\text{A.2})$$

Plusieurs approches permettent d'arriver à ces résultats, mais celle que nous présentons ici est très similaire à celle préconisée par Mallat (1989a) et Härdle *et al.* (1998).

La technique employée par les auteurs mentionnés plus haut utilise plusieurs arguments provenant de la théorie de Fourier. Avant de débiter l'argumentation qui nous mènera aux résultats souhaités, nous commençons donc par un bref rappel des points importants de cette théorie. Pour en savoir plus sur le sujet, le lecteur peut consulter Bracewell (1978), Rudin (1987, chapitre 9), Walker (1988) et Walnut (2002, chapitres 2 et 3).

D'abord, nous définissons la *transformée de Fourier* et donnons quelques unes de ses propriétés.

**Définition A.1.** La transformée de Fourier d'une fonction  $f \in \mathcal{L}^1(\mathbb{R})$ , notée  $\mathcal{F}(f)$ , ou plus simplement  $\hat{f}$ , est donnée par

$$\mathcal{F}(f)(s) = \hat{f}(s) = \int_{\mathbb{R}} f(x) e^{-ixs} dx \quad \forall s \in \mathbb{R}.$$

**Proposition A.1.** La transformée de Fourier  $\hat{f}$  de  $f \in \mathcal{L}^1(\mathbb{R})$  est continue et respecte

$$\hat{f}(0) = \int_{\mathbb{R}} f(x) dx,$$



et

$$\lim_{s \rightarrow \pm\infty} \hat{f}(s) = 0.$$

De plus, dans l'éventualité où  $f \in \mathcal{L}^2(\mathbb{R})$ , nous avons alors aussi que  $\hat{f} \in \mathcal{L}^2(\mathbb{R})$ .

Dans le cas où  $f \in \mathcal{L}^1(\mathbb{R})$  est continue, il est parfois possible de la retrouver entièrement à partir de sa transformée de Fourier. Cette propriété d'inversion est appelée *transformée de Fourier inverse* et est donnée par la proposition suivante.

**Proposition A.2.** Si  $f \in \mathcal{L}^1(\mathbb{R})$  est continue, et si  $\hat{f} \in \mathcal{L}^1(\mathbb{R})$ , alors

$$\frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(s) e^{isx} ds = f(x) \quad \forall x \in \mathbb{R}.$$

Nous notons la transformée de Fourier inverse

$$\mathcal{F}^{-1}(\hat{f})(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(s) e^{isx} ds,$$

et pouvons ainsi écrire simplement

$$\mathcal{F}^{-1}(\hat{f})(x) = f(x) \quad \forall x \in \mathbb{R}.$$

Si nous avons plutôt que  $f$  est continue et  $2\pi$ -périodique (c'est-à-dire périodique et de période  $2\pi$ ), la théorie de Fourier permet d'obtenir un développement en série de  $f$  à l'aide de coefficients très semblables à une transformée de Fourier. La proposition suivante donne la façon exacte de procéder.

**Proposition A.3.** Si  $f \in \mathcal{L}^1(0, 2\pi)$  est continue et  $2\pi$ -périodique, alors elle peut être représentée par

$$f(x) = \sum_{k \in \mathbb{Z}} c_k e^{ikx} \quad \forall x \in \mathbb{R},$$

où les coefficients  $c_k$  sont donnés par

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-ikx} dx.$$

Notons que la série obtenue à partir d'une fonction  $f \in \mathcal{L}^1(0, 2\pi)$  à l'aide du résultat précédent est appelée *série de Fourier* de la fonction  $f$ .

Un dernier concept important de la théorie de Fourier, est celui de *convolution*. En effet, ce concept nous amène au résultat qui est possiblement le plus remarquable de cette branche des mathématiques appliquées, soit ce qu'il est convenu d'appeler le *théorème de convolution*.

**Définition A.2.** La convolution entre deux fonctions  $f$  et  $g$  de  $\mathcal{L}^1(\mathbb{R})$ , notée  $f * g$ , est donnée par

$$f * g(x) = \int_{\mathbb{R}} f(s)g(x-s) ds \quad \forall x \in \mathbb{R}.$$

Notons qu'il est possible de montrer que  $f * g \in \mathcal{L}^1(\mathbb{R})$  et que  $f * g$  est continue (voir Walnut, 2002, section 3.4).

**Théorème A.1.** Si  $f \in \mathcal{L}^1(\mathbb{R})$  et  $g \in \mathcal{L}^1(\mathbb{R})$ , alors

$$\mathcal{F}(f * g)(s) = \hat{f}(s)\hat{g}(s) \quad \forall s \in \mathbb{R}.$$

De plus, si  $f \in \mathcal{L}^2(\mathbb{R})$  et  $g \in \mathcal{L}^2(\mathbb{R})$ , alors

$$\mathcal{F}^{-1}(\hat{f}\hat{g})(x) = f * g(x) \quad \forall x \in \mathbb{R}.$$

Il nous est maintenant possible de passer au coeur du sujet. Les notions précédentes peuvent être employées afin d'arriver aux équations (A.1) et (A.2). Nous obtenons d'abord deux lemmes traitant d'une fonction notée  $m_0$ . Cette fonction permet d'obtenir la construction menant aux résultats souhaités. Le premier de ces lemmes définit la fonction  $m_0$  et donne le lien l'unissant à  $\phi$ . Le second donne le lien unissant  $m_0$  à  $\psi$ .

**Lemme A.1.** Si  $\phi \in \mathcal{L}^1(\mathbb{R}) \cap \mathcal{L}^2(\mathbb{R})$  est une fonction d'échelle, alors il existe une fonction  $m_0$  continue et  $2\pi$ -périodique telle que

$$\hat{\phi}(s) = m_0\left(\frac{s}{2}\right)\hat{\phi}\left(\frac{s}{2}\right) \quad \forall s \in \mathbb{R}.$$

**Démonstration.** D'abord, grâce à l'équation (1.6), nous avons

$$\phi(x) = \sum_{k \in \mathbb{Z}} a_k \phi(2x - k) = \frac{1}{\sqrt{2}} \sum_{k \in \mathbb{Z}} a_k \phi_{1k}(x).$$

En appliquant la transformée de Fourier de chaque côté de cette égalité, nous obtenons

$$\hat{\phi}(s) = \frac{1}{\sqrt{2}} \sum_{k \in \mathbb{Z}} a_k \hat{\phi}_{1k}(s).$$

Par ailleurs,

$$\begin{aligned}
 \hat{\phi}_{1k}(s) &= \sqrt{2} \int_{\mathbb{R}} \phi(2x - k) e^{-ixs} dx \\
 &= \frac{1}{\sqrt{2}} \int_{\mathbb{R}} \phi(x) e^{-i(x+k)\frac{s}{2}} dx \\
 &= \frac{1}{\sqrt{2}} e^{-ik\frac{s}{2}} \int_{\mathbb{R}} \phi(x) e^{-ix\frac{s}{2}} dx \\
 &= \frac{1}{\sqrt{2}} e^{-ik\frac{s}{2}} \hat{\phi}\left(\frac{s}{2}\right).
 \end{aligned} \tag{A.3}$$

Ainsi,

$$\begin{aligned}
 \hat{\phi}(s) &= \frac{1}{2} \hat{\phi}\left(\frac{s}{2}\right) \sum_{k \in \mathbb{Z}} a_k e^{-ik\frac{s}{2}} \\
 &= m_0\left(\frac{s}{2}\right) \hat{\phi}\left(\frac{s}{2}\right).
 \end{aligned}$$

où la fonction  $m_0$  est évidemment donnée par

$$m_0(x) = \frac{1}{2} \sum_{k \in \mathbb{Z}} a_k e^{-ikx}.$$

Cette fonction est clairement continue et  $2\pi$ -périodique.  $\square$

**Lemme A.2.** *Si  $\phi \in \mathcal{L}^1(\mathbb{R}) \cap \mathcal{L}^2(\mathbb{R})$  est une fonction d'échelle, et si  $\psi$  est la mère des ondelettes qui lui est associée, alors*

$$\hat{\psi}(s) = \overline{m_0\left(\pi + \frac{s}{2}\right)} e^{-i\left(\pi + \frac{s}{2}\right)} \hat{\phi}\left(\frac{s}{2}\right) \quad \forall s \in \mathbb{R},$$

où  $\overline{m_0(x)}$  représente le conjugué complexe de  $m_0(x)$ .

**Démonstration.** D'après l'équation (1.11),

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

En procédant comme au lemme A.1, l'équation (A.3) amène

$$\begin{aligned}
 \hat{\psi}(s) &= \frac{1}{\sqrt{2}} \sum_{k \in \mathbb{Z}} (-1)^k a_{1-k} \hat{\phi}_{1k}(s) \\
 &= \frac{1}{2} \hat{\phi}\left(\frac{s}{2}\right) \sum_{k \in \mathbb{Z}} (-1)^k a_{1-k} e^{-ik\frac{s}{2}}.
 \end{aligned}$$

Maintenant, par une simple substitution, nous pouvons écrire

$$\begin{aligned}\hat{\psi}(s) &= \frac{1}{2} \hat{\phi}\left(\frac{s}{2}\right) \sum_{k \in \mathbb{Z}} (-1)^{(1-k)} a_k e^{-i(1-k)\frac{s}{2}} \\ &= \frac{1}{2} \hat{\phi}\left(\frac{s}{2}\right) \sum_{k \in \mathbb{Z}} a_k e^{-i(1-k)(\pi + \frac{s}{2})} \\ &= \frac{1}{2} e^{-i(\pi + \frac{s}{2})} \hat{\phi}\left(\frac{s}{2}\right) \sum_{k \in \mathbb{Z}} a_k e^{ik(\pi + \frac{s}{2})},\end{aligned}$$

puisque  $e^{-i\pi} = -1$ . Finalement, en remarquant que

$$\frac{1}{2} \sum_{k \in \mathbb{Z}} a_k e^{ik(\pi + \frac{s}{2})} = \overline{m_0\left(\pi + \frac{s}{2}\right)},$$

nous obtenons bien le résultat voulu.  $\square$

Le prochain lemme donne une propriété des fonctions d'échelle qui n'a pas été présentée aux chapitres précédents, puisqu'ayant trait à la transformée de Fourier d'une fonction d'échelle. Ce résultat sera employé dans le but de déduire les lemmes A.4 et A.5. Ces lemmes caractérisent le comportement de la fonction  $m_0$ .

**Lemme A.3.** *Si  $\phi \in \mathcal{L}^1(\mathbb{R}) \cap \mathcal{L}^2(\mathbb{R})$  est une fonction d'échelle, alors*

$$\sum_{k \in \mathbb{Z}} \left| \hat{\phi}(s + 2\pi k) \right|^2 = 1 \quad \forall s \in \mathbb{R}.$$

**Démonstration.** D'abord, posons

$$h(x) = \phi * \phi_-(x),$$

où  $\phi_-(x) = \phi(-x)$ . Notons que d'après le théorème de convolution,

$$\hat{h}(s) = \hat{\phi}(s)\hat{\phi}_-(s).$$

Par ailleurs, comme ici  $\phi$  est une fonction réelle,

$$\begin{aligned}\hat{\phi}_-(s) &= \int_{\mathbb{R}} \phi_-(x) e^{-ixs} dx \\ &= \int_{\mathbb{R}} \phi(x) e^{ixs} dx \\ &= \overline{\hat{\phi}(s)},\end{aligned}$$

ce qui amène

$$\hat{h}(s) = \hat{\phi}(s)\overline{\hat{\phi}(s)} = \left| \hat{\phi}(s) \right|^2.$$

Maintenant, posons

$$S(x) = \sum_{k \in \mathbb{Z}} \left| \hat{\phi}(x + 2\pi k) \right|^2 = \sum_{k \in \mathbb{Z}} \hat{h}(x + 2\pi k).$$

Notons qu'en vertu de la proposition A.1,  $S$  est continue. Notons également que  $S$  est  $2\pi$ -périodique, et que  $S \in \mathcal{L}^1(0, 2\pi)$ . En effet,

$$\begin{aligned} \int_0^{2\pi} |S(x)| \, dx &= \int_0^{2\pi} \left| \sum_{k \in \mathbb{Z}} \hat{h}(x + 2\pi k) \right| \, dx \\ &\leq \sum_{k \in \mathbb{Z}} \int_0^{2\pi} \left| \hat{h}(x + 2\pi k) \right| \, dx \\ &= \sum_{k \in \mathbb{Z}} \int_{2\pi k}^{2\pi(k+1)} \left| \hat{h}(x) \right| \, dx \\ &= \int_{\mathbb{R}} \left| \hat{h}(x) \right| \, dx \\ &= \int_{\mathbb{R}} \left| \hat{\phi}(x) \right|^2 \, dx < +\infty, \end{aligned}$$

puisque  $\hat{\phi} \in \mathcal{L}^2(\mathbb{R})$ . Ainsi,  $S$  peut être développée en série de Fourier, ses coefficients étant donnés par

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} S(x) e^{-inx} \, dx.$$

Calculons ces coefficients,

$$\begin{aligned}
 c_n &= \frac{1}{2\pi} \int_0^{2\pi} \left( \sum_{k \in \mathbb{Z}} \hat{h}(x + 2\pi k) \right) e^{-inx} dx \\
 &= \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \int_0^{2\pi} \hat{h}(x + 2\pi k) e^{-inx} dx \\
 &= \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \int_{2\pi k}^{2\pi(k+1)} \hat{h}(s) e^{-in(s-2\pi k)} ds \\
 &= \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} e^{i2\pi nk} \int_{2\pi k}^{2\pi(k+1)} \hat{h}(s) e^{-ins} ds.
 \end{aligned}$$

Or, comme  $e^{i2\pi} = 1$ , alors nous avons aussi

$$e^{i2\pi nk} = 1^{nk} = 1,$$

pour  $n, k \in \mathbb{Z}$ . Les coefficients cherchés peuvent donc s'écrire, d'après le théorème de convolution, comme

$$\begin{aligned}
 c_n &= \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \int_{2\pi k}^{2\pi(k+1)} \hat{h}(s) e^{-ins} ds \\
 &= \frac{1}{2\pi} \int_{\mathbb{R}} \hat{h}(s) e^{-ins} ds \\
 &= \mathcal{F}^{-1}(\hat{h})(-n) \\
 &= h(-n),
 \end{aligned}$$

puisque  $h = \phi * \phi_-$  et que la fonction d'échelle respecte  $\phi \in \mathcal{L}^1(\mathbb{R}) \cap \mathcal{L}^2(\mathbb{R})$ .

Par ailleurs, nous pouvons obtenir, pour  $n \in \mathbb{Z}$ ,

$$\begin{aligned}
 h(-n) &= \phi * \phi_-(-n) \\
 &= \int_{\mathbb{R}} \phi(s) \phi_-(-n-s) ds \\
 &= \int_{\mathbb{R}} \phi(s) \phi(s+n) ds \\
 &= \delta_0(-n),
 \end{aligned}$$

d'après l'équation (1.1). Donc, pour  $n \in \mathbb{Z}$ , il est clair que

$$c_n = \delta_{0(-n)} = \delta_{0n} = \begin{cases} 1 & \text{si } n = 0, \\ 0 & \text{sinon.} \end{cases}$$

Finalement, comme  $S$  est continue, son développement en série de Fourier amène

$$\begin{aligned} S(x) &= \sum_{n \in \mathbb{Z}} c_n e^{-inx} \\ &= \sum_{n \in \mathbb{Z}} \delta_{0n} e^{-inx} = 1, \end{aligned}$$

pour  $x \in \mathbb{R}$ . □

Maintenant, les lemmes qui suivent ont pour but d'étudier le comportement de  $m_0$ . Ils constituent la dernière étape avant d'arriver au but visé ici, soit la démonstration des équations (A.1) et (A.2).

**Lemme A.4.** *Si  $\phi \in \mathcal{L}^1(\mathbb{R}) \cap \mathcal{L}^2(\mathbb{R})$  est une fonction d'échelle, alors la fonction  $m_0$  satisfaisant le lemme A.1 satisfait aussi*

$$|m_0(s)|^2 + |m_0(s + \pi)|^2 = 1 \quad \forall s \in \mathbb{R}.$$

**Démonstration.** D'abord, par le lemme A.1, nous avons

$$\hat{\phi}(2s + 2\pi k) = m_0(s + \pi k) \hat{\phi}(s + \pi k),$$

pour  $s \in \mathbb{R}$ , ce qui nous permet aussi d'écrire

$$\left| \hat{\phi}(2s + 2\pi k) \right|^2 = |m_0(s + \pi k)|^2 \left| \hat{\phi}(s + \pi k) \right|^2.$$

D'autre part, le lemme A.3 amène que, pour  $s \in \mathbb{R}$ ,

$$\sum_{k \in \mathbb{Z}} \left| \hat{\phi}(2s + 2\pi k) \right|^2 = 1.$$

En jumelant ces deux informations, nous avons

$$\begin{aligned}
1 &= \sum_{k \in \mathbb{Z}} |m_0(s + \pi k)|^2 \left| \hat{\phi}(s + \pi k) \right|^2 \\
&= \sum_{k \in \mathbb{Z}} |m_0(s + \pi(2k))|^2 \left| \hat{\phi}(s + \pi(2k)) \right|^2 \\
&\quad + \sum_{k \in \mathbb{Z}} |m_0(s + \pi(2k + 1))|^2 \left| \hat{\phi}(s + \pi(2k + 1)) \right|^2 \\
&= \sum_{k \in \mathbb{Z}} |m_0(s + 2\pi k)|^2 \left| \hat{\phi}(s + 2\pi k) \right|^2 \\
&\quad + \sum_{k \in \mathbb{Z}} |m_0(s + \pi + 2\pi k)|^2 \left| \hat{\phi}(s + \pi + 2\pi k) \right|^2.
\end{aligned}$$

Maintenant, notons que

$$m_0(s + 2\pi k) = m_0(s) \quad \forall k \in \mathbb{Z},$$

et

$$m_0(s + \pi + 2\pi k) = m_0(s + \pi) \quad \forall k \in \mathbb{Z},$$

la fonction  $m_0$  étant  $2\pi$ -périodique. Nous avons donc

$$1 = |m_0(s)|^2 \sum_{k \in \mathbb{Z}} \left| \hat{\phi}(s + 2\pi k) \right|^2 + |m_0(s + \pi)|^2 \sum_{k \in \mathbb{Z}} \left| \hat{\phi}(s + \pi + 2\pi k) \right|^2. \quad (\text{A.4})$$

Finalement, encore une fois grâce au lemme A.3,

$$\sum_{k \in \mathbb{Z}} \left| \hat{\phi}(s + 2\pi k) \right|^2 = \sum_{k \in \mathbb{Z}} \left| \hat{\phi}(s + \pi + 2\pi k) \right|^2 = 1 \quad \forall s \in \mathbb{R},$$

et, en remplaçant ce dernier résultat dans l'équation (A.4), nous obtenons l'égalité recherchée, soit

$$|m_0(s)|^2 + |m_0(s + \pi)|^2 = 1,$$

pour  $s \in \mathbb{R}$ . □

**Lemme A.5.** Si  $\phi \in \mathcal{L}^1(\mathbb{R}) \cap \mathcal{L}^2(\mathbb{R})$  est une fonction d'échelle, alors pour la fonction  $m_0$  satisfaisant le lemme A.1, nous avons

$$|m_0(0)| = 1.$$



**Démonstration.** D'après le lemme A.4,

$$|m_0(0)|^2 + |m_0(\pi)|^2 = 1,$$

et il est donc possible d'affirmer que

$$|m_0(0)| \leq 1. \quad (\text{A.5})$$

Maintenant, supposons que

$$|m_0(0)| < 1, \quad (\text{A.6})$$

et montrons que nous arrivons alors à une contradiction.

D'abord, d'après le lemme A.1, nous avons, pour  $s \in \mathbb{R}$ ,

$$\begin{aligned} \hat{\phi}(s) &= \hat{\phi}\left(\frac{s}{2}\right) m_0\left(\frac{s}{2}\right) \\ &= \hat{\phi}\left(\frac{s}{4}\right) m_0\left(\frac{s}{4}\right) m_0\left(\frac{s}{2}\right) \\ &= \dots = \lim_{n \rightarrow \infty} \hat{\phi}\left(\frac{s}{2^n}\right) \prod_{i=1}^n m_0\left(\frac{s}{2^i}\right) \\ &= \hat{\phi}(0) \prod_{i=1}^{\infty} m_0\left(\frac{s}{2^i}\right), \end{aligned}$$

car  $\hat{\phi}$  est une fonction continue. Par ailleurs, comme  $m_0$  est aussi continue, il existe  $\omega < 1$  et  $\varepsilon > 0$  telles que

$$|m_0(s)| \leq \omega < 1 \quad \forall s \in (-\varepsilon, \varepsilon),$$

et ce, de par l'équation (A.6). Ainsi, pour tout  $s \in \mathbb{R}$ , il existe  $N_s \in \mathbb{N}$  tel que

$$-\varepsilon < \frac{s}{2^i} < \varepsilon \quad \forall i > N_s,$$

et donc tel que

$$\left| m_0\left(\frac{s}{2^i}\right) \right| \leq \omega \quad \forall i > N_s.$$

Alors, pour  $s \in \mathbb{R}$ ,

$$\begin{aligned}
 |\hat{\phi}(s)| &= |\hat{\phi}(0)| \prod_{i=1}^{\infty} \left| m_0 \left( \frac{s}{2^i} \right) \right| \\
 &= |\hat{\phi}(0)| \left( \prod_{i=1}^{N_s} \left| m_0 \left( \frac{s}{2^i} \right) \right| \right) \left( \prod_{i=N_s}^{\infty} \left| m_0 \left( \frac{s}{2^i} \right) \right| \right) \\
 &\leq |\hat{\phi}(0)| \left( \prod_{i=1}^{N_s} \left| m_0 \left( \frac{s}{2^i} \right) \right| \right) \left( \prod_{i=N_s}^{\infty} \omega \right) \\
 &= 0,
 \end{aligned}$$

puisque

$$\hat{\phi}(0) = \int_{\mathbb{R}} \phi(x) dx \leq \int_{\mathbb{R}} |\phi(x)| dx < +\infty, \quad (\text{A.7})$$

et que

$$\prod_{i=1}^{N_s} \left| m_0 \left( \frac{s}{2^i} \right) \right| < +\infty,$$

car  $m_0$  est continue et le nombre de termes apparaissant dans le produit est fini, et que finalement,

$$\prod_{i=N_s}^{\infty} \omega = 0,$$

puisque  $\omega < 1$ ,

Ainsi, l'équation (A.6) amène à la conclusion que

$$\hat{\phi}(s) = 0 \quad \forall s \in \mathbb{R},$$

et donc que

$$\phi(x) = 0 \quad \forall x \in \mathbb{R},$$

qui est un résultat clairement inacceptable. En gardant en vue l'équation (A.5), il est donc nécessaire d'avoir  $|m_0(0)| = 1$ .  $\square$

Finalement, voici les détails plus précis menant aux équations (A.1) et (A.2).

**Théorème A.2.** Si  $\phi \in \mathcal{L}^1(\mathbb{R}) \cap \mathcal{L}^2(\mathbb{R})$  est une fonction d'échelle et, si  $\psi$  est la mère des ondelettes qui lui est associée, alors

$$\int_{\mathbb{R}} \phi(x) dx = \pm 1,$$

et

$$\int_{\mathbb{R}} \psi(x) dx = 0.$$

**Démonstration.** Nous commençons par remarquer que le lemme A.5 permet de conclure que

$$|m_0(2\pi k)| = 1 \quad \forall k \in \mathbb{Z},$$

la fonction  $m_0$  étant  $2\pi$ -périodique, et donc que pour  $j \geq 1$ ,

$$|m_0(2^j \pi k)| = 1 \quad \forall k \in \mathbb{Z}. \quad (\text{A.8})$$

Par ailleurs, d'après le lemme A.1, pour  $j \geq 1$ ,

$$\hat{\phi}(2^{j+1} \pi k) = m_0(2^j \pi k) \hat{\phi}(2^j \pi k) \quad \forall k \in \mathbb{Z},$$

et donc, d'après l'équation (A.8),

$$\begin{aligned} |\hat{\phi}(2^{j+1} \pi k)| &= |m_0(2^j \pi k)| |\hat{\phi}(2^j \pi k)| \\ &= |\hat{\phi}(2^j \pi k)|. \end{aligned}$$

Maintenant, l'équation précédente permet aussi d'écrire, pour  $j \geq 1$ ,

$$|\hat{\phi}(2^j \pi k)| = |\hat{\phi}(2\pi k)| \quad \forall k \in \mathbb{Z}.$$

En passant à la limite, nous obtenons

$$|\hat{\phi}(2\pi k)| = \lim_{j \rightarrow +\infty} |\hat{\phi}(2^j \pi k)| = 0 \quad \forall k \in \mathbb{Z} (k \neq 0), \quad (\text{A.9})$$

puisque des propriétés des transformées de Fourier, nous tirons

$$\lim_{s \rightarrow \pm\infty} \hat{\phi}(s) = 0.$$

Finalement, d'après le lemme A.3,

$$\sum_{k \in \mathbb{Z}} |\hat{\phi}(2\pi k)|^2 = 1,$$

d'où, en enlevant tous les termes nuls donnés par l'équation (A.9),

$$|\hat{\phi}(0)| = 1.$$

Ceci implique finalement que

$$\int_{\mathbb{R}} \phi(x) dx = \hat{\phi}(0) = \pm 1.$$

Passons au deuxième résultat donné par le théorème. Les lemmes A.4 et A.5, permettent de conclure que

$$|m_0(\pi)|^2 = 1 - |m_0(0)|^2 = 0,$$

et donc que

$$m_0(\pi) = 0.$$

Maintenant, en utilisant le lemme A.2 et l'équation précédente, nous avons

$$\hat{\psi}(0) = \overline{m_0(\pi)} e^{-i\pi} \hat{\phi}(0) = 0,$$

puisque  $e^{-i\pi} = -1$  et que d'après l'équation (A.7),  $\hat{\phi}(0) < +\infty$ . Ainsi,

$$\int_{\mathbb{R}} \psi(x) dx = \hat{\psi}(0) = 0,$$

ce qui complète la preuve des résultats d'intérêt. □

## Annexe B

---

Dans cette brève annexe, nous obtenons les équations récursives présentées à la section 1.5. Ces équations relient les divers niveaux de coefficients des ondelettes et des translatés de la fonction d'échelle. Mallat (1989b) a construit son algorithme pyramidal à partir de ces relations récursives. Cet algorithme, aussi appelé algorithme en cascades ("cascade algorithm"), permet de passer d'un niveau de coefficients au suivant ou au précédent. Sa simplicité et son efficacité ont grandement contribué à la popularité des ondelettes.

Le premier résultat présenté ici permet, à partir des coefficients des translatés de la fonction d'échelle d'un niveau donné, d'obtenir tous les coefficients du niveau précédent. Ainsi, il est possible, selon l'équation (1.36), de décomposer toute projection orthogonale  $f_j$  d'une fonction  $f$  en deux parties distinctes. D'abord, les coefficients des translatés de la fonction d'échelle du niveau précédent permettent de construire  $f_{j-1}$ , la projection orthogonale de  $f$  au niveau précédent. Puis, les coefficients des ondelettes du niveau précédent permettent de retrouver les détails de  $f_j$  qui ne sont pas exhibés par  $f_{j-1}$ .

En répétant le procédé plusieurs fois, il devient donc possible de décomposer toute projection orthogonale  $f_j$  en une approximation de base  $f_J$  ( $J < j$ ), et en une série composée de différents niveaux de détails.

**Théorème B.1.** *Si  $\phi \in \mathcal{L}^2(\mathbb{R})$  est une fonction d'échelle, si  $\psi$  est la mère des ondelettes qui lui est associée, si  $f \in \mathcal{L}^2(\mathbb{R})$ , si*

$$c_{jk} = \langle f, \phi_{jk} \rangle \quad \forall j, k \in \mathbb{Z},$$

et si

$$d_{jk} = \langle f, \psi_{jk} \rangle \quad \forall j, k \in \mathbb{Z},$$

alors

$$c_{jk} = \frac{1}{\sqrt{2}} \sum_{m \in \mathbb{Z}} a_{m-2k} c_{(j+1)m} \quad \forall j, k \in \mathbb{Z},$$

et

$$d_{jk} = \frac{1}{\sqrt{2}} \sum_{m \in \mathbb{Z}} (-1)^m a_{1+2k-m} c_{(j+1)m} \quad \forall j, k \in \mathbb{Z},$$

où  $\{a_k\}_{k \in \mathbb{Z}}$  est la suite de filtres associée à  $\phi$ .

**Démonstration.** D'abord, pour les coefficients  $c_{jk}$ , nous avons

$$\begin{aligned} c_{jk} &= \int_{\mathbb{R}} f(x) \phi_{jk}(x) dx \\ &= 2^{j/2} \int_{\mathbb{R}} f(x) \phi(2^j x - k) dx. \end{aligned}$$

Maintenant, l'équation (1.6) amène

$$\begin{aligned} c_{jk} &= 2^{j/2} \int_{\mathbb{R}} f(x) \left( \sum_{m \in \mathbb{Z}} a_m \phi(2(2^j x - k) - m) \right) dx \\ &= 2^{j/2} \sum_{m \in \mathbb{Z}} a_m \int_{\mathbb{R}} f(x) \phi(2^{j+1} x - (2k + m)) dx \\ &= \frac{1}{\sqrt{2}} \sum_{m \in \mathbb{Z}} a_m \int_{\mathbb{R}} f(x) \phi_{(j+1)(2k+m)}(x) dx \\ &= \frac{1}{\sqrt{2}} \sum_{m \in \mathbb{Z}} a_m c_{(j+1)(2k+m)}. \end{aligned}$$

De là, un simple changement de variable permet donc de conclure que

$$c_{jk} = \frac{1}{\sqrt{2}} \sum_{m \in \mathbb{Z}} a_{m-2k} c_{(j+1)m},$$

pour  $j, k \in \mathbb{Z}$ .

Pour les coefficients des ondelettes, nous avons

$$\begin{aligned} d_{jk} &= \int_{\mathbb{R}} f(x) \psi_{jk}(x) dx \\ &= 2^{j/2} \int_{\mathbb{R}} f(x) \psi(2^j x - k) dx, \end{aligned}$$

et l'équation (1.7) amène donc

$$d_{jk} = 2^{j/2} \int_{\mathbb{R}} f(x) \left( \sum_{m \in \mathbb{Z}} b_m \phi(2(2^j x - k) - m) \right) dx,$$

où  $\{b_k\}_{k \in \mathbb{Z}}$  est la suite de filtres associée à  $\psi$ . Maintenant, il suffit de reprendre les mêmes étapes que précédemment pour ainsi obtenir

$$d_{jk} = \frac{1}{\sqrt{2}} \sum_{m \in \mathbb{Z}} b_{m-2k} c_{(j+1)m}.$$

Finalement, grâce à l'équation (1.10), nous pouvons écrire

$$d_{jk} = \frac{1}{\sqrt{2}} \sum_{m \in \mathbb{Z}} (-1)^{m-2k} a_{1+2k-m} c_{(j+1)m},$$

ou, plus simplement,

$$d_{jk} = \frac{1}{\sqrt{2}} \sum_{m \in \mathbb{Z}} (-1)^m a_{1+2k-m} c_{(j+1)m},$$

pour  $j, k \in \mathbb{Z}$ . □

Maintenant, trois nouveaux résultats nous sont indispensables pour obtenir la dernière équation recherchée et donnée par le théorème B.2. Ils sont présentés par les lemmes suivants.

**Lemme B.1.** *Si  $\phi \in \mathcal{L}^2(\mathbb{R})$  est une fonction d'échelle et si  $\{a_k\}_{k \in \mathbb{Z}}$  est la suite de filtres qui lui est associée, alors*

$$\langle \phi_{jk}, \phi_{(j+1)l} \rangle = \frac{1}{\sqrt{2}} a_{l-2k} \quad \forall j, k, l \in \mathbb{Z}.$$

**Démonstration.** Il suffit, encore une fois, d'employer l'équation (1.6). En effet, nous avons alors

$$\begin{aligned} \langle \phi_{jk}, \phi_{(j+1)l} \rangle &= \int_{\mathbb{R}} \phi_{jk}(x) \phi_{(j+1)l}(x) dx \\ &= 2^{j/2} \int_{\mathbb{R}} \phi(2^j x - k) \phi_{(j+1)l}(x) dx \\ &= 2^{j/2} \int_{\mathbb{R}} \left( \sum_{m \in \mathbb{Z}} a_m \phi(2(2^j x - k) - m) \right) \phi_{(j+1)l}(x) dx \\ &= 2^{j/2} \sum_{m \in \mathbb{Z}} a_m \int_{\mathbb{R}} \phi(2^{j+1} x - (2k + m)) \phi_{(j+1)l}(x) dx \\ &= \frac{1}{\sqrt{2}} \sum_{m \in \mathbb{Z}} a_m \int_{\mathbb{R}} \phi_{(j+1)(2k+m)}(x) \phi_{(j+1)l}(x) dx. \end{aligned}$$

Ainsi, nous obtenons

$$\langle \phi_{jk}, \phi_{(j+1)l} \rangle = \frac{1}{\sqrt{2}} \sum_{m \in \mathbb{Z}} a_m \langle \phi_{(j+1)(2k+m)}, \phi_{(j+1)l} \rangle,$$

et l'équation (1.1) amène

$$\langle \phi_{jk}, \phi_{(j+1)l} \rangle = \frac{1}{\sqrt{2}} \sum_{m \in \mathbb{Z}} a_m \delta_{(2k+m)l} = \frac{1}{\sqrt{2}} a_{l-2k},$$

pour  $j, k, l \in \mathbb{Z}$ . □

**Lemme B.2.** Si  $\phi \in \mathcal{L}^2(\mathbb{R})$  est une fonction d'échelle et si  $\psi$  est la mère des ondelettes qui lui est associée, alors

$$\langle \psi_{jk}, \phi_{(j+1)l} \rangle = \frac{1}{\sqrt{2}} (-1)^l a_{1+2k-l} \quad \forall j, k, l \in \mathbb{Z},$$

où  $\{a_k\}_{k \in \mathbb{Z}}$  est la suite de filtres associée à  $\phi$ .

**Démonstration.** Nous devons cette fois plutôt employer l'équation (1.7). Nous avons effectivement

$$\begin{aligned} \langle \psi_{jk}, \phi_{(j+1)l} \rangle &= \int_{\mathbb{R}} \psi_{jk}(x) \phi_{(j+1)l}(x) dx \\ &= 2^{j/2} \int_{\mathbb{R}} \psi(2^j x - k) \phi_{(j+1)l}(x) dx \\ &= 2^{j/2} \int_{\mathbb{R}} \left( \sum_{m \in \mathbb{Z}} b_m \phi(2(2^j x - k) - m) \right) \phi_{(j+1)l}(x) dx, \end{aligned}$$

où  $\{b_k\}_{k \in \mathbb{Z}}$  correspond à la suite de filtres de  $\psi$ . Il suffit maintenant de procéder comme au lemme B.1 pour obtenir finalement

$$\langle \psi_{jk}, \phi_{(j+1)l} \rangle = \frac{1}{\sqrt{2}} b_{l-2k}.$$

Ceci est équivalent, d'après l'équation (1.10), à

$$\begin{aligned} \langle \psi_{jk}, \phi_{(j+1)l} \rangle &= \frac{1}{\sqrt{2}} (-1)^{l-2k} a_{1+2k-l} \\ &= \frac{1}{\sqrt{2}} (-1)^l a_{1+2k-l} \end{aligned}$$

pour  $j, k, l \in \mathbb{Z}$ . □



**Lemme B.3.** Si  $\phi \in \mathcal{L}^2(\mathbb{R})$  est une fonction d'échelle, si  $f \in \mathcal{L}^2(\mathbb{R})$ , si

$$c_{jk} = \langle f, \phi_{jk} \rangle \quad \forall j, k \in \mathbb{Z},$$

et si  $f_j$  est la projection orthogonale de  $f$  dans  $V_j$ , c'est-à-dire

$$f_j(x) = \sum_{k \in \mathbb{Z}} c_{jk} \phi_{jk}(x),$$

alors

$$\langle f_j, \phi_{jk} \rangle = c_{jk} \quad \forall j, k \in \mathbb{Z}.$$

**Démonstration.** Ce résultat s'obtient directement de la définition de  $f_j$  et de l'orthonormalité de la suite  $\{\phi_{jk}\}_{k \in \mathbb{Z}}$ , puisque

$$\begin{aligned} \langle f_j, \phi_{jk} \rangle &= \int_{\mathbb{R}} f_j(x) \phi_{jk}(x) dx \\ &= \int_{\mathbb{R}} \left( \sum_{m \in \mathbb{Z}} c_{jm} \phi_{jm}(x) \right) \phi_{jk}(x) dx \\ &= \sum_{m \in \mathbb{Z}} c_{jm} \int_{\mathbb{R}} \phi_{jm}(x) \phi_{jk}(x) dx \\ &= \sum_{m \in \mathbb{Z}} c_{jm} \langle \phi_{jm}, \phi_{jk} \rangle, \end{aligned}$$

et donc, d'après l'équation (1.1),

$$\langle f_j, \phi_{jk} \rangle = \sum_{m \in \mathbb{Z}} c_{jm} \delta_{mk} = c_{jk},$$

pour  $j, k \in \mathbb{Z}$ . □

Passons finalement à la dernière relation récursive sur laquelle repose l'algorithme de Mallat. Cette relation permet de trouver les coefficients des translatés de la fonction d'échelle d'un niveau donné en regroupant les coefficients du niveau précédent. Ainsi, toujours selon l'équation (1.36), le prochain résultat rend possible la reconstruction d'une projection orthogonale  $f_j$  à partir d'une approximation initiale  $f_J$  et des niveaux de détails appropriés, soit de  $J$  à  $j - 1$ .

**Théorème B.2.** Si  $\phi \in \mathcal{L}^2(\mathbb{R})$  est une fonction d'échelle, si  $\psi$  est la mère des ondelettes qui lui est associée, si  $f \in \mathcal{L}^2(\mathbb{R})$ , si

$$c_{jk} = \langle f, \phi_{jk} \rangle \quad \forall j, k \in \mathbb{Z},$$

et si

$$d_{jk} = \langle f, \psi_{jk} \rangle \quad \forall j, k \in \mathbb{Z},$$

alors

$$c_{(j+1)k} = \frac{1}{\sqrt{2}} \left[ \sum_{m \in \mathbb{Z}} a_{k-2m} c_{jm} + (-1)^k \sum_{m \in \mathbb{Z}} a_{1+2m-k} d_{jm} \right],$$

où  $\{a_k\}_{k \in \mathbb{Z}}$  est la suite de filtres associée à  $\phi$ .

**Démonstration.** D'abord, d'après le lemme B.3 et l'équation (1.36), nous avons

$$\begin{aligned} c_{(j+1)k} &= \langle f_{j+1}, \phi_{(j+1)k} \rangle \\ &= \int_{\mathbb{R}} f_{j+1}(x) \phi_{(j+1)k}(x) dx \\ &= \int_{\mathbb{R}} \left( \sum_{m \in \mathbb{Z}} c_{jm} \phi_{jm}(x) + \sum_{m \in \mathbb{Z}} d_{jm} \psi_{jm}(x) \right) \phi_{(j+1)k}(x) dx \\ &= \sum_{m \in \mathbb{Z}} c_{jm} \int_{\mathbb{R}} \phi_{jm}(x) \phi_{(j+1)k}(x) dx + \sum_{m \in \mathbb{Z}} d_{jm} \int_{\mathbb{R}} \psi_{jm}(x) \phi_{(j+1)k}(x) dx \\ &= \sum_{m \in \mathbb{Z}} c_{jm} \langle \phi_{jm}, \phi_{(j+1)k} \rangle + \sum_{m \in \mathbb{Z}} d_{jm} \langle \psi_{jm}, \phi_{(j+1)k} \rangle. \end{aligned}$$

Il ne reste plus maintenant qu'à appliquer les lemmes B.1 et B.2 pour obtenir finalement

$$c_{(j+1)k} = \frac{1}{\sqrt{2}} \left[ \sum_{m \in \mathbb{Z}} a_{k-2m} c_{jm} + (-1)^k \sum_{m \in \mathbb{Z}} a_{1+2m-k} d_{jm} \right],$$

pour  $j, k \in \mathbb{Z}$ . □

## Annexe C

---

Dans cette annexe, nous présentons quelques lemmes très fortement inspirés de Davis et Rabinowitz (1984, sections 2.9 et 3.4). Ces lemmes sont essentiels puisqu'ils permettent d'obtenir les théorèmes 2.2, 2.10 et 2.11.

**Lemme C.1.** *Si  $f \in \mathcal{L}^1(\mathbb{R}) \cap C^1(\mathbb{R})$  et si  $f' \in \mathcal{L}^1(\mathbb{R})$ , alors*

$$\left| \int_{\mathbb{R}} f(s) ds - h \sum_{k \in \mathbb{Z}} f(kh) \right| \leq \frac{h}{2} \|f'\|_1,$$

pour tout  $h > 0$ .

**Démonstration.** Pour commencer, considérons

$$I_n = \int_{-nh}^{nh} f(s) ds,$$

et notons que

$$\lim_{n \rightarrow \infty} I_n = \int_{\mathbb{R}} f(s) ds,$$

ce qui correspond à l'intégrale que nous souhaitons ici approximer. L'approximation de  $I_n$  obtenue par la méthode du trapèze est donnée par

$$T_n = h \left[ \frac{1}{2} f(-nh) + \sum_{k=-(n-1)}^{n-1} f(kh) + \frac{1}{2} f(nh) \right],$$

et, d'après Davis et Rabinowitz (1984, section 2.9), elle satisfait, en utilisant un développement de type Euler-MacLaurin,

$$T_n = \int_{-nh}^{nh} f(s) ds + h \int_{-nh}^{nh} P_1\left(\frac{s}{h}\right) f'(s) ds,$$

où

$$P_1(s) = s - [s] - \frac{1}{2}, \tag{C.1}$$

où  $[s]$  correspond à la partie entière de  $s$ . Notons que  $P_1$  respecte

$$|P_1(s)| \leq \frac{1}{2} \quad \forall s \in \mathbb{R}.$$

L'erreur d'approximation associée à  $T_n$  peut donc être bornée selon

$$\begin{aligned} |I_n - T_n| &= \left| h \int_{-nh}^{nh} P_1\left(\frac{s}{h}\right) f'(s) ds \right| \\ &\leq h \int_{-nh}^{nh} \left| P_1\left(\frac{s}{h}\right) f'(s) \right| ds \\ &\leq \frac{h}{2} \int_{-nh}^{nh} |f'(s)| ds. \end{aligned}$$

En procédant comme Davis et Rabinowitz (1984, section 3.4), un passage à la limite amène finalement

$$\begin{aligned} \left| \int_{\mathbb{R}} f(s) ds - h \sum_{k \in \mathbb{Z}} f(kh) \right| &= \lim_{n \rightarrow +\infty} |I_n - T_n| \\ &\leq \frac{h}{2} \|f'\|_1, \end{aligned}$$

quelque soit  $h > 0$ . □

**Lemme C.2.** Si  $f \in \mathcal{L}^1(\mathbb{R}) \cap C^2(\mathbb{R})$ , si  $f' \in \mathcal{L}^\infty$  et si  $f'' \in \mathcal{L}^1(\mathbb{R})$ , alors

$$\left| \int_{\mathbb{R}} f(s) ds - h \sum_{k \in \mathbb{Z}} f(kh) \right| \leq \frac{h^2}{6} \|f'\|_\infty + \frac{h^2}{12} \|f''\|_1,$$

pour tout  $h > 0$ .

**Démonstration.** Comme au lemme C.1, considérons

$$I_n = \int_{-nh}^{nh} f(s) ds,$$

et son approximation obtenue par la méthode du trapèze, soit

$$T_n = h \left[ \frac{1}{2} f(-nh) + \sum_{k=-(n-1)}^{n-1} f(kh) + \frac{1}{2} f(nh) \right].$$

D'après Davis et Rabinowitz (1984, section 2.9), comme  $f \in C^2(\mathbb{R})$ , alors il est possible de montrer, en utilisant encore une fois un développement de type Euler-MacLaurin, que  $T_n$  satisfait

$$T_n = \int_{-nh}^{nh} f(s) ds + \frac{h^2}{12} [f'(nh) - f'(-nh)] - h^2 \int_{-nh}^{nh} P_2\left(\frac{s}{h}\right) f''(s) ds, \quad (\text{C.2})$$

où

$$P_2(s) = \frac{1}{2\pi^2} \sum_{j=1}^{\infty} \frac{\cos 2\pi j s}{j^2}. \quad (\text{C.3})$$

Notons qu'il est clair que  $P_2$  respecte

$$|P_2(s)| \leq \frac{1}{2\pi^2} \sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{1}{12},$$

pour  $s \in \mathbb{R}$ , puisque

$$\sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6}.$$

Il est maintenant possible de borner l'erreur d'approximation associée à  $T_n$ . En effet, d'après l'équation (C.2), nous avons

$$\begin{aligned} |I_n - T_n| &\leq \frac{h^2}{12} |f'(nh) - f'(-nh)| + \left| h^2 \int_{-nh}^{nh} P_2\left(\frac{s}{h}\right) f''(s) ds \right| \\ &\leq \frac{h^2}{6} \|f'\|_{\infty} + h^2 \int_{-nh}^{nh} \left| P_2\left(\frac{s}{h}\right) f''(s) \right| ds \\ &\leq \frac{h^2}{6} \|f'\|_{\infty} + \frac{h^2}{12} \int_{-nh}^{nh} |f''(s)| ds, \end{aligned}$$

et il ne reste qu'à passer à la limite pour obtenir le résultat recherché. Nous avons alors

$$\begin{aligned} \left| \int_{\mathbb{R}} f(s) ds - h \sum_{k \in \mathbb{Z}} f(kh) \right| &= \lim_{n \rightarrow +\infty} |I_n - T_n| \\ &\leq \frac{h^2}{6} \|f'\|_{\infty} + \frac{h^2}{12} \|f''\|_1, \end{aligned}$$

quelque soit  $h > 0$ . □

**Lemme C.3.** Si  $f \in \mathcal{L}^1(\mathbb{R}) \cap C^1(\mathbb{R})$  et si  $f' \in \mathcal{L}^1(\mathbb{R})$ , alors

$$\left| \int_a^\infty f(s) ds - h \left[ \frac{1}{2}f(a) + \sum_{k=1}^{\infty} f(a+kh) \right] \right| \leq \frac{h}{2} \|f'\|_1,$$

pour  $h > 0$  et quelque soit  $a \in \mathbb{R}$ .

De même,

$$\left| \int_{-\infty}^a f(s) ds - h \left[ \frac{1}{2}f(a) + \sum_{k=-\infty}^{-1} f(a+kh) \right] \right| \leq \frac{h}{2} \|f'\|_1,$$

pour  $h > 0$  et  $a \in \mathbb{R}$ .

**Démonstration.** Nous démontrons uniquement le premier résultat. Pour cela, nous considérons cette fois

$$I_n = \int_a^{a+nh} f(s) ds,$$

et notons que

$$\lim_{n \rightarrow \infty} I_n = \int_a^\infty f(s) ds.$$

Comme au lemme C.1, nous approximations  $I_n$  par la méthode du trapèze, soit ici

$$T_n = h \left[ \frac{1}{2}f(a) + \sum_{k=1}^{n-1} f(a+kh) + \frac{1}{2}f(a+nh) \right].$$

Cette approximation satisfait l'égalité suivante, obtenue encore une fois par un développement de type Euler-MacLaurin,

$$T_n = \int_a^{a+nh} f(s) ds + h \int_a^{a+nh} P_1 \left( \frac{s-a}{h} \right) f'(s) ds,$$

où la fonction  $P_1$  est toujours donnée par l'équation (C.1). Ainsi, nous avons

$$\begin{aligned} |I_n - T_n| &\leq \left| h \int_a^{a+nh} P_1 \left( \frac{s-a}{h} \right) f'(s) ds \right| \\ &\leq h \int_a^{a+nh} \left| P_1 \left( \frac{s-a}{h} \right) f'(s) \right| ds \\ &\leq \frac{h}{2} \int_a^{a+nh} |f'(s)| ds, \end{aligned}$$

ce qui nous amène à écrire

$$\left| \int_a^\infty f(s) ds - h \left[ \frac{1}{2} f(a) + \sum_{k=1}^{\infty} f(a + kh) \right] \right| \leq \frac{h}{2} \int_a^\infty |f'(s)| ds$$

$$\leq \frac{h}{2} \|f'\|_1,$$

pour tout  $h > 0$  et indépendamment de  $a \in \mathbb{R}$ . □

**Lemme C.4.** Si  $f \in \mathcal{L}^1(\mathbb{R}) \cap C^2(\mathbb{R})$ , si  $f' \in \mathcal{L}^\infty$  et si  $f'' \in \mathcal{L}^1(\mathbb{R})$ , alors

$$\left| \int_a^\infty f(s) ds - h \left[ \frac{1}{2} f(a) + \sum_{k=1}^{\infty} f(a + kh) \right] \right| \leq \frac{h^2}{6} \|f'\|_\infty + \frac{h^2}{12} \|f''\|_1$$

pour tout  $a \in \mathbb{R}$  et  $h > 0$ .

De même, nous avons également

$$\left| \int_{-\infty}^a f(s) ds - h \left[ \frac{1}{2} f(a) + \sum_{k=-\infty}^{-1} f(a + kh) \right] \right| \leq \frac{h^2}{6} \|f'\|_\infty + \frac{h^2}{12} \|f''\|_1$$

pour  $a \in \mathbb{R}$  et  $h > 0$ .

**Démonstration.** Nous ne démontrons, encore une fois, que le premier résultat. D'abord, comme au lemme C.3, nous nous intéressons à

$$I_n = \int_a^{a+nh} f(s) ds,$$

et à son approximation par la méthode du trapèze donnée par

$$T_n = h \left[ \frac{1}{2} f(a) + \sum_{k=1}^{n-1} f(a + kh) + \frac{1}{2} f(a + nh) \right].$$

Un dernier développement de type Euler-MacLaurin permet d'écrire

$$T_n = \int_a^{a+nh} f(s) ds + \frac{h^2}{12} [f'(a + nh) - f'(a)]$$

$$+ h^2 \int_a^{a+nh} P_2 \left( \frac{s-a}{h} \right) f''(s) ds,$$

la fonction  $P_2$  ayant déjà été présentée au lemme C.2. Ceci permet de borner l'erreur d'approximation de  $T_n$  par

$$\begin{aligned} |I_n - T_n| &\leq \frac{h^2}{12} |f'(a+nh) - f'(a)| + \left| h^2 \int_a^{a+nh} P_2 \left( \frac{s-a}{h} \right) f''(s) ds \right| \\ &\leq \frac{h^2}{6} \|f'\|_\infty + \frac{h^2}{12} \int_a^{a+nh} |f''(s)| ds. \end{aligned}$$

En passant à la limite, nous obtenons finalement

$$\left| \int_a^\infty f(s) ds - h \left[ \frac{1}{2} f(a) + \sum_{k=1}^\infty f(a+kh) \right] \right| \leq \frac{h^2}{6} \|f'\|_\infty + \frac{h^2}{12} \|f''\|_1,$$

soit la même borne qu'au lemme C.2, et ce indépendamment de la valeur de  $a \in \mathbb{R}$ .  $\square$



## Bibliographie

---

Abramovich, F. et Benjamini, Y. (1995), *Thresholding of wavelet coefficients as a multiple hypothesis testing procedure*, Wavelets and statistics (A. Antoniadis et G. Oppenheim eds.), p. 5-14, Springer-Verlag, New York.

Abramovich, F. et Benjamini, Y. (1996), *Adaptive thresholding of wavelet coefficients*, Computational Statistics and Data Analysis, Vol. 22, p. 351-361.

Abramovich, F. et Sapatinas, T. (1999), *Bayesian approach to wavelet decomposition and shrinkage*, Bayesian inference in wavelet-based models (P. Müller et B. Vidakovic eds.), p. 33-50, Springer-Verlag, New York.

Abramovich, F., Sapatinas, T. et Silverman, B.W. (1998), *Wavelet thresholding via a Bayesian approach*, Journal of the Royal Statistical Society, Série B, Vol. 60, p. 725-749.

Abry, P., Veitch, D. et Flandrin, P. (1998), *Long-range dependence: revisiting aggregation with wavelets*, Journal of Time Series Analysis, Vol. 19, p. 253-266.

Alpert, B.K. (1992), *Wavelets and other bases for fast numerical linear algebra*, Wavelets-A tutorial in theory and applications (C.K. Chui ed.), p. 181-216, Academic Press, Boston.

Angers, J.-F. (1992), *Use of the Student-t prior for the estimation of normal means: a computational approach*, Bayesian Statistics 4 (J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith eds.), p. 567-575, Oxford University Press.

Angers, J.-F. (1996a), *Fourier transforms and Bayes estimator of a location parameter*, Statistics and Probability Letters, Vol. 29, p. 353-359.

Angers, J.-F. (1996b), *Protection against outliers using a symmetric stable law prior*, IMS Lecture Notes, Vol. 29, p. 273-283.

Angers, J.-F. (1997), *Bayesian estimator of the location parameter of a Student-t density*, rapport technique 97-07, Nottingham University Statistics Group,

Angers, J.-F. et Berger, J.O. (1991), *Robust hierarchical Bayes estimation of exchangeable means*, The Canadian Journal of Statistics, Vol. 19, p. 39-56.

Antoniadis, A., Gijbels, I. et Grégoire, G. (1997), *Model selection using wavelet decomposition and applications*, Biometrika, Vol. 84, p. 751-763.

- Belsley, D.A., Kuh, E. et Welsch, R.E. (1980), *Regression diagnostics : identifying influential data and sources of collinearity*, Wiley, New York.
- Berger, J.O. (1984), *The robust Bayesian viewpoint*, Robustness of Bayesian analysis (J.B. Kadane ed.), p. 63-144, North-Holland, Amsterdam.
- Berger, J.O. (1985), *Statistical decision theory and Bayesian analysis*, 2<sup>ème</sup> édition, Springer-Verlag, New York.
- Berger, J.O. et Berliner, L.M. (1986), *Robust Bayes and empirical Bayes analysis with  $\varepsilon$ -contaminated priors*, Annals of Statistics, Vol. 14, p. 461-486.
- Berger, J.O. et Bernardo, J.M. (1992), *On the development of the reference prior method*, Bayesian Statistics 4 (J.M. Bernardo, J.O. Berger, A.P. Dawid et A.F.M. Smith eds.), p. 35-60, Oxford University Press.
- Bernardo, J.M. (1979), *Reference posterior distributions for Bayesian inference*, Journal of the Royal Statistical Society, Série B, Vol. 41, p. 113-147.
- Beylkin, G., Coifman, R. et Rokhlin, V. (1991), *Fast wavelet transforms and numerical algorithms*, Communications on Pure and Applied Mathematics, Vol. 44, p. 141-183.
- Bickel, P.J. et Ritov, Y. (1987), *Efficient estimation in the errors-in-variables model*, Annals of Statistics, Vol. 15, p. 513-540.
- Bracewell, R.N. (1978), *The Fourier transform and its applications*, 2<sup>ème</sup> édition, McGraw-Hill, New York.
- Carroll, R.J. et Hall, P. (1988), *Optimal rates of convergence for deconvolving a density*, Journal of the American Statistical Association, Vol. 83, p. 1184-1186.
- Chipman, H.A., Kolaczyk, E.D. et McCulloch, R.E. (1997), *Adaptive Bayesian wavelet shrinkage*, Journal of the American Statistical Association, Vol. 92, p. 1413-1421.
- Chipman, H.A. et Wolfson, L.J. (1999), *Prior elicitation in the wavelet domain*, Bayesian inference in wavelet-based models (P. Müller et B. Vidakovic eds.), p. 83-94, Springer-Verlag, New York.
- Chui, C.K. (1992), *An introduction to wavelets*, Academic Press, Boston.
- Clyde, M.A. et George, E.I. (1999), *Empirical Bayes estimation in wavelet non-parametric regression*, Bayesian inference in wavelet-based models (P. Müller et B. Vidakovic eds.), p. 309-322, Springer-Verlag, New York.
- Clyde, M.A., Parmigiani, G. et Vidakovic, B. (1998), *Multiple shrinkage and subset selection in wavelets*, Biometrika, Vol. 85, p. 391-402.

- Dalal, S.R. et Hall, W.J. (1983), *Approximating priors by mixtures of natural conjugate priors*, Journal of the Royal Statistical Society, Série B, Vol. 45, p. 278-286.
- Daubechies, I. (1988), *Orthonormal bases of compactly supported wavelets*, Communications on Pure and Applied Mathematics, Vol. 41, p. 909-996.
- Daubechies, I. (1992), *Ten lectures on wavelets*, SIAM, Philadelphie.
- Daubechies, I. (1993), *Orthonormal bases of compactly supported wavelets II: variations on a theme*, SIAM Journal on Mathematical Analysis, Vol. 24, p. 499-519.
- Daubechies, I. et Lagarias, J. (1991), *Two-scale difference equations I: existence and global regularity of solutions*, SIAM Journal on Mathematical Analysis, Vol. 22, p. 1388-1410.
- Daubechies, I. et Lagarias, J. (1992), *Two-scale difference equations II: local regularity, infinite products of matrices and fractals*, SIAM Journal on Mathematical Analysis, Vol. 23, p. 1031-1079.
- Davis, P.J. et Rabinowitz, P. (1984), *Methods of numerical integration*, 2<sup>ième</sup> édition, Academic Press, Orlando.
- Delyon, B. et Juditsky, A. (1995), *Estimating wavelet coefficients*, Wavelets and statistics (A. Antoniadis et G. Oppenheim eds.), p. 151-168, Springer-Verlag, New York.
- Delyon, B. et Juditsky, A. (1997), *On the computation of wavelet coefficients*, Journal of Approximation Theory, Vol. 88, p. 47-79.
- Donoho, D.L. (1995), *De-noising by soft thresholding*, IEEE Transactions on Information Theory, Vol. 41, p. 613-627.
- Donoho, D.L. et Johnstone, I.M. (1994), *Ideal spatial adaptation by wavelet shrinkage*, Biometrika, Vol. 81, p. 425-455.
- Donoho, D.L. et Johnstone, I.M. (1995), *Adapting to unknown smoothness via wavelet shrinkage*, Journal of the American Statistical Association, Vol. 90, p. 1200-1224.
- Donoho, D.L. et Johnstone, I.M. (1998), *Minimax estimation via wavelet shrinkage*, Annals of statistics, Vol. 26, p. 879-921.
- Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. et Picard, D. (1995), *Wavelet shrinkage: asymptopia?*, Journal of the Royal Statistical Society, Série B, Vol. 57, p. 301-369.

- Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. et Picard, D. (1996), *Density estimation by wavelet thresholding*, Annals of Statistics, Vol. 24, p. 508-539.
- Efron, B. et Morris, C.N. (1972), *Limiting the risk of Bayes and empirical Bayes estimators - Part II: the empirical Bayes case*, Journal of the American Statistical Association, Vol. 67, p. 130-139.
- Eggermont, P.P.B. et LaRiccia, V.N. (1995), *Maximum smoothed likelihood density estimation for inverse problems*, Annals of Statistics, Vol. 23, p. 199-220.
- Fan, J. (1991a), *Global behavior of deconvolution kernel estimates*, Statistica Sinica, Vol. 1, p. 541-551.
- Fan, J. (1991b), *On the optimal rates of convergence for nonparametric deconvolution problems*, Annals of Statistics, Vol. 19, p. 1257-1272.
- Fan, J. (1996), *Test of significance based on wavelet thresholding and Neyman's truncation*, Journal of the American Statistical Association, Vol. 91, p. 674-688.
- Fan, J., Hall, P., Martin, M. et Patil, P. (1999), *Adaptation to high spatial inhomogeneity using wavelet methods*, Statistica Sinica, Vol. 9, p. 85-102.
- Fan, J. et Truong, Y.K. (1993), *Nonparametric regression with errors in variables*, Annals of Statistics, Vol. 21, p. 1900-1925.
- Fan, T.-H. et Berger, J.O. (1990), *Exact convolution of  $T$  distributions, with applications to Bayesian inference for a normal mean with  $T$  prior distributions*, Journal of Statistical Computation and Simulation, Vol. 36, p. 209-228.
- Foucart, T. (1996), *Analyse de la collinéarité et classification de variables*, Revue de Statistique Appliquée, Vol. 44, No. 4, p. 41-57.
- Gao, H.-Y. (1997), *Choice of thresholds for wavelet shrinkage estimate of the spectrum*, Journal of Time Series Analysis, Vol. 18, p. 231-251.
- Gelfand, A.E. et Smith, A.F.M. (1990), *Sampling based approaches to calculating marginal densities*, Journal of the American Statistical Association, Vol. 85, p. 398-409.
- Goel, P.K. (1983), *Information measures and Bayesian hierarchical models*, Journal of the American Statistical Association, Vol. 78, p. 408-410.
- Goel, P.K. et DeGroot, M.H. (1981), *Information about hyperparameters in hierarchical models*, Journal of the American Statistical Association, Vol. 76, p. 140-147.
- Good, I.J. (1980), *Some history of the hierarchical Bayesian methodology*, Bayesian Statistics 2 (J.M. Bernardo, M.H. DeGroot, D.V. Lindley et A.F.M. Smith eds.), p. 489-519, University Press, Valencia.

- Goutis, C. (1997), *Nonparametric estimation of a mixing density via the kernel method*, Journal of the American Statistical Association, Vol. 92, p. 1445-1450.
- Graybill, F.A. (1969), *Introduction to matrices with applications in statistics*, Wadsworth, Belmont.
- Grossman, A. et Morlet, J. (1984), *Decomposition of Hardy functions into square integrable wavelets of constant shape*, Siam Journal on Mathematical Analysis, Vol. 15, p. 723-736.
- Haar, A. (1910), *Zur theorie der orthogonalen funktionensysteme*, Mathematische Annalen, Vol. 69, p. 331-371.
- Hall, P. et Patil, P. (1996a), *Effect of the thresholding rules on performance of wavelet-based curve estimators*, Statistica Sinica, Vol. 6, p. 331-345.
- Hall, P. et Patil, P. (1996b), *On the choice of smoothing parameter, threshold and truncation in nonparametric regression by nonlinear wavelet methods*, Journal of the Royal Statistical Society, Série B, Vol. 58, p. 361-377.
- Hall, P. et Wehrly, T.E. (1991), *A geometrical method for removing edge effects from kernel-type nonparametric regression estimators*, Journal of the American Statistical Association, Vol. 86, p. 665-672.
- Härdle, W., Kerkycharian, G., Picard, D. et Tsybakov, A.B. (1998), *Wavelets, approximations, and statistical applications*, Springer-Verlag, New York.
- Healy, D.M. et Kim, P.T. (1996), *An empirical Bayes approach to directional data and efficient computation on the sphere*, Annals of Statistics, Vol. 24, p. 232-254.
- Hill, R.C. et Judge, G.G. (1990), *Improved estimation under collinearity and squared error loss*, Journal of Multivariate Analysis, Vol. 32, p. 296-312.
- Holmes, C.C. et Denison, D.G.T. (1999), *Bayesian wavelet analysis with a model complexity prior*, Bayesian Statistics 6 (J.M. Bernardo, J.O. Berger, A.P. Dawid et A.F.M. Smith eds.), p. 769-776, Oxford University Press.
- Huang, S.-Y. (1997), *Wavelet based empirical Bayes estimation for the uniform distribution*, Statistics and Probability Letters, Vol. 32, p. 141-146.
- Huang, S.-Y. (1999), *Density estimation by wavelet based reproducing kernels*, Statistica Sinica, Vol. 9, p. 137-151.
- Izenman, A.J. (1991), *Recent developments in nonparametric density estimation*, Journal of the American Statistical Association, Vol. 86, p. 205-224.
- Jaynes, E.T. (1968), *Prior probabilities*, IEEE Transactions on systems, science and cybernetics, Vol. 4, p. 227-241.

- Jeffreys, H. (1961), *Theory of probability*, 3<sup>ième</sup> édition, Oxford University Press, Londres.
- Kashin, B.S. et Saakyan, A.A. (1989), *Orthogonal series*, American Mathematical Society, Providence.
- Kass, R.E. et Wasserman, L. (1996), *The selection of prior distributions by formal rules*, *Journal of the American Statistical Association*, Vol. 91, p. 1343-1370.
- Kerkyacharian, G. et Picard, D. (1992), *Density estimation in Besov spaces*, *Statistics and Probability Letters*, Vol. 13, p. 15-24.
- Kerkyacharian, G. et Picard, D. (1993), *Density estimation by kernel and wavelet methods: optimality of Besov spaces*, *Statistics and Probability Letters*, Vol. 18, p. 327-336.
- Laird, N.M. (1978), *Nonparametric maximum likelihood estimation of a mixing distribution*, *Journal of the American Statistical Association*, Vol. 73, p. 805-811.
- Laird, N.M. et Louis, T.A. (1991), *Smoothing the nonparametric estimate of a prior distribution by roughening: a computational study*, *Computational Statistics and Data Analysis*, Vol. 12, p. 27-37.
- Lancaster, P. et Tismenetsky, M. (1985), *The theory of matrices*, 2<sup>ième</sup> édition, Academic Press, Orlando.
- Leblanc, A. (1995), *Utilisation de la transformée de Fourier rapide en estimation bayésienne*, mémoire de maîtrise, Université de Montréal.
- Leblanc, A. et Angers, J.-F. (1995), *Fast Fourier transforms and Bayesian estimation of location parameters*, rapport technique DMS-380, Département de mathématiques et de statistique, Université de Montréal.
- Li, Y. et Xie, Z. (1997), *The wavelet detection of hidden periodicities in time series*, *Statistics and Probability Letters*, Vol. 35, p. 9-23.
- Lindley, D.V. et Smith, A.F.M. (1972), *Bayes estimates for the linear model*, *Journal of the Royal Statistical Society, Série B*, Vol. 34, p. 1-41.
- Madger, L.S. et Zeger, S.L. (1996), *A smooth nonparametric estimate of a mixing distribution using mixtures of gaussians*, *Journal of the American Statistical Association*, Vol. 91, p. 1141-1151.
- Mallat, S. (1989a), *Multiresolution approximations and wavelet orthonormal bases of  $\mathcal{L}^2(\mathbb{R})$* , *Transactions of the American Mathematical Society*, Vol. 315, p. 69-88.

- Mallat, S. (1989b), *A theory for multiresolution signal decomposition: the wavelet transform*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, p. 674-693.
- Mallat, S. (1998), *A wavelet tour of signal processing*, Academic Press, San Diego.
- Mandel, J. (1982), *Use of the singular value decomposition in regression analysis*, American Statistician, Vol. 36, p. 15-24.
- Marron, J.S., Adak, S., Johnstone, I.M., Neumann, M.H. et Patil, P. (1998), *Exact risk analysis of wavelet regression*, Journal of Computational and Graphical Statistics, Vol. 7, p. 278-309.
- McCoy, E.J. et Walden A.T. (1996), *Wavelet analysis and synthesis of stationary long-memory processes*, Journal of Computational and Graphical Statistics, Vol. 5, p. 26-56.
- Meyer, Y. (1990), *Ondelettes et opérateurs I: Ondelettes*, Hermann, Paris.
- Meyer, Y. (1992), *Wavelets and operators*, Cambridge University Press, Cambridge.
- Morris, C.N. (1983), *Parametric empirical Bayes inference: theory and applications*, Journal of the American Statistical Association, Vol. 78, p. 47-65.
- Müller, P. et Vidakovic, B. (1998), *Bayesian inference with wavelets: density estimation*, Journal of Computational and Graphical Statistics, Vol. 7, p. 456-468.
- Müller, P. et Vidakovic, B. (1999), *MCMC methods in wavelet shrinkage: non-equally spaced regression, density and spectral density estimation*, Bayesian inference in wavelet-based models (P. Müller et B. Vidakovic eds.), p. 187-202, Springer-Verlag, New York.
- Nason, G. (1995), *Choice of the threshold parameter in wavelet function estimation*, Wavelets and Statistics (A. Antoniadis et G. Oppenheim eds.), p. 261-280, Springer-Verlag, New York.
- Nason, G. (1996), *Wavelet shrinkage using cross-validation*, Journal of the Royal Statistical Society, Série B, Vol. 58, p. 463-479.
- Naylor, J.C. et Smith, A.F.M. (1982), *Applications of a method for the efficient computation of posterior distributions*, Applied Statistics, Vol. 31, p. 214-225.
- Neumann, M.H. (1996), *Spectral density estimation via nonlinear wavelet methods for stationary non-gaussian time series*, Journal of Time Series Analysis, Vol. 17, p. 601-633.

- Ogden, R.T. (1997), *Essential wavelets for statistical applications and data analysis*, Birkhäuser, Boston.
- Ogden, R.T. et Parzen, E. (1996), *Data dependent wavelet thresholding in non-parametric regression with change-point applications*, Computational Statistics and Data Analysis, Vol. 22, p. 53-70.
- O'Hagan, A. (1990), *Outliers and credence for location parameter inference*, Journal of the American Statistical Association, Vol. 85, p. 172-176.
- O'Hagan, A. (1994), *Kendall's advanced theory of statistics (volume 2B): Bayesian inference*, Edward Arnold, Londres.
- Pensky, M. (1999), *Nonparametric empirical Bayes estimation via wavelets*, Bayesian inference in wavelet-based models (P. Müller et B. Vidakovic eds.), p. 323-340, Springer-Verlag, New York.
- Pinheiro, A. et Vidakovic, B. (1997), *Estimating the square root of a density via compactly supported wavelets*, Computational Statistics and Data Analysis, Vol. 25, p. 399-415.
- Priestley, M.B. (1996), *Wavelets and time-dependent spectral analysis*, Journal of Time Series Analysis, Vol. 17, p. 85-104.
- Robert, C.P. (2001), *The bayesian choice*, 2<sup>ième</sup> édition, Springer-Verlag, New York.
- Robbins, H. (1955), *An empirical Bayes approach to statistics*, Proceedings of the third Berkeley symposium on mathematical statistics and probabilities, Vol. 1, p. 157-164, University of California Press, Berkeley.
- Robbins, H. (1964), *The empirical Bayes approach to statistical decision problems*, Annals of Mathematical Statistics, Vol. 35, p. 1-20.
- Roeder, K., Carroll, R.J. et Lindsay, B.G. (1996), *A semiparametric mixture approach to case-control studies with errors in covariables*, Journal of the American Statistical Association, Vol. 91, p. 722-732.
- Rosenblatt, M. (1956), *Remarks on some nonparametric estimates of a density function*, Annals of Mathematical Statistics, Vol. 27, p. 832-837.
- Rudin, W. (1987), *Real and complex analysis*, 3<sup>ième</sup> édition, McGraw-Hill, New York.
- Ruggeri, F. (1999), *Robust Bayesian and Bayesian decision theoretic wavelet shrinkage*, Bayesian inference in wavelet-based models (P. Müller et B. Vidakovic eds.), p. 139-154, Springer-Verlag, New York.



- Ruggeri, F. et Vidakovic, B. (1999), *A Bayesian decision theoretic approach to wavelet thresholding*, *Statistica Sinica*, Vol. 9, p. 183-197.
- Searle, S.R. (1982), *Matrix algebra useful for statistics*, Wiley, New York.
- Serfling, R.J. (1980), *Approximation theorems of mathematical statistics*, Wiley, New York.
- Spiegelhalter, D.J. (1985), *Exact Bayesian inference on the parameters of a Cauchy distribution with vague prior information*, *Bayesian Statistics 2* (J.M. Bernardo, M.H. DeGroot, D.V. Lindley et A.F.M. Smith eds.), p. 743-750, North-Holland, Amsterdam.
- Stein, C.S. (1981), *Estimation of the mean of a multivariate normal distribution*, *Annals of Statistics*, Vol. 9, p. 1135-1151.
- Stewart, G.W. (1973), *Introduction to matrix computations*, Academic Press, New York.
- Stewart, G.W. (1987), *Collinearity and least squares regression*, *Statistical Science*, Vol. 2, p. 68-100.
- Strang, G. (1989), *Wavelets and dilation equations: a brief introduction*, *SIAM Review*, Vol. 31, p. 614-627.
- Tanner, M.A. (1993), *Tools for statistical inference*, 2<sup>ième</sup> édition, Springer-Verlag, New York.
- Tao, T. (1996), *On the almost everywhere convergence of wavelet summation methods*, *Applied Computational Harmonic Analysis*, Vol. 3, p. 384-387.
- Tao, T. et Vidakovic, B. (1998), *Almost everywhere convergence of general wavelet shrinkage estimators*, rapport technique 98-31, ISDS, Duke University.
- Tierney, L. et Kadane, J.B. (1986), *Accurate approximations for posterior moments and marginal densities*, *Journal of the American Statistical Association*, Vol. 81, p. 82-86.
- Tribouley, K. (1995a), *Adaptive density estimation*, *Wavelets and Statistics* (A. Antoniadis et G. Oppenheim eds.), p. 385-395, Springer-Verlag, New York.
- Tribouley, K. (1995b), *Practical estimation of multivariate densities using wavelet methods*, *Statistica Neerlandica*, Vol. 49, p. 41-62.
- Vannucci, M. et Corradi, F. (1999), *Modeling dependence in the wavelet domain*, *Bayesian inference in wavelet-based models* (P. Müller et B. Vidakovic eds.), p. 173-186, Springer-Verlag, New York.
- Vannucci, M. et Vidakovic, B. (1995), *Preventing the Dirac disaster: wavelet based density estimation*, rapport technique 95-27, ISDS, Duke University.

- Vidakovic, B. (1996), *A note on random densities via wavelets*, Statistics and Probability Letters, Vol. 26, p. 315-321.
- Vidakovic, B. (1998), *Nonlinear wavelet shrinkage with Bayes rules and Bayes factors*, Journal of the American Statistical Association, Vol. 93, p. 173-179.
- Vidakovic, B. (1999), *Statistical modelling by wavelets*, Wiley, New York.
- Vinod, H.D. et Ullah, A. (1981), *Recent advances in regression methods*, Dekker, New York.
- Walker, J.S. (1988), *Fourier analysis*, Oxford University Press, New York.
- Walnut, D.F. (2002), *An introduction to wavelet analysis*, Birkäuser, Boston.
- Wang, Y. (1995), *Jump and sharp cusp detection by wavelets*, Biometrika, Vol. 82, p. 385-397.
- Wang, Y. (1996), *Function estimation via wavelet shrinkage for long-memory data*, Annals of Statistics, Vol. 24, p. 466-484.
- Zellner, A. (1977), *Maximal data information prior distributions*, New developments in the applications of Bayesian methods (A. Aykac et C. Brumat eds.), p. 201-215, North-Holland, Amsterdam.
- Zhang, C.H. (1990), *Fourier methods for estimating mixing densities and distributions*, Annals of Statistics, Vol. 18, p. 806-831.
- Zhang, Q. (1995), *Wavelets and regression analysis*, Wavelets and statistics (A. Antoniadis et G. Oppenheim eds.), p. 397-407, Springer-Verlag, New York.