



HAL
open science

Étude des fonctions B-splines pour la fusion d'images segmentées par approche bayésienne

Atizez Hadrich Ben Arab

► **To cite this version:**

Atizez Hadrich Ben Arab. Étude des fonctions B-splines pour la fusion d'images segmentées par approche bayésienne. Statistiques [math.ST]. Université du Littoral Côte d'Opale; Université de Sfax. Faculté des sciences, 2015. Français. NNT : 2015DUNK0385 . tel-01306291

HAL Id: tel-01306291

<https://theses.hal.science/tel-01306291v1>

Submitted on 22 Apr 2016

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

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



Université du Littoral Côte d'Opale
Ecole Doctorale de Lille

Thèse de DOCTORAT
N° d'ordre:



Université de Sfax
Faculté des Sciences de Sfax

THÈSE

Présentée

en vue de l'obtention du

DOCTORAT
(En Cotutelle)

Spécialité : Traitement du signal et de l'image et Mathématiques Appliquées
par

Atizez Hadrich Ben Arab

**ÉTUDE DES FONCTIONS B-SPLINES POUR LA FUSION
D'IMAGES SEGMENTÉES PAR APPROCHE BAYÉSIENNE**

Soutenue le 02 décembre 2015, devant le jury composé de :

Président : M. Slim Chaabane, Professeur de la Faculté des Sciences de Sfax, Tunisie

Rapporteurs : M. Stéphane Derrode, Professeur à l'École Centrale de Lyon, France
M. Hassine Maatoug, Maître de Conférences HDR à la Faculté des Sciences de Monastir, Tunisie

Examinateur : M. Yassine Ruichek, Professeur de l'Université de Technologie de Belfort Montbéliard, France

Directeurs : M. Afif Masmoudi, Professeur à la Faculté des Sciences de Sfax, Tunisie
M. Mourad Zribi, Maître de Conférences HDR à l'Université du Littoral Côte d'Opale, Calais, France

Invité : M. Mahdi Louati, Maître Assistant à l'École Nationale d'Électronique et des Télécommunications de Sfax,
Tunisie



DÉDICACES

*A mon cher père **Mustapha** et ma chère mère **Naima**, en leur disant que je vous dois tout ce que j'ai et j'aurai. Mes parents, nulle expression ne pourrait exprimer ce que je vous dois pour vos efforts déployés et sacrifices pour ma formation. Que dieu leur assure bonne santé et longue vie.*

*A mon très chère époux **Taher** et mes adorables fils **Mohamed** et **Mahdi**. En témoignage de leur soutien, leur amour, leur compréhension, leur encouragement et leurs aides infinies. Que Dieu vous garde pour moi et m'accorde santé et longue vie et m'aide à vous rendre heureux autant que vous méritez.*

*A mes chers beaux parents **Mohamed** et **Najwa** En témoignage de ma grande affection et mes attachements, je vous présente l'expression de ma gratitude et mon grand amour. Que dieu vous procure bonne santé et longue vie.*

*A mon frère **Fakreddine** et sa femme **Olfa** et leurs adorables enfants **Aimna**, **Mohamed Aimin** et **Ahmed**.*

*A mon frère **Mohamed Fouazi** et sa femme **Maroi** et leurs adorable enfants **Yassine** et **Yesmine**.*

*A mon frère **Sami** et son femme **Siwar** et leurs adorables filles **Mariam** et **Maram**.*

*A ma sœur **Inès** et son mari **Houssem** et leurs adorables enfants **Mariam**, **Mohamed** et **Syrine**.*

Que Dieu vous préserve la bonne santé, heureuse vie et un avenir professionnel très brillant. A tous les membres de ma famille sans oublier personne, je dis que je suis fier de vous, espérant que vous êtes fiers de moi.

*A mes beaux frères **Bilal**, **Haïkal** et **Morsi**, Que dieu vous procure une vie pleine de bonheur et de prospérité. Que Dieu vous préserve la bonne santé, heureuse vie et un avenir professionnel très brillant.*

REMERCIEMENTS

Ce travail de recherche est le résultat d'une convention en cotutelle entre la Faculté des Sciences de Sfax (FSS) Tunisie et l'Université du Littoral Côte d'Opale (ULCO) France. Ce travail a été réalisé au sein du Laboratoire de Probabilités et Statistique (LPS-FSS) du coté Tunisien et au sein du Laboratoire d'Informatique Signal et Image de la Côte d'Opale (LISIC-ULCO) du coté Français.

*J'adresse mes plus vifs remerciements à Monsieur **Affif Masmoudi**, Professeur à la faculté des Sciences de Sfax, ainsi qu'à Monsieur **Mourad Zribi**, Maître de conférences HDR à l'Université du Littoral Côte d'Opale pour avoir assuré la direction scientifique de ce travail de thèse, pour les nombreux et fructueuses discussions, ainsi que ses encouragements qui m'ont été d'une grande aide.*

*J'adresse ensuite mes plus sincères remerciements à Monsieur **Slim Chaabane**, Professeur à la faculté des Sciences de Sfax d'avoir accepté de présider le jury de cette thèse.*

*Je tiens à remercier Monsieur **Stéphane Derrode**, Professeur à l'École Centrale de Lyon, ainsi qu'à Monsieur **Hassine Maatoug**, Maître de conférences HDR à la Faculté des Sciences de Monastir pour avoir accepté la lourde charge de rapporteur, ainsi que pour les remarques constructives qu'ils m'ont fournies.*

*Je remercie Monsieur **Yassine Ruichek**, Professeur de l'Université de Technologie de Belfort Montbéliard, d'avoir accepté de juger ce travail en tant qu'examinateur.*

*Mes vifs remerciements s'adressent également à Monsieur **Mahdi Louati** Maître Assistant à l'École Nationale d'Electronique et des Télécommunications de Sfax, pour m'avoir fait l'honneur de participer à mon jury de thèse.*

Mes sincères remerciements s'adressent également à tous les membres du laboratoire de Probabilités et Statistique de la faculté des Sciences de Sfax, Tunisie, et du laboratoire d'Informatique Signal et Image de la Côte d'Opale de l'Université du Littoral Côte d'Opale, France.

A tous les enseignants qui ont contribué à l'enrichissement de ma formation, qu'ils acceptent mes remerciements profonds et qu'ils trouvent dans cet écrit le fruit de leur engagement et leur travail persévérant.

Trouvez ici l'expression de ma reconnaissance et mes remerciements les plus profonds.

Table des matières

Introduction	1
1 Estimation non paramétrique par les fonctions B-splines	5
1.1 Introduction	5
1.2 Rappel de quelques méthodes d'estimation non paramétrique	6
1.2.1 Méthode d'histogramme	7
1.2.2 Méthode de noyau	8
1.2.3 Méthode des fonctions orthogonales	9
1.3 Proposition d'un nouvel estimateur non paramétrique	11
1.3.1 Les fonctions B-Splines	12
1.3.2 Estimation par les fonctions B-splines	14
1.3.3 Description de l'algorithme proposé	15
1.3.4 Performance de l'estimateur proposé	18
1.4 Généralisation de l'estimateur proposé	24
1.4.1 Présentation et interprétation	24
1.4.2 Estimation du mélange de lois par les fonctions B-splines	25
1.4.3 Description de l'algorithme EMGB	26

1.4.4	Existence, Unicité et Propriétés de l'algorithme EMGB	29
1.4.5	Performance de l'estimateur du mélange proposé	36
1.5	Apport de l'approche Bayésienne en estimation non paramétrique	43
1.5.1	Estimation Bayésienne par les fonctions B-Splines d'une seule loi	45
1.5.2	Estimation Bayésienne par les fonctions B-Splines d'un mélange de lois	53
1.5.3	Existence, Unicité et Propriétés de l'estimateur Bayésien pour le mélange fini de lois	56
1.6	Conclusion	65
2	Segmentation statistique non paramétrique des images	67
2.1	Introduction	67
2.2	Modélisation probabiliste en classification	69
2.2.1	Différentes types de classifications	70
2.2.2	Modèle aléatoire de l'image	71
2.3	Modèle Markovien en analyse des images	76
2.3.1	Description de l'image	76
2.3.2	Champs aléatoire Markovien MRF	77
2.3.3	Distribution de Gibbs	78
2.3.4	Théorème de Hammersley-Clifford	79
2.4	Modèles de Markov cachés dans un cadre non paramétrique	82
2.4.1	Rappel sur les champs de Markov cachés	83
2.4.2	Estimation non paramétrique	85

2.5	Application à la segmentation des images	91
2.5.1	Segmentation non paramétrique par approche aveugle	91
2.5.2	Segmentation non paramétrique dans un cadre Markovien	97
2.6	Conclusion	110
3	Fusion d'images segmentées par l'approche Bayésienne	111
3.1	Introduction	111
3.2	Fusion en traitement des images	112
3.2.1	Systèmes de fusion	113
3.2.2	Principe de fusion	115
3.3	Fusion par l'approche Bayésienne	116
3.3.1	Modélisation	117
3.3.2	Estimation	118
3.3.3	Combinaison	119
3.3.4	Décision	121
3.4	Fusion Bayésienne dans un cadre non paramétrique	122
3.4.1	Principe de la fusion d'images segmentées	122
3.4.2	Estimation non paramétrique	126
3.5	Analyse quantitative	131
3.5.1	Critères objectifs non pondérés	131
3.5.2	Résultats expérimentaux	133
3.5.3	Fusion d'images segmentées	138
3.6	Conclusion	144

Conclusion et perspectives	146
Annexe A	148
3.6.1 Mélange de lois	148
Annexe B	153
3.6.2 L'échantillonneur de Gibbs	153
3.6.3 L'algorithme de Métropolis	154
3.6.4 Quelques MRF fondamentaux	156
Bibliographie	158

Introduction

Ce travail s'inscrit dans une thématique de recherche commune au laboratoire de Probabilités et Statistique de l'université de Sfax (Tunisie) et au laboratoire d'Informatique, Signal et Image de l'université du Littoral Côte d'Opale (France). Il présente une synthèse des travaux de recherche que nous avons réalisés dans les domaines de probabilités et statistique et de traitement de signal et d'images.

Le traitement d'images voit ses débuts dans les années 1920 dans la transmission de données [51] par câble mais ne connaît de vrai essor que dans les années 1960 avec le développement des ordinateurs. Au départ, les techniques de traitement d'images [22] sont essentiellement des méthodes de restauration et de compression d'images. Ensuite ces techniques ont été développées avec les progrès de l'informatique, des techniques de détection de primitives (contours, point d'intérêt, lignes d'intérêt, . . .) et de nombreux autres traitements dans les domaines aussi variés que le médical, la télévision, l'imagerie satellitaire, le multimédia. C'est dans les années 2000 que l'image numérique et par conséquent le traitement d'images devient omni-présent. Que cela soit sur internet, au cinéma, à la télévision, sur les téléphones, dans le domaine médical, l'image est partout. Aujourd'hui il ne s'agit plus uniquement de traiter les images pour les améliorer mais aussi de les comprendre et

de les interpréter.

Le présent travail consiste à étudier les fonctions B-splines en traitement d'images. Ces fonctions ont en effet reçu une attention soutenue d'un point de vue théorique dans les communautés des probabilités et des statistiques, mais elles sont peu étudiées en traitement d'images. Nous cherchons donc ici à évaluer l'apport de ces fonctions B-splines en fusion d'images segmentées dans un cadre probabiliste.

L'approche statistique la plus répandue en segmentation d'images repose sur les modèles mélanges. Généralement les images sont complexes et elles ne peuvent pas être modélisées par une loi classique. D'où le modèle de mélange [41] est un moyen intéressant pour modéliser une image. L'application du modèle mélange en segmentation introduit une étape d'estimation des lois. Dans la littérature plusieurs algorithmes ont été proposés pour estimer le mélange de lois. Nous pouvons citer l'algorithme EM (Estimation, Maximisation) [37].

L'algorithme EM découle naturellement des équations de vraisemblance et possède de bonnes propriétés statistiques en plus de sa simplicité de mise en oeuvre et de son efficacité. Cependant, cet algorithme possède un certain nombre de limitations. D'abord, la valeur de l'estimateur à la convergence peut être fortement dépendante de son initialisation. Une autre limitation de l'algorithme EM est que la convergence peut être lente et l'algorithme peut même se trouver bloqué dans un point selle de la vraisemblance. Ces limitations ont donné lieu à des développements complémentaires. Pour pallier la limitation de la dépendance de la solution vis à vis de l'initialisation, il est courant dans la pratique de lancer plusieurs EM pour quelques itérations depuis des initialisations aléatoires et de choisir la valeur des paramètres associés à la plus grande vraisemblance comme initialisa-

tion d'un autre EM qui lui itérera jusqu'à la convergence. Parmi les variantes existantes de cet algorithme nous citons l'algorithme GEM (Generalized EM), proposé par Dempster et al [37], qui permet de simplifier le problème de l'étape maximisation, l'algorithme CEM (Classification EM), proposé par Celeux et Govaert [26], permettant de prendre en compte l'aspect classification lors de l'estimation, ainsi que l'algorithme SEM (Stochastic EM), proposé par Celeux et Diebolt [25], dont l'objectif est de réduire le risque de tomber dans un optimum local de vraisemblance. Kumar et Koller [76] ont proposé une stratégie pour obtenir une bonne initialisation de l'algorithme EM.

Dans notre travail nous nous sommes intéressés aux méthodes non paramétriques pour l'estimation des densités de probabilités et plus particulièrement la méthode qui utilise les fonctions splines [66].

Ce mémoire est composé de trois chapitres. Le chapitre 1 est consacré au problème de l'estimation non paramétrique des lois de probabilités par les fonctions B-splines [62][63][65][66].

Dans un premier temps, nous rappelons quelques méthodes d'estimation non paramétrique [27][42][43]. Dans un deuxième temps, d'abord nous proposons un nouvel estimateur non paramétrique basé sur les fonctions B-splines [62]. Puis, nous généralisons l'estimateur proposé pour le mélange de lois par les fonctions B-splines [65]. Ensuite, nous améliorons notre résultat en utilisant l'estimation Bayésienne par les fonctions B-Splines [66]. Enfin, nous généralisons l'estimation Bayésienne par les fonctions B-Splines pour le cas d'un mélange de lois [66]. Des simulations illustrent la performance des estimateurs proposés.

Le chapitre 2 traite la segmentation statistique non paramétrique des images. D'abord, nous présentons la modélisation probabiliste en classification. Ensuite, nous décrivons les modèles de Markov cachés dans un cadre non paramétrique [64]. Enfin, nous appliquons

les estimateurs proposés à la segmentation non supervisée des images [64].

Le chapitre 3 traite le problème de fusion d'images segmentées par l'approche Bayésienne.

D'abord, nous rappelons la fusion par l'approche Bayésienne. Ensuite, nous décrivons la fusion Bayésienne dans un cadre non paramétrique. Enfin, nous illustrons ce chapitre par des résultats expérimentaux.

La conclusion générale fera l'objet d'un bilan critique des diverses approches proposées pour l'estimation des lois, la segmentation et la fusion des images et suggère quelques perspectives de recherche.

Chapitre 1

Estimation non paramétrique par les fonctions B-splines

1.1 Introduction

Lorsque l'on cherche à étudier une suite de mesures provenant de la répétition d'une expérience, une méthode de modélisation consiste à supposer que ces mesures sont des réalisations indépendantes d'une variable aléatoire X . Comprendre ces mesures et la façon dont elles sont distribuées revient à étudier la loi de probabilité de la variable aléatoire X . Lorsque l'on n'a pas d'idée a priori sur la forme particulière que peut prendre la densité f de X , construire un estimateur de f ne se résume pas à l'estimation d'une moyenne et d'une variance, comme c'est le cas pour des lois Gaussiennes. Il s'agit de reconstruire une fonction densité. Le problème est alors dit non-paramétrique. La particularité de la statistique non-paramétrique [20][27][41] est que le paramètre inconnu qu'on cherche à estimer ou à classifier est supposé ne pas appartenir à une famille indexée par un petit nombre de

paramètres réels. En général, dans la théorie non-paramétrique on suppose que le nombre des paramètres qui décrivent la loi des observations est une fonction croissante du nombre d'observations.

Notre étude réside dans l'adoption de la méthode des fonctions splines pour l'estimation des densités de probabilités et en particulier les fonctions B-splines quadratiques. La structure flexible des splines et leurs dérivées faciles à manipuler favorisent l'imposition de contraintes sur le modèle désiré. Nous montrons dans ce chapitre la possibilité de mettre en place un algorithme itératif dans le but d'estimer les paramètres du mélange de lois en utilisant les fonctions B-splines quadratiques. D'abord, nous rappelons quelques méthodes d'estimation non paramétrique [20][27][41][48][68]. Les plus connues sont l'estimation par la méthode du noyau introduite par Rosenblatt [99], l'estimation par la méthode de l'histogramme [40], l'estimation par la méthode des K-plus proches voisins [70] et l'estimation par la méthode des fonctions orthogonales [68][69]. Puis, nous introduisons un nouvel estimateur de la densité de probabilités en se basant sur les fonctions B-splines quadratiques [62][63]. Ensuite, nous montrons la généralisation de ce nouvel estimateur pour le cas de mélange de lois [65]. Enfin, nous décrivons l'apport de l'approche Bayésienne à ce nouvel estimateur [66].

1.2 Rappel de quelques méthodes d'estimation non paramétrique

Dans l'inférence statistique classique, on suppose souvent que la forme fonctionnelle de la densité de probabilités d'une variable aléatoire est connue, et on s'intéresse à l'estimation

des paramètres inconnus. Lorsque la forme fonctionnelle de la densité est inconnue, on peut soit considérer la sélection parmi plusieurs alternatives possibles, soit envisager des méthodes non-paramétriques d'estimation et de sélection des densités. Le problème de l'estimation se pose de la façon suivante.

On dispose de n réalisations (x_1, \dots, x_n) d'un échantillon aléatoire (X_1, \dots, X_n) de variables aléatoires indépendantes et de même loi de densité de probabilité inconnue f . Notre objectif consiste à estimer cette densité de probabilités à partir de l'échantillon aléatoire (X_1, \dots, X_n) . Plusieurs méthodes ont été développées pour estimer cette densité de probabilités [20][27][41] [48][68]. Dans la suite de cette section, nous citons quelques méthodes d'estimation non paramétrique en précisant la condition nécessaire et suffisante pour la convergence en moyenne quadratique [52] pour chaque méthode.

1.2.1 Méthode d'histogramme

Dans sa plus grande généralité, un histogramme [40][43] se définit à partir d'une suite double de valeurs croissantes $\{\dots, a_{-i}, \dots, a_{-1}, a_0, a_1, \dots, a_i, \dots\}$ constituant un découpage en intervalles de la droite réelle. Soit n_k la fréquence des observations situées dans l'intervalle $]a_k, a_{k+1}]$ pour un échantillon de taille n , alors l'histogramme est la fonction constante par morceaux \hat{f}_n telle que, pour tout $k \in \mathbb{Z}$:

$$\hat{f}_n(x) = \frac{\frac{n_k}{n}}{(a_{k+1} - a_k)} \text{ pour } x \in]a_k, a_{k+1}],$$

Dans le cadre d'un n -échantillon aléatoire de densité f définie sur \mathbb{R} , l'estimateur \hat{f}_n s'écrit sous la forme suivante :

$$\hat{f}_n(x) = \frac{N_h}{nh_n} = \frac{1}{nh_n} \sum_{k=1}^n \mathbf{1}_{]a_k, a_{k+1}]}(x),$$

où h_n est la largeur de chaque intervalle et N_h est le nombre des observations appartenant à $]a_k, a_{k+1}]$, x étant fixé dans cet intervalle.

La condition nécessaire et suffisante pour que la densité estimée \hat{f} converge en moyenne quadratique vers f , lorsque le pas h_n converge vers 0 et nh_n tend vers l'infini quand n tend vers l'infini [24].

L'histogramme convient bien pour des analyses relativement grossières [40][43]. Néanmoins, ses discontinuités (c'est une fonction en escalier) n'apparaissent pas très naturelles.

1.2.2 Méthode de noyau

Cette méthode a été introduite par Rosenblatt en 1956 [99] puis repris par Parzen en 1962 [92]. L'idée principale consiste à estimer la densité en un point x en comptant le nombre d'observations situées dans un certain voisinage de x ; sur \mathbb{R} on peut choisir un voisinage de la forme $]x - \frac{h_n}{2}, x + \frac{h_n}{2}]$, où h_n est un réel strictement positif, ce qui conduit à l'estimateur :

$$\hat{f}_n(x) = \frac{F_n(x+h) - F_n(x-h)}{2h_n}$$

où F_n désigne la fonction de répartition empirique associée à un échantillon de taille n .

Cette expression peut encore s'écrire :

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right),$$

où (X_1, \dots, X_n) est un n-échantillon aléatoire et $K = \mathbf{1}_{[-1/2, 1/2[}$ est appelé noyau. La densité estimée \hat{f}_n converge en moyenne quadratique vers f [70], si et seulement si le pas h_n est choisi de telle manière à ce que h_n tend vers zéro et nh_n tend vers l'infini lorsque n tend vers l'infini.

Exemples

$K(u) = \frac{1}{2}\mathbf{1}_{(|u|\leq 1)}$ (noyau rectangulaire).

$K(u) = (1 - |u|)\mathbf{1}_{(|u|\leq 1)}$ (noyau triangulaire).

$K(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)$ (noyau Gaussien).

La méthode de noyau se révèle inefficace quand la densité a un support confondu à \mathbb{R}_+ , notamment dans le cas d'irrégularité de la densité de probabilité à l'origine. De plus, l'estimateur par la méthode de noyau [99] est difficile à calculer car il met en oeuvre une réestimation itérée à chaque mise à jour de l'échantillon. Dans ce cas, la méthode des fonctions orthogonales offre des avantages par sa commodité de calcul, notamment pour de grandes tailles d'échantillons. Cette méthode d'estimation utilise des familles de fonctions orthogonales différentes lorsque la nature du support de la densité n'est pas le même.

1.2.3 Méthode des fonctions orthogonales

La littérature statistique regorge des travaux sur l'estimation de la densité utilisant les séries orthogonales [4][42][102][108]. La théorie des fonctions orthogonales utilisées pour l'estimation des densités de probabilités a l'intérêt d'être adaptée aux différents types d'espaces de représentation. Cette méthode d'estimation, utilise des systèmes de fonctions orthogonales différents lorsque la nature du support de la densité n'est pas la même. Aussi, le calcul de l'estimateur de la densité peut être facilité par l'utilisation de la formule récursive qui existe entre les coefficients. On suppose que la fonction $f(x)$ est une densité strictement positive (relativement à la mesure de Lebesgue) sur l'intervalle compact $[a, b]$ et $\{e_m(x)\}_{m \in \mathbb{N}}$ est une base orthonormale sur l'espace hilbertien $L^2([a, b])$, c'est à dire :

$\{e_m(x)\}_{m \in \mathbb{N}}$ est une base orthonormée sur l'espace hilbertien $L^2([a, b])$,

$$\int_a^b e_m(x)e_n(x)dx = \delta_{m,n}, 0 \leq m, n < \infty$$

où $\delta_{m,n}$ est le symbole de Kronecker défini par :

$$\delta_{m,n} = \begin{cases} 1, & \text{si } m = n, \\ 0, & \text{sinon.} \end{cases}$$

Dans ce cas, la densité $f(x)$ peut être représentée sous forme d'une série orthogonale donnée par :

$$f(x) = \sum_{m=0}^{\infty} a_m e_m(x), a \leq x \leq b.$$

Par commodité, on suppose habituellement que $\sum_{m=0}^{\infty} a_m e_m(x)$ converge uniformément sur $[a, b]$, chaque fonction est continue, et est une fonction constante.

Soient X_1, \dots, X_n , n variables aléatoires indépendantes et identiquement distribuées de même densité f définie sur $[a, b]$.

Les coefficients de Fourier a_m sont définis par :

$$a_m = \int_a^b e_m(x)f(x)dx, m = 0, 1, 2, \dots$$

on peut estimer ces derniers par

$$\begin{aligned} \hat{a}_{m,N} &= \frac{1}{N} \sum_{j=1}^N e_m(X_j), m = 0, 1, 2, \dots \\ &= \frac{1}{N} [(N-1)\hat{a}_{m,N-1} + e_m(X_N)] \end{aligned}$$

$\hat{a}_{m,N}$ est un estimateur sans biais de a_m . Donc on peut définir l'estimateur de la fonction f par l'expression suivante :

$$\hat{f}_{K_N}(x) = \sum_{m=0}^{K_N} \hat{a}_{m,N} e_m(x).$$

Puisque $e_0(x)$ est constant, $\widehat{a}_{0,N} = a_0$ est connu. L'entier K_N est appelé point de troncature. La convergence en moyenne quadratique est obtenue par le choix d'un K_N (sélectionner une valeur de K_N qui minimise la moyenne quadratique) [115].

Cependant, cette approche présente quelques inconvénients, puisque d'une part, le modèle mathématique à l'origine tend à ignorer le fait que la fonction à estimer est une densité. En effet, l'approximation fonctionnelle par une somme partielle d'une série de fonctions orthogonales est parfois à valeurs négatives. D'autre part, cette méthode a été développée au prix d'un biais aux bords dû au phénomène de Gibbs [68]. Pour résoudre ce problème, nous proposons un nouvel estimateur basé sur les fonctions B-Splines quadratiques [62][63][65].

1.3 Proposition d'un nouvel estimateur non paramétrique

Dans cette section, afin d'améliorer la qualité de l'estimation des densités de probabilités, nous proposons les B-splines en tant qu'estimateurs des densités de probabilités. Dans Daub et al. [36], les auteurs ont comparé leur méthode avec d'autres techniques d'estimation de probabilités. Ils ont montré, sur des données artificiellement générées, que l'estimation par les fonctions B-splines offre des résultats en moyenne deux fois meilleurs qu'avec des histogrammes. Ils ont conclu d'utiliser des degrés de splines trop élevés ne donnent pas forcément de meilleurs résultats qu'un ordre $k = 2$ ou $k = 3$.

Dans nos travaux de recherche, nous nous sommes intéressés plus particulièrement par les fonctions B-splines quadratiques (l'ordre $k = 2$).

1.3.1 Les fonctions B-Splines

Les fonctions B-splines font partie des splines de régression [24]. Le terme B-spline introduit en 1966 par Curry et Schoenberg [34] qui signifie tout simplement une base spline. En effet, Curry et Schoenberg ont montré que tous les splines de degré d ($d = 1, 2, \dots$) avec m noeuds ($m = 1, 2, \dots$) ont une extension unique,

$$S(x) = \sum_{l=1}^{m+d} \frac{b_l}{h_l} B_l^d(x), \text{ avec } a < x < b \text{ et } h_l = \int_{-\infty}^{+\infty} B_l^d(x) dx,$$

où $a, b \in \mathbb{R}$ et b_l sont des paramètres inconnus qui ont besoin d'être estimés. Notons que, $b_l \geq 0$, et $\sum_{l=1}^{m+d} b_l = 1$ est une exigence particulière. En général, cette exigence particulière n'est pas vraie pour les fonctions splines. Le cas $d = 1$, correspond à une approximation linéaire par morceaux. Cette approximation ne présente pas une bonne allure. Pour $d = 2$ et 3, on obtient des fonctions B-splines quadratiques et cubiques respectivement.

Les fonctions B-splines de degré d sont définies par

$$B_l^d(x) = \frac{x - x_l}{x_{l+d} - x_l} B_l^{d-1}(x) + \frac{x_{l+d+1} - x}{x_{l+d+1} - x_{l+1}} B_{l+1}^{d-1}(x)$$

où

$$B_l^0(x) = \begin{cases} 1, & \text{si } x \in [x_l, x_{l+1}) \\ 0, & \text{sinon.} \end{cases}$$

En termes plus techniques, une fonction B-spline [10][88][105] de degré d avec m noeuds, $a = x_0 < x_1 < \dots < x_{m+1} = b$ est une fonction $S(x)$ de classe C^{d-1} sur (a, b) de telle sorte que $S(x)$ est une fonction de polynômes de degré au plus d dans chacun des intervalles $(a, x_1), (x_1, x_2), \dots, (x_m, b)$. C'est-à-dire $S(x)$ une fonction $(d - 1)$ dérivable et sa dérivée d'ordre $(d - 1)$ est continue sur (a, b) .

Dans la littérature plusieurs travaux ont traité le cas des B-splines cubiques ($d = 3$) dans

divers applications [6][35][58]. Ces derniers sont des fonctions de classes $C^2(a, b)$. En pratique, cette condition pose beaucoup de problèmes au niveau de la manipulation. Pour éviter ce problème, nous proposons dans notre travail d'utiliser les B-splines quadratiques qui sont des fonctions de classe $C^1(a, b)$.

Une fonction B-spline quadratique notée $(B_l^2)_{l=1\dots m+2}$ est définie par :

$$B_l^2(x) = \begin{cases} \frac{(x - x_l)^2}{(x_{l+1} - x_l)(x_{l+2} - x_l)}, & x \in [x_l, x_{l+1}) \\ \frac{1}{x_{l+2} - x_{l+1}} \left[\frac{(x - x_l)(x_{l+2} - x)}{x_{l+2} - x_l} + \frac{(x - x_{l+1})(x_{l+3} - x)}{x_{l+3} - x_{l+1}} \right], & x \in [x_{l+1}, x_{l+2}) \\ \frac{(x_{l+3} - x)^2}{(x_{l+3} - x_{l+1})(x_{l+3} - x_{l+2})}, & x \in [x_{l+2}, x_{l+3}) \\ 0, & \text{sinon.} \end{cases}$$

Cette fonction vérifie les conditions suivantes :

$$0 \leq B_l^2(x) \leq 1 \text{ et } \sum_{l=1}^{m+2} B_l^2(x) = 1, \text{ pour tout } x \in \mathbb{R}.$$

Soit $x \in [x_{l+1}, x_{l+2})$, alors

$$\begin{aligned} \sum_{l=1}^{m+2} B_l^2(x) &= B_1^2(x) + \dots + B_{l-1}^2(x) + B_l^2(x) + B_{l+1}^2(x) + \dots + B_{m+2}^2(x) \\ &= B_{l-1}^2(x) + B_l^2(x) + B_{l+1}^2(x) \\ &= \frac{(x_{l+2} - x)^2}{(x_{l+2} - x_l)(x_{l+2} - x_{l+1})} + \frac{1}{x_{l+2} - x_{l+1}} \left[\frac{(x - x_l)(x_{l+2} - x)}{x_{l+2} - x_l} + \frac{(x - x_{l+1})(x_{l+3} - x)}{x_{l+3} - x_{l+1}} \right] + \\ &\quad \frac{(x - x_{l+1})^2}{(x_{l+2} - x_{l+1})(x_{l+3} - x_{l+1})} \\ &= \left[\frac{(x_{l+2} - x)^2}{(x_{l+2} - x_l)(x_{l+2} - x_{l+1})} + \frac{(x - x_l)(x_{l+2} - x)}{(x_{l+2} - x_l)(x_{l+2} - x_{l+1})} \right] + \left[\frac{(x - x_{l+1})(x_{l+3} - x)}{(x_{l+3} - x_{l+1})(x_{l+2} - x_{l+1})} + \right. \\ &\quad \left. \frac{(x - x_{l+1})^2}{(x_{l+2} - x_{l+1})(x_{l+3} - x_{l+1})} \right] \\ &= \frac{(x - x_{l+1})}{(x_{l+2} - x_{l+1})(x_{l+3} - x_{l+1})} [x - x_{l+1} + x_{l+3} - x] + \\ &\quad \frac{(x_{l+2} - x)}{(x_{l+2} - x_l)(x_{l+2} - x_{l+1})} [x_{l+2} - x + x - x_l] \\ &= \frac{(x_{l+2} - x)}{(x_{l+2} - x_{l+1})} + \frac{(x - x_{l+1})}{(x_{l+2} - x_{l+1})} \\ &= 1 \end{aligned}$$

Le support de chaque spline couvre trois intervalles dont chacun est constitué d'un polynôme quadratique. Nous présentons dans la figure FIG. 1.1. les parties non-nuls d'une fonction B-spline de degré 2. D'après cette figure, nous remarquons que les valeurs des sommets de cette fonction sont inférieures à 1.

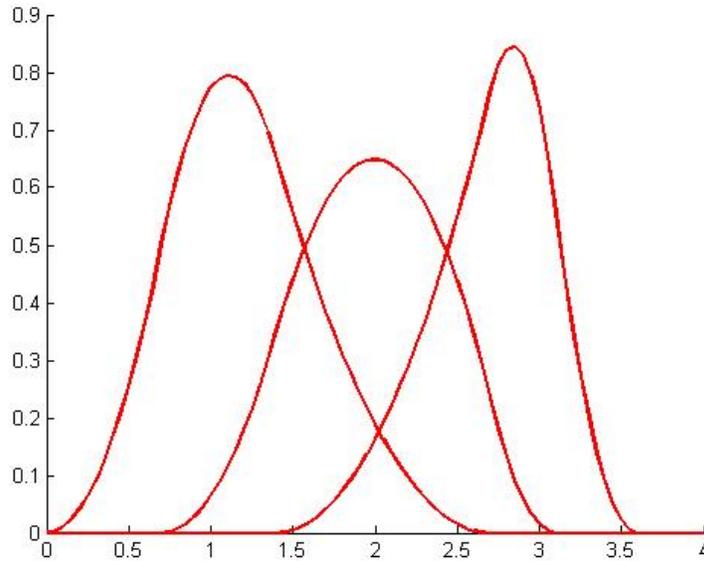


FIG. 1.1 – Représentation d'une fonction B-spline d'ordre 2.

1.3.2 Estimation par les fonctions B-splines

La théorie des fonctions B-splines utilisée pour l'estimation des densités de probabilités a l'intérêt d'être adaptée aux différents types d'espaces de représentation. Le but de ce paragraphe est de montrer comment nous pouvons estimer une densité de probabilités f en utilisant les fonctions B-splines quadratiques. Soit n réalisations (x_1, \dots, x_n) d'un échantillon (X_1, \dots, X_n) de variables aléatoires indépendantes et de même loi de densité de probabilités inconnue f . En utilisant les fonctions B-splines quadratiques, nous pouvons exprimer la densité f sous la forme d'un mélange de B-splines donnée par la relation

suivante :

$$f(x) = \sum_{l=1}^{m+2} b_l B_{l,h_l}^2(x), \tag{1.1}$$

avec $b_l \geq 0$, $\sum_{l=1}^{m+2} b_l = 1$, $B_{l,h_l}^2(x) = \frac{B_l^2}{h_l}$ et $h_l = \int_{-\infty}^{+\infty} B_{l,h_l}^2(x) dx = \int_{x_l}^{x_{l+3}} B_l^2(x) dx$.

D'après l'équation (1.1), nous remarquons que l'estimation de la densité f se réduit à l'estimation du vecteur des paramètres composé par les coefficients des B-spline $b = (b_1, b_2, \dots, b_l)$. L'estimation par la méthode du maximum de vraisemblance de ces paramètres s'obtiendra à partir d'un algorithme itératif basé sur l'utilisation de l'algorithme EM [39] décrit dans le paragraphe suivant.

1.3.3 Description de l'algorithme proposé

D'après l'expression de la fonction densité f donnée par l'équation (1.1), nous remarquons que cette dernière s'écrit sous la forme d'un mélange des fonctions B-splines de $m+2$ termes [105] où les b_l jouent le rôle des probabilités à priori. L'estimation de f revient à estimer ces coefficients b_l . Pour estimer ces coefficients nous avons utilisé le principe du maximum de vraisemblance. Toutefois, dans le cas des problèmes complexes, un estimateur du maximum de vraisemblance peut devenir non explicite [11][24]. Il faut avoir recours à des méthodes itératives comme l'algorithme Espérance-Maximisation (EM) [39]. En se basant sur cet algorithme pour estimer ces coefficients, nous introduisons notre nouvel algorithme appelé l'algorithme EMB (Espérance Maximisation B-spline) [65].

Soit un échantillon X_1, \dots, X_n issu d'une loi de densité de probabilités f donnée par l'équation (1.1) et soit (x_1, \dots, x_n) n réalisations associées à cet échantillon aléatoire. On

définit la fonction de vraisemblance par :

$$L(X_1, \dots, X_n | b) = \prod_{i=1}^n f(X_i)$$

et la fonction du log-vraisemblance par

$$l(X_1, \dots, X_n | b) = \log L(X_1, \dots, X_n | b) = \sum_{i=1}^n \log f(X_i) = \sum_{i=1}^n \log \left(\sum_{l=1}^{m+2} b_l B_{l,h_l}^2(X_i) \right).$$

L'algorithme EMB consiste à résoudre de manière itérative les équations de vraisemblance.

Cet algorithme comporte trois étapes : une étape d'initialisation, une étape d'espérance

"E" pour le calcul des probabilités a posteriori et une étape "MB" consiste à une maximisation de la vraisemblance.

Algorithme EMB

Étape d'initialisation : Les paramètres peuvent être initialisés comme suit

$$b_l^{(0)} = \frac{\text{nombre d'observations } X_j \in [x_l, x_{l+3})}{n}.$$

Pour chaque X_i , nous considérons une variable aléatoire $Z_i = (Z_{1,i}, \dots, Z_{m+2,i})$ de telle sorte que Z_i suit une distribution multinomiale $Mul(1, b_1, \dots, b_{m+2})$ (i-e; $\mathbb{P}(Z_{l,i} = 1) = b_l$).

Etant donnée $X = (X_1, \dots, X_n)$ et $Z = (Z_1, \dots, Z_n)$, à l'itération q , la fonction du log-vraisemblance s'écrit sous la forme

$$l(X, Z | b^{(q)}) = \sum_{i=1}^n \left[\sum_{l=1}^{m+2} Z_{l,i} \log(b_l^{(q)} B_{l,h_l}^2(X_i)) \right].$$

Soit $\tau_{l,i}^{(q)} = \mathbb{E}(Z_{l,i} | X, b^{(q)})$, alors

$$\begin{aligned} Q(b, b^{(q)}) &= \mathbb{E}(l(X, Z | b) | b^{(q)}) \\ &= \sum_{i=1}^n \sum_{l=1}^{m+2} \mathbb{E}(Z_{l,i} | X, b^{(q)}) \log(b_l B_{l,h_l}^2(X_i)) \\ &= \sum_{i=1}^n \left[\sum_{l=2}^{m+2} \tau_{l,i}^{(q)} \log(b_l B_{l,h_l}^2(X_i)) + \tau_{1,i}^{(q)} \log \left(\left(1 - \sum_{l=2}^{m+2} b_l \right) B_{1,h_1}^2(X_i) \right) \right]. \end{aligned}$$

Étape d'espérance (E) : À l'itération (q) , nous calculons la quantité intermédiaire $Q(b, b^{(q)})$ et la probabilités a postériori $\tau_{l,i}^{(q)}$ est estimée par

$$\tau_{l,i}^{(q)} = \frac{b_l^{(q)} B_{l,h_l}^2(X_i)}{\sum_{l=1}^{m+2} b_l^{(q)} B_{l,h_l}^2(X_i)}, l = 1, \dots, m + 2, i = 1, \dots, n.$$

Étape de maximisation (MB) : À l'itération $(q + 1)$, nous calculons les coefficients du B-splines $b^{(q+1)}$ qui maximise la fonction Q :

$$b^{(q+1)} = \arg \max_b Q(b, b^{(q)}).$$

Comme $b = (b_1, \dots, b_{m+2}) \mapsto Q(b, b^{(q)})$ est une fonction concave alors le maximum de Q est atteint en $b_l^{(q+1)}$ lorsque

$$\frac{\partial Q}{\partial b_l} = \sum_{i=1}^n \left(\tau_{l,i}^{(q)} \frac{1}{b_l} - \tau_{1,i}^{(q)} \frac{1}{b_1} \right) = 0. l = 2, \dots, m + 2.$$

Cela signifie que

$$\frac{1}{b_l} \left[\sum_{i=1}^n \left(\tau_{l,i}^{(q)} \right) \right] = \frac{1}{b_1} \left[\sum_{i=1}^n \left(\tau_{1,i}^{(q)} \right) \right],$$

ce qui implique

$$b_l = b_1 \frac{\sum_{i=1}^n \tau_{l,i}^{(q)}}{\sum_{i=1}^n \tau_{1,i}^{(q)}}; \text{ avec } l = 2, \dots, m + 2.$$

Comme $\sum_{l=1}^{m+2} b_l = 1$, nous obtenons à l'itération $(q + 1)$:

$$b_l^{(q+1)} = \frac{1}{n} \sum_{i=1}^n \tau_{l,i}^{(q)} \quad l = 1, \dots, m + 2,$$

avec $b^{(q+1)} = (b_1^{(q+1)}, b_2^{(q+1)}, \dots, b_{m+2}^{(q+1)})$.

Arrêt de l'algorithme

En pratique, nous arrêtons cet algorithme au bout d'un certain nombre d'itérations quand les coefficients b_l ne varient presque plus, c'est à dire pour un nombre ϵ petit :

$$|b_l^{(q+1)} - b_l^{(q)}| < \epsilon.$$

1.3.4 Performance de l'estimateur proposé

Dans cette section, nous étudions la performance des différents estimateurs \hat{f} d'une densité de probabilités f obtenus par les différentes méthodes d'estimation non paramétrique présentées dans ce chapitre. Les estimateurs peuvent être comparés sur la base de différents critères : le biais, la variance, l'erreur quadratique moyenne. Le biais mesure l'écart entre la vraie valeur de la quantité à estimer et la valeur que l'estimateur prend en moyenne. La variance quantifie la variabilité autour de l'espérance. Idéalement, un bon estimateur possède un petit biais voire nul et une petite variance. Il n'est pas possible de réduire simultanément le biais et la variance d'un estimateur, ce qui amène aux définitions de l'erreur quadratique moyenne (EQM)[47], le critère d'information Akaike (AIC) [2] et la distance de Kullback-Leibler (DKL) [47]. On définit respectivement l'EQM, l'AIC et la DKL de l'estimateur \hat{f} d'une densité de probabilités inconnue f par :

$$\text{EQM}(\hat{f}) = \frac{1}{n} \sum_{j=1}^n (f(x_j) - \hat{f}(x_j))^2$$

$$\text{AIC} = 2nb - 2 \log(L)$$

$$\text{DKL} = \frac{1}{n} \sum_{j=1}^n f(x_j) \log \left[\frac{f(x_j)}{\hat{f}(x_j)} \right]$$

avec (x_1, \dots, x_n) est une réalisation d'un échantillon aléatoire (X_1, \dots, X_n) de densité inconnue f , n représente le nombre d'observations, nb représente le nombre de paramètres

à estimer du modèle et L est le maximum de la fonction de vraisemblance du modèle. Ces critères nous ont permis de faire une étude comparative des performances des différentes méthodes d'estimation non paramétriques adaptées. La mesure de performance est obtenue en calculant l'EQM, AIC et DKL de l'estimateur de la densité inconnue f .

Afin d'apprécier le comportement de ces estimateurs, nous présentons dans la figure FIG.1.2 quelques résultats de simulations, pour un nombre d'échantillons $N = 5000$ et pour un pas $h = 1$, sur les méthodes d'estimation non paramétriques telles que la méthode d'histogramme, la méthode de noyau (noyau Gaussien), la méthode orthogonale (base hermite) et la méthode proposée basée sur les B-splines quadratique.

Dans ce cas, nous procédons à faire quelques simulations sur des lois connues (Gaussienne, Bêta, Gamma).

Cas d'une loi Gaussienne

Une variable aléatoire X suit une loi Gaussienne $N(\mu, \sigma^2)$ sur \mathbb{R} de moyenne μ et de variance σ^2 si sa densité de probabilités est définie par :

$$f_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right].$$

Nous présentons dans la figure FIG.1.2 l'estimation d'une densité de la loi Gaussienne de moyenne $\mu = 4$ et de variance $\sigma^2 = 2$ par les différentes méthodes d'estimation non paramétrique présentées dans ce chapitre.

Les courbes en ligne discontinue et en ligne continue représentent respectivement la vraie densité f et la densité estimée \hat{f} .

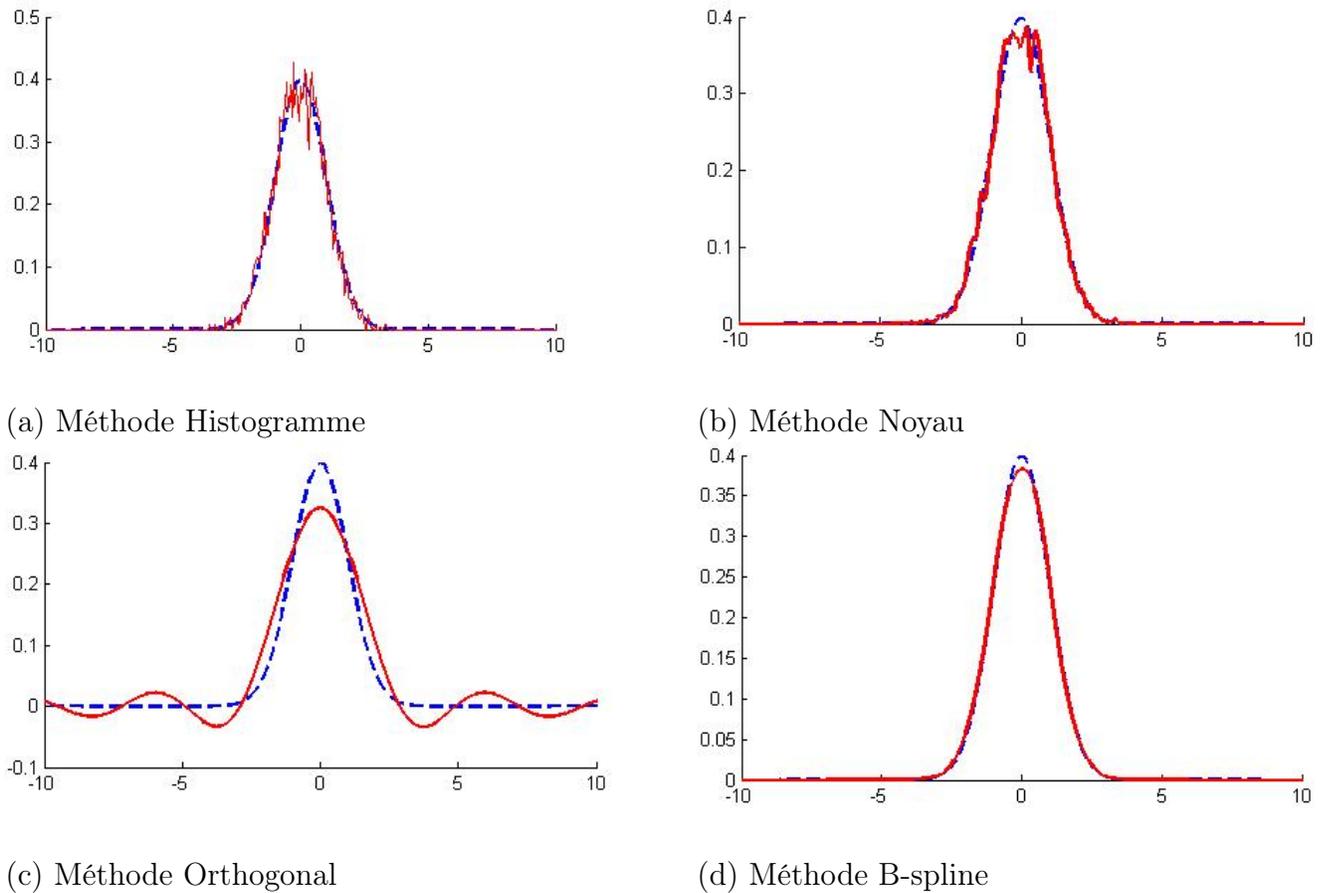


FIG. 1.2 – Estimation non paramétrique de la distribution Gaussienne.

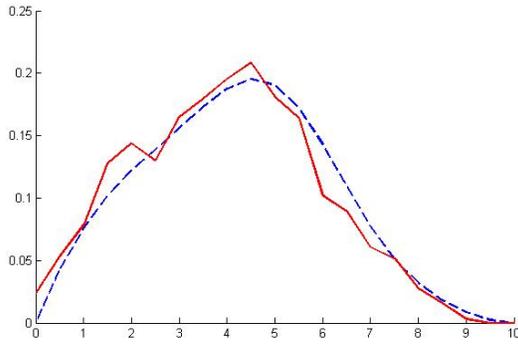
Cas d'une loi Bêta

L'expression analytique de la densité de probabilités de la loi Bêta $B(p, q)$ de paramètres strictement positifs p et q est donnée par :

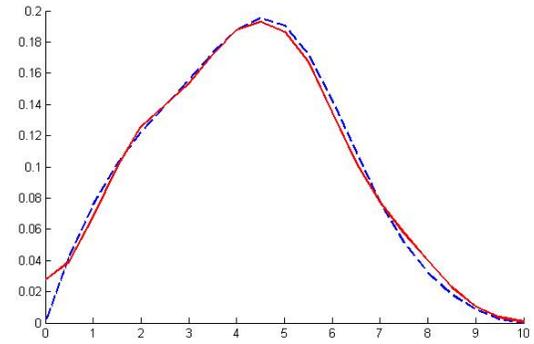
$$f_{p,q}(x) = \frac{1}{B(p, q)} (x - a)^{p-1} (b - x)^{q-1} \mathbf{1}_{[a,b]}(x),$$

avec $B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p + q)}$, où $\Gamma(\alpha) = \int_0^{+\infty} \exp(-x)x^{\alpha-1}dx, \alpha > 0$.

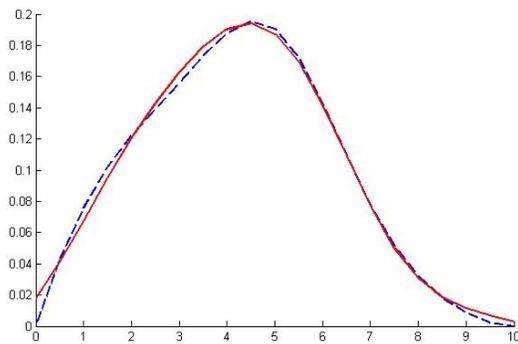
Nous présentons dans la figure FIG.1.3, l'estimation de la densité de la loi Bêta $B(p, q)$ de paramètres $p = 2$ et $q = 3$ par les différentes méthodes d'estimation non paramétrique présentées dans ce chapitre.



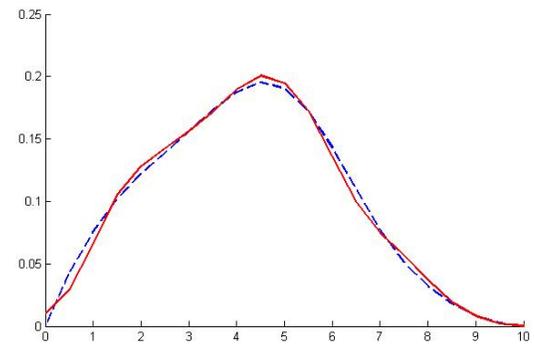
(a) Méthode Histogramme



(b) Méthode Noyau



(c) Méthode Orthogonal



(d) Méthode B-spline

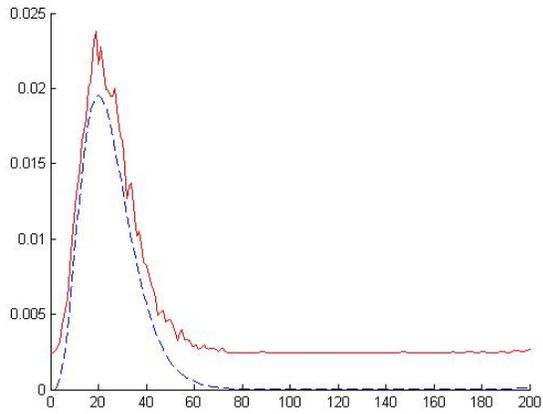
FIG. 1.3 – Estimation non paramétrique de la distribution Bêta.

Cas d'une loi Gamma

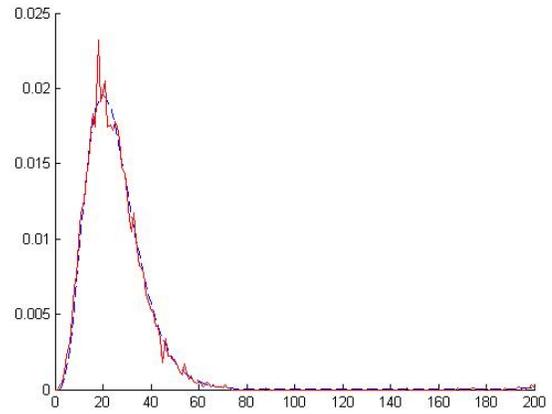
Une variable aléatoire X suit une loi Gamma $\gamma(a, \lambda)$ de paramètres $(a, \lambda) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$, notée $\gamma(a, \lambda)$, si sa densité de probabilités est définie par

$$f_{a,\lambda}(x) = \frac{1}{\Gamma(a)} x^{a-1} \exp\left(-\frac{x}{\lambda}\right) \mathbf{1}_{[0,+\infty[}(x).$$

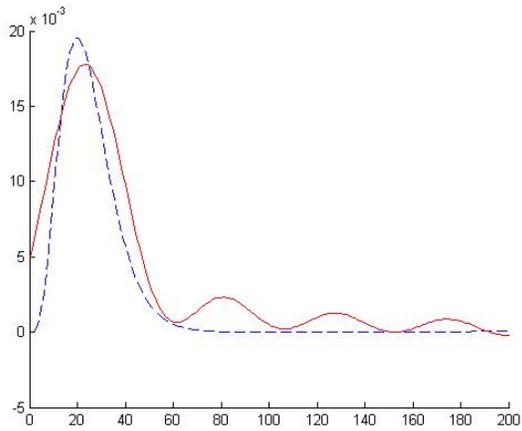
Nous présentons dans la figure Fig. 1.4 l'estimation d'une densité de la loi Gamma $\gamma(a, \lambda)$ de paramètres $a = 5$ et $\lambda = 5$ par les différentes méthodes d'estimation non paramétrique présentées dans ce chapitre.



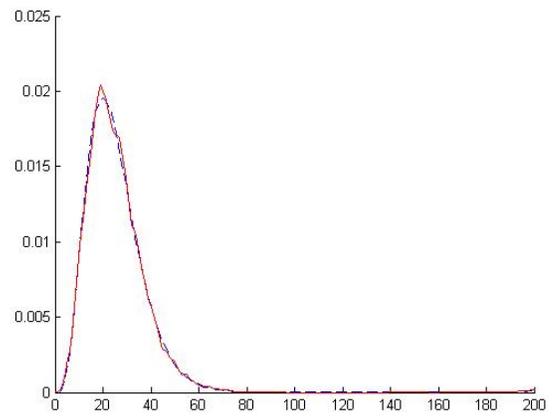
(a) Méthode Histogramme



(b) Méthode Noyau



(c) Méthode Orthogonal



(d) Méthode B-spline

FIG. 1.4 – Estimation non paramétrique de la distribution Gamma.

Nous présentons dans les Tableaux TAB. 1.1, TAB. 1.2 et TAB. 1.3, l'évaluation des différents estimateurs de ces trois exemples de lois en calculant l'EQM, AIC et DKL entre la vraie densité f et la densité estimée \hat{f} .

Vraie densité f EQM(\hat{f})	Histogramme	Noyau	Orthogonale	B-spline
N(4, 2)	0.143	0.04	0.0094	0.00299
B(2, 3)	0.079	0.022	0.00119	0.001115
G(5, 5)	0.072	0.019	0.0027	0.0012

TAB. 1.1 – EQM des différentes méthodes d’estimation.

Vraie densité f AIC(\hat{f})	Histogramme	Noyau	Orthogonale	B-spline
N(4, 2)	0.083	0.156	0.035	0.205
B(2, 3)	0.033	0.089	0.138	0.198
G(5, 5)	0.134	0.043	0.098	0.274

TAB. 1.2 – AIC des différentes méthodes d’estimation.

Vraie densité f DKL(\hat{f})	Histogramme	Noyau	Orthogonale	B-spline
N(4, 2)	0.226	0.152	0.0853	0.0239
B(2, 3)	0.356	0.035	0.097	0.0108
G(5, 5)	0.179	0.098	0.072	0.039

TAB. 1.3 – DKL des différentes méthodes d’estimation.

D’après TAB. 1.1, TAB. 1.2 et TAB. 1.3, nous remarquons pour ces différents exemples que l’EQM est minimale lorsque la densité de probabilités est estimée par l’algorithme proposé. L’AIC est élevé en utilisant les B-splines ceci est justifié du grand nombre de paramètres à estimer. Par contre le DKL qui dépend du taille de l’échantillon est minimal en utilisant les B-splines. D’après ces résultats, nous remarquons que l’utilisation des fonctions B-splines a permis d’améliorer l’estimation des densités de probabilités. Dans la suite de ce chapitre,

nous montrons comment nous pouvons généraliser la méthode proposée pour le cas des mélanges de lois.

1.4 Généralisation de l'estimateur proposé

Les mélanges de lois de probabilités ont fourni un ensemble d'approches mathématiques pour la modélisation statistique d'une grande variété de phénomènes aléatoires. Vue leur utilité comme méthodes de modélisation exibles, les modèles de mélange ont continué à avoir un succès et un intérêt croissant du point de vue pratique et théorique [11][90][114]. Les domaines dans lesquels les modèles mélange ont été appliqués avec succès comprennent l'astronomie, la biologie, la génétique, la médecine, la psychiatrie, l'économie, l'ingénierie et le marketing.

L'objet de cette section est la généralisation de l'estimateur proposé dans la section 1.3 pour le cas de mélange fini de lois de probabilités.

1.4.1 Présentation et interprétation

Le modèle de mélange suppose que chaque composante est caractérisée par une distribution de probabilités. Dans un modèle mélange, on suppose que les données x_1, \dots, x_n constituent un échantillon de n réalisations indépendantes du vecteur aléatoire X à valeurs dans \mathbb{R}^p dont la fonction densité f s'écrit comme combinaison convexe de K densités f_1, \dots, f_K .

$$f(x) = \sum_{k=1}^K \pi_k f_k(x), \quad (1.2)$$

où K est le nombre de composantes, f_k est la densité de la distribution de X conditionnellement à la composante k et les π_k sont les proportions du mélange qui vérifient les conditions suivantes ($\pi_k \in (0, 1)$ et $\sum_{k=1}^K \pi_k = 1$).

La définition générale de la loi mélange recouvre une famille très large de lois de probabilités. Dans la littérature, on suppose que ces lois sont des lois paramétriques [52][97]. Dans notre travail, nous supposons que les composantes f_k ($k = 1, \dots, K$) du mélange sont paramétrées par les coefficients B-splines $b_{l,k}$ ($k = 1, \dots, K; l = 1, \dots, m_k + 2$) avec $m_k + 2$ représente le nombre des éléments de la k -ème composante. La densité de probabilités de la loi mélange donnée par l'équation (1.2) s'écrit alors :

$$f(x, \theta) = \sum_{k=1}^K \pi_k f(x/b_{l,k}); l = 1, \dots, m_k + 2$$

où $\theta = (\pi_1, \dots, \pi_K, b_{l,1}, \dots, b_{l,K})$ désigne le vecteur des paramètres du mélange et $f(x/b_{l,k} = \sum_{l=1}^{m_k+2} b_{l,k} B_{l,h_l}^2(x)$. L'estimation de la densité mélange revient à estimer ce vecteur des paramètres θ , d'où l'introduction d'un nouvel algorithme d'estimation basée sur l'algorithme EM noté l'algorithme EMGB (Espérance Maximisation Généralisé des coefficients B-splines) [65].

1.4.2 Estimation du mélange de lois par les fonctions B-splines

Dans ce paragraphe, nous montrons l'intérêt du nouvel estimateur pour les cas de mélange fini de lois de probabilités. En utilisant ce nouvel estimateur pour chaque composante du mélange de lois, nous obtenons l'expression suivante :

$$f(x) = \sum_{k=1}^K \pi_k \sum_{l=1}^{m_k+2} b_{l,k} B_{l,h_l}^2(x)$$

où $\pi_k \geq 0$, $\sum_{k=1}^K \pi_k = 1$ et $\sum_{l=1}^{m_k+2} b_{l,k} = 1$.

Nous définissons l'ensemble Θ des paramètres par

$$\Theta = \left\{ \theta = (\pi_1, \dots, \pi_K, b_{.,1}, \dots, b_{.,K}); \pi_k > 0, \sum_{k=1}^K \pi_k = 1; k = 1, \dots, K; b_{.,k} \in]0, 1[^{m_k+2}; \sum_{l=1}^{m_k+2} b_{l,k} = 1 \right\},$$

nous nottons $b_{.,k} = (b_{l,k})_{l=1, \dots, m_k+2}$.

L'estimation de la densité mélange revient à estimer le vecteur des paramètres $\theta = (\pi_1, \dots, \pi_K, b_{.,1}, \dots, b_{.,K}) \in \Theta$. L'estimation de ce vecteur des paramètres s'obtiendra à partir d'un algorithme itératif appelé algorithme EMGB décrit dans le paragraphe suivant.

1.4.3 Description de l'algorithme EMGB

Cet algorithme est basé sur la méthode du maximum de vraisemblance en utilisant l'algorithme EM dans un cadre non paramétrique. Ceci revient à estimer les coefficients des B-splines et les probabilités à priori de façon itérative. Cet algorithme est composé de trois étapes qui sont :

- **Étape d'initialisation** : Les paramètres du mélange peuvent être initialisé comme suit

$$\pi_k^{(0)} = \frac{n_k}{n}, \quad b_{l,k}^{(0)} = \frac{\text{nombre d'observation } X_j \in [x_l, x_{l+3})}{n_k},$$

où n_k est le nombre total d'observations dans la $k^{\text{ème}}$ composante.

Pour chaque X_i , on associe une variable aléatoire $Z_i = (Z_{1,i}, \dots, Z_{K,i})$ de telle sorte que Z_i suit une distribution multinomiale $Mul(1, \pi_1, \dots, \pi_K)$ (i-e; $\mathbb{P}(Z_{j,i} = 1) = \pi_j$).

Soient $X = (X_1, \dots, X_n)$ et $Z = (Z_1, \dots, Z_n)$ puis à l'itération q , nous avons

$$l(X, Z/\theta^{(q)}) = \sum_{i=1}^n \left[\sum_{k=1}^K Z_{k,i} \log(\pi_k^{(q)} f_k(X_i/b_{.,k}^{(q)}) \right].$$

$$Q(\theta, \theta^{(q)}) = \mathbb{E}(l(X, Z/X, \theta^{(q)})) = \sum_{i=1}^n \left[\sum_{k=1}^K \mathbb{E}(Z_{k,i}/X, \theta^{(q)}) \log(\pi_k f_k(X_i/b_{.,k})) \right]$$

Si $\tau_{k,i}^{(p)} = \mathbb{E}(Z_{k,i}/X, \theta^{(q)})$, alors

$$\begin{aligned} Q(\theta, \theta^{(q)}) &= \sum_{i=1}^n \left[\sum_{k=1}^K \tau_{k,i}^{(q)} \log(\pi_k f_k(X_i/b_{.,k})) \right] \\ &= \sum_{i=1}^n \left[\sum_{k=2}^K \tau_{k,i}^{(q)} \log(\pi_k f_k(X_i/b_{.,k})) + \tau_{1,i}^{(q)} \log\left(1 - \sum_{k=2}^K \pi_k\right) f_1(X_i/b_{.,1}) \right] \end{aligned}$$

- **Étape d'espérance (E)** : À l'itération (q) , nous calculons la quantité intermédiaire $Q(b, b^{(q)})$ et nous estimons la probabilité a posteriori $\tau_{k,i}^{(q)}$ pour l'observation X_i appartenant à la $k^{\text{ème}}$ composante par :

$$\tau_{k,i}^{(q)} = \frac{\pi_k^{(q)} f_k(X_i/b_{.,k}^{(q)})}{\sum_{k=1}^K \pi_k^{(q)} f_k(X_i/b_{.,k}^{(q)})}$$

- **Étape de maximisation (MGB)** : À l'itération $(q+1)$, nous calculons le paramètre $\theta^{(q+1)}$ qui maximise :

$$\theta^{(q+1)} = \arg \max_{\theta} Q(\theta, \theta^{(q)})$$

Lemme 1.4.1. *L'application définie sur Θ par $\theta \mapsto Q(\theta, \theta^{(q)})$ est concave.*

Preuve :

$$Q(\theta, \theta^{(q)}) = \sum_{i=1}^n \sum_{k=1}^K \tau_{k,i}^{(q)} \log(\pi_k) + \sum_{i=1}^n \sum_{k=1}^K \tau_{k,i}^{(q)} \log(f_k(X_i/b_{.,k})).$$

Il est clair que les applications

$$\pi = (\pi_1, \dots, \pi_K) \mapsto \sum_{i=1}^n \sum_{k=1}^K \tau_{k,i}^{(q)} \log(\pi_k)$$

et

$$b_{.,k} = (b_{1,k}, \dots, b_{m_k+2,k}) \mapsto \log(f_k(X_i/b_{.,k})) = \log\left(\sum_{l=1}^{m_k+2} b_{l,k} B_{l,h_l}^2(X_i)\right)$$

sont concaves (car la fonction log est concave). Comme la somme des fonctions concaves est concave alors Q est aussi concave.

Pour déterminer le maximum de $\theta \mapsto Q(\theta, \theta^{(q)})$ il suffit de résoudre $\frac{\partial Q(\theta, \theta^{(q)})}{\partial \theta} = 0$

1) Calcul des probabilités a priori à l'itération $(q + 1)$

$$\frac{\partial Q(\theta, \theta^{(q)})}{\partial \pi_k} = \sum_{i=1}^n \left(\frac{\tau_{k,i}^{(q)}}{\pi_k} - \frac{\tau_{1,i}^{(q)}}{1 - \sum_{k=2}^K \pi_k} \right) = 0.$$

Ce qui implique,

$$\pi_k = \frac{\sum_{i=1}^n \tau_{k,i}^{(q)}}{\sum_{i=1}^n \tau_{1,i}^{(q)}} \left(1 - \sum_{k=2}^K \pi_k \right)$$

D'où,

$$\sum_{k=1}^K \pi_k = \sum_{k=1}^K \frac{\sum_{i=1}^n \tau_{k,i}^{(q)}}{\sum_{i=1}^n \tau_{1,i}^{(q)}} \left(1 - \sum_{k=2}^K \pi_k \right)$$

Alors,

$$\pi_k = \frac{\sum_{i=1}^n \tau_{k,i}^{(q)}}{\sum_{k=1}^K \sum_{i=1}^n \tau_{k,i}^{(q)}} = \frac{1}{n} \sum_{i=1}^n \tau_{k,i}^{(q)}.$$

Donc, à l'itération $(q + 1)$, la probabilité a priori $\pi_k^{(q+1)}$ s'écrit sous la forme

$$\pi_k^{(q+1)} = \frac{1}{n} \sum_{i=1}^n \tau_{k,i}^{(q)}$$

2) Calcul des coefficients B-splines à l'itération $(q + 1)$

$$\frac{\partial Q(\theta, \theta^{(q)})}{\partial b_{l,k}} = \sum_{i=1}^n \tau_{k,i} \left(\frac{B_{l,h_l}^2(X_i) - B_{1,h_l}^2(X_i)}{f(X_i)} \right) = 0.$$

Soit

$$\frac{\partial Q(\theta, \theta^{(q)})}{\partial b_{l,k}} = F_l(b_{.,k}^{(q)}).$$

Pour résoudre $F_l(b_{.,k}^{(p)}) = 0$, nous utilisons la méthode de Newton Raphson [111]

$$F'_l(b_{.,k}^{(q)})(b_{l,k}^{(q+1)} - b_{l,k}^{(q)}) = -F_l(b_{.,k}^{(q)}),$$

où $F'_l(b_{.,k}^{(q)})$ est la différentielle de F_l en $b_{.,k}^{(q)}$. Après quelques calculs, nous avons

$$b_{l,k}^{(q+1)} = b_{l,k}^{(q)} + \frac{\sum_{i=1}^n \tau_{k,i}^{(q)} \left(\frac{B_{l,h_l}^2(X_i) - B_{1,h_1}^2(X_i)}{f_k(X_i/b_{.,k}^{(q)})} \right)}{\sum_{i=1}^n \tau_{k,i}^{(q)} \left(\frac{B_{l,h_l}^2(X_i) - B_{1,h_1}^2(X_i)}{f_k(X_i/b_{.,k}^{(q)})} \right)^2}, \quad (1.3)$$

pour $l = 2, \dots, m_k + 2$ et $k = 1, \dots, K$.

Il est à noter qu'une solution trouvée peut ne pas représenter un vrai maximum global. Elle peut représenter un maximum ou un maximum local ou encore un point d'inflexion de la fonction log-vraisemblance. Une vérification de chaque solution trouvée est nécessaire pour s'assurer qu'il s'agit d'un vrai maximum, ou un calcul des dérivées secondes peut confirmer la nature de l'optimum trouvée. L'existence et l'unicité de la solution obtenue par notre algorithme va être donnée dans le paragraphe suivant sous forme de propositions.

1.4.4 Existence, Unicité et Propriétés de l'algorithme EMGB

Dans cette sous section, dans un premier temps, nous prouvons l'existence et l'unicité de l'estimateur $(\widehat{\pi}_k, \widehat{b}_{l,k})$ obtenu par l'algorithme EMGB et dans un deuxième temps, nous présentons quelques propriétés asymptotique de cet estimateur.

Existence et Unicité

Soit $\theta = (\pi_1, \dots, \pi_K, b_{.,1}, \dots, b_{.,K}) \in \Theta$. La fonction log-vraisemblance du mélange est définie par :

$$l(X_1, \dots, X_n/\theta) = \log L(X_1, \dots, X_n/\theta) = \sum_{j=1}^n \log \sum_{k=1}^K \pi_k f_k(X_j/b_{.,k}) = \sum_{j=1}^n \log \sum_{k=1}^K \pi_k \left(\sum_{l=1}^{m_k+2} b_{l,k} B_{l,h_l}^2(X_j) \right).$$

Dans la proposition suivante et sous certaines conditions, nous prouvons l'existence de l'estimateur du maximum de vraisemblance obtenu par l'algorithme EMGB.

Proposition 1.4.2. *Si $\theta^{(q+1)}$ converge presque sûrement vers $\hat{\theta}$, quand $q \rightarrow +\infty$, alors $\hat{\theta}$ satisfait les équations de vraisemblance.*

Preuve : D'une part, pour k fixé, la dérivée de la fonction de log-vraisemblance par rapport à $b_{l,k}$ est donnée par

$$\frac{\partial l(X_1, \dots, X_n/\theta)}{\partial b_{l,k}} = \sum_{j=1}^n \pi_k \frac{(B_{l,h_l}^2(X_j) - B_{1,h_1}^2(X_j))}{f(X_j)}; \quad l = 2, \dots, m_k + 2.$$

Quand $q \rightarrow +\infty$ dans (1.3) et en utilisant le fait que $\hat{b}_{l,k}^{(q)} \rightarrow \hat{b}_{l,k}$ et $\tau_{k,j}^{(q)} \rightarrow \hat{\tau}_{k,j} = \frac{\hat{\pi}_k \hat{f}_k(X_j/b_{.,k})}{\sum_{k=1}^K \hat{\pi}_k \hat{f}_k(X_j/b_{.,k})}$, avec $\hat{f}(X_j) = \sum_{k=1}^K \hat{\pi}_k \hat{f}_k(X_j/b_{.,k})$ et $\hat{f}_k(X_j/b_{.,k}) = \sum_{l=1}^{m_k+2} \hat{b}_{l,k} B_{l,h_l}^2(X_j)$. Ainsi

$$\sum_{j=1}^n \hat{\tau}_{k,j} \frac{(B_{l,h_l}^2(X_j) - B_{1,h_1}^2(X_j))}{\hat{f}_k(X_j/b_{.,k})} = 0; \quad \text{avec } l = 2, \dots, m_k + 2.$$

Puisque

$$\tau_{k,j} = \frac{\hat{\pi}_k \hat{f}_k(X_j/b_{.,k})}{\sum_{k=1}^K \hat{\pi}_k \hat{f}_k(X_j/b_{.,k})},$$

alors

$$\sum_{j=1}^n \hat{\pi}_k \frac{(B_{l,h_l}^2(X_j) - B_{1,h_1}^2(X_j))}{\hat{f}(X_j)} = 0.$$

D'où $\widehat{b}_{l,k}$ satisfait la dérivée de l'équation de vraisemblance.

D'autre part, on sait que

$$\widehat{\pi}_k = \frac{1}{n} \sum_{i=1}^n \tau_{k,i}^{(q)} = \frac{1}{n} \sum_{j=1}^n \frac{\widehat{\pi}_k \widehat{f}_k(X_j/b_{.,k})}{\sum_{k=1}^K \widehat{\pi}_k \widehat{f}_k(X_j/b_{.,k})},$$

ce qui implique

$$\begin{aligned} 0 &= n - \sum_{j=1}^n \frac{\widehat{f}_k(X_j/b_{.,k})}{\sum_{k=1}^K \widehat{\pi}_k \widehat{f}_k(X_j/b_{.,k})} \\ &= \sum_{j=1}^n \frac{\widehat{\pi}_1 \widehat{f}_1(X_j/b_{.,1}) + \widehat{\pi}_2 \widehat{f}_2(X_j/b_{.,2}) + \dots + \widehat{\pi}_{K-1} \widehat{f}_{K-1}(X_j/b_{.,K-1})}{\sum_{k=1}^K \widehat{\pi}_k \widehat{f}_k(X_j/b_{.,k})} + \\ &\quad \sum_{j=1}^n \frac{(1 - \widehat{\pi}_1 - \widehat{\pi}_2 - \dots - \widehat{\pi}_{K-1}) \widehat{f}_K(X_j/b_{.,K}) - \widehat{f}_k(X_j/b_{.,k})}{\sum_{k=1}^K \widehat{\pi}_k \widehat{f}_k(X_j/b_{.,k})} \\ &= \sum_{j=1}^n \frac{\widehat{\pi}_1 (\widehat{f}_1(X_j/b_{.,1}) - \widehat{f}_K(X_j/b_{.,K})) + \dots + \widehat{\pi}_{K-1} (\widehat{f}_{K-1}(X_j/b_{.,K-1}) - \widehat{f}_K(X_j/b_{.,K}))}{\sum_{k=1}^K \widehat{\pi}_k \widehat{f}_k(X_j/b_{.,k})} + \\ &\quad \sum_{j=1}^n \frac{\widehat{f}_K(X_j/b_{.,K}) - \widehat{f}_k(X_j/b_{.,k})}{\sum_{k=1}^K \widehat{\pi}_k \widehat{f}_k(X_j/b_{.,k})}. \end{aligned} \tag{1.4}$$

Posons

$$H = \sum_{j=1}^n \frac{\widehat{\pi}_1 (\widehat{f}_1(X_j/b_{.,1}) - \widehat{f}_K(X_j/b_{.,K})) + \dots + \widehat{\pi}_{K-1} (\widehat{f}_{K-1}(X_j/b_{.,K-1}) - \widehat{f}_K(X_j/b_{.,K}))}{\sum_{k=1}^K \widehat{\pi}_k \widehat{f}_k(X_j/b_{.,k})}. \tag{1.5}$$

En multipliant (1.4) par $\widehat{\pi}_k$, pour $k = 1, \dots, K - 1$ et en additionnant, nous obtenons

$$(\widehat{\pi}_1 + \widehat{\pi}_2 + \dots + \widehat{\pi}_{K-1})H = H.$$

Comme $\sum_{k=1}^K \hat{\pi}_k = 1$, alors $(1 - \hat{\pi}_K)H = H$ et $\hat{\pi}_K H = 0$.

Puisque nous avons choisi $\hat{\pi}_K$ d'une façon arbitraire dans (1.5), nous aurions pu choisi l'un des $\hat{\pi}_k, k = 1, 2, \dots, K$. Donc, nous obtenons

$$\hat{\pi}_k H = 0, \forall k = 1, \dots, K.$$

D'où, $H = 0$.

Cela signifie, à partir de (1.4) que

$$\sum_{j=1}^n \frac{f_k(X_j/b_{.,k}) - f_K(X_j/b_{.,K})}{\sum_{k=1}^K \hat{\pi}_k \hat{f}_k(X_j/b_{.,k})} = 0 \text{ pour } k = 1, \dots, K - 1$$

Comme

$$\frac{\partial l(X_1, \dots, X_n/\theta)}{\partial \pi_k} = \sum_{j=1}^n \frac{f_k(X_j/b_{.,k}) - f_K(X_j/b_{.,K})}{\sum_{k=1}^K \pi_k f_k(X_j/b_{.,k})} \text{ pour } k = 1, \dots, K - 1$$

Par conséquent, les estimateurs $\hat{\pi}_1, \hat{\pi}_2, \dots, \hat{\pi}_K$ obtenus par l'algorithme EMGB satisfont les équations de vraisemblance.

Dans les deux propositions suivantes, nous prouvons l'unicité de l'estimateur obtenu par l'algorithme EMGB.

Proposition 1.4.3. *Supposons que chaque intervalle de la subdivision contient au moins trois observations alors : La matrice hessienne $\left(\frac{\partial^2 l(X_1, \dots, X_n/\theta)}{\partial b_{l,k} \partial b_{p,k}}\right)_{1 \leq l, p \leq m_k + 2}$ est définie négative.*

Preuve : Les dérivées d'ordre un et deux de la fonction log-vraisemblance sont données par

$$\frac{\partial l(X_1, \dots, X_n/b, \pi)}{\partial b_{l,k}} = \sum_{j=1}^n \pi_k \frac{B_{l,h_l}^2(X_j) - B_{1,h_1}^2(X_j)}{f(X_j)}, \quad l = 2, \dots, m_k + 2$$

et

$$\frac{\partial^2 l(X_1, \dots, X_n/b, \pi)}{\partial b_{l,k} \partial b_{p,k}} = - \sum_{j=1}^n \pi_k^2 \frac{(B_{l,h_l}^2(X_j) - B_{1,h_1}^2(X_j))(B_{p,h_p}^2(X_j) - B_{1,h_1}^2(X_j))}{f^2(X_j)} \quad l = 2, \dots, m_k+2.$$

Ceci impliquent que

$$\left(\frac{\partial^2 l(X_1, \dots, X_n/b, \pi)}{\partial b_{l,k} \partial b_{p,k}} \right)_{1 \leq l, k \leq m_k+2} = - \sum_{j=1}^n T_j \otimes T_j,$$

où $T_j = {}^t (t_{2,j}, t_{3,j}, \dots, t_{m_k+2,j})$, $t_{k,j} = \left(\pi_k \frac{B_{l,h_l}^2(X_j) - B_{1,h_1}^2(X_j)}{f(X_j)} \right)_{2 \leq l \leq m_k+2} \in \mathbb{R}^{m+1}$ et

$T_j \otimes T_j = T_j \cdot T_j^t$ avec T_j^t est le transposé de T_j .

Maintenant, nous allons prouver que $\sum_{j=1}^n T_j \otimes T_j$ est définie positive.

$$\sum_{j=1}^n y'(T_j \otimes T_j)y = \sum_{j=1}^n (y_2 t_{2,j} + y_3 t_{3,j} + \dots + y_{m_k+2} t_{m_k+2,j})^2 \geq 0, \quad \forall y = {}^t (y_2, y_3, \dots, y_{m_k+2}) \neq 0.$$

Supposons que $y \neq 0$ et $\sum_{j=1}^n y'(T_j \otimes T_j)y = 0$. Cela signifie que

$$y_2 t_{2,j} + y_3 t_{3,j} + \dots + y_{m_k+2} t_{m_k+2,j} = 0, \quad \forall j = 1, \dots, n.$$

Alors

$$- \left(\sum_{i=2}^{m_k+2} y_i \right) B_{1,h_1}^2(X_j) + y_2 B_{2,h_2}^2(X_j) + \dots + y_{m_k+2} B_{m_k+2,h_{m_k+2}}^2(X_j) = 0, \quad \forall j = 1, \dots, n.$$

En outre, la fonction définie par

$$x \longmapsto \left(\sum_{i=2}^{m_k+2} y_i \right) B_{1,h_1}^2(x) + y_2 B_{2,h_2}^2(x) + \dots + y_{m_k+2} B_{m_k+2,h_{m_k+2}}^2(x)$$

est une fonction polynômiale et $\left(B_{l,h_l}^2(x) \right)_{1 \leq l \leq m_k+2}$ sont linéairement indépendantes, alors

$y_2 = \dots = y_{m_k+2} = 0$. D'où la fonction hessienne $\left(\frac{\partial^2 l(X_1, \dots, X_n/\theta)}{\partial b_{l,k} \partial b_{p,k}} \right)_{1 \leq l, p \leq m_k+2}$ est

définie négative.

Proposition 1.4.4. *Supposons que chaque intervalle de la subdivision contient au moins trois observations alors : La matrice hessienne $\left(\frac{\partial^2 l(X_1, \dots, X_n/\theta)}{\partial \pi_k \partial \pi_s}\right)_{1 \leq k, s \leq K}$ est définie négative.*

Preuve : Les dérivées d'ordre un et deux de la fonction log-vraisemblance sont données par

$$\frac{\partial l(X_1, \dots, X_n/\theta)}{\partial \pi_k} = \sum_{j=1}^n \frac{f_k(X_j/b_{.,k}) - f_1(X_j/b_{.,1})}{f(X_j)}, \quad k = 2, \dots, K$$

et

$$\begin{aligned} \frac{\partial^2 l(X_1, \dots, X_n/\theta)}{\partial \pi_k \partial \pi_s} &= - \sum_{j=1}^n \frac{(f_k(X_j/b_{.,k}) - f_1(X_j/b_{.,1}))(f_s(X_j/b_{.,s}) - f_1(X_j/b_{.,1}))}{f^2(X_j)}, \quad k = 2, \dots, K \\ &= - \sum_{j=1}^n \frac{f_k(X_j/b_{.,k}) - f_1(X_j/b_{.,1})}{f(X_j)} \frac{f_s(X_j/b_{.,s}) - f_1(X_j/b_{.,1})}{f(X_j)}, \end{aligned}$$

Ceci impliquent que

$$\left(\frac{\partial^2 l(X_1, \dots, X_n/\theta)}{\partial \pi_k \partial \pi_s}\right)_{1 \leq k, s \leq K} = - \sum_{j=1}^n T_j \otimes T_j,$$

où $T_j = {}^t(t_{2,j}, t_{3,j}, \dots, t_{K,j})$, $t_{k,j} = \frac{f_k(X_j/b_{.,k}) - f_1(X_j/b_{.,1})}{f(X_j)}$.

Comme

$$\sum_{j=1}^n y'(T_j \otimes T_j)y = \sum_{j=1}^n (y_2 t_{2,j} + y_3 t_{3,j} + \dots + y_K t_{K,j})^2 \geq 0, \text{ pour tout } y = {}^t(y_2, y_3, \dots, y_K) \neq 0,$$

alors $\sum_{j=1}^n T_j \otimes T_j$ est définie positive. Supposons maintenant que $y \neq 0$ et $\sum_{j=1}^n y'(T_j \otimes T_j)y = 0$.

Cela signifie que

$$y_2 t_{2,j} + y_3 t_{3,j} + \dots + y_K t_{K,j} = 0 \text{ pour tout } j = 1, \dots, n.$$

Alors,

$$- \left(\sum_{i=2}^K y_i \right) f_1(X_j/b_{.,1}) + y_2 f_2(X_j/b_{.,2}) + \dots + y_K f_K(X_j/b_{.,K}) = 0 \quad \forall j = 1, \dots, n$$

Étant donné que chaque composante k contient au moins une observation X_j , alors $y_k F_k(X_j) = 0$. Donc, $y_2 = y_3 = \dots = y_K = 0$.

D'où, la hessienne $\left(\frac{\partial^2 l(X_1, \dots, X_n/\theta)}{\partial \pi_k \partial \pi_s}\right)_{1 \leq k, s \leq K}$ est une matrice définie négative.

Proposition 1.4.5. Soit $\theta = (\pi_1, \dots, \pi_K, b_{.,1}, \dots, b_{.,K}) \in \Theta$. Si $\theta^{(q+1)}$ converge vers $\hat{\theta}$ quand $q \mapsto +\infty$ alors $\hat{\theta}$ est l'unique estimateur de la fonction de vraisemblance.

Étude de simulation

Dans ce paragraphe, nous présentons quelques résultats de simulation afin d'évaluer et de comparer la performance de l'algorithme EMGB et l'algorithme EM (avec loi gaussienne).

Nous avons simulé un échantillon de taille $n = 5000$ selon le vrai modèle de B-splines pour un pas $h = 1$. Pour l'estimation des coefficients de B-splines b_l , nous avons utilisé l'algorithme EMGB.

Dans les figures FIG.1.5 et FIG.1.6, nous présentons l'estimation de la densité (vrai modèle de B-splines) par le mélange de fonctions B-splines et mélange Gaussiens. A partir des deux courbes l'estimation par B-splines est meilleur que celle obtenue par le modèle Gaussien.

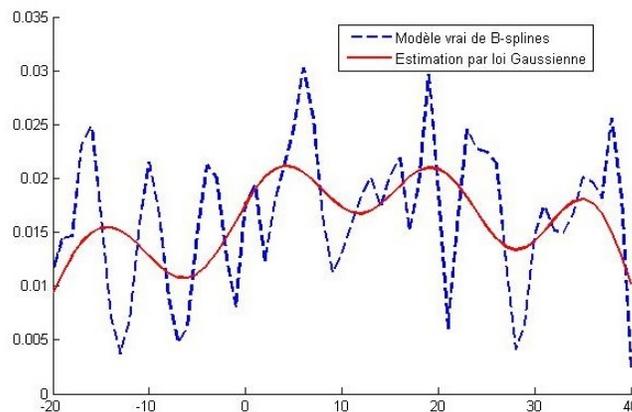


FIG. 1.5 – Densité du vrai modèle B-splines et densité estimée selon le modèle Gaussien.

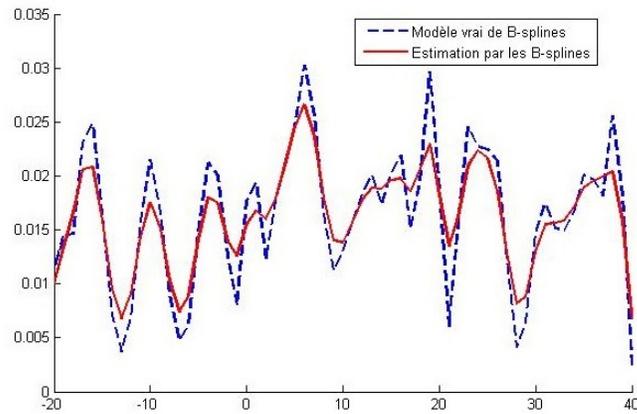


FIG. 1.6 – Densité du vrai modèle B-splines et densité estimée selon le modèle B-splines.

Afin de comparer les deux cas, nous avons calculé l’erreur quadratique moyenne (EQM) entre la densité réelle f et la densité estimée \hat{f} définie par :

$$EQM(f, \hat{f}) = \frac{1}{n} \sum_{j=1}^n (fx_j - \hat{f}x_j)^2.$$

	lois Gaussiennes	B-splines
EQM	0.3015	0.0147

TAB. 1.4 – EQM entre la vraie densité et la densité estimé

D’après le tableau TAB 1.4, nous remarquons que l’EQM calculée selon notre modèle est nettement plus faible que celle calculée avec le modèle Gaussien.

1.4.5 Performance de l’estimateur du mélange proposé

Afin de montrer la performance de l’algorithme EMGB, nous présentons quelques résultats de simulations. D’abord, nous simulons un mélange de lois connues pour un nombre

d'échantillons $N = 5000$ et pour un pas $h = 1$. Puis, nous estimons la densité de ce mélange en utilisant les différentes méthodes d'estimation non paramétrique présentées dans ce chapitre (méthodes histogramme, noyau (noyau Gaussien) et orthogonale (base hermite)). Enfin, nous comparons les différents estimateurs obtenus par ces différentes méthodes en calculant l'EQM, l'AIC et la DKL entre la densité mélange simulée et celle estimée.

Mélange des lois Gaussiennes

Dans la figure FIG.1.7, nous présentons l'estimation de mélange de lois Gaussiennes de moyennes $\mu_1 = 4$ et $\mu_2 = 15$ et de variances $\sigma_1^2 = 2$ $\sigma_2^2 = 10$ par les différentes méthodes d'estimation non paramétrique présentées dans ce chapitre.

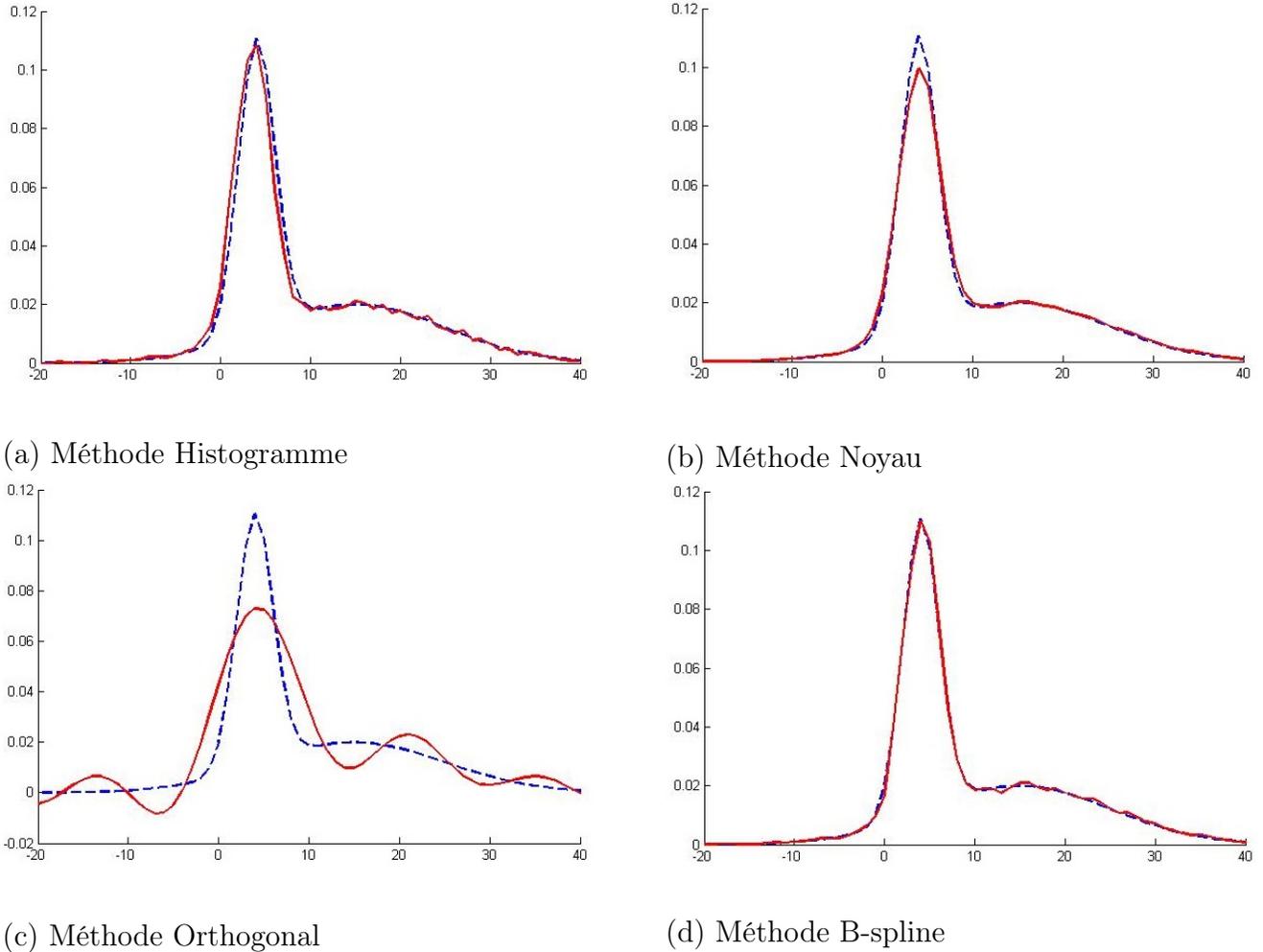


FIG. 1.7 – Estimation de mélange de lois Gaussiennes $N(4, 2)$ et $N(15, 10)$ par les méthodes non paramétriques avec $\pi_1 = 0.5$ et $\pi_2 = 0.5$.

Mélange des lois Gaussiennes et Bêta

Dans la figure FIG.1.8, nous présentons l'estimation de mélange Gaussien-Bêta de paramètres $\mu = 8$, $\sigma^2 = 12$, $p = 9$ et $q = 9$ par les différentes méthodes d'estimation non paramétrique présentées dans ce chapitre.

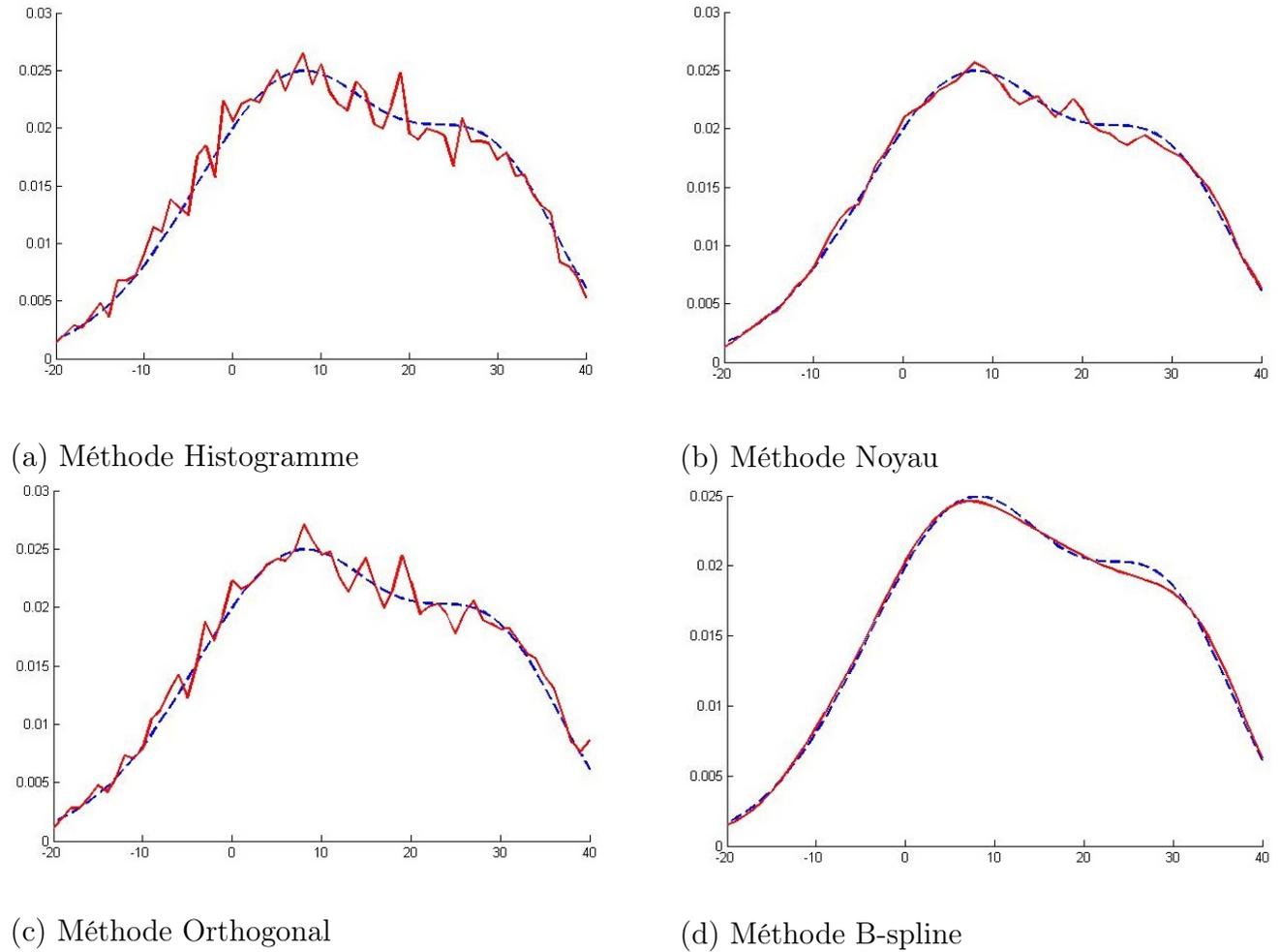


FIG. 1.8 – Estimation de mélange de lois Gaussienne $N(8, 12)$ et Bêta $B(9, 9)$ par les méthodes non paramétriques avec $\pi_1 = 0.75$ et $\pi_2 = 0.25$.

Mélange des lois Bêta Gaussienne et Gamma

Dans la figure FIG.1.9, nous présentons l'estimation de mélange des lois Gaussienne Bêta et Gamma de paramètres $\mu = 8$, $\sigma^2 = 12$, $p = 9$, $q = 9$, $a = 8$ et $\lambda = 12$ par les différentes méthodes d'estimation non paramétrique présentées dans ce chapitre.

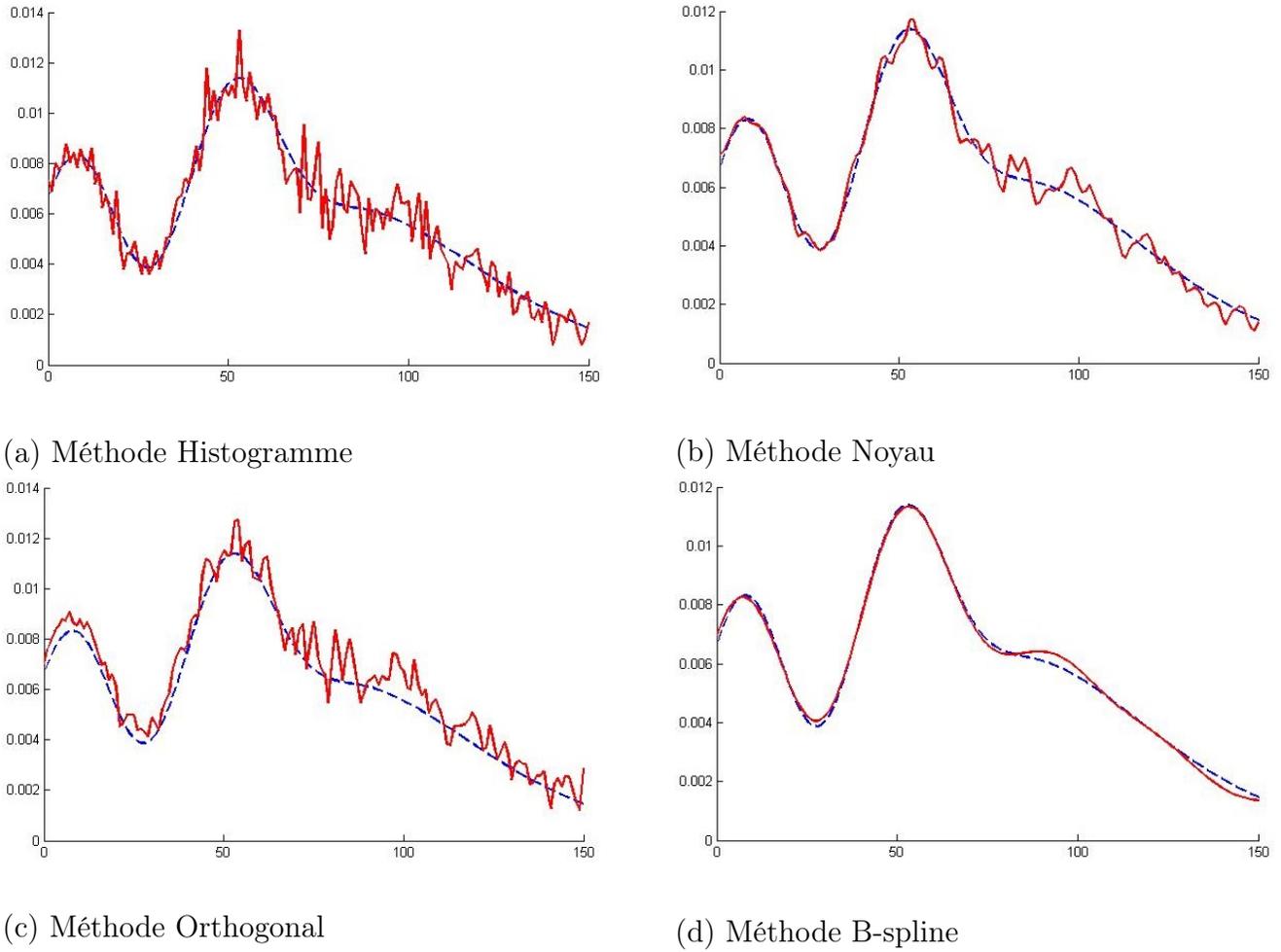


FIG. 1.9 – Estimation de mélange de lois Bêta $B(9,9)$, Gaussienne $N(8,12)$ et Gamma $G(8,12)$ par les méthodes non paramétriques avec $\pi_1 = 0.25$, $\pi_2 = 0.25$ et $\pi_3 = 0.5$.

Mélange de lois Bêtas et une loi Gamma

Dans la figure FIG.1.10, nous présentons l'estimation de mélange de lois Bêtas et une Gamma de paramètres $p_1 = 2$, $q_1 = 3$, $p_2 = 9$, $q_2 = 9$, $a = 8$ et $\lambda = 12$ par les différentes méthodes d'estimation non paramétrique présentées dans ce chapitre.

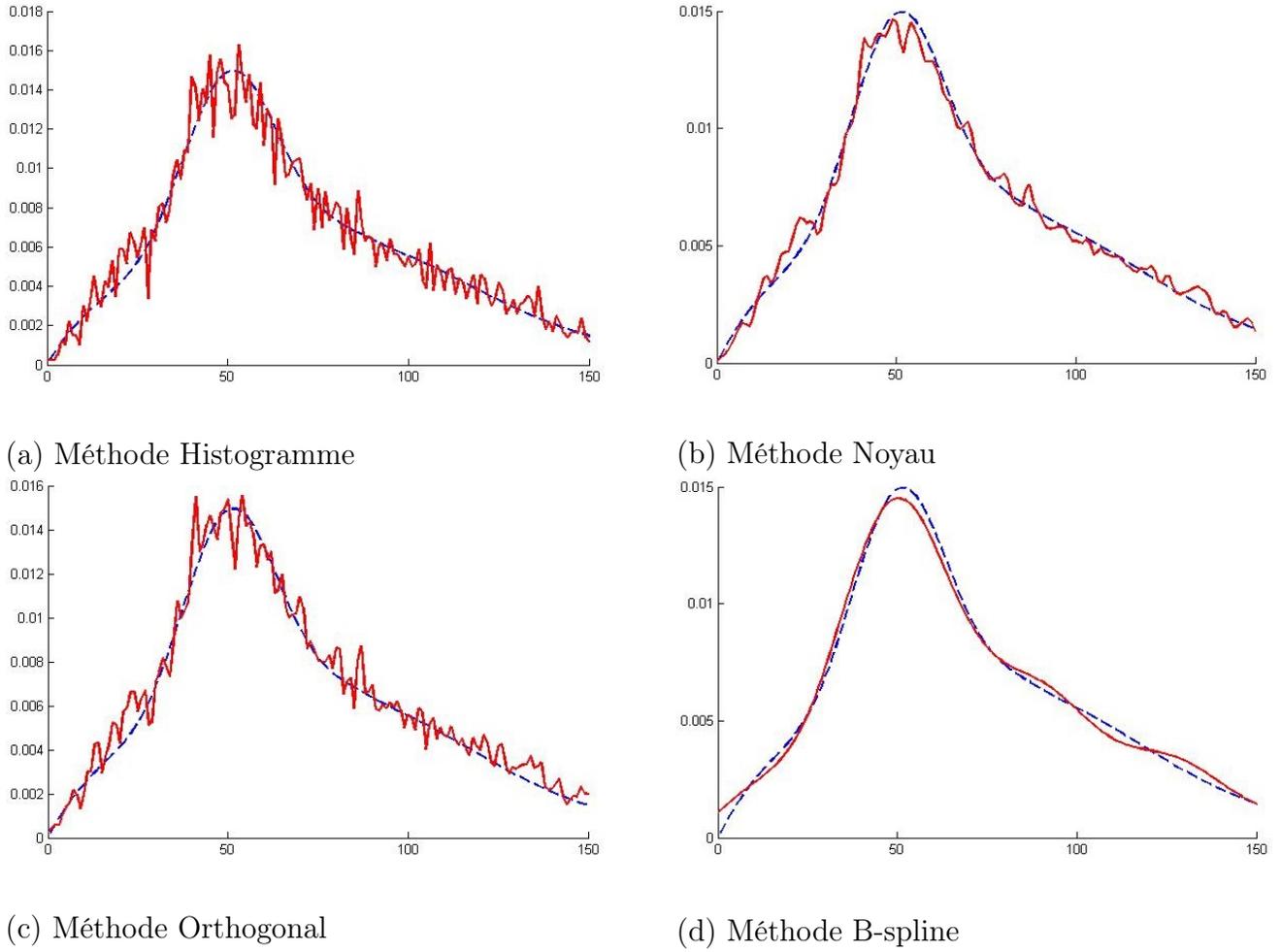


FIG. 1.10 – Estimation de mélange de lois Bêta $B(2, 3)$ Bêta $B(9, 9)$ et Gamma $G(8, 12)$ par les méthodes non paramétriques avec $\pi_1 = 0.25$, $\pi_2 = 0.25$ et $\pi_3 = 0.5$.

Nous présentons dans les Tableaux TAB 1.5, TAB 1.6 et TAB 1.7, l'évaluation des différents estimateurs de ces trois exemples de mélange des lois en calculant l'EQM, AIC et DKL entre la vraie densité f et la densité estimée \hat{f} .

Vraie densité f EQM(\hat{f})	Histogramme	Noyau	Orthogonale	B-spline
Mélange $N(4, 2)$ et $N(15, 10)$	1.49×10^{-5}	5.139×10^{-6}	1.148×10^{-4}	7.2666×10^{-7}
Mélange $N(8, 12)$ et $B(9, 9)$	1.89×10^{-6}	4.063×10^{-7}	1.314×10^{-6}	1.706×10^{-7}
Mélange $G(8, 12)$, $N(8, 12)$ et $B(9, 9)$	5.35×10^{-7}	1.3629×10^{-7}	5.449×10^{-7}	2.591×10^{-8}
Mélange $B(2, 3)$, $B(9, 9)$ et $G(8, 12)$	9.191×10^{-7}	2.509×10^{-7}	5.903×10^{-7}	1.307×10^{-7}

TAB. 1.5 – EQM de différentes méthodes d’estimation

Vraie densité f AIC(\hat{f})	Histogramme	Noyau	Orthogonale	B-spline
Mélange $N(4, 2)$ et $N(15, 10)$	0.045	0.124	0.0145	0.218
Mélange $N(8, 12)$ et $B(9, 9)$	0.0245	0.0471	0.0421	0.148
Mélange $G(8, 12)$, $N(8, 12)$ et $B(9, 9)$	0.0214	0.178	0.084	0.294
Mélange $B(2, 3)$, $B(9, 9)$ et $G(8, 12)$	0.0762	0.0250	0.0874	0.0845

TAB. 1.6 – AIC de différentes méthodes d’estimation

Vraie densité f DKL(\hat{f})	Histogramme	Noyau	Orthogonale	B-spline
Mélange $N(4, 2)$ et $N(15, 10)$	0.345	0.178	0.127	0.0198
Mélange $N(8, 12)$ et $B(9, 9)$	0.247	0.125	0.0987	0.0147
Mélange $G(8, 12)$, $N(8, 12)$ et $B(9, 9)$	0.145	0.207	0.0984	0.0147
Mélange $B(2, 3)$, $B(9, 9)$ et $G(8, 12)$	0.142	0.978	0.852	0.0147

TAB. 1.7 – DKL des différentes méthodes d’estimation

D’après TAB 1.5, TAB 1.6 et TAB 1.7, nous remarquons pour les différents exemples que l’EQM est minimale lorsque la densité de mélange est estimée par l’algorithme proposé. L’AIC est élevé en utilisant les B-splines ceci est justifié du grand nombre de paramètres à estimer. Par contre le DKL qui dépend du taille de l’échantillon est minimal en utilisant les B-splines.

Dans la littérature, il existe autre méthode d'estimation, qui est l'analyse Bayésienne [14][83][98] qui a gagné en popularité au cours des dernières années. Dans le formalisme Bayésien, il y a une introduction de connaissance a priori qui est susceptible d'améliorer l'estimation des paramètres. Dans ce sens, nous allons intégrer dans la suite de notre travail ce formalisme pour enrichir notre méthode d'estimation.

1.5 Apport de l'approche Bayésienne en estimation non paramétrique

Dans l'approche classique de la statistique, les paramètres à estimer, bien qu'inconnus sont supposés fixes. L'idée dans ce cas est de déterminer les paramètres qui cadrent le mieux avec des données observées, étant donnée la fonction vraisemblance. On obtient alors des valeurs pour différents paramètres et on peut directement procéder aux prévisions sans tenir compte de l'incertitude qui existe dans les paramètres ainsi estimés. Dans une optique différente, l'approche Bayésienne [53][54][83] cherche à formaliser autrement l'incertitude relative aux paramètres inconnus du modèle. Le statisticien part d'une loi a priori [54] des paramètres et obtient une loi a posteriori plus précise à la lumière des données. Les prévisions sont alors basées sur cette loi a posteriori, prenant ainsi en compte l'incertitude qui plane sur les paramètres. Cette approche nous semble assez naturelle dans la mesure où elle reflète l'idée d'apprentissage constatée dans la vie quotidienne. En effet, le statisticien part avec une idée d'un phénomène et révisé ses anticipations au fur et à mesure qu'il observe les tendances du phénomène. Ce qui différencie une situation statistique d'une situation Bayésienne est qu'on tient compte d'une information a priori sur un paramètre

inconnu b .

Principe

Considérons un modèle statistique paramétré :

$F = \{P_\theta(dx) = f(x|\theta)\nu(dx); \theta \in \Theta \subset \mathbb{R}^K\}$ dominé par une mesure σ -finie ν . Soit $X = (X_1, \dots, X_n)$ un n -échantillon aléatoire de loi P_θ . Soit $x = (x_1, \dots, x_n)$ une réalisation de l'échantillon aléatoire X . On note $f(x_i|\theta)$ la densité conditionnelle de X_i sachant θ . La fonction vraisemblance associée à l'échantillon aléatoire $X = (X_1, \dots, X_n)$ est définie par $L(x_1, \dots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta)$. La loi conjointe des observations et du vecteur des paramètres est donnée par $L(x|\theta)\pi(\theta)$, avec $\pi(\theta)$ est la densité a priori de θ . Par application directe du théorème de Bayes, la loi conditionnelle de θ sachant $(X = x)$ est appelée distribution a posteriori du vecteur des paramètres θ qui est définie par

$$P(\theta|x) = \frac{L(x|\theta)\pi(\theta)}{q(x)},$$

avec $q(x) = \int_{\Theta} \pi(\theta)L(x|\theta)d\theta$.

Cette distribution a postérieure est donc proportionnelle au produit de la vraisemblance par la densité a priori (*i-e*, $P(\theta|x) \propto L(x|\theta)\pi(\theta)$).

Cette distribution a posteriori contient toute l'information que l'on peut connaître sur le vecteur des paramètres inconnus θ . Mais, en générale, pour définir un estimateur ponctuel, on définit une fonction coût $C(\hat{\theta}, \theta)$ et on minimise le coût moyen. Suivant le choix de la fonction coût on obtient, par exemple, l'estimation en moyenne quadratique ou à maximum a posteriori (MAP) défini par :

$$\hat{\theta} = \arg \max_{\theta \in \Theta} P(\theta|x).$$

Dans notre travail, nous nous sommes intéressés par l'estimateur MAP pour estimer les paramètres du modèle F .

1.5.1 Estimation Bayésienne par les fonctions B-Splines d'une seule loi

Dans le cadre de l'estimation par les fonctions B-splines en utilisant l'approche Bayésienne, nous supposons que les coefficients de B-splines sont des variables aléatoires.

Soit X une variable aléatoire de densité $f(x) = \sum_{l=1}^{m+2} b_l B_l^2(x)$, $b_l \geq 0$ et $\sum_{l=1}^{m+2} b_l = 1$.

Nous supposons que $b = (b_1, \dots, b_{m+2})$ est un vecteur aléatoire suivant une distribution de Dirichlet $Dir(\alpha_1, \dots, \alpha_{m+2})$ [53][54][106].

Rappelons que La densité de Dirichlet est définie par

$$Dir(\alpha_1, \dots, \alpha_{m+2})(b) \propto b_1^{\alpha_1-1} b_2^{\alpha_2-1} \dots b_{m+2}^{\alpha_{m+2}-1} \mathbf{1}_{(\sum_{l=1}^{m+2} b_l=1)},$$

où " \propto " signifie "est proportionnelle à".

Dans le cas particulier où $\alpha_1 = \dots = \alpha_{m+2} = 1$, on obtient une distribution a priori uniforme et dans le cas $m = 1$ on retrouve la distribution a priori Bêta de première espèce de paramètres (α_1, α_2) .

Avec l'hypothèse de nos coefficients de B-splines qui vérifient les conditions de Dirichlet

(c'est à dire $b_l \geq 0$ et $\sum_{l=1}^{m+2} b_l = 1$), nous choisissons dans notre étude d'estimation comme loi a priori du vecteur des paramètres $b = (b_1, \dots, b_{m+2})$ la loi de Dirichlet $Dir(\alpha_1, \dots, \alpha_{m+2})$

du vecteur des paramètres $(\alpha_1, \dots, \alpha_{m+2})$, le choix est justifié car $b_l \geq 0$ et $\sum_{l=1}^{m+2} b_l = 1$.

L'estimation Bayésienne du vecteur des paramètres $b = (b_1, \dots, b_{m+2})$ s'obtiendra par

l'introduction d'un algorithme itératif basé sur la méthode du maximum de vraisemblance

appelé l'algorithme EMBB (Espérance Maximisation B-Spline Bayésienne) [66].

Description de l'algorithme EMBB

Soient X_1, \dots, X_n , n variables aléatoires indépendantes ayant la même densité $f(x) = \sum_{l=1}^{m+2} b_l B_{l,h_l}^2(x)$. Nous définissons la fonction de vraisemblance comme

$$L(X_1, \dots, X_n | b) = \prod_{i=1}^n f(X_i)$$

et la fonction log-vraisemblance par

$$l(X_1, \dots, X_n | b) = \log L(X_1, \dots, X_n | b) = \sum_{i=1}^n \log f(X_i) = \sum_{i=1}^n \log \left(\sum_{l=1}^{m+2} b_l B_{l,h_l}^2(X_i) \right).$$

La distribution a posteriori $P(b | X_1, \dots, X_n)$ du vecteur $b = (b_1, \dots, b_{m+2})$ est proportionnelle à

$$P(b | X_1, \dots, X_n) \propto L(X_1, \dots, X_n | b) \text{Dir}(\alpha_1, \dots, \alpha_{m+2})(b).$$

Ensuite, la fonction log a posteriori est donnée par

$$\begin{aligned} \log(P(b | X_1, \dots, X_n)) &\propto l(X_1, \dots, X_n | b) + \log(\text{Dir}(\alpha_1, \dots, \alpha_{m+2})(b)) \\ &\propto \sum_{i=1}^n \log \left(\sum_{l=1}^{m+2} b_l B_{l,h_l}^2(X_i) \right) + \log(\text{Dir}(\alpha_1, \dots, \alpha_{m+2})(b)). \end{aligned}$$

En se basant sur l'algorithme EM, nous introduisons l'algorithme EMBB pour estimer nos paramètres. Cet algorithme est composé de trois étapes qui sont :

Étape d'initialisation : $b_l^{(0)} = \frac{\text{nombre d'observation } X_j \in [x_l, x_{l+3})}{n}$.

$$\begin{aligned}
 Q(b, b^{(q)}) &= \mathbb{E}(l(X, Z|b)|b^{(q)}, X) \\
 &= \log(Dir(\alpha_1, \dots, \alpha_{m+2})(b)) + \sum_{j=1}^n \sum_{l=1}^{m+2} \mathbb{E}(Z_{l,j}|X, b^{(q)}) \log(b_l B_{l,h_l}^2(X_j)) \\
 &= \log(Dir(\alpha_1, \dots, \alpha_{m+2})(b)) + \sum_{j=1}^n \sum_{l=2}^{m+2} \tau_{l,j}^{(q)} \log(b_l B_{l,h_l}^2(X_j)) + \tau_{1,j}^{(q)} \log\left[\left(1 - \sum_{l=2}^{m+2} b_l\right) B_{1,h_1}^2(X_j)\right] \\
 &= (\alpha_1 - 1) \log\left(1 - \sum_{l=2}^{m+2} b_l\right) + \sum_{s=2}^{m+2} (\alpha_s - 1) \log b_s + \sum_{j=1}^n \left[\sum_{l=2}^{m+2} \tau_{l,j}^{(q)} \log(b_l B_{l,h_l}^2(X_j)) \right] + \\
 &\quad \tau_{1,j}^{(q)} \log\left[\left(1 - \sum_{l=2}^{m+2} b_l\right) B_{1,h_1}^2(X_j)\right].
 \end{aligned}$$

Étape d'espérance (E) : Cette étape permet de calculer la quantité intermédiaire $Q(b, b^{(q)})$ et d'estimer les probabilités a posteriori $\tau_{l,j}^{(q)}$ à l'itération q pour $l = 1, \dots, m + 2, j = 1, \dots, n$ par

$$\tau_{l,j}^{(q)} = \frac{b_l^{(q)} B_{l,h_l}^2(X_j)}{\sum_{s=1}^{m+2} b_s^{(q)} B_{s,h_s}^2(X_j)};$$

Étape de maximisation "MBB" : À l'itération $(q + 1)$, nous calculons le paramètre $b_l^{(q+1)}$ qui maximise :

$$b_l^{(q+1)} = \arg \max_b Q(b, b^{(q)})$$

Nous obtenons $b_l^{(q+1)}$ l'unique solution de l'équation suivante, pour $l = 2, \dots, m + 2$, définie par

$$\frac{\partial Q}{\partial b_l} = \sum_{i=1}^n \left(\tau_{l,i}^{(q)} \frac{1}{b_l} - \tau_{1,i}^{(q)} \frac{1}{b_1} \right) + \frac{\alpha_l - 1}{b_l} - \frac{\alpha_1 - 1}{b_1} = 0.$$

Cela signifie que

$$\frac{1}{b_l} \left[\sum_{i=1}^n \left(\tau_{l,i}^{(q)} \right) + (\alpha_l - 1) \right] = \frac{1}{b_1} \left[\sum_{i=1}^n \left(\tau_{1,i}^{(q)} \right) + (\alpha_1 - 1) \right]$$

Ce qui implique

$$b_l = b_1 \frac{\sum_{i=1}^n \tau_{l,i}^{(q)} + \alpha_l - 1}{\sum_{i=1}^n \tau_{1,i}^{(q)} + \alpha_1 - 1}; \quad l = 2, \dots, m + 2.$$

Puisque $\sum_{l=1}^{m+2} b_l = 1$, alors

$$b_l^{(q+1)} = \frac{1}{n - m + \alpha - 2} \left(\left(\sum_{j=1}^n \tau_{l,j}^{(q)} \right) + \alpha_l - 1 \right); \quad l = 1, \dots, m + 2 \text{ et } \alpha = \sum_{l=1}^{m+2} \alpha_l.$$

Remarque

Si $\alpha_1 = \alpha_2 = \dots = \alpha_{m+2} = 1$ alors $\alpha = \sum_{i=1}^{m+2} \alpha_i = m + 2$ et $b_l^{(q+1)} = \frac{1}{n} \sum_{j=1}^n \tau_{l,j}^{(q)}$ correspond à l'estimateur du maximum de vraisemblance (voir page 16).

Dans la continuité de notre travail, nous notons par \hat{b}_l et \hat{f} les estimateurs des coefficients b_l et de la densité de probabilité f respectivement. La liaison entre ces estimateurs est donnée par l'expression suivante :

$$\hat{f}(x) = \sum_{l=1}^{m+2} \hat{b}_l B_{l,h_l}^2(x).$$

Étude de simulation

Dans cette section, nous présentons quelques résultats numériques afin d'évaluer et de comparer la performance de l'algorithme EMBB et l'algorithme EMB. Nous avons simulé N échantillons de taille n (où $n = 10, 100, 200, \dots, 1000$) d'un mélange de B-splines. Pour l'estimation des coefficients de B-splines b_l , nous avons utilisé les deux algorithmes EMBB et EMB.

Nous avons choisi comme critère de comparaison le calcul de l'EQM entre les valeurs réelles

des coefficients b_l et les valeurs estimées \widehat{b}_l donné par l'expression suivante :

$$\text{EQM}(b, \widehat{b}) = \frac{1}{N(m+2)} \sum_{j=1}^N \sum_{l=1}^{m+2} (b_l - \widehat{b}_l^{(j)})^2.$$

où $\widehat{b}_l^{(j)}$ représente l'estimateur du coefficient b_l obtenu à partir de l'échantillon j .

Dans la figure FIG.1.11, nous présentons la variation des valeurs EQM en fonction de la taille n de l'échantillon simulé pour les deux cas. A partir des deux courbes représentées dans la figure FIG. 1.11, nous remarquons, à partir de la taille $n = 500$, les deux courbes deviennent presque similaire. Il est clair que la méthode bayésienne estime mieux les coefficients pour les échantillons de petits taille. Par contre, cette méthode donne des résultats qui ressemblent à celles non bayésien pour des échantillons de grande taille.

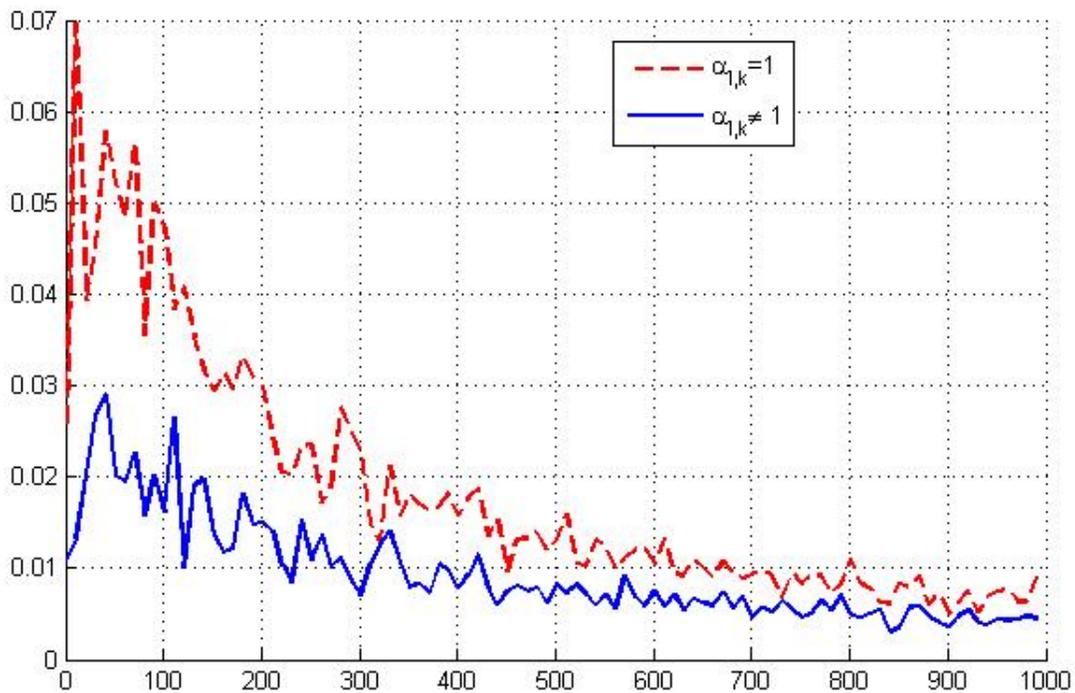


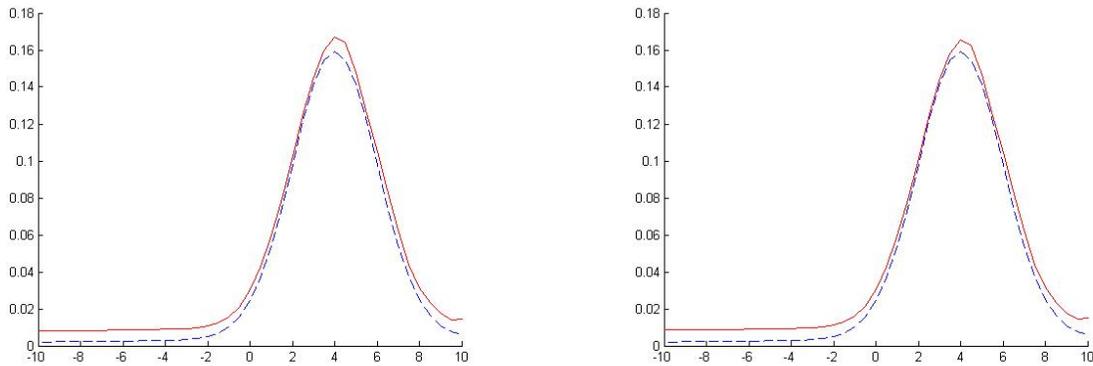
FIG. 1.11 – Variation de l'EQM en fonction de la taille de l'échantillon n .

Performance de l'algorithme EMBB

Nous présentons dans ce paragraphe quelques résultats de simulations en utilisant comme critère de performance le calcul de l'EQM, de l'AIC et de la DKL entre la vraie densité et celle estimée en utilisant respectivement les deux algorithmes EMB et EMBB.

Cas Gaussien

Nous présentons dans la figure FIG.1.12 l'estimation d'une densité de la loi Gaussienne de moyenne $\mu = 4$ et de variance $\sigma^2 = 2$ en utilisant les deux algorithmes EMB et EMBB.



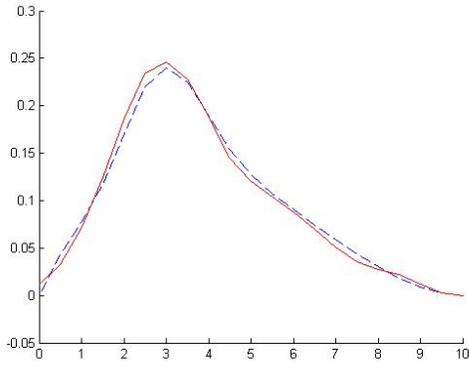
(a) EMBB

(b) EMB

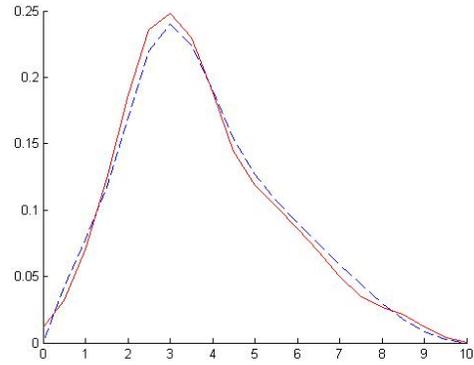
FIG. 1.12 – Estimation de la densité Gaussienne.

Cas Bêta

Nous présentons dans la figure FIG.1.13 l'estimation de la densité de la loi Bêta de paramètres $p = 3$ et $q = 5$ en utilisant les deux algorithmes EMB et EMBB.



(a) EMBB

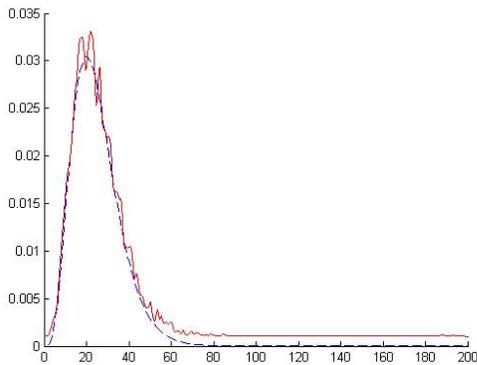


(b) EMB

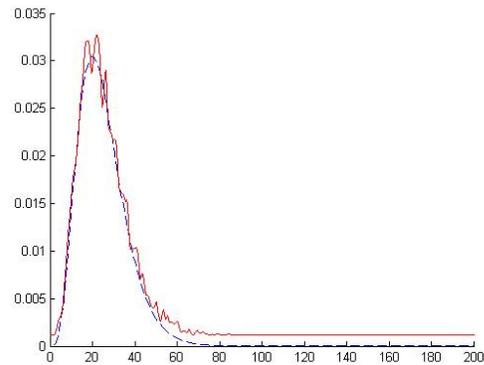
FIG. 1.13 – Estimation de la densité Bêta.

Cas Gamma

Nous présentons dans la figure FIG.1.14 l'estimation d'une densité de la loi Gamma de paramètres $a = 2$ et $\lambda = 10$ en utilisant les deux algorithmes EMB et EMBB.



(a) EMBB



(b) EMB

FIG. 1.14 – Estimation de la densité Gamma.

Nous présentons dans les Tableaux TAB.1.8, TAB.1.9 et TAB.1.10, les valeurs des EQM,

AIC et DKL pour les différentes lois obtenues par l'utilisation de ces deux algorithmes EMB et EMBB au bout de 25 itérations.

vraie densité f EQM(\hat{f})	B-spline Bayesian	B-spline
N(0, 1)	0.00299	0.0094
B(3, 5)	0.001115	0.00119
G(2, 3)	0.0012	0.0027

TAB. 1.8 – Estimation de l'EQM par les deux algorithmes EMB et EMBB.

vraie densité f AIC(\hat{f})	B-spline Bayesian	B-spline
N(0, 1)	0.125	0.136
B(3, 5)	0.638	0.755
G(2, 3)	0.332	0.562

TAB. 1.9 – Estimation du AIC par les deux algorithmes EMB et EMBB.

vraie densité f DKL(\hat{f})	B-spline Bayesian	B-spline
N(0, 1)	0.0036	0.0058
B(3, 5)	0.0029	0.0036
G(2, 3)	0.0081	0.0088

TAB. 1.10 – Estimation du DKL par les deux algorithmes EMB et EMBB.

Nous constatons, d'après les Tableaux TAB.1.8, TAB.1.9 et TAB.1.10, l'amélioration de l'estimateur de la densité de probabilités en utilisant l'algorithme EMBB. Ceci montre bien l'intérêt de l'apport de l'approche Bayésienne au niveau de l'estimation par les fonctions

B-splines.

Dans la suite, nous montrons la généralisation de ce nouvel estimateur pour le cas de mélange de lois de probabilités.

1.5.2 Estimation Bayésienne par les fonctions B-Splines d'un mélange de lois

Dans ce paragraphe, nous allons montrer l'apport de l'approche Bayésienne dans un cadre non paramétrique en utilisant les fonctions B-splines quadratique pour les cas de mélange fini de lois. Considérons la densité du mélange suivante :

$$f(x) = \sum_{k=1}^K \pi_k f(x|b_{.,k}) = \sum_{k=1}^K \pi_k \sum_{l=1}^{m_k+2} b_{l,k} B_{l,h_l}^2(x).$$

Où

$$\Theta = \left\{ (\pi_1, \dots, \pi_K, b_{.,1}, \dots, b_{.,K}); \pi_k > 0, \sum_{k=1}^K \pi_k = 1; k = 1, \dots, K; b_{.,k} \in]0, 1[^{m_k+2}; \sum_{l=1}^{m_k+2} b_{l,k} = 1 \right\}$$

et $b_{.,k} = (b_{l,k})_{l=1, \dots, m_k+2}$.

L'intégration de l'approche bayésienne dans l'estimation du mélange de lois dans un cadre non paramétrique en utilisant les fonctions B-splines, nous a permis de définir un nouveau vecteur de paramètres $\theta = (\pi_1, \dots, \pi_K, b_{.,1}, \dots, b_{.,K}) \in \Theta$. Comme par hypothèse les coefficients des B-splines $b_{1,k}, \dots, b_{m_k+2,k}, k = 1 \dots, K$, vérifient les deux conditions suivantes : $b_{l,k} \geq 0$ et $\sum_{l=1}^{m_k+2} b_{l,k} = 1$, pour $k = 1 \dots, K$. Nous pouvons supposer que le vecteur des paramètres formé par ces coefficients suit la loi de Dirichlet $Dir(\alpha_{1,k}, \dots, \alpha_{m_k+2,k})$ de vecteur des paramètres $(\alpha_{1,k}, \dots, \alpha_{m_k+2,k})$.

Pour estimer ce vecteur de paramètres, nous introduisons un nouvel algorithme itératif noté EMBGB (Estimation Maximisation Bayésienne Généralisé B-splines).

Description de l'algorithme EMBGB

Soient X_1, X_2, \dots, X_n , une suite de variables aléatoires i.i.d de même densité f . Pour chaque X_i , on associe une variable aléatoire $Z_i = (Z_{1,i}, \dots, Z_{K,i})$ de telle sorte que Z_i suit une distribution multinomiale $Mul(1, \pi_1, \dots, \pi_K)$.

Soient $X = (X_1, \dots, X_n)$ et $Z = (Z_1, \dots, Z_n)$ alors, dans la $q^{\text{ème}}$ itération, nous avons

$$l(X, Z|\theta) = \sum_{i=1}^n \left[\sum_{k=1}^K Z_{k,i} \log(\pi_k f(X_i|b_{.,k})) \right] + \sum_{k=1}^K \log Dir(\alpha_{1,k}, \dots, \alpha_{m_k+2,k})(b_{.,k}),$$

où $\theta = (\pi_1, \dots, \pi_K, b_{.,1}, \dots, b_{.,K}) \in \Theta$ et $\alpha_{l,k} \geq 1 \forall l = 1, \dots, K$.

L'algorithme de EMBGB cherche à trouver l'estimateur du maximum de vraisemblance en appliquant de manière itérative les étapes suivantes :

Étape d'initialisation : Les paramètres du mélange peuvent être initialisé comme suit

$$\pi_k = \frac{n_k}{n} \text{ et } b_{l,k} = \frac{\text{nombre d'observations } X_j \in [x_l, x_{l+3})}{n_k}$$

où n_k est le nombre total d'observations dans la $k^{\text{ème}}$ composante.

$$Q(\theta, \theta^{(q)}) = \mathbb{E}(l(X, Z|\theta)|\theta^{(q)}) = \sum_{i=1}^n \left[\sum_{k=1}^K \mathbb{E}(Z_{k,i}|X, \theta^{(q)}) \log(\pi_k f(X_i|b_{.,k})) \right] + \sum_{k=1}^K \log Dir(\alpha_{1,k}, \dots, \alpha_{m_k+2,k})(b_{.,k}).$$

Si $\tau_{k,i}^{(q)} = \mathbb{E}(Z_{k,i}|X, \theta^{(q)})$, alors

$$\begin{aligned} Q = Q(\theta, \theta^{(q)}) &= \sum_{i=1}^n \left[\sum_{k=1}^K \tau_{k,i}^{(q)} \log(\pi_k f(X_i|b_{.,k})) \right] + (\alpha_{1,k} - 1) \log b_{1,k} + \dots + (\alpha_{m_k+2,k} - 1) \log b_{m_k+2,k} \\ &= \sum_{i=1}^n \left[\sum_{k=2}^K \tau_{k,i}^{(q)} \log(\pi_k f(X_i|b_{.,k})) + \tau_{1,i}^{(q)} \log\left(1 - \sum_{k=2}^K \pi_k\right) f(X_i|b_{.,1}) \right] + \sum_{l=1}^{m_k+2} (\alpha_{l,k} - 1) \log b_{l,k}. \end{aligned}$$

Étape d'espérance (E) : Dans cette étape, et pour l'observation X_i , nous calculons la quantité intermédiaire $Q(b, b^{(q)})$ et nous estimons la probabilité a posteriori $\tau_{k,i}^{(q)}$ apparte-

nant à la $k^{\text{ème}}$ composante à la $q^{\text{ème}}$ itération par :

$$\begin{aligned}
 \tau_{k,i}^{(q)} &= \mathbb{E}(Z_{k,i} | X_i, \theta^{(q)}) \\
 &= P[Z_{k,i} = 1 | X_i, \theta^{(q)}] \\
 &= \frac{P[X_i | Z_{k,i} = 1, \theta^{(q)}] P[Z_{k,i} = 1 | \theta^{(q)}]}{\sum_{k=1}^K P[X_i | Z_{k,i} = 1, \theta^{(q)}] P[Z_{k,i} = 1 | \theta^{(q)}]} \\
 &= \frac{\pi_k^{(q)} f(X_i | b_{\cdot,k}^{(q)})}{\sum_{k=1}^K \pi_k^{(q)} f(X_i | b_{\cdot,k}^{(q)})}.
 \end{aligned}$$

Étape de maximisation (MBGB) : Nous calculons le paramètre $\theta^{(q+1)}$ qui maximise

$$\theta^{(q+1)} = \arg \max_{\theta \in \Theta} Q(\theta, \theta^{(q)}).$$

1) Calcul des probabilités a priori à l'itération $(q + 1)$.

$$\frac{\partial Q}{\partial \pi_k} = \sum_{i=1}^n \left(\frac{\tau_{k,i}^{(q)}}{\pi_k} - \frac{\tau_{1,i}^{(q)}}{1 - \sum_{k=2}^K \pi_k} \right) = 0.$$

Cela signifie que

$$\pi_k = \frac{\sum_{i=1}^n \tau_{k,i}^{(q)}}{\sum_{i=1}^n \tau_{1,i}^{(q)}} \left(1 - \sum_{k=2}^K \pi_k \right).$$

Ce qui implique

$$\sum_{k=1}^K \pi_k = \sum_{k=1}^K \frac{\sum_{i=1}^n \tau_{k,i}^{(q)}}{\sum_{i=1}^n \tau_{1,i}^{(q)}} \left(1 - \sum_{k=2}^K \pi_k \right).$$

Donc,

$$\pi_k^{(q+1)} = \frac{\sum_{i=1}^n \tau_{k,i}^{(q)}}{\sum_{k=1}^K \sum_{i=1}^n \tau_{k,i}^{(q)}} = \frac{1}{n} \sum_{i=1}^n \tau_{k,i}^{(q)}.$$

2) Calcul des coefficients B-splines à l'itération $(q + 1)$.

Nous avons

$$\frac{\partial Q}{\partial b_{l,k}} = \sum_{i=1}^n \tau_{k,i}^{(q)} \left(\frac{B_{l,h_l}^2(X_i) - B_{1,h_l}^2(X_i)}{f(X_i|b_{\cdot,k})} \right) + \frac{\alpha_{l,k} - 1}{b_{l,k}} - \frac{\alpha_{1,k} - 1}{1 - \sum_{l=2}^{m_k+2} b_{l,k}}; \quad l = 2, \dots, m_k + 2$$

et

$$\frac{\partial^2 Q}{\partial b_{l,k}^2} = - \sum_{i=1}^n \tau_{k,i}^{(q)} \left(\frac{B_{l,h_l}^2(X_i) - B_{1,h_l}^2(X_i)}{f(X_i|b_{\cdot,k})} \right)^2 - \frac{\alpha_{l,k} - 1}{(b_{l,k})^2} - \frac{\alpha_{1,k} - 1}{(1 - \sum_{l=2}^{m_k+2} b_{l,k})^2}; \quad l = 2, \dots, m_k + 2.$$

Afin de résoudre $\frac{\partial Q}{\partial b_{l,k}} = 0$, nous utilisons la méthode de Newton Raphson, d'où l'équation récursive suivante

$$b_{l,k}^{(q+1)} = b_{l,k}^{(q)} + \frac{\sum_{i=1}^n \tau_{k,i}^{(q)} \left(\frac{B_{l,h_l}^2(X_i) - B_{1,h_l}^2(X_i)}{f(X_i|b_{\cdot,k}^{(q)})} \right) + \frac{\alpha_{l,k} - 1}{b_{l,k}^{(q)}} - \frac{\alpha_{1,k} - 1}{b_{1,k}^{(q)}}}{\sum_{i=1}^n \tau_{k,i}^{(q)} \left(\frac{B_{l,h_l}^2(X_i) - B_{1,h_l}^2(X_i)}{f(X_i|b_{\cdot,k}^{(q)})} \right)^2 + \frac{\alpha_{l,k} - 1}{(b_{l,k}^{(q)})^2} + \frac{\alpha_{1,k} - 1}{(b_{1,k}^{(q)})^2}} \quad (1.6)$$

pour $l = 2, \dots, m_k + 2$ et $k = 1, \dots, K$, où $b_{1,k}^{(q)} = 1 - \sum_{l=2}^{m_k+2} b_{l,k}$.

Les étapes de calcul sont de façon itérative jusqu'à ce que nous réalisons la convergence.

1.5.3 Existence, Unicité et Propriétés de l'estimateur Bayésien pour le mélange fini de lois

Dans cette sous section, dans un premier temps, nous prouvons l'existence et l'unicité de l'estimateur $\hat{\theta} = (\hat{\pi}_1, \dots, \hat{\pi}_K, \hat{b}_{\cdot,1}, \dots, \hat{b}_{\cdot,K})$ obtenu par l'algorithme EMBGB et dans un deuxième temps, nous présentons quelques propriétés asymptotique de cet estimateur.

Existence et Unicité

Dans les propositions (1.5.1) et (1.5.2), nous montrons que l'algorithme EMBGB converge presque sûrement vers l'estimateur du maximum de vraisemblance $\hat{\theta}$.

Proposition 1.5.1. *Si $\theta^{(q+1)} = (\pi_1^{(q+1)}, \dots, \pi_K^{(q+1)}, b_{.,1}^{(q+1)}, \dots, b_{.,K}^{(q+1)})$ converge presque sûrement vers $\hat{\theta}$, quand $q \rightarrow +\infty$, alors $\hat{\theta}$ satisfait les équations de vraisemblance a posteriori.*

Preuve : D'une part, pour k fixé, la dérivée de la fonction a posteriori log-vraisemblance par rapport à $b_{l,k}$ est donnée par

$$\frac{\partial \log(p(\theta|X_1, \dots, X_n))}{\partial b_{l,k}} = \sum_{j=1}^n \pi_k \frac{(B_{l,h_l}^2(X_j) - B_{1,h_1}^2(X_j))}{f(X_j)} + \frac{\alpha_{l,k} - 1}{b_{l,k}} - \frac{\alpha_{1,k} - 1}{1 - \sum_{l=2}^{m_k+2} b_{l,k}};$$

pour tout $l = 2, \dots, m_k + 2$.

Quand $q \rightarrow +\infty$ dans l'équation (1.6) et en utilisant le fait que $b_{l,k}^{(q)} \rightarrow \hat{b}_{l,k}$ et

$$\tau_{k,j}^{(q)} \rightarrow \hat{\tau}_{k,j} = \frac{\hat{\pi}_k \hat{f}_k(X_j/b_{.,k})}{\sum_{k=1}^K \hat{\pi}_k \hat{f}_k(X_j/b_{.,k})}, \text{ où } \hat{f}(X_j) = \sum_{k=1}^K \hat{\pi}_k \hat{f}_k(X_j/b_{.,k}) \text{ et } \hat{f}_k(X_j/b_{.,k}) = \sum_{j=1}^n \hat{b}_{l,k} B_{l,h_l}^2(X_j).$$

Nous obtenons

$$\sum_{j=1}^n \hat{\tau}_{k,j} \frac{(B_{l,h_l}^2(X_j) - B_{1,h_1}^2(X_j))}{\hat{f}_k(X_j/b_{.,k})} + \frac{\alpha_{l,k} - 1}{\hat{b}_{l,k}} - \frac{\alpha_{1,k} - 1}{1 - \sum_{l=2}^{m_k+2} \hat{b}_{l,k}} = 0; \quad l = 2, \dots, m_k + 2.$$

Donc

$$\sum_{j=1}^n \hat{\pi}_k \frac{(B_{l,h_l}^2(X_j) - B_{1,h_1}^2(X_j))}{\hat{f}(X_j)} + \frac{\alpha_{l,k} - 1}{\hat{b}_{l,k}} - \frac{\alpha_{1,k} - 1}{1 - \sum_{l=2}^{m_k+2} \hat{b}_{l,k}} = 0.$$

D'où, $\widehat{b}_{l,k}$ satisfait l'équation de vraisemblance a posteriori.

D'autre part,

$$\widehat{\pi}_k = \frac{1}{n} \sum_{j=1}^n \frac{\widehat{\pi}_k \widehat{f}_k(X_j/b_{.,k})}{\sum_{k=1}^K \widehat{\pi}_k \widehat{f}_k(X_j/b_{.,k})}.$$

Ceci implique que

$$\begin{aligned} 0 &= n - \sum_{j=1}^n \frac{\widehat{f}_k(X_j/b_{.,k})}{\sum_{k=1}^K \widehat{\pi}_k \widehat{f}_k(X_j/b_{.,k})} \\ &= \sum_{j=1}^n \frac{\widehat{\pi}_1 \widehat{f}_1(X_j/b_{.,1}) + \widehat{\pi}_2 \widehat{f}_2(X_j/b_{.,2}) + \dots + \widehat{\pi}_{K-1} \widehat{f}_{K-1}(X_j/b_{.,K-1})}{\sum_{k=1}^K \widehat{\pi}_k \widehat{f}_k(X_j/b_{.,k})} + \\ &\quad \sum_{j=1}^n \frac{(1 - \widehat{\pi}_1 - \widehat{\pi}_2 - \dots - \widehat{\pi}_{K-1}) \widehat{f}_K(X_j/b_{.,K}) - \widehat{f}_k(X_j/b_{.,k})}{\sum_{k=1}^K \widehat{\pi}_k \widehat{f}_k(X_j/b_{.,k})} \\ &= \sum_{j=1}^n \frac{\widehat{\pi}_1 (\widehat{f}_1(X_j/b_{.,1}) - \widehat{f}_K(X_j/b_{.,K})) + \dots + \widehat{\pi}_{K-1} (\widehat{f}_{K-1}(X_j/b_{.,K-1}) - \widehat{f}_K(X_j/b_{.,K}))}{\sum_{k=1}^K \widehat{\pi}_k \widehat{f}_k(X_j/b_{.,k})} + \\ &\quad \sum_{j=1}^n \frac{\widehat{f}_K(X_j) - \widehat{f}_k(X_j/b_{.,k})}{\sum_{k=1}^K \widehat{\pi}_k \widehat{f}_k(X_j/b_{.,k})}. \end{aligned} \tag{1.7}$$

Soit

$$H = \sum_{j=1}^n \frac{\widehat{\pi}_1 (\widehat{f}_1(X_j/b_{.,1}) - \widehat{f}_K(X_j/b_{.,K})) + \dots + \widehat{\pi}_{K-1} (\widehat{f}_{K-1}(X_j/b_{.,K-1}) - \widehat{f}_K(X_j/b_{.,K}))}{\sum_{k=1}^K \widehat{\pi}_k \widehat{f}_k(X_j/b_{.,k})}. \tag{1.8}$$

Ensuite, multipliant l'équation (1.7) par $\widehat{\pi}_k$, pour $k = 1, \dots, K - 1$, et en additionnant les résultats des $(K - 1)$ multiplications successives, nous obtenons

$$(\widehat{\pi}_1 + \widehat{\pi}_2 + \dots + \widehat{\pi}_{K-1})H = H.$$

Comme $\sum_{k=1}^K \widehat{\pi}_k = 1$, alors $(1 - \widehat{\pi}_K)H = H$ et $\widehat{\pi}_K H = 0$.

Notons que nous avons choisi $\widehat{\pi}_K$ arbitrairement dans l'équation (1.8). Nous aurons pu choisir un des $\widehat{\pi}_k$. Donc, nous pouvons dire que $\widehat{\pi}_k H = 0, \forall k = 1, \dots, K$. Ceci donne, $H = 0$. Ainsi,

$$\sum_{j=1}^n \frac{\widehat{f}_k(X_j/b_{.,k}) - \widehat{f}_K(X_j/b_{.,K})}{\sum_{k=1}^K \widehat{\pi}_k \widehat{f}_k(X_j/b_{.,k})} = 0, \quad \forall k = 1, \dots, K - 1.$$

Comme

$$\frac{\partial \log(p(\theta|X_1, \dots, X_n))}{\partial \pi_k} = \sum_{j=1}^n \frac{f_k(X_j/b_{.,k}) - f_K(X_j/b_{.,K})}{\sum_{k=1}^K \pi_k f_k(X_j/b_{.,k})}, \quad \forall k = 1, \dots, K - 1.$$

Donc, les paramètres $\widehat{\pi}_1, \widehat{\pi}_2, \dots, \widehat{\pi}_K$ estimés par l'algorithme EMBGB satisfont les équations de vraisemblance a posteriori. \square

Proposition 1.5.2. *Supposons qu'une subdivision de l'intervalle de la densité contenant au moins trois observations, alors la matrice Hessienne $\left(\frac{\partial^2 \log(p(\theta|X_1, \dots, X_n))}{\partial b_{l,k} \partial b_{p,k}} \right)_{2 \leq l, p \leq m_k + 2}$ est définie négative.*

La preuve de la proposition 1.5.2 nécessite le lemme suivant.

Lemme 1.5.3. *la matrice Hessienne $\left(\frac{\partial^2 l(X_1, \dots, X_n|\theta)}{\partial b_{l,k} \partial b_{p,k}} \right)_{2 \leq l, p \leq m_k + 2}$ est définie négative.*

Preuve du Lemme 1.5.3. Les première et deuxième dérivations de la fonction log-vraisemblance sont données par

$$\frac{\partial l(X_1, \dots, X_n|\theta)}{\partial b_{l,k}} = \sum_{j=1}^n \pi_k \frac{B_{l,h}^2(X_j) - B_{1,h_1}^2(X_j)}{f(X_j)}$$

et

$$\frac{\partial^2 l(X_1, \dots, X_n|\theta)}{\partial b_{l,k} \partial b_{p,k}} = - \sum_{j=1}^n \pi_k^2 \frac{(B_{l,h_l}^2(X_j) - B_{1,h_1}^2(X_j))(B_{p,h_p}^2(X_j) - B_{1,h_1}^2(X_j))}{(f(X_j))^2}; \quad l, p = 2, \dots, m_k + 2.$$

Ceci impliquent que

$$\left(\frac{\partial^2 l(X_1, \dots, X_n | \theta)}{\partial b_{l,k} \partial b_{p,k}} \right)_{2 \leq l, p \leq m_k + 2} = - \sum_{j=1}^n T_j \cdot T_j^t,$$

où $T_j = (t_{2,j}, t_{3,j}, \dots, t_{m_k+2,j})^t$, $t_{l,j} = \pi_k \frac{B_{l,h_l}^2(X_j) - B_{1,h_1}^2(X_j)}{f(X_j)}$; $l = 2, \dots, m_k + 2$

et où T_j^t désigne la transposée de T_j .

Maintenant, nous allons prouver que $\sum_{j=1}^n T_j \cdot T_j^t$ est définie positive. Puisque

$$\sum_{j=1}^n y'(T_j \cdot T_j^t)y = \sum_{j=1}^n (y_2 t_{2,j} + y_3 t_{3,j} + \dots + y_{m_k+2} t_{m_k+2,j})^2 \geq 0, \quad \forall y = (y_2, y_3, \dots, y_{m_k+2})^t \neq 0.$$

Supposons que $y \neq 0$ et $\sum_{j=1}^n y'(T_j \cdot T_j^t)y = 0$, ceci implique

$$y_2 t_{2,j} + y_3 t_{3,j} + \dots + y_{m_k+2} t_{m_k+2,j} = 0, \quad \text{pour } j = 1, \dots, n.$$

Puis, par l'introduction de l'expression de $t_{l,j}$, nous avons

$$- \left(\sum_{i=2}^{m_k+2} y_i \right) B_{1,h_1}^2(X_j) + y_2 B_{2,h_2}^2(X_j) + \dots + y_{m_k+2} B_{m_k+2,h_{m_k+2}}^2(X_j) = 0, \quad \forall j = 1, \dots, n.$$

Par hypothèse chaque intervalle de la subdivision contient au moins 3 observations. Alors, $x \mapsto \left(\sum_{i=2}^{m_k+2} y_i \right) B_{1,h_1}^2(x) + y_2 B_{2,h_2}^2(x) + \dots + y_{m_k+2} B_{m_k+2,h_{m_k+2}}^2(x)$ est un polynôme ayant au moins trois racines. En outre, le degré de cette fonction polynomiale est inférieur à 2.

Nécessairement,

$$\left(\sum_{i=2}^{m_k+2} y_i \right) B_{1,h_1}^2(x) + y_2 B_{2,h_2}^2(x) + \dots + y_{m_k+2} B_{m_k+2,h_{m_k+2}}^2(x) = 0; \quad \forall x \in \mathbb{R}.$$

Puisque $\left(B_{l,h_l}^2(x) \right)_{1 \leq l \leq m_k+2}$ sont linéairement indépendants, alors $y_2 = \dots = y_{m_k+2} = 0$.

Par conséquent, la matrice Hessienne $\left(\frac{\partial^2 l(X_1, \dots, X_n | \theta)}{\partial b_{l,k} \partial b_{p,k}} \right)_{2 \leq l, p \leq m_k+2}$ est définie négative.

□

Preuve de la proposition 1.5.2. Tout d'abord, la fonction définie par

$$b_{.,k} \longmapsto \log(\text{Dir}(\alpha_1, \dots, \alpha_{m_k+2})(b_{.,k}))$$

est une fonction concave pour tous $\alpha_{l,k} \geq 1$, $k = 1, \dots, K$. Puis, en utilisant le Lemme 1.5.3., la fonction log a postériori

$$b \longmapsto \log(p(\theta|X_1, \dots, X_n)) = l(X_1, \dots, X_n|\theta) + \sum_{k=1}^K \log(\text{Dir}(\alpha_1, \dots, \alpha_{m_k+2})(b_{.,k})),$$

est une fonction concave. \square

Proposition 1.5.4. *Supposons qu'il existe un intervalle de la subdivision contenant au moins trois observations, alors la matrice Hessienne $\left(\frac{\partial^2 \log(p(\theta|X_1, \dots, X_n))}{\partial \pi_k \partial \pi_s}\right)_{2 \leq k, s \leq K}$ est définie négative.*

Preuve. En calculant les dérivées partielles du premier ordre et second ordre de la fonction log a postériori, nous obtenons

$$\frac{\partial \log(p(\theta|X_1, \dots, X_n))}{\partial \pi_k} = \sum_{j=1}^n \frac{f_k(X_j/b_{.,k}) - f_1(X_j/b_{.,1})}{f(X_j)}; \quad k, s = 2, \dots, K,$$

et

$$\frac{\partial^2 \log(p(\theta|X_1, \dots, X_n))}{\partial \pi_k \partial \pi_s} = - \sum_{j=1}^n \frac{(f_k(X_j/b_{.,k}) - f_1(X_j/b_{.,1}))(f_s(X_j/b_{.,s}) - f_1(X_j/b_{.,1}))}{f^2(X_j)}; \quad k = 2, \dots, K.$$

Ceci impliquent que

$$\left(\frac{\partial^2 \log(p(\theta|X_1, \dots, X_n))}{\partial \pi_k \partial \pi_s}\right)_{2 \leq k, s \leq K} = - \sum_{j=1}^n U_j \cdot U_j^t,$$

où $U_j = (u_{2,j}, u_{3,j}, \dots, u_{K,j})^t$ et $u_{k,j} = \left(\frac{f_k(X_j/b_{.,k}) - f_1(X_j/b_{.,1})}{f(X_j)}\right)_{2 \leq k \leq K}$.

Nous remarquons que :

$$\sum_{j=1}^n y'(U_j \cdot U_j^t)y = \sum_{j=1}^n (y_2 u_{2,j} + y_3 u_{3,j} + \dots + y_K u_{K,j})^2 \geq 0, \text{ for any } y = (y_2, y_3, \dots, y_K)^t \neq 0.$$

Maintenant, supposons que $y \neq 0$ et $\sum_{j=1}^n y'(U_j.U_j^t)y = 0$. Cela signifie que

$$y_2u_{2,j} + y_3u_{3,j} + \dots + y_Ku_{K,j} = 0 \text{ for each } j = 1, \dots, n.$$

Alors

$$-\left(\sum_{i=2}^K y_i\right) f_1(X_j/b_{.,1}) + y_2 f_2(X_j/b_{.,2}) + \dots + y_K f_K(X_j/b_{.,K}) = 0, \forall j = 1, \dots, n.$$

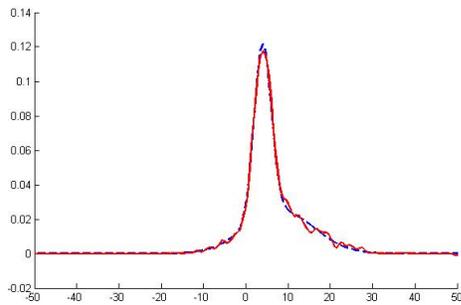
Étant donné que chaque intervalle de la subdivision contient au moins une observation X_j , alors $y_k f_k(X_j/b_{.,k}) = 0$. Donc, $y_k = 0$. Cela signifie que $y_2 = y_3 = \dots = y_K = 0$. Par conséquent, la matrice Hessienne $\left(\frac{\partial^2 \log(p(\theta|X_1, \dots, X_n))}{\partial \pi_k \partial \pi_s}\right)_{2 \leq k, s \leq K}$ est définie négative. \square

Performance de l'estimateur du mélange proposé

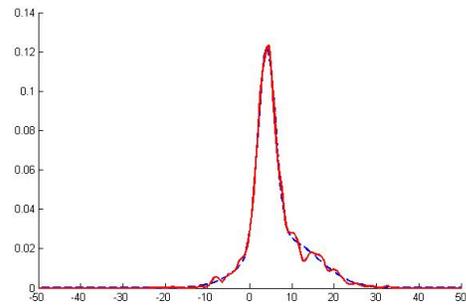
Afin de montrer la performance de l'algorithme EMGBB, nous présentons quelques résultats de simulations. D'abord, nous simulons un mélange de lois connues. Puis, nous estimons la densité de ce mélange en utilisant les deux algorithmes (EMGB, EMGBB). Dans le cas de l'approche EMGBB. Dans le cas de l'approche bayésienne, les hyperparamètres α_k de la loi à priori de Dirichlet vont être estimés à partir de la méthode d'histogramme. Enfin, nous comparons les estimateurs de la densité mélange obtenus par ces deux algorithmes. Cette comparaison est réalisée tout en calculant l'EQM, l'AIC et la DKL entre la densité de mélange simulée et celle estimée.

Mélange de lois Gaussiennes

Nous présentons dans la figure FIG.1.15 l'estimation de la densité de mélange de deux lois Gaussiennes de paramètres $\mu_1 = 4$, $\mu_2 = 8$, $\sigma_1^2 = 2$ et $\sigma_2^2 = 2$ pondérées par les probabilités a priori $\pi_1 = 0.5$, $\pi_2 = 0.5$ par les deux algorithmes EMGB et EMBGB.



(a) EMBB

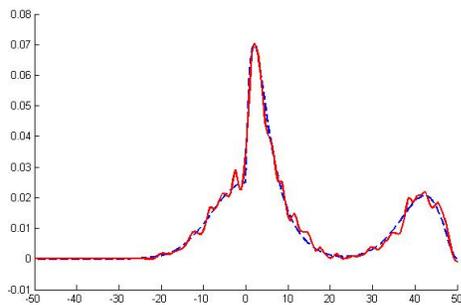


(b) EMB

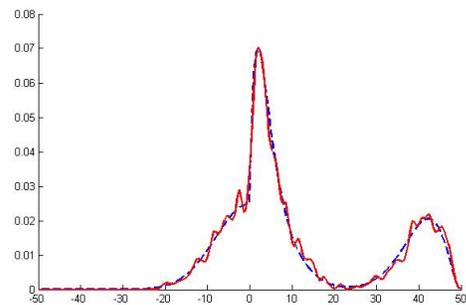
FIG. 1.15 – Estimation non paramétrique de la densité de mélange des lois.

Mélange de lois Bêta, Gaussienne et Gamma

Nous présentons dans la figure FIG.1.16 l'estimation de la densité de mélange d'une loi Bêta de paramètre $p = 12$ et $q = 3$, d'une loi Gaussienne de paramètres $\mu = 0$ et de variance $\sigma^2 = 8$ et d'une loi Gamma de paramètres $a = 2$ et $\lambda = 2$ pondérées par les probabilités a priori $\pi_1 = 0.25$, $\pi_2 = 0.5$ et $\pi_3 = 0.25$ par les deux algorithmes EMGB et EMBGB.



(a) EMBB



(b) EMB

FIG. 1.16 – Estimation non paramétrique de la densité mélange des lois.

Mélange de lois Bêtas et Gamma

Nous présentons dans la figure FIG.1.17 l'estimation de la densité de mélange de deux

lois Bêtas de paramètres $p_1 = 12$, $p_2 = 9$, $q_1 = 3$ et $q_2 = 320$ et d'une loi Gamma de paramètres $a = 2$ et $\lambda = 2$ pondérées par les probabilités a priori $\pi_1 = 0.25$, $\pi_2 = 0.5$ et $\pi_3 = 0.25$ par les deux algorithmes EMGB et EMBGB.

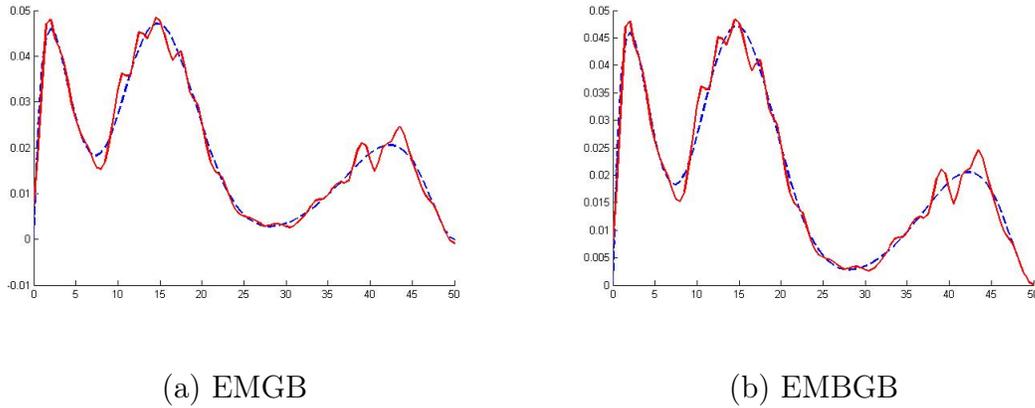


FIG. 1.17 – Estimation non paramétrique de la densité mélange des lois.

Nous présentons dans les Tableaux TAB.1.11, TAB 1.12 et TAB 1.13 les valeurs des EQMs, AIC et DLK pour les différentes lois obtenues par l'utilisation de ces deux algorithmes EMGB et EMBGB au bout de 25 itérations.

Vraie densité de mélange f EQM(\hat{f})	EMGB	EMBGB
$N(4, 2)$ et $N(8, 8)$	3.0935×10^{-6}	3.0813×10^{-6}
$B(12, 3)$, $N(0, 8)$ et $G(2, 2)$	5.5947×10^{-6}	5.5523×10^{-6}
Mélange $B(12, 3)$, $B(9, 20)$ et Gamma $G(2, 2)$	8.9558×10^{-6}	8.8895×10^{-6}

TAB. 1.11 – Estimation de l'EQM par les deux algorithmes EMGB et EMBGB.

Nous constatons d'après les tableaux TAB.1.11, TAB.1.12 et TAB.1.13 l'amélioration de l'estimateur de la densité de probabilités en utilisant l'algorithme EMBGB. Ceci prouve l'intérêt de l'apport de l'approche Bayésienne au niveau de l'estimation par les fonctions B-splines.

Vraie densité de mélange f $AIC(\hat{f})$	EMGB	EMBGB
$N(4, 2)$ et $N(8, 8)$	3.079×10^{-4}	3.078×10^{-4}
$B(12, 3)$, $N(0, 8)$ et $G(2, 2)$	8.61×10^{-5}	8.59×10^{-5}
Mélange $B(12, 3)$, $B(9, 20)$ et Gamma $G(2, 2)$	1.633×10^{-5}	1.628×10^{-5}

TAB. 1.12 – Estimation du AIC par les deux algorithmes EMGB et EMBGB.

Vraie densité de mélange f $DKL(\hat{f})$	EMGB	EMBGB
$N(4, 2)$ et $N(8, 8)$	0.0245	0.0128
$B(12, 3)$, $N(0, 8)$ et $G(2, 2)$	0.0384	0.0294
Mélange $B(12, 3)$, $B(9, 20)$ et Gamma $G(2, 2)$	0.0127	0.0095

TAB. 1.13 – Estimation du DKL par les deux algorithmes EMGB et EMBGB.

1.6 Conclusion

Ce chapitre a traité le problème de l'estimation non paramétrique des lois de probabilités. Dans un premier temps, nous avons rappelé quelques méthodes d'estimation non paramétrique telles que la méthode d'histogramme, la méthode des noyaux, la méthode des fonctions orthogonales. Dans un deuxième temps, nous avons rappelé les fonctions B-splines ainsi la description des B-splines quadratiques. Dans un troisième temps, d'abord nous avons supposé que la densité inconnue f a été approchée par un mélange des fonctions B-spline quadratique. Puis, nous avons proposé un nouvel estimateur de la densité inconnue f basé sur les fonctions B-splines quadratiques, en appliquant deux méthodes d'estimation. La première est basée sur la méthode du maximum de vraisemblance et la deuxième est basée sur la méthode d'estimation Bayésienne MAP. Ensuite, nous avons généralisé notre étude d'estimation dans le cadre du mélange et nous avons proposé un

nouvel estimateur du mélange de lois inconnues basé sur les deux méthodes d'estimation adaptées.

Chapitre 2

Segmentation statistique non paramétrique des images

2.1 Introduction

L'analyse d'une image est souvent envisagée comme une procédure de décomposition en régions homogènes qu'il s'agit ensuite d'identifier ou de reconnaître. De nombreuses techniques de segmentation s'appuient sur des décisions prises au niveau des points image (pixels) sur la base des propriétés de ces points eux-mêmes, de leur voisinage et/ou de leur arrangement spatial. Ces méthodes, qui font souvent appel à des modèles décisionnels de nature statistique, permettant de regrouper les points d'une même image en classes homogènes vis-à-vis de certains critères visuels. Il s'agit ensuite de simplifier l'image en la partitionnant en régions distinctes. Dans le cas simple, chaque région est constituée de points d'une image appartenant à la même classe et adjacents les uns aux autres.

Dans ce chapitre, nous montrons l'intérêt que peuvent apporter les fonctions B-splines en

segmentation statistique des images, avec un accent particulier mis sur l'utilisation des modèles Markoviens. La segmentation d'images est une opération de traitement d'images qui a pour but de rassembler des pixels entre eux suivant des critères pré-définis. La segmentation est une étape primordiale en traitement d'image. La segmentation statistique des images trouve de nombreuses applications dans les domaines de traitement d'images les plus divers [15][11][21]. Dans le cadre de notre travail de recherche, nous nous plaçons dans le cas de la segmentation statistique non supervisée des images où nous n'avons aucune connaissance a priori sur les régions qui représentent les différentes classes de l'image [29]. Les algorithmes les plus utilisés dans le cadre de la segmentation non supervisée sont issus de la famille d'algorithmes basés sur l'algorithme EM [39]. En effet, le niveau de gris de l'image à segmenter est souvent supposé suivre un mélange de lois connues [12][97]. Cependant, lorsque les distributions du mélange ne peuvent pas être approchées par des lois connues, des approches non paramétriques peuvent résoudre le problème [62][63][64][65][66][114][115]. Notre étude réside dans l'utilisation des fonctions B-splines en segmentation statistique non supervisée à fin de régulariser et d'assurer une meilleure identification du mélange. Dans ce chapitre, nous présentons tout d'abord la modélisation probabiliste en classification. Puis, nous rappelons le modèle Markovien en analyse d'images. Ensuite, nous décrivons le modèle de Markov caché dans un cadre non paramétrique en utilisant les fonctions B-splines. Enfin, nous montrons l'apport de cette hybridation entre le modèle de Markov caché et les fonctions B-splines en segmentation statistique semi supervisée des images.

2.2 Modélisation probabiliste en classification

Les premières approches qui ont été proposées en classification sont des méthodes heuristiques ou géométriques et reposaient essentiellement sur la dissimilarité entre les objets à classer. L'approche statistique, récente, se base sur des modèles probabilistes qui formalisent l'idée de classe. Les méthodes statistiques en classification ont pour but de construire des partitions d'ensembles d'objets décrits par des variables de telle sorte que les classes soient les plus homogènes possibles. L'application des méthodes statistiques à la reconnaissance des formes a été formalisée par Chow en 1965 [30]. Plusieurs systèmes existent de nos jours s'inspirant de ces méthodes. Le domaine des méthodes statistiques est très vaste, on se limite dans ce chapitre à l'étude de la classification statistique Bayésienne. Cette méthode suppose que le problème de représentation des formes admet un modèle probabiliste. Elle définit l'appartenance d'une forme à une classe avec un minimum d'erreur et permet d'évaluer le risque de la décision à prendre. Dans ce type d'approche, les modèles de distributions des observations sont supposés connus a priori. Sous cette hypothèse paramétrique, le problème de l'analyse des données peut être ramené à celui de la détermination des paramètres d'un mélange de fonctions de densité représentant les distributions des observations provenant de chacune des classes en présence dans l'échantillon analysé. Le problème de l'estimation des paramètres d'un mélange de fonctions de densité à partir d'un échantillon d'observations représentatif de ce mélange a simultanément été abordé par des techniques d'estimation par le maximum de vraisemblance.

2.2.1 Différentes types de classifications

Dans la littérature, il existe trois types de classifications : supervisée, non supervisée et semi-supervisée.

Classification supervisée : Dans cette méthode de classification, le classifieur nécessite une phase d'apprentissage. L'apprentissage est dit supervisé si les différentes classes des objets sont connues a priori et si la tâche d'apprentissage est guidée par un superviseur ou un professeur. L'apprentissage se déroule de la façon suivante : on choisit un sous-ensemble de l'ensemble des objets à reconnaître. Chacun de ces objets est analysé, puis la description de chaque objet est prise en compte par la tâche d'apprentissage. Le superviseur indique le nom de l'objet entré, c'est à dire de la classe dans laquelle il souhaite le voir rangé. Le travail d'apprentissage consiste alors à analyser les ressemblances entre les objets d'une même classe et les dissemblances entre les objets de classes différentes pour en réduire les classes d'apprentissage avec les meilleures séparatrices possibles.

Classification non supervisée : Il s'agit, à partir d'échantillons de référence et de règles de regroupement ou de modélisation, de construire automatiquement les classes ou les modèles sans intervention de l'opérateur. Ce mode d'apprentissage nécessite un nombre élevé d'échantillons et des règles de construction précises et non contradictoires pour bien assurer la formation des classes. Il évite l'assistance d'un opérateur mais n'assure pas toujours une classification correspondante à la réalité.

Classification semi-supervisée : L'apprentissage semi-supervisé est une classe de techniques d'apprentissage automatique qui utilise un ensemble de données étiquetées et non étiquetés. Il se situe ainsi entre l'apprentissage supervisé qui n'utilise que des données

étiquetées et l'apprentissage non-supervisé qui n'utilise que des données non-étiquetées. Il a été démontré que l'utilisation de données non-étiquetées, en combinaison avec des données étiquetées, permet d'améliorer significativement la qualité de l'apprentissage. Un autre intérêt provient du fait que l'étiquetage des données nécessite l'intervention d'un utilisateur humain. Lorsque les jeux de données deviennent très grand, cette opération peut s'avérer fastidieuse. Dans ce cas, l'apprentissage semi-supervisé, qui ne nécessite que quelques étiquettes, revêt un intérêt pratique évident.

2.2.2 Modèle aléatoire de l'image

Le modèle statistique le plus couramment utilisé en segmentation est le suivant : une image est modélisée par un couple $(X = (X_s)_{s \in S}, (Y_s)_{s \in S})$ de champs aléatoires, où S représente le plus souvent l'ensemble des pixels d'une image : $S = \{s(i, j); 1 \leq i \leq N_c \text{ et } 1 \leq j \leq N_l\}$. L'ensemble $\{(X_s)_{s \in S}\}$ représente le champ d'observations où chaque X_s est une variable aléatoire réelle. L'ensemble $\{(Y_s)_{s \in S}\}$ définit le champ de classes où chaque Y_s est à valeur dans un ensemble fini de classes $\Omega = \{1, \dots, K\}$ avec K le nombre de classes. Ce modèle est un modèle hiérarchique [74] composé de deux niveaux de champs aléatoires, c'est à dire, le processus de la région cachée Y et le processus observable X . A chaque classe k de Ω sont associés les quantités suivantes :

- π_k appelée la probabilité a priori d'appartenance à la classe k :

$$P(Y = k) = \pi_k, 1 \leq k \leq K. \quad (2.1)$$

• f_k : la densité de probabilités conditionnelle à la classe k et qui est donnée par la formule suivante :

$$f_k(x) = f_{(X/Y=k)}(x). \quad (2.2)$$

• La probabilité d'appartenance à la classe k conditionnelle à une observation x est connue sous le nom de la loi a posteriori du numéro de la forme, est notée par :

$$\pi_x(k) = p(Y = k/X = x) = \frac{\pi_k f_k(x)}{f(x)} \quad (2.3)$$

où f représente la densité mélange donnée par :

$$f(x) = \sum_{k=1}^K \pi_k f_k(x). \quad (2.4)$$

• $d : \mathbb{R} \rightarrow \Omega$, la fonction de décision qui, à chaque représentation x de \mathbb{R} , associe une classe $d(x)$ de Ω . La fonction de décision idéale notée d_0 est celle qui, à toute forme, associe sa vraie classe : $d_0(x) = \omega$; $\omega \in \Omega$.

Décision Bayésienne : En pratique, on trouve rarement une fonction de décision idéale.

Une solution consiste alors à choisir la fonction de décision Bayésienne g^* définie comme suit :

$$g^*(x) = \arg\left\{ \max_{1 \leq k \leq K} [\pi_k f_k(x)] \right\}. \quad (2.5)$$

En analyse discriminante, le problème revient à minimiser la probabilité d'erreur notée e [70]. L'erreur minimale e^* étant alors égale à :

$$e^* = 1 - \int \sup_{1 \leq k \leq K} \left[\frac{\pi_k f_k(x)}{f(x)} \right] dx. \quad (2.6)$$

La décision Bayésienne nécessite la connaissance des probabilités a priori π_k des classes et les densités de probabilités f_k de chaque classe. En pratique, ces quantités ne sont pas connues et doivent être estimées.

Application à l'image

Nous avons testé les deux algorithmes EM et EMBGB présentés dans le chapitre 1 pour estimer la densité mélange d'une image numérique. Nous présentons dans les figures (FIG. 2.1 et FIG. 2.2) les résultats de l'estimation des densités mélanges des deux images de taille 256×256 pixels obtenu au bout de 15 itérations. Nous constatons à partir de ces résultats que la densité mélange qui caractérise l'image est mieux estimée en utilisant l'algorithme EMBGB que l'algorithme EM (voir Fig. 2.1 et Fig. 2.2). (voir figures FIG. 2.1 et FIG. 2.2).

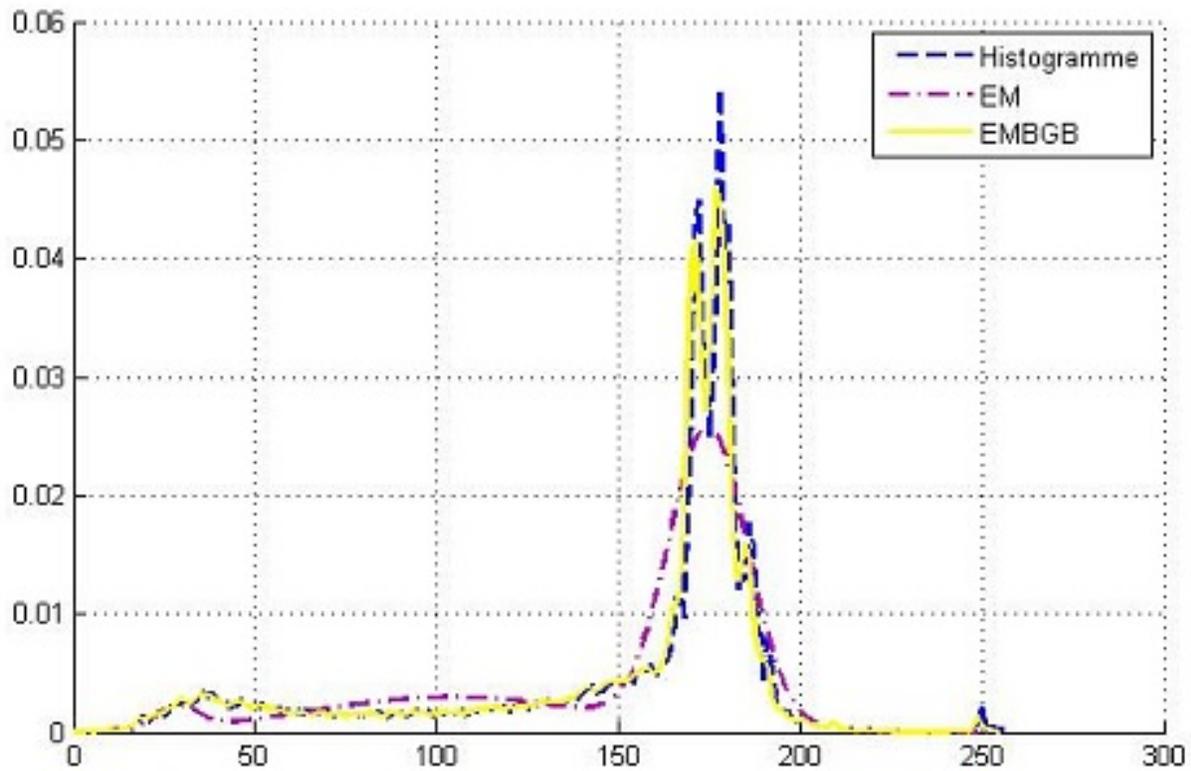


FIG. 2.1 – Histogramme et densités mélanges estimées en utilisant les algorithmes EM et EMBGB pour l'image Oiseau.

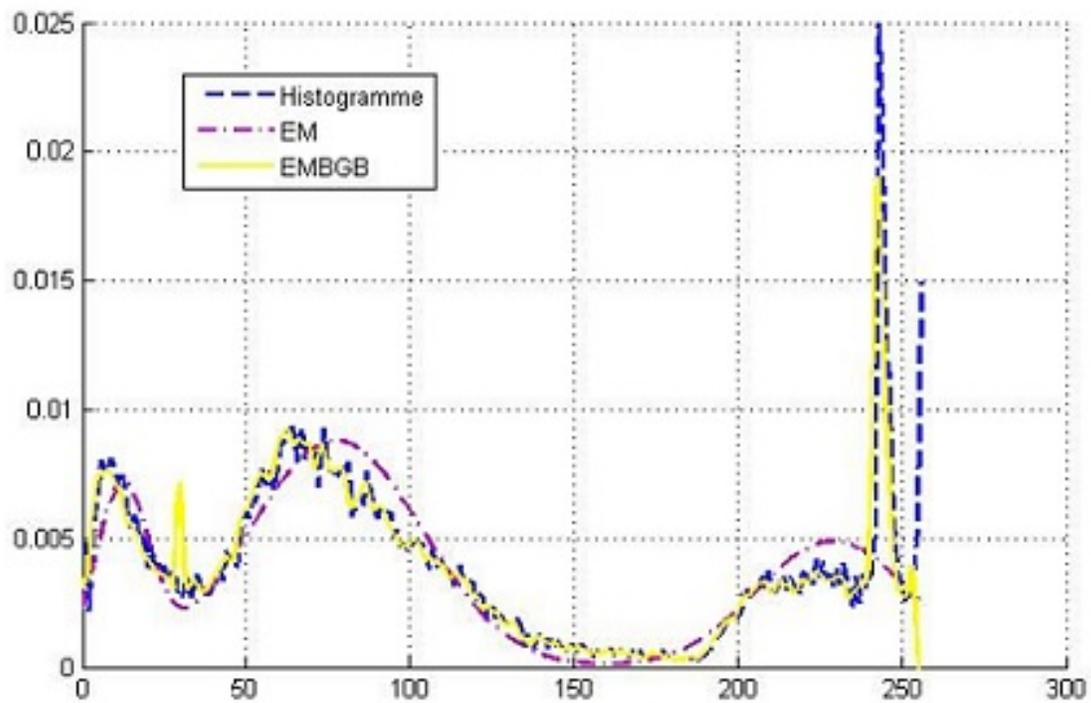


FIG. 2.2 – Histogramme et densités mélanges estimées en utilisant les algorithmes EM et EMBGB pour l'image Laq.

2.3 Modèle Markovien en analyse des images

Les champs de Markov sont utilisés depuis maintenant une vingtaine d'années en traitement d'images et font désormais partie des techniques de base de cette discipline. Nés à l'origine dans le cadre de la physique statistique pour étudier les phénomènes de transition de phase [71], ils sont rapidement appliqués aux réseaux bidimensionnels que constituent les images. Les premiers modèles sont restreints par des contraintes de causalité [1][73] et donc limités jusqu'à l'article fondateur de Geman et Geman en 84 [59] qui ouvre réellement leur utilisation en traitement d'images. Cette utilisation ira croissant avec les années et donnera lieu à de nombreuses recherches sur les problèmes que suscite leur utilisation (calcul des paramètres des modèles, techniques de simulation et de recherche de solutions rapides, etc...). A côté de ces travaux toujours d'actualité, se sont ouvertes de nouvelles recherches dédiées à des tâches de plus haut niveau et visant à l'interprétation des images. L'utilisation des techniques Markoviennes en traitement d'images demeure aujourd'hui un point fondamentale dans le domaine de la segmentation, la restauration ou encore de la modélisation stochastique.

2.3.1 Description de l'image

L'image est formée d'un ensemble fini S de sites s_i correspondant aux pixels. S est donc essentiellement un réseau discret fini, partie de \mathbb{Z}^d , si on note d la dimension de l'espace ($d = 2$ le plus classiquement, $d = 3$ pour les volumes, etc...). A chaque site est associée un descripteur, représentant l'état du site et qui peut être son niveau de gris, une étiquette, ou une information plus complexe, et prenant ses valeurs dans Ω .

La notion d'interactions locales nécessite de structurer les relations spatiales entre les différents sites du réseau. Pour ce faire, on munit S d'un système de voisinage V défini de la façon suivante :

$$\begin{cases} s \notin V_s \\ t \in V_s \implies s \in V_t \end{cases} \quad (2.7)$$

A partir d'un système de voisinage, un système de cliques peut être déduit : une clique est soit un singleton de S , soit un ensemble de sites tous voisins les uns des autres. On notera C l'ensemble des cliques relatif à V , et C_k l'ensemble des cliques de cardinal k . En fonction du système de voisinage utilisé, le système de cliques sera différent et fera intervenir plus ou moins de sites comme illustré dans la figure FIG.2.3.

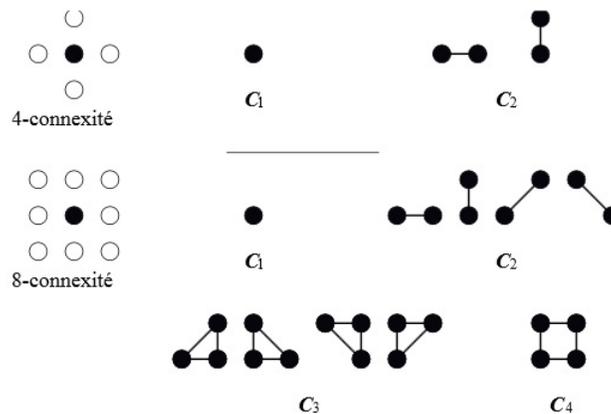


FIG. 2.3 – Les cliques associées à deux systèmes de voisinage en dimension 2.

2.3.2 Champs aléatoire Markovien MRF

Soit s un site de l'image, on peut en effet lui associer une variable aléatoire prenant ses valeurs dans E (espace de valeurs des descripteurs). Le niveau de gris en s n'est ainsi qu'une réalisation de la v.a X_s . On définit alors le champ aléatoire prenant ses valeurs

dans $\Omega = E^{\text{card}(S)}$. Dans ce cadre probabiliste, l'image est simplement une réalisation x du champ. La probabilité globale de x , $P(X = x)$, permet d'accéder en quelque sorte à la vraisemblance de l'image, et les probabilités conditionnelles locales d'une valeur en un site permettent de mesurer le lien statistique entre un niveau de gris et le reste de l'image. L'hypothèse Markovienne permet d'évaluer ces quantités.

Définition 2.3.1. *On dit que le processus $X = (X_t, t \in S)$ est un champ aléatoire Markovien (Markov random field : MRF) en regard du système de voisinage V si : $\forall s \in S, \forall t \neq s, P(X = x) > 0$ et $\forall s \in S, P(X_s = x_s / X_t = x_t, t \in S) = P(X_s = x_s / X_t = x_t, t \in V_s)$ où V_s est un voisinage de s .*

Cette définition traduit le fait que la connaissance locale ramenée à l'ensemble de la connaissance peut se réduire à la conditionnalité sur le voisinage. Une propriété ou un traitement de l'image dans sa globalité va donc se ramener à un ensemble de propriétés ou traitements locaux (un pixel et son voisinage). De plus, on doit à Geman et Geman [59] la mise en évidence du lien entre ce modèle et la distribution de Gibbs.

2.3.3 Distribution de Gibbs

Une distribution de Gibbs relative à une image et à son système de voisinage v est une mesure de probabilités $P(X = x)$ sur obéissant à :

$$P(X = x) = \frac{1}{Z} \exp(-U(x)) \quad (2.8)$$

où Z est une constante et U , appelée fonction d'énergie, est de la forme $U(x) = \sum_{c \in \mathcal{C}} U_c(x)$. U_c est une fonction de v ne dépendant que de la configuration de la clique. La constante Z

est une constante de normalisation appelée fonction de partition de Gibbs, qui permet de normaliser la distribution $P(X = x)$. La constante de normalisation Z est obtenue par :

$$Z = \sum_{x \in \Omega} \exp(-U(x))$$

Dans le cas de l'image, l'énergie $U(x)$ a une faible valeur lorsque les pixels ressemblent beaucoup à leurs voisins. La mesure de Gibbs donne donc une préférence aux configurations x dans lesquelles de nombreux pixels voisins sont de même type (niveau de gris).

2.3.4 Théorème de Hammersley-Clifford

Le champ X est un champ de Gibbs relativement au système de voisinage v si et seulement si la distribution de X est une distribution de Gibbs. Par ailleurs, Hammersley-Clifford [13] a établi l'équivalence entre la notion d'un champ de Markov et celle d'un champ de Gibbs sous forme d'un théorème.

Théorème 2.3.2. *Soit X un champ de variables aléatoires indexées par un ensemble de sites S et à valeurs dans Ω . Alors X est un champ de Markov relativement à un système de voisinage v de S si et seulement si X est un champ de Gibbs.*

Le théorème de Hammersley-Clifford est la forme bien spécifique de probabilités de X qui en résulte, va permettre de mieux exploiter la Markovienité du champ de classes d'une image en donnant une expression simple de l'énergie à partir de celle des cliques associées aux voisinages, ça va permettre aussi de lier les probabilités globales et locales d'une image.

En effet, on définit l'énergie locale U_s par :

$$U_s(x_s/x_t, t \in V_s) = \sum_{c \in C/s \in c} U_c(x_s, x_t, t \in V_s) = \sum_{c \in C/s \in c} U_c(x_s, V_s). \quad (2.9)$$

où $V_s = (x_t, t \in V_s)$. Cette énergie locale ne fait donc intervenir que les voisins de s . On peut alors écrire l'énergie globale $U(x)$ sous la forme :

$$\begin{aligned} U(x) &= \sum_{c \in C/s \notin c} U_c(x) + \sum_{c \in C/s \in c} U_c(x) \\ &= \sum_{c \in C/s \notin c} U_c(x) + U_s(x_s/V_s). \end{aligned} \quad (2.10)$$

En simplifiant l'expression de la probabilité conditionnelle locale en supprimant les termes communs qui font intervenir les cliques ne contenant pas le site s au numérateur et au dénominateur, on a :

$$P(X_s = x_s / X_t = x_t, t \neq s) = \frac{\exp(-U_s(x_s/V_s))}{\sum_{x_s \in E} \exp(-U_s(x_s/V_s))}. \quad (2.11)$$

En effet, le calcul de la distribution du champ Markovien X nécessite la connaissance de la constante de normalisation Z qui est une somme sur l'ensemble de toutes les configurations possibles du champ X . Malheureusement, la taille de l'ensemble des configurations Ω ne permet pas un calcul explicite de cette constante (par exemple pour une image à deux classes et de taille 256×256 , le calcul de cette constante nécessitera une somme sur 2^{65536} configurations). Il reste cependant possible de simuler une réalisation de ce champ à partir de ses caractéristiques locales. Il faut alors avoir recours à des techniques itératives stochastiques qui permettent de simuler le champ X (c'est-à-dire de construire des configurations suivant la loi $P(X = x)$). Parmi ces techniques de simulation, on peut citer, l'échantillonneur de Gibbs et l'algorithme de Métropolis [89]. Des détails de ses deux algorithmes sont présentés en annexe (voir Annexe B). Nous présentons quelques exemples de réalisation de champs de Markov obtenu à travers l'échantillonneur de Gibbs (voir Figure FIG.2.4.a et Figure FIG.2.5.a) et l'algorithme de Métropolis (voir Figure

FIG.2.4.b et Figure FIG.2.5.b) au cours de 50 itérations en utilisant le modèle de l'Ising décrit en annexe B.

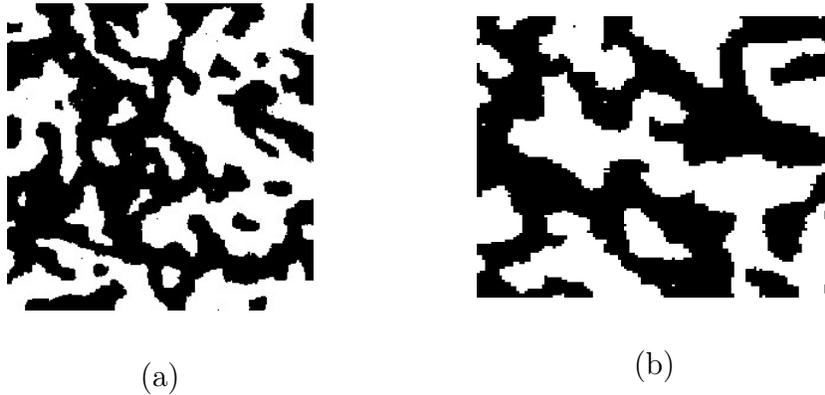


FIG. 2.4 – Exemple de champs de Markov simulés avec deux classes.

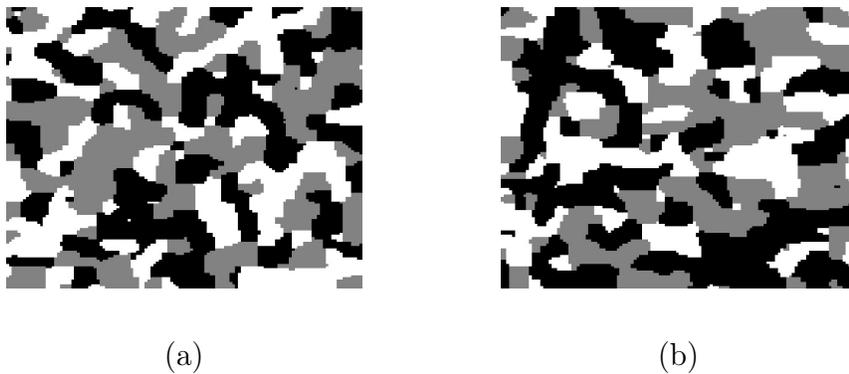


FIG. 2.5 – Exemple de champs de Markov simulés avec trois classes.

Dans la Figure FIG.2.4, les champs de Markov comportent deux classes et pour paramètre de régularisation $\beta = 1$ et un système de voisinage de 4-connexité. Dans la Figure FIG.2.5, les champs de Markov comportent trois classes et pour paramètre de régularisation $\beta = 3$ et un système de voisinage de 4-connexité.

2.4 Modèles de Markov cachés dans un cadre non paramétrique

Un modèle de Markov caché (MMC) (en anglais Hidden Markov Models (HMM)) est un modèle statistique dans lequel le système est modélisé par un processus Markovien de paramètres inconnus. Les modèles de Markov cachés sont massivement utilisés notamment en reconnaissance de formes, en intelligence artificielle ou encore en traitement automatique du langage naturel. Les HMM ont été décrits pour la première fois dans une série de publications de statistiques par Leonard E. Baum et d'autres auteurs après 1965 [79][80]. Ils ont été appliqués dès la fin des années 1970 à la reconnaissance vocale [78]. Dans la 2^{ème} moitié des années 1980, les HMM ont commencé à être appliqués à l'analyse de séquences biologique, en particulier l'ADN [16]. À partir des années 1990, les premières applications à la reconnaissance d'images [101] et de l'écriture apparaissent [46][77][91]. Récemment, les modèles de Markov cachés ont même été utilisés pour l'ordonnancement de tâches [103][104] et les technologies de l'information [3][113].

De nombreuses variantes des modèles de Markov cachés ont été développées au fil des années tels que : les modèles de Markov cachés à densité de probabilités continues [96], les modèles de Markov cachés hiérarchiques [55], les modèles de Markov cachés multidimensionnels [23] ou les modèles de Markov cachés à substitution de symboles [7][8]. Dans notre travail, notre objectif est de montrer comment nous pouvons intégrer les fonctions B-splines dans les champs de Markovs cachés.

2.4.1 Rappel sur les champs de Markov cachés

Dans cette sous section, nous considérons le problème de la segmentation statistique d'image. Soit S un ensemble fini correspondant à une image de N pixels. Nous considérons deux ensembles de variables aléatoires $X = (X_s)_{s \in S}$ et $Y = (Y_s)_{s \in S}$, appelés 'champs aléatoires'. Y représente l'image observée et X représente l'échantillon de classe. Chaque X_s prend ses valeurs dans un ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_K\}$ et chaque Y_s prend ses valeurs dans \mathbb{R} . Nous notons la réalisation de X , et Y par $x = \{x_1, \dots, x_N\}$ et $y = \{y_1, \dots, y_N\}$ respectivement. Comme nous sommes dans un cadre de segmentation, nous considérons la variable cachée à valeurs discrètes. Les processus $X = (X_s)_{s \in S}$ et $Y = (Y_s)_{s \in S}$ peuvent comporter chacun un nombre important de composantes. Donc, il est possible d'adopter pour la loi du couple (X, Y) le modèle de champ de Markov caché. La loi jointe du couple (X, Y) est définie par la donnée de $p(x)$, la loi de X et de $p(x/y)$, la loi de Y conditionnelle à X . Le champ aléatoire X est un champ de Markov par rapport à un système de voisinage $V = (V_s)_{s \in S}$ si sa loi est de la forme :

$$p(x) = \frac{1}{Z} \exp\left\{-\sum_{c \in C} U_c(x_c)\right\} \quad (2.12)$$

avec C désigne l'ensemble des cliques et U_c désigne la fonction potentiel. La loi du champ X vérifie les hypothèses de Markov suivants :

$$\begin{cases} p(x) > 0 \\ p(X_s = x_s \mid X_t = x_t, t \neq s) = p(X_s = x_s \mid X_t = x_t, t \in V_s) \end{cases} \quad (2.13)$$

La loi de X_s conditionnelle à toutes les autres réalisations ne dépend que des réalisations sur le voisinage. En d'autres termes, X_s et $(X_t)_{t \notin V_s}$ sont indépendantes conditionnellement à $(X_t)_{t \in V_s}$. La loi $p(x)$ étant définie par (2.12), il reste à définir les lois $p(x/y)$. Leur forme

la plus simple, couramment utilisée, est (on suppose les f_{x_s} , qui sont les densités des lois de Y_s conditionnelles à $X_s = x_s$, strictement positives) :

$$p(y | x) = \prod_{s \in S} f_{x_s}(y_s) = \exp\left[\sum_{s \in S} \log(f_{x_s}(y_s))\right] \quad (2.14)$$

La loi jointe de (X, Y) est alors

$$\begin{aligned} p(x, y) = p(x)p(y | x) &= \frac{1}{Z} \exp[-U(x, y)] \\ &= \frac{1}{Z} \exp\left[-\sum_{c \in C} U_c(x_c) + \sum_{s \in S} \log(f_{x_s}(y_s))\right] \end{aligned} \quad (2.15)$$

et la loi de a posteriori $p(x | y)$, qui est proportionnelle, à une constante k dépendant de y près, à $p(x | y) = kp(x)p(y | x)$, s'écrit

$$\begin{aligned} p(x | y) = kp(x)p(y | x) &= \frac{k}{Z'} \exp[-U(x | y)] \\ &= \frac{1}{Z'} \exp\left[-\sum_{c \in C} U_c(x_c) + \sum_{s \in S} \log(f_{x_s}(y_s))\right] \end{aligned} \quad (2.16)$$

avec $Z' = \frac{Z}{k}$.

On constate que la relation (2.16) diffère de la relation (2.12) par la présence d'une somme sur les pixels; les singletons étant toujours des cliques, la structure Markovienne donnée par la relation (2.12) est ainsi préservée.

Pour définir ces différentes probabilités, nous avons besoin d'estimer les fonctions f_{x_s} . Dans la littérature, plusieurs auteurs ont estimé ces fonctions par des distributions Gaussiennes [85]. Dans notre travail, afin de résoudre ce problème, nous proposons d'utiliser une méthode d'estimation basée sur les fonctions B-splines définie dans le chapitre 1.

2.4.2 Estimation non paramétrique

D'après la définition de l'estimateur de la densité de probabilités donnée dans le chapitre 1, nous pouvons exprimer la densité de probabilités f_{x_s} de la variable aléatoire $Y_s | X_s = x_s$ par :

$$f_{x_s}(y_s) = p(y_s | x_s) = \sum_{l=1}^{m+2} b_{l,x_s} B_{l,h_l}^2(y_s) \quad (2.17)$$

avec b_{l,x_s} $l = 1, \dots, m+2$ les coefficients qui vérifient les conditions suivantes : $b_{l,x_s} \geq 0$, $\sum_{l=1}^{m+2} b_{l,x_s} = 1$. En utilisant les expressions (2.14) et (2.17), nous pouvons exprimer la fonction de vraisemblance par

$$\begin{aligned} p(y | x) &= \prod_{s \in S} f_{x_s}(y_s) \\ &= \prod_{s \in S} \left(\sum_{l=1}^{m+2} b_{l,x_s} B_{l,h_l}^2(y_s) \right) \\ &= \exp \sum_{s \in S} \log \left(\sum_{l=1}^{m+2} b_{l,x_s} B_{l,h_l}^2(y_s) \right) \\ &= \exp \left[- \left(- \sum_{s \in S} \log \left(\sum_{l=1}^{m+2} b_{l,x_s} B_{l,h_l}^2(y_s) \right) \right) \right] \end{aligned} \quad (2.18)$$

Nous pouvons écrire cette dernière relation sous la forme

$$p(y | x) = \exp(-U(y | x)) = \exp \left(- \sum_{s \in S} U(y_s | x_s) \right) \quad (2.19)$$

avec $U(y_s | x_s) = -\log \left(\sum_{l=1}^{m+2} b_{l,x_s} B_{l,h_l}^2(y_s) \right)$ et $U(y | x) = \sum_{s \in S} U(y_s | x_s)$.

En utilisant le fait que $\sum_{l=1}^{m+2} B_{l,h_l}^2(y) = 1$ et $\sum_{l=1}^{m+2} b_l B_{l,h_l}^2(y) \leq \sum_{l=1}^{m+2} B_{l,h_l}^2(y) = 1$,

alors $\log \left(\sum_{l=1}^{m+2} b_{l,x_s} B_{l,h_l}^2(y_s) \right) \leq 0$.

Ceci implique que $U(y_s | x_s) = -\log \left(\sum_{l=1}^{m+2} b_{l,x_s} B_{l,h_l}^2(y_s) \right) \geq 0$.

D'après les relations (2.12) (2.16) et (2.19), nous pouvons définir la probabilité a posteriori

par :

$$\begin{aligned}
p(x | y) &= k p(y | x)p(x) \\
&= \frac{k}{Z} \exp\left(-\sum_{c \in C} U_c(x)\right) \exp\left[-\left(-\sum_{s \in S} \log\left(\sum_{l=1}^{m+2} b_{l,x_s} B_{l,h_l}^2(y_s)\right)\right)\right] \\
&= \frac{1}{Z'} \exp\left(-\left[\sum_{c \in C} U_c(x) - \sum_{s \in S} \log\left(\sum_{l=1}^{m+2} b_{l,x_s} B_{l,h_l}^2(y_s)\right)\right]\right) \\
&= \frac{1}{Z'} \exp[-U(x | y)]
\end{aligned} \tag{2.20}$$

avec $U(x | y) = -\sum_{s \in S} \log\left(\sum_{l=1}^{m+2} b_{l,x_s} B_{l,h_l}^2(y_s)\right) + \sum_{c \in C} U_c(x)$.

D'après l'expression (2.19), nous remarquons que la probabilité a postérieure $p(x | y)$ s'écrit sous la forme d'une distribution de Gibbs. D'après le théorème d'équivalence d'Hammersley-Clifford, nous pouvons dire que le champ X conditionnelle à Y est un champ de Markov. De même nous constatons que cette distribution de Gibbs s'exprime en fonction des coefficients de B-splines. Donc le problème de l'estimation de cette densité revient à estimer ces coefficients. Afin de résoudre ce problème, nous proposons un algorithme itératif que nous l'appelons EMMB (Estimation maximisation Markov B-splines).

Description de l'algorithme EMMB

L'application de cet algorithme à notre problème se décompose en trois étapes qui sont les suivantes :

Etape d'initialisation : Les coefficients de B-splines peuvent être initialisé comme suit

$$b_{l,x}^{(0)} = \frac{1}{S} \sum_{s=1}^S B_{l,h_l}^2(y_s), x \in \Omega = \{1, \dots, K\}$$

Etape d'espérance : A l'itération (q) , nous calculons la quantité intermédiaire $Q(b, b^{(q)})$

et nous estimons la probabilité a posteriori par

$$\begin{aligned} p^{(q)}(x_s | y_s, b^{(q)}) &= \frac{p^{(q)}(y_s | x_s, b^{(q)})p(x_s)}{p(y_s)} \\ &\propto \sum_{l=1}^{m+2} (b_{l,x_s}^{(q)} B_{l,h_l}^2(y_s)) p(x_s). \end{aligned}$$

$$Q = Q(b, b^{(q)}) = \sum_{s \in S} p^{(q)}(x_s | y_s, b^{(q)}) \log p(x_s | y_s, b). \quad (2.21)$$

où

$$\log(p(x_s | y_s)) = \log\left(\sum_{l=1}^{m+2} b_{l,x_s} B_{l,h_l}^2(y_s)\right) + \log(p(x_s)),$$

Etape de maximisation : Les estimateurs des coefficients b_{l,x_s} à l'itération $(q + 1)$ sont définis par :

$$b^{(q+1)} = \arg \max_b Q(b, b^{(q)}). \quad (2.22)$$

En utilisant le fait que Q est concave alors son maximum est atteint lorsque :

$$\frac{\partial Q}{\partial b_{l,x_s}} = \sum_{s \in S} p^{(q)}(x_s | y_s, b^{(q)}) \left(\frac{B_{l,h_l}^2(y_s) - B_{k,h_l}^2(y_s)}{p(x_s | y_s, b)} \right) = 0$$

et

$$\frac{\partial^2 Q}{\partial b_{l,x_s}^2} = - \sum_{s \in S} p^{(q)}(x_s | y_s, b^{(q)}) \left(\frac{B_{l,h_l}^2(y_s) - B_{k,h_l}^2(y_s)}{p(x_s | y_s, b)} \right)^2;$$

pour $l = 1, \dots, m + 2$ et $x_s \in \Omega$.

Posant

$$\frac{\partial Q}{\partial b_{l,x_s}} = F_l(b) \text{ et } \frac{\partial^2 Q}{\partial b_{l,x_s}^2} = F'_l(b);$$

$l = 1, \dots, m + 2; x_s \in \Omega$.

Pour résoudre $F_l(b) = 0$, nous utilisons la méthode de Newton Raphson

$$F'_l(b^{(q)})(b^{(q+1)} - b^{(q)}) = -F_l(b^{(q)}).$$

Par conséquent, nous obtenons notre estimateur à l'itération $(q+1)$

$$b_{l,x_s}^{(q+1)} = b_{l,x_s}^{(q)} + \frac{\sum_{s \in S} p^{(q)}(x_s | y_s, b^{(q)}) \left(\frac{B_{l,h_l}^2(y_s) - B_{1,h_1}^2(y_s)}{p(x_s | y_s, b)} \right)}{\sum_{s \in S} p^{(q)}(x_s | y_s, b^{(q)}) \left(\frac{B_{l,h_l}^2(y_s) - B_{1,h_1}^2(y_s)}{p(x_s | y_s, b)} \right)^2} \quad (2.23)$$

pour $l = 1, \dots, m + 2, x_s \in \Omega$.

Arrêt de l'algorithme

En pratique, nous arrêtons cet algorithme au bout d'un certain nombre d'itérations quand

les coefficients b_{l,x_s} ne varient presque plus, c'est à dire pour un nombre ϵ petit :

$$| b_{l,x_s}^{q+1} - b_{l,x_s}^q | < \epsilon.$$

Exemples de synthèse

Nous présentons ici quelques champs de Markov obtenus en utilisant le modèle de potts défini en Annexe B et la méthode des B-splines. Pour synthétiser une réalisation d'un champ de Markov, nous utilisons l'échantillonneur de Gibbs (voir Annexe B).

Dans la Figure FIG.2.6, les champs de Markov comportent trois classes et pour différentes valeurs de paramètre de régularisation β et un système de voisinage de 4-connexité.

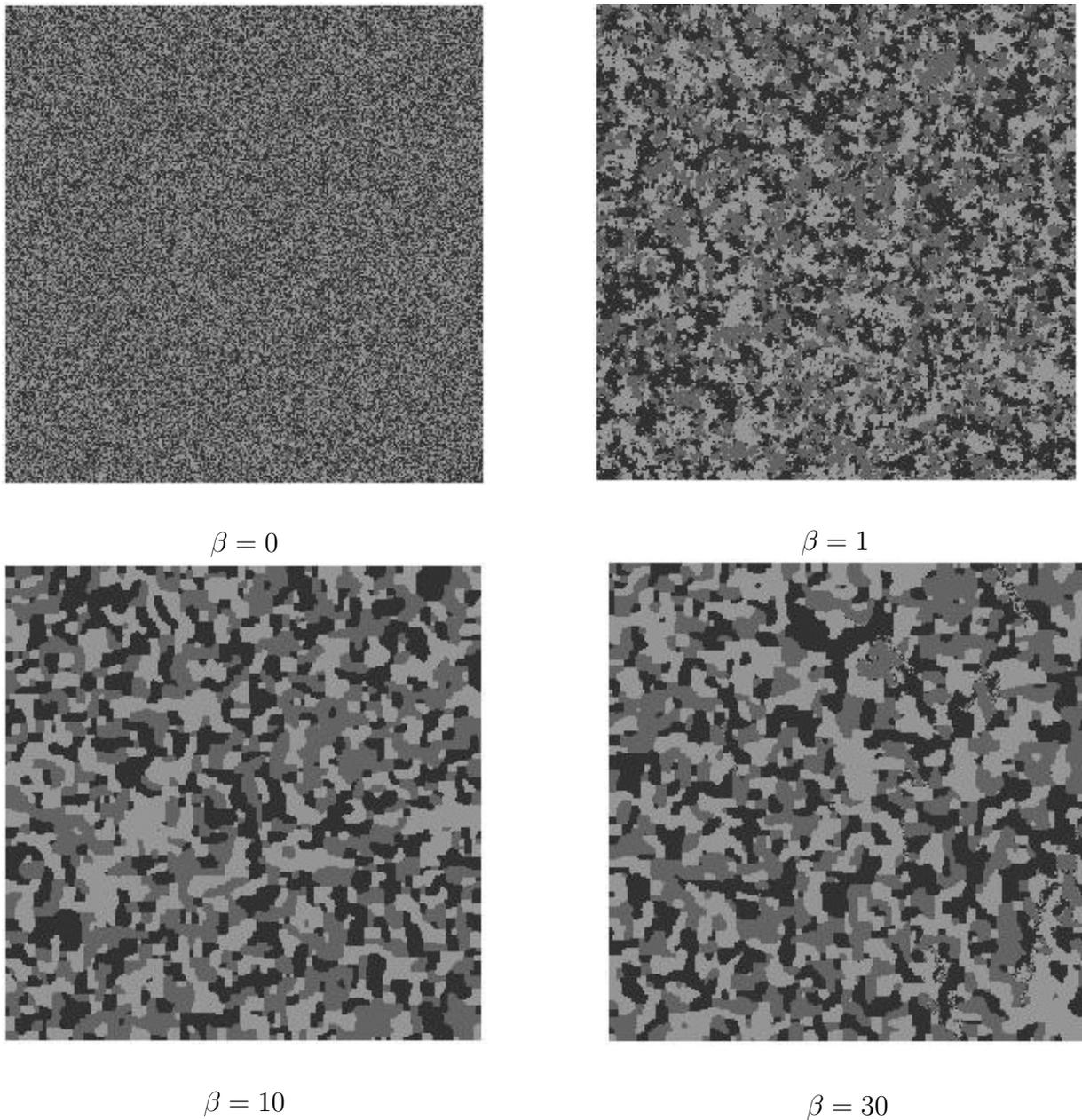


FIG. 2.6 – Champs de Markov simulé par l'échantillonneur de Gibbs (Voisinage : 4-connexité, 3 classes, 20 itérations).

Dans la Figure FIG.2.7, les champs de Markov comportent trois classes et pour différentes valeurs de paramètre de régularisation β et un système de voisinage de 8-connexité.

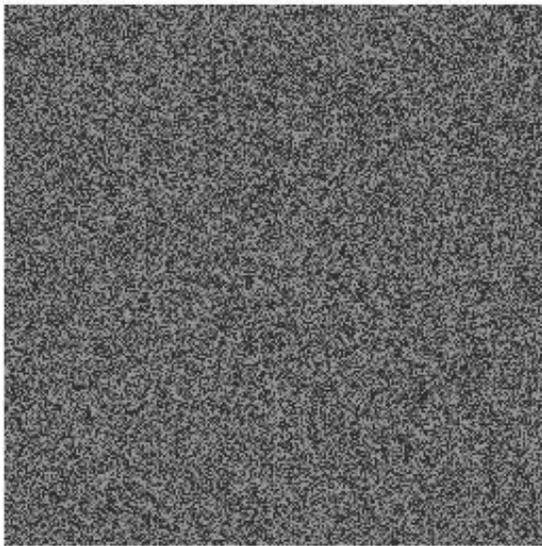
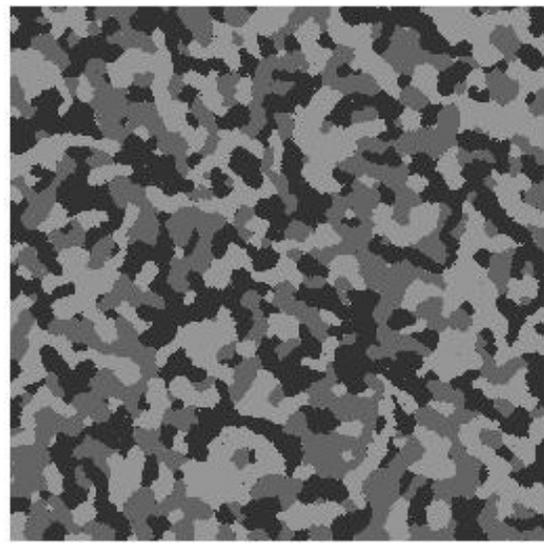
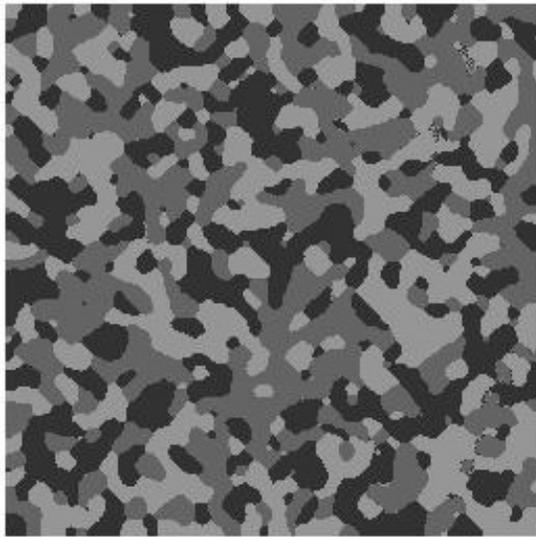
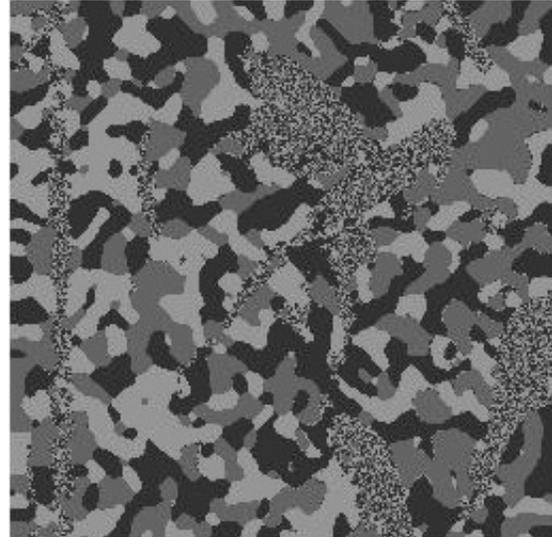
 $\beta = 0$  $\beta = 1$  $\beta = 10$  $\beta = 30$

FIG. 2.7 – Champs de Markov simulé par l'échantillonneur de Gibbs (Voisinage : 8-connexité, 3 classes, 20 itérations).

2.5 Application à la segmentation des images

La segmentation est une étape importante pour l'analyse de nombreux types d'images, puisqu'elle conditionne les opérations ultérieures dans un système de vision ou de reconnaissance de formes (extraction de contours, description de formes, compression des données, . . . , etc) [22][81]. Cette opération consiste à subdiviser une image en ses différentes régions significatives à l'aide d'une règle de décision.

Différentes méthodes de segmentation d'images ont fait l'objet de nombreux travaux telles que les méthodes locales et les méthodes globales de segmentation Bayésienne. Les méthodes locales de segmentation sont une application directe d'une règle de classification Bayésienne définie sur l'espace d'une observation restreinte à un contexte du plan de l'image, tandis que les méthodes globales consistent en une approximation de la solution de la règle de classification Bayésienne définie sur l'espace d'une observation globale [82]. Dans les méthodes locales il y a deux approches, l'approche aveugle qui ignore la corrélation spatiale de voisinage en faisant l'hypothèse d'indépendance des observations définies sur les différents pixels de l'image et l'approche contextuelle qui consiste à classer chaque pixel de l'image en vu de l'observation sur son voisinage. Dans ce chapitre, nous nous plaçons dans le cas des méthodes locales de segmentation bayésienne non supervisée.

2.5.1 Segmentation non paramétrique par approche aveugle

Dans le cas de la segmentation statistique non supervisée, nous n'avons aucune connaissance a priori sur les régions qui représentent les différentes classes de l'image [75][109]. L'image que nous observons est considérée comme une réalisation du champ X et le pro-

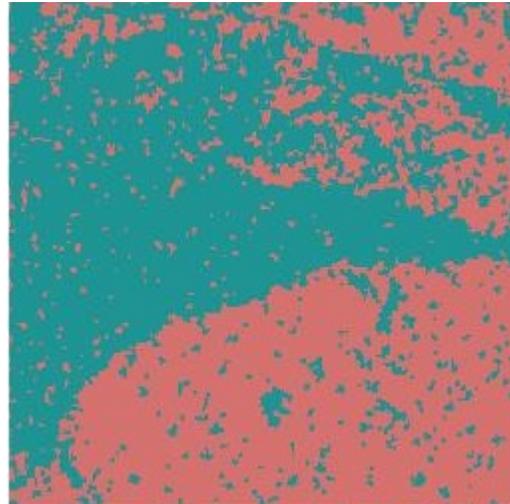
blème de segmentation est celui de l'estimation du champ inobservable Y à partir d'une observation du champ X [50][87]. Ceci revient à estimer les paramètres du mélange, cette opération est connue sous le nom d'identification du mélange. Une fois le mélange est identifié, la classification dans le cas aveugle est effectuée pixel par pixel grâce à la règle de décision de Bayes qui consiste à attribuer au pixel de l'image de niveau de gris x , la classe $j(x)$ vérifiant la relation ci-dessous :

$$j(x) = \arg[\max_{1 \leq i \leq K} \{\pi_i f(x, \theta_i)\}]$$

Nous avons appliqué l'algorithme EM Gaussien, l'algorithme EMGB et l'algorithme EMBGB proposés dans le chapitre 1 dans le cadre de la segmentation des deux images réelles. L'image présentée dans la figure FIG.2.8.a représente une image captée par un satellite, de résolution $256 \times 256 \times 3$ pixels et pour laquelle chaque pixel est codé sur huit éléments binaires. L'image donnée dans la figure FIG.2.9.a représente une photographie composée de différents objets de formes variées "Lena". Cette image est de taille 256×256 pixels. Dans ces deux cas d'images, la segmentation est réalisée à l'aide de trois classes. Les images segmentées sont présentées dans les figures FIG.2.8.b, FIG.2.8.c, FIG.2.8.d et FIG.2.9.b, FIG.2.9.c, FIG.2.9.d.



(a) Image satellitaire



(b) EM



(c) EMGB



(d) EMBGB

FIG. 2.8 – Image satellitaire segmentée avec EM classique, EMGB et EMBGB



(a) Image Lena



(b) EM



(c) EMGB



(d) EMBGB

FIG. 2.9 – Image Lena segmentée avec EM, EMGB et EMBGB

L'interprétation d'une image naturelle segmentée est subjective, tout dépend de l'information recherchée dans l'image. Par exemple, dans le cas d'image satellitaire, nous remarquons dans les figures FIG.2.8.c et FIG.2.8.d que la tache bleue de l'image segmentée est plus nette que dans l'image originale. De même dans les figures FIG.2.9.c et FIG.2.9.d, nous constatons que le regard des yeux sur ces figures est plus net par rapport à la figure FIG.2.9.a. D'après ces images segmentées, nous remarquons l'amélioration qu'a pu apporter notre approche non paramétrique.

Nous comparons les performances respectives de l'algorithme EM et les algorithmes non paramétriques proposés dans le chapitre 1 sur les deux images en calculant l'erreur quadratique moyenne (EQM)[47], le critère d'information Akaike (AIC) [2] et la distance de Kullback-Leibler (DKL) [47]. On définit respectivement l'EQM, l'AIC et la DKL par :

$$\text{EQM} = \frac{1}{N} \sum_{s=1}^N (f(y_s) - \hat{f}(y_s))^2,$$

$$\text{AIC} = 2nb - 2 \log(L)$$

et

$$\text{DKL} = \frac{1}{N} \sum_{s=1}^N f(y_s) \log \left[\frac{f(y_s)}{\hat{f}(y_s)} \right]$$

avec N représente le nombre d'observations, nb représente le nombre de paramètres à estimer du modèle et L est le maximum de la fonction de vraisemblance du modèle.

Nous présentons dans le tableau TAB.2.1, les valeurs de l'EQM, l'AIC et la DKL obtenues par l'utilisation de ces trois algorithmes EM, EMGB et EMBGB sur les deux images.

EQM Images	EQM de l'image 1	EQM de l'image 2
EM	0.257	0.514
EMGB	0.0406	0.145
EMBGB	0.0386	0.120
DKL Images	DKL de l'image 1	DKL de l'image 2
EM	0.147	0.354
EMGB	0.0524	0.0965
EMBGB	0.0412	0.00895
AIC Images	AIC de l'image 1	AIC de l'image 2
EM	0.0871	0.0754
EMGB	0.2179	0.379
EMBGB	0.2084	0.381

TAB. 2.1 – EQM, AIC et DKL de la distribution de mélange obtenu par EM, EMGB de EMBGB.

D'après le tableau TAB.2.1, nous constatons que l'utilisation des algorithmes de types EMBGB et EMGB en segmentation donnent des valeurs d'EQM et de DKL plus faible que celles obtenues par l'utilisation de l'algorithme EM pour les deux images. Par contre l'utilisation des algorithmes de types EMBGB et EMGB en segmentation donnent des valeurs d'AIC plus élevé que celles obtenues par l'utilisation de l'algorithme EM pour les deux images.

Dans la suite de notre travail, nous supposons la dépendance entre les pixels de l'images ce qui ramène à la modélisation Markovienne. Les méthodes locales contextuelles ont été

utilisées et donnent dans certaines situations des résultats de qualité suffisante ; cependant, on se heurte rapidement à des problèmes de calcul lorsque la taille de voisinage croît [95]. Pour deux classes, il est ainsi difficile de prendre en compte des voisinages de taille supérieure à huit. Le nombre de pixels étant couramment de l'ordre de cinquante mille, on peut imaginer qu'une partie non négligeable de l'information disponible ne peut être exploitée par les méthodes locales. L'introduction des champs de Markov cachés [13][59] a permis de contourner, de manière particulièrement élégante, les problèmes de calcul.

2.5.2 Segmentation non paramétrique dans un cadre Markovien

Les champs de Markov introduisent la prise en compte de la notion de dépendance spatiale dans l'image et permettent de pondérer la portée de l'influence de ces dépendances, offrant ainsi un cadre mathématique cohérent pour divers traitements. L'intérêt de l'utilisation des champs de Markov en segmentation d'images est de mieux modéliser l'image. Le processus d'une segmentation que nous proposons est donné dans la figure suivante :

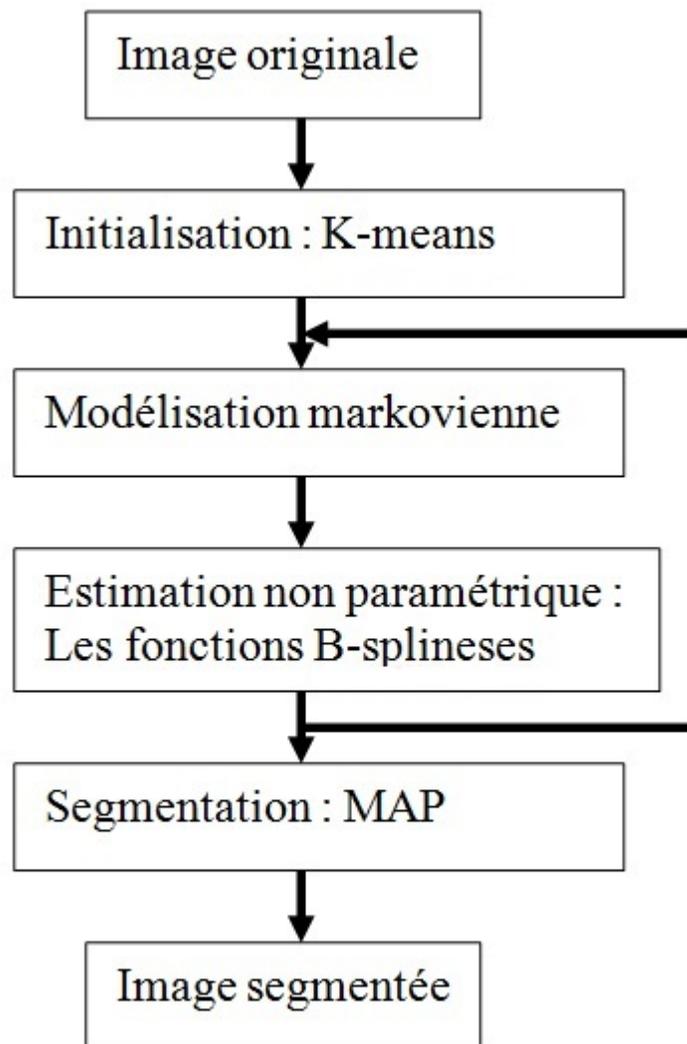


FIG. 2.10 – Organigramme de la méthode proposée de segmentation.

Cadre bayésien

Nous disposons d'une certaine donnée (image) que nous noterons y et que nous pouvons considérer comme une réalisation d'un champ aléatoire Y . Nous cherchons une réalisation x de l'image segmentée, que nous pouvons modéliser comme un champ de Markov X qui représente le champ des étiquettes (labels). Les espaces de configurations ne sont donc pas nécessairement les mêmes pour X et Y . Ces deux champs sont liés par le processus d'acquisition de l'image, qui conduit du champ idéal X , le processus image original que nous cherchons, au champ Y que nous observons. La segmentation a pour objectif d'inverser le processus et donc de remonter à une réalisation de X à partir de l'observation des données y . On parle dans ce contexte de champ de Markov "caché" pour X , ou de données incomplètes puisque y n'est pas une réalisation de X . Nous pouvons par exemple utiliser le critère du maximum a posteriori et rechercher la configuration \hat{x} maximisant la probabilité de X conditionnellement à la donnée y , i.e $p(X = x/Y = y)$. Or la règle de Bayes permet d'écrire :

$$p(X = x/Y = y) = \frac{p(Y = y/X = x)p(X = x)}{p(Y = y)}$$

Expression dans laquelle il s'agit alors d'analyser chacun des termes $p(Y = y/X = x)$ et $p(X = x)$, sachant que $p(Y = y)$ est une constante (indépendante de la réalisation x). Le premier terme $p(Y = y/X = x)$ décrit justement le processus d'observation et d'acquisition des données. L'hypothèse la plus courante (dont la validité reste à justifier) consiste à supposer l'indépendance conditionnelle des pixels (bruit non corrélé par exemple) :

$$p(Y = y/X = x) = \prod_s p(Y_s = y_s/X_s = x_s)$$

Cette écriture n'est plus valable lorsqu'il y a une convolution par la fonction de transfert du système d'acquisition, mais nous pouvons montrer que le champ a posteriori reste Markovien. Par ailleurs, nous faisons sur le champ X recherché une hypothèse Markovienne selon un voisinage V et un modèle donné dépendant de l'application. Nous pouvons alors écrire :

$$p(X = x) = \frac{\exp(-U(x))}{Z}.$$

Si on revient maintenant à la distribution a posteriori, celle-ci s'exprime par :

$$\begin{aligned} p(X = x/Y = y) &\propto p(Y = y/X = x)p(X = x) \\ &\propto e^{-\mathcal{U}(x/y)} \end{aligned}$$

$$\text{avec } \mathcal{U}(x/y) = \sum_{s \in S} -\log p(y_s/x_s) + \sum_{c \in C} U_c(x).$$

Application à la segmentation

Dans le cadre de la segmentation, le champ Markovien X est défini sur un autre espace de configurations que Y car seulement quelques étiquettes sont considérées : $E = \{1, \dots, m - 1\}$. Dans ce cas, le processus de passage de X (champ des labels) à Y ne décrit pas tant le processus d'acquisition que l'apparence des classes dans l'image. Le terme $p(Y = y/X = x)$ traduit donc la probabilité de réalisation d'une configuration donnée connaissant son étiquetage (i.e connaissant la classe de chaque pixel). En supposant l'indépendance des sites les uns par rapport aux autres, et en supposant que le niveau de gris y_s en un site s ne dépend que de l'étiquette x_s en ce site, nous avons :

$$p(Y = y/X = x) = \prod_{s=1}^S p(Y_s = y_s/X_s = x_s).$$

Les valeurs des probabilités conditionnelles sont données par l'histogramme conditionnel des niveaux de gris pour une classe donnée. Par exemple, si nous supposons que chaque classe i a une distribution B-splines de coefficients b_{l,x_s} nous avons :

$$p(y_s/x_s = i) = \sum_{l=1}^{m+2} b_{l,i} B_{l,h_l}^2(y_s).$$

Si comme précédemment nous faisons une hypothèse Markovienne sur X et que nous se limitons aux cliques d'ordre 2, nous avons :

$$p(X = x) = \frac{1}{Z} \exp(-\beta \sum_{(s,t) \in C_2} \phi(x_s, x_t)).$$

D'où l'énergie a posteriori :

$$U(x/y) = \beta \sum_{(s,t) \in C_2} \phi(x_s, x_t) - \sum_{s \in S} \log\left(\sum_{l=1}^{m+2} b_{l,x_s} B_{l,h_l}^2(y_s)\right).$$

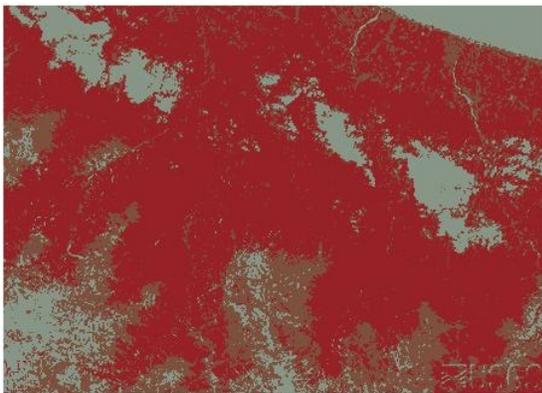
Le champ des labels conditionnellement à y est Markovien et d'énergie de Gibbs $U(x/y)$.

Nous choisissons souvent un modèle de Potts pour X , ce qui donne une image segmentée avec de larges zones homogènes. Les figures FIG. 2.11, FIG. 2.12, FIG. 2.13 et FIG. 2.14 représentent des exemples d'images segmentées avec EMMB et EM.

Exemple 1 :



Image Sat1



EM



EMMB

FIG. 2.11 – Image segmentée avec EMMB ($\beta = 0.2$) et EM.

Exemple 2 :



Image sat2



EM



EMMB

FIG. 2.12 – Image segmentée avec EMMB ($\beta = 10$) et EM.

Exemple 3 :



Image sat3



EM



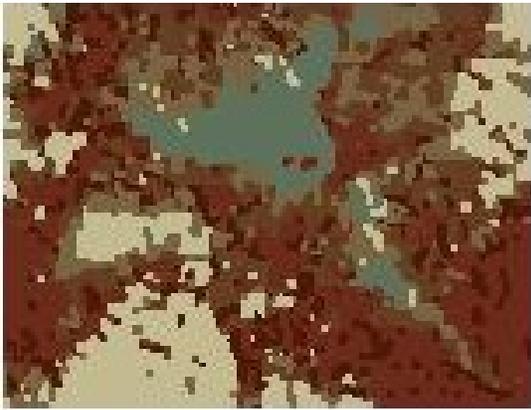
EMMB

FIG. 2.13 – Image segmentée avec EMMB ($\beta = 2$) et EM.

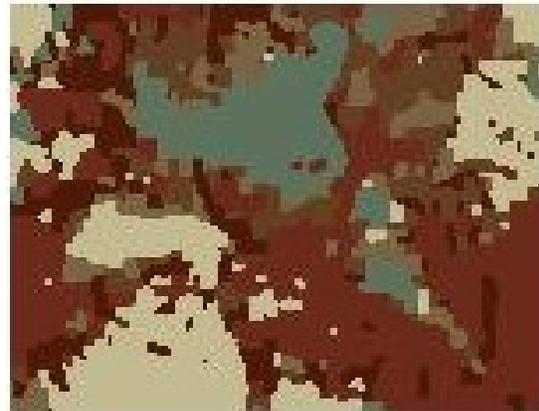
Exemple 4 :



Image sat4



EM



EMMB

FIG. 2.14 – Image segmentée avec EMMB ($\beta = 5$) et EM.

Résultats et discussion

Nous présentons dans le tableau TAB.2.2, l'évaluation de la densité mélange estimée pour les trois composantes en utilisant les deux algorithmes EM et EMMB pour l'image Sat1 en calculant l'EQM, l'AIC et le DKL entre la densité de l'histogramme de l'image présentée et la densité mélange estimée.

EQM	Image Sat1	img R	img G	img B
EM	0.247	0.185	0.457	0.247
EMMB	0.0745	0.0654	0.0143	0.0478
DKL	Image Sat1	img R	img G	img B
EM	0.274	0.0984	0.197	0.148
EMMB	0.0845	0.0784	0.0851	0.0047
AIC	Image Sat1	img R	img G	img B
EM	0.194	0.0741	0.2804	0.0557
EMMB	0.284	0.174	0.2991	0.1821

TAB. 2.2 – EQM, AIC et DKL pour les distributions mélanges obtenu avec EM et EMMB.

Nous présentons dans le tableau TAB.2.3, l'évaluation de la densité mélange estimé en utilisant les deux algorithmes EM et EMMB pour l'image sat2 en calculant l'EQM, l'AIC et le DKL entre la densité de l'histogramme de l'image présentée et la densité mélange estimée.

EQM	Image Sat2	img R	img G	img B
EM	0.199	0.377	0.347	0.241
EMMB	0.0142	0.0441	0.0198	0.047
DKL	Image Sat2	img R	img G	img B
EM	0.377	0.159	0.325	0.354
EMMB	0.0145	0.0145	0.125	0.158
AIC	Image Sat2	img R	img G	img B
EM	0.0988	0.1543	0.1984	0.1842
EMMB	0.2875	0.2345	0.2845	0.2742

TAB. 2.3 – EQM, AIC et DKL pour les distributions mélanges obtenu avec EM et EMMB.

Nous présentons dans le tableau TAB.2.4, l'évaluation de la densité mélange estimée en utilisant les deux algorithmes EM et EMMB pour l'image sat3 en calculant l'EQM, l'AIC et le DKL entre la densité de l'histogramme de l'image présentée et la densité mélange estimée.

EQM	Image Sat3	img R	img G	img B
EM	0.145	0.215	0.141	0.327
EMMB	0.0217	0.085	0.047	0.199
DKL	Image Sat3	img R	img G	img B
EM	0.184	0.214	0.148	0.213
EMMB	0.0214	0.145	0.0145	0.0145
AIC	Image Sat3	img R	img G	img B
EM	0.0221	0.1512	0.1274	0.0976
EMMB	0.1184	0.284	0.2147	0.1664

TAB. 2.4 – EQM, AIC et DKL pour les distributions mélanges obtenu avec EM et EMMB.

Nous présentons dans le tableau TAB.2.5, l'évaluation de la densité mélange estimée en utilisant les deux algorithmes EM et EMMB pour l'image sat4 en calculant l'EQM, l'AIC et le DKL entre la densité de l'histogramme de l'image présentée et la densité mélange estimée.

EQM	Image Sat4	img R	img G	img B
EM	0.325	0.214	0.145	0.147
EMMB	0.214	0.123	0.124	0.047
DKL	Image Sat4	img R	img G	img B
EM	0.219	0.125	0.298	0.298
EMMB	0.197	0.085	0.145	0.145
AIC	Image Sat4	img R	img G	img B
EM	0.0551	0.0882	0.1274	0.1668
EMMB	0.1227	0.1574	0.2446	0.2358

TAB. 2.5 – EQM, AIC et DKL pour les distributions mélanges obtenu avec EM et EMMB.

A partir de ces résultats, nous constatons que l'algorithme EMMB est plus performant par rapport à l'algorithme EM et ça est prouvé à partir du calcul de l'EQM et DKL.

2.6 Conclusion

Ce chapitre a traité le problème de la segmentation statistique non supervisée des images en se basant sur les fonctions B-splines. Dans un premier temps, nous avons présenté la modélisation probabiliste en classification. Dans un deuxième temps, nous avons rappelé le modèle Markovien en analyse d'images et nous avons décrit le modèle de Markov caché dans un cadre non paramétrique en utilisant les fonctions B-splines. Dans un troisième temps, nous avons montré l'apport de l'hybridation du modèle de Markov caché et les fonctions B-splines en segmentation statistique bayésienne semi supervisée des images. D'après les résultats de la segmentation, nous avons remarqué quelques améliorations obtenues par l'approche non paramétrique proposée. Nous allons essayer dans le troisième chapitre d'insérer dans notre approche une phase de fusion d'images et assurons ainsi une meilleure qualité de segmentation.

Chapitre 3

Fusion d'images segmentées par l'approche Bayésienne

3.1 Introduction

La fusion n'est ni une nouvelle théorie mathématique, ni une nouvelle discipline du traitement de l'information : simplement une nouvelle orientation de toutes ces disciplines "classiques", qui leur ajoute une composante multi-méthodes. C'est ainsi que le spécialiste du traitement des images pourra envisager l'étude de la détection par fusion, de l'estimation par fusion, de la reconstruction par fusion, de la classification par fusion, ect. L'idée de la fusion (de capteurs, de données, de méthodes) vient du domaine militaire et plus précisément de la surveillance du champ de bataille. Le commandement en chef dispose théoriquement de multiples sources d'informations. Le but est de prendre la meilleure décision possible, en utilisant les différentes sources de façon optimale et sans saturer les moyens de communication et de calcul. D'où la fusion de données, qui prend le meilleur

de chaque source d'information, le mélange aux autres et fournit une information plus riche [19][107]. L'utilisation de plusieurs sources permet grâce à la complémentarité et à la redondance des informations disponible de faire face à l'imperfection des données [72]. La modélisation des informations s'effectue dans le cadre de la théorie mathématique bayésienne [100], de Dempster-Shafer [38], des ensembles flous ou des possibilités [112]. Dans ce chapitre, nous allons présenter une méthode de fusion d'images segmentées fondée sur la modélisation de la connaissance par l'approche bayésienne, qui présente l'avantage de pouvoir modéliser de manière simple l'information par une probabilité conditionnelle. Une telle méthode va être en fait appliquée sur des images segmentées par notre approche développée au cours du deuxième chapitre.

Le paragraphe 3.2 présente la notion de fusion telle que décrite en traitement d'images. La méthode de fusion choisie, son principe et son apport, ainsi que la procédure adoptée dans le cadre de notre approche de segmentation statistique non paramétrique sont donnés au paragraphe 3.3. Dans le cadre de notre application aux images aériennes, le paragraphe 3.4 présente les résultats que nous avons obtenu. Il procède, de plus, à l'évaluation des performances de la fusion sur une image de synthèse.

3.2 Fusion en traitement des images

La fusion d'images est l'utilisation conjointe d'images hétérogènes pour l'aide à la prise de décision [17]. Elle intervient le plus souvent en analyse d'images issues de plusieurs capteurs : photographies, radiographies, infra-rouge, radar, etc. Il s'agit alors de fusion des données purement numériques. Ce paragraphe vise à présenter les systèmes de fusion,

définir le principe de la fusion, et rappeler les modèles de fusion.

3.2.1 Systèmes de fusion

La fusion d'informations s'est largement développée à travers différents domaines allant de l'imagerie satellitaire [5] à l'imagerie industrielle [9], en passant par l'imagerie médicale [57][67]. Les systèmes de fusion offrent la possibilité de combiner différents types d'informations, pouvant être des observations directes ou des résultats du traitement de ces observations. Dans notre travail, il s'agit des images segmentées. En effet, les informations fusionnées en traitement des images sont fortement entachées d'imperfections [19], dont l'origine se trouve à différents niveaux, depuis les phénomènes observés jusqu'aux traitements. La fusion d'informations a pour objectif de palier les imperfections associées à chaque observation. Parmi les principales imperfections sont regroupées les notions suivantes :

- *l'incertitude* : elle caractérise le degré de conformité à la réalité d'une information, en ce qui concerne sa nature, sa qualité, son essence ou son occurrence.
 - *l'imprécision* : elle décrit le contenu de l'information et mesure le défaut quantitatif d'information apportée par celui-ci. Elle fait référence au manque d'exactitude en quantité, en taille et en durée. Elle ne doit pas être confondue avec l'incertitude, même si l'un peut conduire à l'autre.
 - *l'incomplétude* : elle caractérise l'absence d'informations apportées par la source sur certains aspects du problème. Elle est justement la raison principale de la fusion d'informations fournies par plusieurs sources.
 - *l'ambiguïté* : elle fait référence à la capacité pour une information à conduire à plusieurs
-

interprétations. L'objectif de la fusion est alors de lever cette ambiguïté en combinant cette information avec d'autres.

- *le conflit* : il caractérise deux ou plusieurs informations conduisant à des interprétations contradictoires. Celui-ci est assez délicat à manipuler au sein du processus de fusion. Ainsi, certaines méthodes l'ignorent totalement, d'autres en tiennent compte en cherchant à l'éliminer (en éliminant les sources apportant ce conflit, en prenant en compte des informations supplémentaires,...) ou en le conservant afin d'attirer l'attention de l'opérateur sur sa présence.

D'autres propriétés des informations, plus positives, sont exploitées pour limiter les imperfections lors de la fusion :

- la redondance : elle caractérise les sources qui apportent plusieurs fois la même information. Elle est utilisée pour réduire les incertitudes et les imprécisions.

- la complémentarité : elle décrit des sources qui apportent des informations sur des grandeurs différentes. Elle est exploitée lors de la fusion pour avoir une information plus complète et pour lever les ambiguïtés.

Traditionnellement, les systèmes de fusion se divisent en plusieurs étapes de modélisation, d'estimation, de combinaison et de décision [19][38][45]. L'étape de modélisation comporte le choix du formalisme adopté, ainsi que des expressions à fusionner. Une étape d'estimation est également nécessaire dans les systèmes de fusion. Les informations sont ensuite regroupées lors d'une étape de combinaison, dans laquelle l'opérateur de fusion doit être compatible au formalisme adopté. L'étape finale du processus de fusion consiste à passer des informations observées au choix d'une décision selon un critère sélectionné. Par conséquent, cette phase est appelée étape de décision. Les méthodes de fusion d'infor-

mations furent développées selon différentes approches numériques, allant des méthodes probabilistes et bayésiennes à la théorie des fonctions de croyances.

3.2.2 Principe de fusion

On considère N images I_j issues d'une même scène vue au travers de systèmes différents (caméras, capteurs, experts, systèmes d'interprétation, etc). On note x l'élément (pixel ou objet plus complexe extrait des images) pour lequel on souhaite prendre une décision c'est à dire à qui l'on souhaite affecter un élément c_i de l'espace de décision $D = \{c_1, c_2, \dots, c_n\}$. Typiquement c_i correspond à une classe. Cette décision est prise à partir de la fusion d'informations sur x données par l'ensemble des I_j .

La fusion s'effectue généralement par l'intermédiaire de mesures sur x extraites de I_j vis-à-vis de c_i , notées $M_i^j(x)$ [17]. Toutes les mesures $M_i^j(x)$ sont rassemblées dans la matrice M (de taille $N \times n$) (voir TAB. 3.1).

	I_1	I_2	\dots	I_N
c_1	M_1^1	M_1^2	\dots	M_1^N
c_2	M_2^1	M_2^2	\dots	M_2^N
\dots	\dots	\dots	\dots	\dots
c_n	M_n^1	M_n^2	\dots	M_n^N

TAB. 3.1 – Matrice M des informations à fusionner : mesures issues des N images relatives aux n classes (extrait de [17]).

Quatre principes de fusion sont envisageables [17] :

- le principe global qui prend en compte la matrice M complète ;
- le principe vertical qui prend " des décisions locales sur chaque image I_j séparément, c'est-à-dire sur chaque colonne de la matrice M , puis une décision globale sur x ;
- le principe horizontal qui est basé sur la combinaison de toutes les mesures relatives à une même classe c_i , c'est-à-dire à une même ligne de la matrice M , puis une décision est prise sur le résultat de la combinaison ;
- le principe intermédiaire qui choisit de manière adaptative les informations nécessaires pour un problème donné en fonction des spécificités des images (adapté aux systèmes à base de règles ou multi-agents).

Le principe horizontal est le plus classique. C'est celui qui va nous intéresser dans le cadre de notre application.

3.3 Fusion par l'approche Bayésienne

L'approche bayésienne et plus généralement les approches probabilistes reposent sur un cadre mathématique rigoureux : la théorie des probabilités. Les probabilités offrent le plus ancien formalisme permettant de gérer de façon événementielle itérative l'incertitude dans les données. Dans ce cadre, la relation entre l'information des données et les différentes hypothèses envisagées est représentée par une distribution de probabilités conditionnelle. Les probabilités bénéficient de quatre siècles de travaux et reposent donc sur des fondements mathématiques et une expérience solides, ce qui explique pourquoi c'est encore la théorie la plus utilisée pour représenter l'incertain.

La fusion par la théorie bayésienne de la décision a été l'une des premières techniques utilisées pour faire de la combinaison d'images avec prise de décision. La littérature est très abondante en ce qui concerne la segmentation des images multispectrales avec l'aide de la théorie des probabilités. Cette technique a d'abord été développée pour l'imagerie visible [31] et aérienne [28][86]. De nombreux auteurs ont choisi cette modélisation, car elle possède un cadre mathématique très bien défini, et dont les propriétés mathématiques sont bien connues. L'expertise nécessaire dans certains cas est aussi plus facilement abordable par le biais de probabilités, et on connaît des méthodes efficaces pour modéliser un phénomène par une densité de probabilités à partir par exemple d'un ensemble d'apprentissage. Dans la suite de cette section, nous rappelons les quatre étapes de l'approche bayésienne : la modélisation, l'estimation, la combinaison et la décision en s'appuyant sur [18][84].

3.3.1 Modélisation

La théorie la plus exploitée dans la littérature est de loin la théorie des probabilités, associée à la théorie bayésienne de la décision. Un pixel x appartient à une classe ($i = 1, \dots, n$) particulière, étant donné les N images disponibles est modélisé par une probabilité conditionnelle :

$$M_i^j(x) = p(x \in c_i | I_j) \quad (3.1)$$

Cette probabilité est calculée à partir de certaines caractéristiques de l'information extraites des images. Il peut s'agir, dans les cas les plus simples, du niveau de gris du pixel considéré, ou d'informations plus complexes nécessitant des traitements préliminaires. Les probabilités $p(I_j | x \in c_i)$ (qui représente la probabilité, conditionnelle à la classe c_i , de

l'information fournie par l'image I_j), sont apprises par dénombrement sur des zones de test et on en déduit les probabilités de l'équation (3.1) par application de la règle de Bayes [17].

3.3.2 Estimation

A partir du moment où l'étape de modélisation fait intervenir des distributions de probabilités, l'étape d'estimation est indispensable. En effet, les distributions $p(x \in c_i | I_j)$ sont rarement connues en traitement du signal et des images. Pour l'estimation de ces distributions, deux cas se présentent discret ou continu. Dans notre travail, nous nous sommes intéressés au cas continu. Dans ce cas, il faut chercher à approcher les distributions $p(x \in c_i | I_j)$ par des distributions connues. Les distributions Gaussiennes sont très couramment employées et trouvent leur justification dans l'application du théorème centrale limite. Pour les distributions Gaussiennes il s'agit donc d'estimer les moyennes et variances.

D'après la règle de Bayes, la probabilité $p(x \in c_i | I_j)$ est définie par la relation suivante :

$$p(x \in c_i | I_j) = \frac{p(I_j | x \in c_i)p(x \in c_i)}{\sum_{k=1}^n p(I_j | x \in c_k)p(x \in c_k)} \quad (3.2)$$

où $p(x \in c_i)$ est la probabilité a priori de la classe c_i et $p(I_j | x \in c_i)$ représente la probabilité d'observer l'image I_j lorsque la classe c_i est réalisée.

Les lois $p(I_j | x \in c_i)$ et $p(x \in c_i)$ sont en pratique rarement connues. Elles sont souvent estimées à partir des données. Les probabilités $p(x \in c_i)$ sont déterminées par l'expérience ou par une analyse d'exemples (base d'apprentissage) et les probabilités conditionnelles $p(I_j | x \in c_i)$ sont estimées par des lois de probabilités. Ces dernières peuvent être para-

métriques, et dans ce cas une forme est choisie pour $p(I_j | x \in c_i)$ et ses paramètres sont estimés par des algorithmes très connus et fiables (EM, SEM,...), ou non paramétriques.

3.3.3 Combinaison

Une fois les informations issues de chaque capteur sont représentées selon le modèle décrit précédemment, elles peuvent être combinées, selon un critère propre au cadre théorique probabiliste. Cette étape constitue à proprement parler la fusion des images. Dans la suite, nous prenons l'exemple de la classification à partir d'images multi-sources. Les c_i sont donc ici des classes. Dans le modèle bayésien, la fusion peut être effectuée de manière équivalente à deux niveaux :

- Soit au niveau de la modélisation, et on calcule alors des probabilités de la forme $p(x \in c_i | I_1, \dots, I_N)$ à l'aide de la règle de Bayes :

$$p(x \in c_i | I_1, \dots, I_N) = \frac{p(I_1, \dots, I_N | x \in c_i)p(x \in c_i)}{p(I_1, \dots, I_N)} \quad (3.3)$$

où les différents termes sont estimés par apprentissage.

- Soit par la règle de Bayes elle-même, où l'information issue d'un capteur vient mettre à jour l'information sur x estimée d'après les capteurs précédents (c'est la seule forme utilisable si les informations sont disponibles successivement et non simultanément).

Très souvent, étant donné la complexité de l'apprentissage à partir de plusieurs capteurs et la difficulté d'obtenir des statistiques suffisantes, ces équations sont simplifiées sous

l'hypothèse d'indépendance. Les formules précédentes deviennent alors :

$$p(x \in c_i | I_1, \dots, I_N) = \frac{p(x \in c_i) \prod_{j=1}^m p(I_j | x \in c_i)}{\prod_{j=1}^m p(I_j)} \quad (3.4)$$

où les différentes probabilités sont à estimer dans l'étape d'estimation.

L'équation (3.4) fait apparaître clairement le type de combinaison des informations, sous la forme d'un produit, donc une fusion conjonctive. Il est notable que la probabilité a priori joue exactement le même rôle dans la combinaison que chacune des sources, auxquelles elle est combinée également par un produit.

L'avantage de cette théorie, du point de vue de la combinaison, est qu'elle repose sur de solides bases mathématiques, et peut être utilisée pour la mise à jour de réseaux complexes de connaissances [93][94]. Elle permet d'introduire des informations qui s'expriment facilement sous forme de probabilités, telles que le contexte spatial dans le cadre des champs de Markov (voir chapitre 2) ou la qualité des informations exprimée comme la probabilité pour qu'une mesure soit fiable [61].

Une autre manière de voir la fusion probabiliste consiste à considérer que chaque source donne une probabilité (d'appartenance à une classe) par exemple, et que la fusion consiste à combiner ces probabilités pour trouver la probabilité globale d'appartenance à la classe. Cette vision revient à considérer la fusion comme un problème d'estimation, et permet d'utiliser des opérateurs de combinaison différents du produit. En particulier les méthodes de moyenne ou moyenne pondérée, de médiane et de consensus sont souvent employées [32][33][56].

3.3.4 Décision

La dernière étape concerne la décision, par exemple le choix de la classe à laquelle appartient un point. Cette décision binaire peut être assortie d'une mesure de la qualité de cette décision, pouvant éventuellement conduire à la rejeter. La règle la plus utilisée pour la décision probabiliste et bayésienne est le maximum a posteriori :

$$x \in c_i \text{ si } p(x \in c_i | I_1, \dots, I_N) = \max\{p(x \in c_k | I_1, \dots, I_N), 1 \leq k \leq n\} \quad (3.5)$$

De nombreux critères ont été développés par les probabilistes et les statisticiens, pour qu'ils s'adaptent au mieux aux besoins de l'utilisateur et au contexte de sa décision : maximum de vraisemblance, maximum d'entropie, marginale maximale, risque minimal, etc. Cependant, la grande variété de ces critères laisse l'utilisateur à nouveau diminué devant la justification d'un choix et l'éloigne de l'objectivité recherchée initialement par ces méthodes. Nous présentons dans la suite quelques un de ces critères :

-la règle du maximum de vraisemblance est définie comme suit :

$$x \in c_i \text{ si } L(I_1, \dots, I_N; x \in c_i) = \max\{L(I_1, \dots, I_N; x \in c_k), 1 \leq k \leq n\} \quad (3.6)$$

avec $L(I_1, \dots, I_N; x \in c_i) = p(x \in c_i) \prod_{j=1}^m p(I_j | x \in c_i)$

-la règle du maximum d'entropie est définie comme suit :

$$x \in c_i \text{ si } H(x \in c_i | I_1, \dots, I_N) = \max\{H(x \in c_k | I_1, \dots, I_N), 1 \leq k \leq n\} \quad (3.7)$$

avec $H(x \in c_i | I_1, \dots, I_N) = -p(x \in c_i | I_1, \dots, I_N) \log p(x \in c_i | I_1, \dots, I_N)$.

3.4 Fusion Bayésienne dans un cadre non paramétrique

Bien que la théorie des probabilités soit plus orientée vers un objectif d'estimation, il est possible d'envisager une fusion purement probabiliste. Le fait de manipuler des densités de probabilités étant plus familier que d'utiliser des notions de masse ou de distributions de possibilités. Nous proposons dans la suite la méthode de fusion d'images segmentées par approche probabiliste dans un cadre non paramétrique. L'aspect non paramétrique provient de l'utilisation des fonctions B-Splines décrites dans le chapitre 1 pour l'estimation des densités de probabilités pour notre méthode de fusion.

3.4.1 Principe de la fusion d'images segmentées

Pour mettre en évidence les étapes de fusion et évaluer la méthode que nous proposons, nous présentons dans ce paragraphe le principe de la fusion dans le cas des images segmentées provenant des sources différentes qui observent le même phénomène.

Nous avons vu au paragraphe 3.2.2 qu'il existe plusieurs façons de combiner les informations provenant de plusieurs sources, la plus classique étant celle du principe horizontal : combinaison de toutes les mesures relatives à une même classe, puis prise de décision sur le résultat de la combinaison. C'est le principe que nous adoptons pour notre système de fusion d'images segmentées (voir figure FIG.3.1).

La matrice M des informations à fusionner (paragraphe 3.2.2) est alors donnée par la matrice des degrés d'appartenance. Soit x , un pixel appartenant à la classe d'une image.

En reprenant l'équation (3.1), nous obtenant : $M_i^j(x) = p(x \in c_i | I_j) = p_j^i(x)$.

Bien que la théorie des probabilités soit plus orientée vers un objectif d'estimation, il est

possible d'envisager une fusion purement probabiliste. Le fait de manipuler des densités de probabilités étant plus familier que d'utiliser des notions de masse ou de distribution de possibilité, cette théorie est encore largement employée [44].

Etape de fusion

En reprenant les notations du paragraphe 3.3.2, nous considérons ici que les capteurs fournissent N images I_j d'une classe c_i . La règle de Bayes (3.2) permet de calculer la probabilité d'obtenir la classe c_i en ayant effectivement les N images :

$$p_i(x) = p(x \in c_i | I_1, \dots, I_N) = \frac{p(x \in c_i)p(I_1, \dots, I_N | x \in c_i)}{\sum_{k=1}^n p(x \in c_k)p(I_1, \dots, I_N | x \in c_k)} \quad (3.8)$$

La fonction de vraisemblance $p(I_1, \dots, I_N | x \in c_i)$ décrit la probabilité que le premier capteur observe I_1 et que le N ème observe I_N , étant donnée la vraie valeur de $x \in c_i$.

Sous l'hypothèse d'indépendance conditionnelle (i.e $p(I_j | x \in c_i) : p(I_1, \dots, I_N | x \in c_i) = \prod_{j=1}^m p(I_j | x \in c_i)$) l'equation (3.8) devient :

$$p_i(x) = p(x \in c_i | I_1, \dots, I_N) = \frac{p(x \in c_i) \prod_{j=1}^N p(I_j | x \in c_i)}{\sum_{k=1}^n p(x \in c_k) \prod_{j=1}^N p(I_j | x \in c_k)} \quad (3.9)$$

Les lois $p(x \in c_i)$ et $p(I_j | x \in c_i)$ sont en pratique rarement connues. Elles sont souvent estimées à partir des données. Les probabilités $p(x \in c_i)$ sont déterminées par l'expérience ou par une analyse d'exemples (base d'apprentissage) et les probabilités conditionnelles $p(I_j | x \in c_i)$ sont estimées par des lois statistiques. Ces dernières peuvent être paramétriques ou non paramétriques. Dans notre travail nous nous sommes intéressés par le non paramétrique en se basant sur les fonctions B-Splines introduites dans le chapitre 1.

Décision Bayésienne

Bien que de nombreux critères de décision bayésienne existent (paragraphe 3.3.4), la règle la plus utilisée reste celle du maximum a posteriori. Elle consiste à privilégier l'événement ayant la plus forte probabilité a posteriori à l'issue de l'étape de fusion (3.9). Le résultat pour chaque pixel est une affectation à la classe c_i le plus probable :

$$p_i(x) = \max_{j=1,\dots,N} p_j^i(x) \quad (3.10)$$

Nous présentons dans la figure FIG.3.1 sous forme d'un organigramme les étapes à suivre pour effectuer une fusion d'images segmentées dans un cadre bayésien. Nous notons par I_j l'image segmentée et par C_i^j la classe i sachant l'image j .

Le maximum (max) gère la complémentarité : un fort degré d'appartenance désigne une région bien segmentée, qui doit être prise en compte même si cette région existe seulement sur une seule segmentation.

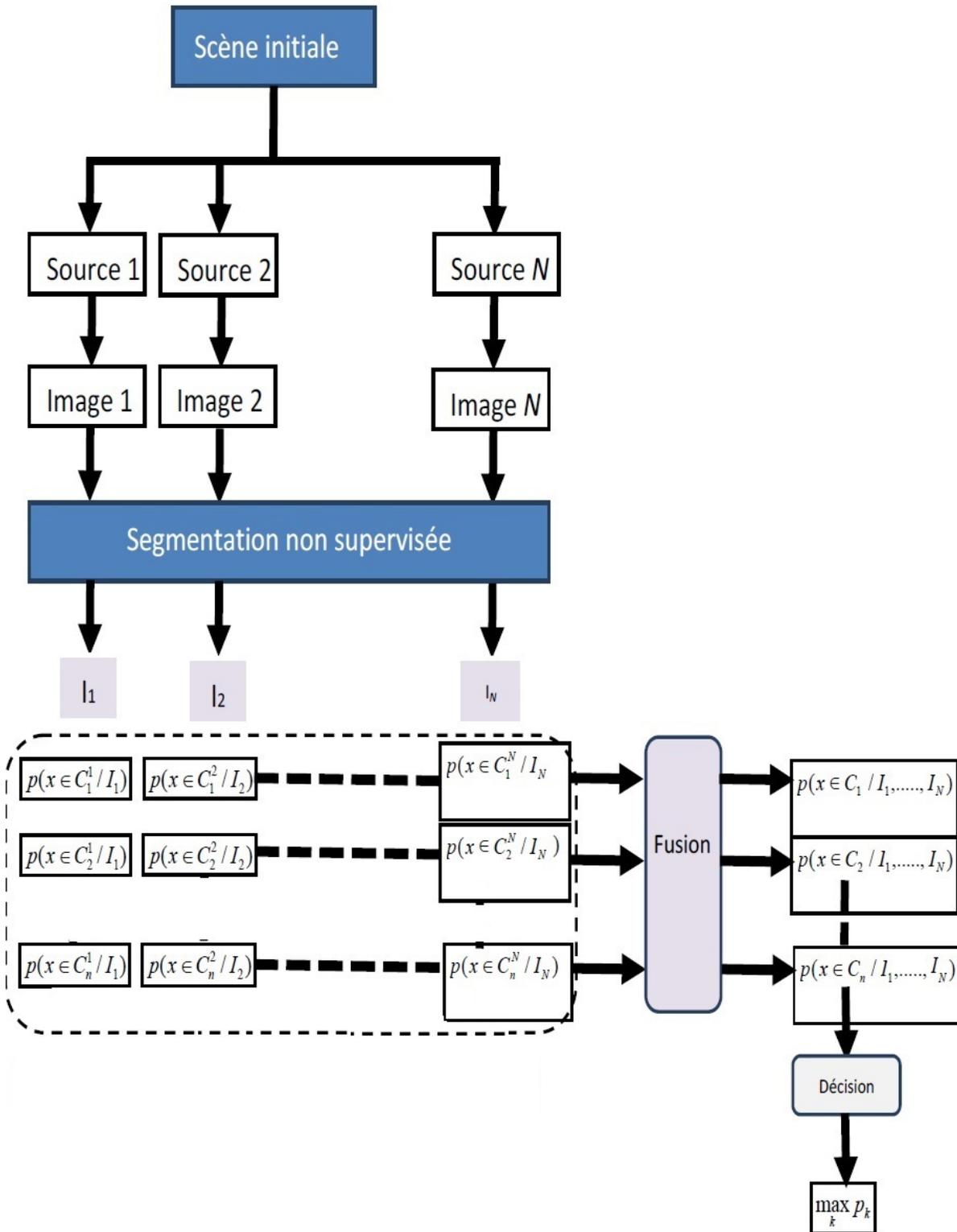


FIG. 3.1 – Fusion d'images segmentées.

3.4.2 Estimation non paramétrique

Afin d'établir une étape de fusion entre les images segmentées, les probabilités $p(x \in c_i)$ et $p(I_j | x \in c_i)$ doivent être estimées. Pour résoudre ce problème nous avons fait appel à la méthode d'estimation non paramétrique introduite dans le chapitre 1. Dans la suite de ce paragraphe nous décrivons les différentes étapes d'estimation de ces probabilités.

Descriptions de la méthode

Pour modéliser les probabilités à priori et les fonctions de vraisemblance, nous supposons que les données obéissent la distribution B-splines définie dans le chapitre 1. En l'absence de toute information sur la répartition des classes. Si on appelle n le nombre de classes supposées présentes dans l'image, la probabilité à priori de $x \in c_i$ est :

$$P(x \in c_i) = \pi_i; \quad i = 1, \dots, n. \quad (3.11)$$

Supposons que les fonctions de vraisemblance, pour chaque classe c_i , $P(I_j | x \in c_i)$ dans le cas non paramétrique sont des fonctions de B-splines et nous pouvons utiliser la formule suivante :

$$P(I_j | x \in c_i) = \sum_{l=1}^{m_i+2} b_{l,c_i} B_{l,h_l}^2(I_j) \quad j = 1, \dots, N; \quad i = 1, \dots, n \quad (3.12)$$

où b_{l,c_i} sont les paramètres des B-splines. La densité mélange $p(I_j)$ qui représente l'image I_j est définie comme suit

$$p(I_j) = \sum_{i=1}^n \pi_i \sum_{l=1}^{m_i+2} b_{l,c_i} B_{l,h_l}^2(I_j)$$

avec π_i désigne la proportion des pixels appartenant à la classe c_i ($\sum_{i=1}^n \pi_i = 1$).

En utilisant les suppositions précédemment, nous exprimons la relation de l'équation (3.4)

par :

$$p(x \in c_i \mid I_1, \dots, I_N) = \frac{\pi_i \prod_{j=1}^N \sum_{l=1}^{m_i+2} b_{l,c_i} B_{l,h_l}^2(I_j)}{\sum_{i=1}^n \pi_i \prod_{j=1}^N \sum_{l=1}^{m_i+2} b_{l,c_i} B_{l,h_l}^2(I_j)} \quad (3.13)$$

Le problème de l'estimation de ces différentes probabilités revient à estimer ces coefficients de B-splines b_{l,c_i} de chaque classe. Pour estimer ces coefficients, nous proposons d'utiliser l'algorithme EMBB défini dans le chapitre 1. Dans la suite, nous présentons ces différentes étapes de cet algorithme.

Étape d'initialisation : Les coefficients b_{l,c_i} peuvent être initialisés comme suit

$$b_{l,c_i}^{(0)} = \frac{\text{nombre d'observations de } I_j \in [x_l, x_{l+3}) \text{ sachant } c_i}{N}.$$

Étape d'espérance (E) : A l'itération (q) , la probabilité a posteriori $\tau_{l,j}^{(q)}$ est donnée par

$$\tau_{l,j}^{(q)} = \frac{b_{l,c_i}^{(q)} B_{l,h_l}^2(I_j)}{\sum_{l=1}^{m_i+2} b_{l,c_i}^{(q)} B_{l,h_l}^2(I_j)}, \quad l = 1, \dots, m_i + 2, \quad j = 1, \dots, N.$$

Étape de maximisation (MBB) : A l'itération $(q + 1)$, nous calculons le paramètre $b^{(q+1)}$.

$$b_{l,c_i}^{(q+1)} = \frac{1}{N - m + \alpha - 2} \left(\left(\sum_{j=1}^N \tau_{l,j}^{(q)} \right) + \alpha_l - 1 \right); \quad l = 1, \dots, m_i + 2 \text{ et } \alpha = \sum_{l=1}^{m+2} \alpha_l.$$

Avec $b_{l,c_i}^{(q+1)} = (b_{1,c_i}^{(q+1)}, b_{2,c_i}^{(q+1)}, \dots, b_{m_i+2,c_i}^{(q+1)})$.

Illustration

Dans ce paragraphe nous allons décrire et illustrer les différentes démarches de notre algorithme dans le cas de deux images (deux capteurs). A partir des modélisations des classes, nous pouvons appliquer la règle de Bayes pour toutes les hypothèses et tous les couples

de niveaux de gris. Nous obtenons ainsi deux probabilités a posteriori, et nous choisissons l'hypothèse correspondante à la probabilité a posteriori maximal. Nous représentons alors notre prise de décision à cette étape intermédiaire à l'aide d'une « table de fusion », qui est construite comme une image d'étiquettes. Les coordonnées d'une telle image représentent les deux valeurs (I_1, I_2) . Ainsi, la table de fusion représente la répartition des classes (ou hypothèses) assignées à un couple de valeurs (I_1, I_2) mesurées à partir des pixels des images initiales. La table de fusion est créée comme décrit ci-dessous. La probabilité a priori et la fonction de vraisemblance de chaque hypothèse sont d'abord calculées. Puis, les probabilités a posteriori de chaque couple d'observations (I_1, I_2) sont évaluées à partir de l'équation 3.14. Nous introduisons ainsi une matrice de probabilités a posteriori pour chaque hypothèse. Chaque matrice de probabilités est définie comme image à niveau de gris avec des valeurs comprises entre 0 et 1. La valeur du pixel de coordonnées (I_1, I_2) dans l'image j représente la probabilité a posteriori $P(c_i/I_1, I_2) = \frac{\pi_i \sum_{l=1}^{m_i+2} b_{l,c_i} B_{l,h_l}^2(I_1) \sum_{l=1}^{m_i+2} b_{l,c_i} B_{l,h_l}^2(I_2)}{\sum_{k=1}^n \pi_k \sum_{l=1}^{m_k+2} b_{l,c_k} B_{l,h_l}^2(I_1) \sum_{l=1}^{m_k+2} b_{l,c_k} B_{l,h_l}^2(I_2)}$.

L'hypothèse à affecter à chaque couple de valeurs (I_1, I_2) correspond, selon le principe du MAP, à l'hypothèse dont la probabilité a posteriori est la plus élevée. Une fois que nous avons obtenu une première segmentation, et une première image de fusion, nous avons obtenu une certaine classification des pixels. Nous pouvons alors actualiser les paramètres des différentes classes à partir de cette classification. Les paramètres de chaque classe sont donc remis à jour après chaque segmentation des images. Le processus est réitéré jusqu'à la convergence de l'algorithme, c'est-à-dire jusqu'à ce qu'aucun pixel ne change de classe d'une itération à l'autre.

Dans la suite nous présentons dans la figure FIG.3.2 quelques résultats de simulations. La méthode de fusion proposée est d'abord testée sur un exemple de simulation d'un mélange de loi Gaussienne $N(8, 1)$, $N(15, 10)$ et $N(0, 4)$ avec $\pi_1 = 0.2$, $\pi_2 = 0.4$ et $\pi_3 = 0.4$. Deux courbes bruitées par un bruit Gaussien à partir de l'exemple. Chaque courbe contient trois classes ($n = 3$). La courbe (FIG.3.2 (b)) est atteinte à un bruit Gaussien de moyenne 0.2 et variance 0, 004 à partir de la courbe (FIG.3.2 (a)). La courbe (FIG.3.2 (c)) est atteinte à un bruit Gaussien de moyenne 0 et variance 0, 007 à partir de la courbe (FIG.3.2 (a)). Le but ici est d'exploiter, par l'utilisation de la méthode de fusion proposée, afin d'estimer correctement la loi en trois classes. Le maximum à postériori est utilisé comme une règle de décision.

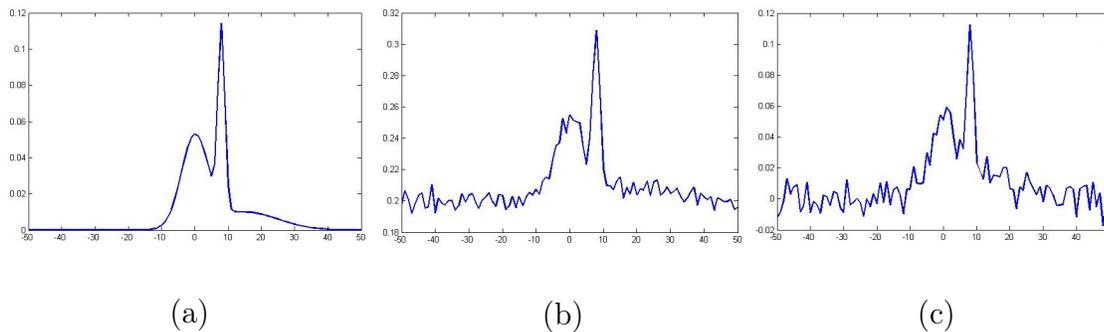
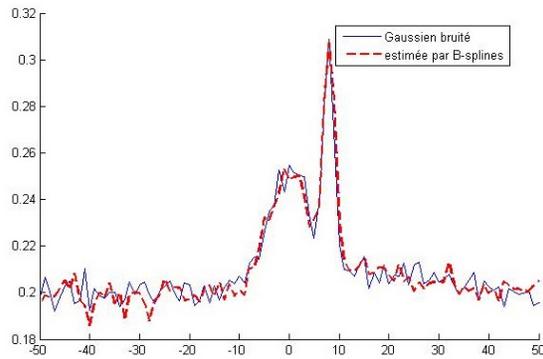
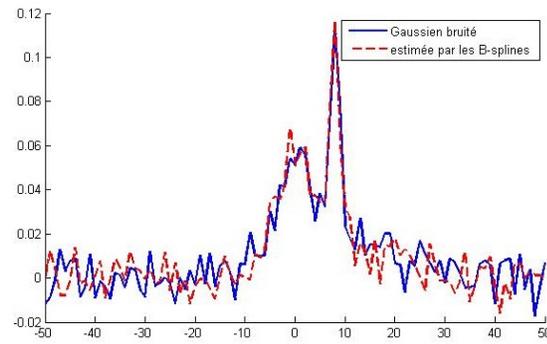


FIG. 3.2 – Trois courbes de simulation avec $n = 3$.

Nous avons appliqué à chacune de ces courbes notre algorithme. Les résultats de la simulation sont présentés dans la figure FIG.3.3.



Courbe (b) estimée en utilisant les
B-splines



Courbe (c) estimée en utilisant les
B-splines

FIG. 3.3 – Résultats de l'estimation par les B-splines.

Les deux courbes estimées vont être ensuite fusionnées en utilisant l'approche décrite dans le paragraphe 3.4. Le résultat de cette fusion est illustré dans la figure FIG.3.4.

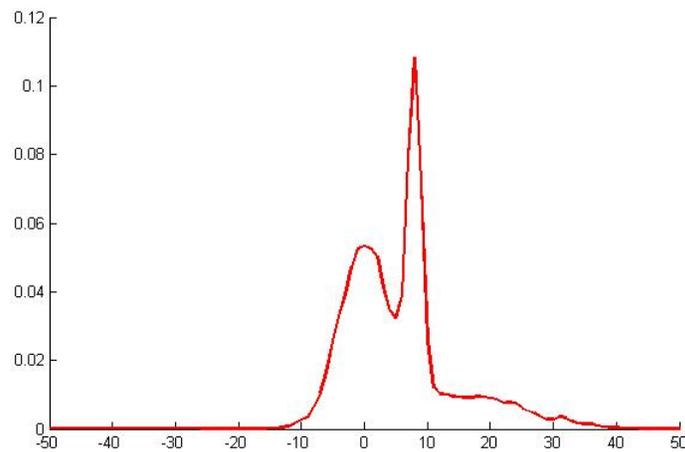


FIG. 3.4 – Résultat de la fusion par les B-splines.

La courbe fusionnée par notre approche montre bien une meilleure approximation des différentes lois de mélanges de départ. En effet, la complémentarité de nos deux courbes

a assuré une meilleure estimation.

3.5 Analyse quantitative

Dans ce paragraphe, nous allons faire une étude comparative entre deux méthodes de fusion d'images segmentées une dans un cadre paramétrique (les probabilités sont des Gaussiennes) et l'autre dans un cadre non paramétrique (les fonctions B-splines). L'analyse visuelle du résultat et des différences est un outil essentiel. Le recours à l'analyse visuelle peut paraître subjectif car l'appréciation d'une personne sur le résultat de segmentation d'une image va dépendre de ses connaissances, de ses attentes et de l'application considérée. Pour cette raison, nous faisons appelle à l'analyse quantitative. Dans notre travail, nous nous sommes intéressés plus particulièrement aux critères objectifs non pondérés. Dans la suite de ce paragraphe, nous présentons quelques critères et quelques résultats expérimentaux.

3.5.1 Critères objectifs non pondérés

Les mesures des distorsions permettent de comparer les qualités des images segmentées. Elles influent aussi sur les performances des algorithmes de traitement misent en oeuvre. Or, dans la phase de comparaison, il est nécessaire de disposer de métriques de fidélité pour évaluer la différence perceptible entre l'image fusionnée segmentée (dans un cadre paramétrique ou dans un cadre non paramétrique) I_{fus} et l'image réelle I_{re} . Les deux images sont de taille $L_1 \times L_2$ pixels. Ces mesures se basent le plus souvent sur des formules mathématiques simples.

Somme totale et distance euclidienne

Nous définissons la somme totale de la valeur absolue des différences entre I_{re} et I_{fus} que nous notons S_t et S_e la distance euclidienne par les expressions suivantes :

$$S_t = \sum_{i=1}^{L_1} \sum_{j=1}^{L_2} |I_{re}(i, j) - I_{fus}(i, j)| \quad (3.14)$$

$$D_e = \left[\sum_{i=1}^{L_1} \sum_{j=1}^{L_2} |I_{re}(i, j) - I_{fus}(i, j)|^2 \right]^{\frac{1}{2}} \quad (3.15)$$

Différence moyenne

La différence moyenne notée D_m est obtenue par la division de la somme totale S_t par la taille totale de l'image $L_1 \times L_2$. Cette différence est donnée par la formule suivante :

$$D_m = \frac{1}{L_1 \times L_2} \sum_{i=1}^{L_1} \sum_{j=1}^{L_2} |I_{re}(i, j) - I_{fus}(i, j)| \quad (3.16)$$

Erreur quadratique moyenne

L'erreur quadratique moyenne notée EQM n'est rien d'autre que la moyenne des différences quadratiques sur l'ensemble des pixels de l'image. Elle peut évaluer la qualité de l'image fusionnée segmentée. Nous définissons l' EQM par la formule suivante :

$$EQM = \frac{1}{L_1 \times L_2} \sum_{i=1}^{L_1} \sum_{j=1}^{L_2} |I_{re}(i, j) - I_{fus}(i, j)|^2 \quad (3.17)$$

Racine de l'erreur quadratique moyenne

La racine de l'erreur quadratique moyenne notée $REQM$ est un indicateur statistique qui mesure l'information spectrale. Il est moins sensible aux grandes différences que l' EQM .

Nous définissons la $REQM$ par la formule suivante :

$$REQM = \sqrt{\frac{1}{L_1 \times L_2} \sum_{i=1}^{L_1} \sum_{j=1}^{L_2} |I_{re}(i, j) - I_{fus}(i, j)|^2} \quad (3.18)$$

3.5.2 Résultats expérimentaux

Nous proposons dans ce paragraphe l'étude de la performance de la méthode proposée de fusion présentée dans la section 3.4. L'évaluation de la performance peut être effectuée à partir du calcul des métriques présentées dans le paragraphe précédent à travers des images synthétiques.

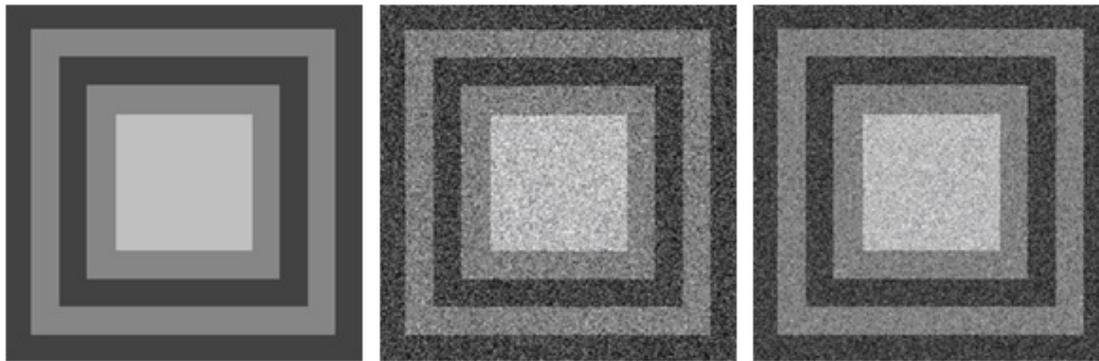
Images synthétiques

Afin de connaître l'influence du bruit sur la fusion d'images, nous avons simulé une image de taille 256×256 composée de trois classes. Les paramètres de chaque classe sont présentés dans le Tableau 3.2.

classe	1	2	3
proportion	0.473	0.383	0.143
moyenne	40	130	200
nombre de pixels	31007	25120	9409

TAB. 3.2 – Paramètres de l'image simulée.

A partir de cette image simulée, nous avons construit deux images bruitées à l'aide d'un bruit gaussien centré tout en faisant varier la valeur de la variance. Pour les images présentées dans FIG 3.5.b et FIG 3.5.C, nous avons choisi respectivement les variances du bruit gaussien 0.01 et 0.005.



(a)

(b)

(c)

FIG. 3.5 – Trois images de synthèse avec $n = 3$.

Nous identifions le mélange de l'image (FIG.3.5 (a)) respectivement en utilisant la méthode paramétrique gaussienne et la méthode non paramétrique (les fonctions B-splines). Les courbes bleu, rouge et jaune dans la figure FIG.3.6 représentent respectivement l'histogramme normalisé, le mélange de lois de l'image dans un cadre paramétrique et le mélange de lois de l'image dans un cadre non paramétrique.

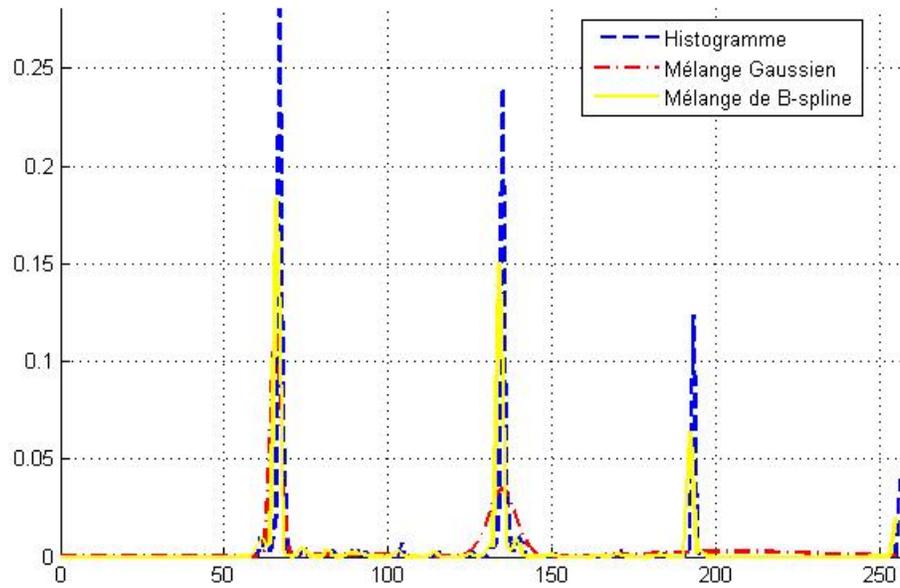
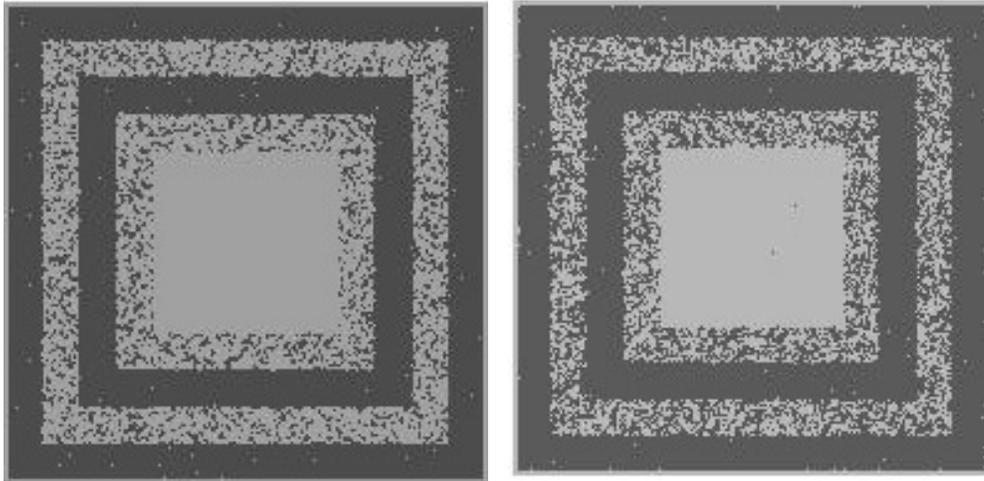
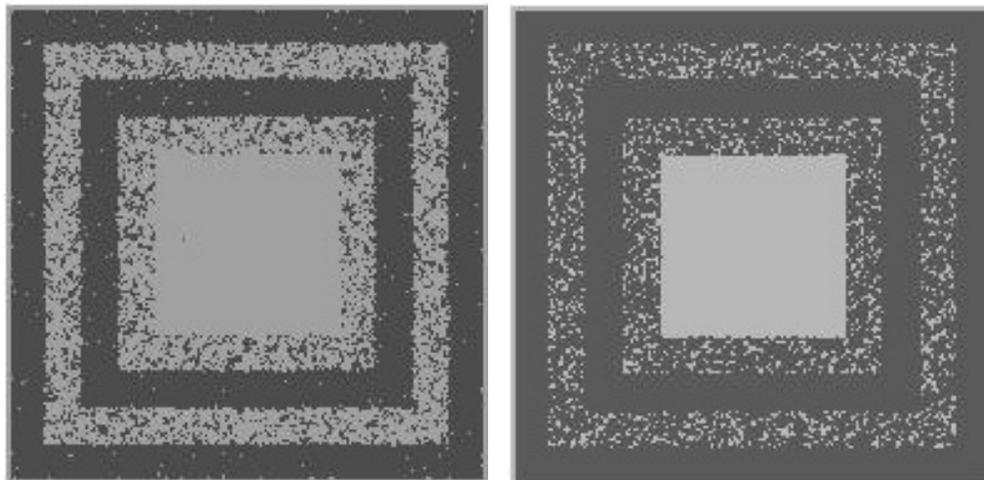


FIG. 3.6 – Identification du mélange.

Nous avons segmenté les deux images bruitées présentées dans la figure FIG.3.5 en utilisant les deux algorithmes EM et EMGB. Les résultats de la segmentation sont présentés dans la figure FIG. 3.7.



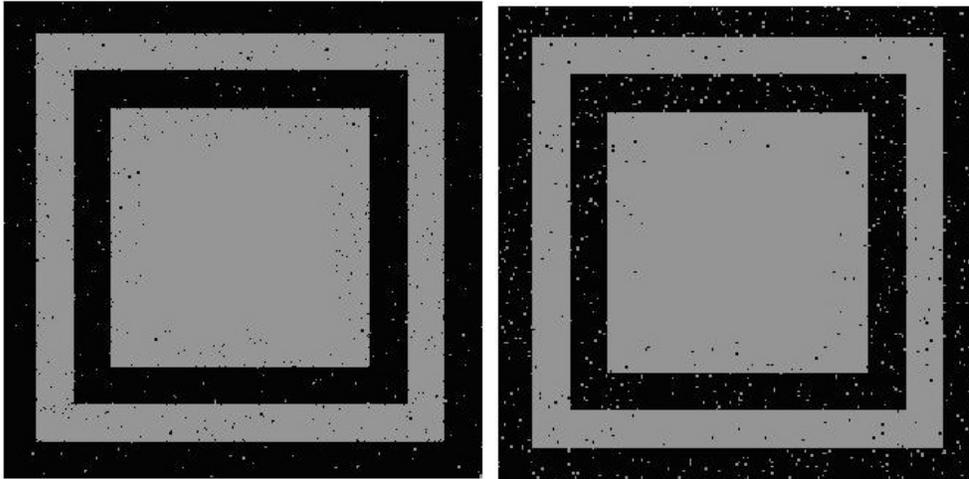
Segmentation non paramétrique des images (b) et (c)



Segmentation paramétrique des images (b) et (c)

FIG. 3.7 – Résultats de la segmentation de l'image test.

Les quatre images segmentées vont être ensuite fusionnées deux à deux en utilisant l'approche bayésienne paramétrique et non paramétrique. Nous présentons dans la figure FIG.3.8 le résultat de fusion d'images segmentées.



Avec les B-splines

Avec la loi Gaussienne

FIG. 3.8 – Résultats de fusion d'images segmentées.

L'image fusionnée par notre approche montre bien une meilleure représentation des différentes régions de notre image de départ. En effet, la complémentarité de nos deux images a assuré une meilleure définition des classes de l'image.

Dans le tableau suivant nous présentons les différentes valeurs de S_t , D_e , D_m , EQM et $REQM$.

	S_t	D_e	D_m	EQM	$REQM$
Fusion paramétrique	30.1673	26.9713	4.603×10^{-4}	0.0054	0.0735
Fusion non paramétrique	21.3521	18.8121	3.258×10^{-4}	0.0001	0.1054

TAB. 3.3 – Evaluation des résultats de fusion d'images segmentées.

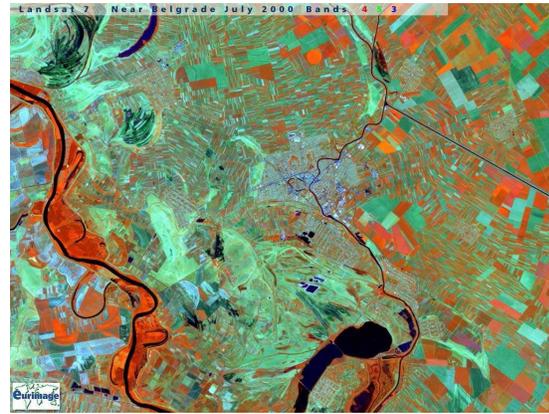
D'après les résultats présentés dans le tableau TAB 3.3, nous remarquons que les résultats de fusion d'images segmentées dans un cadre non paramétrique est plus performant que dans un cadre paramétrique.

3.5.3 Fusion d'images segmentées

Nous illustrons dans cette section la fusion par l'approche Bayésienne sur des images segmentées à l'aide de l'approche proposée dans le chapitre 2 définie par l'algorithme EMBB et cela à partir de données réelles multi-spectrales ou encore multi-temporelles. La fusion est effectuée au niveau des pixels, à partir de l'information de niveaux de gris.

Image multi-spectrale

Notre première image représente une région de Belgrade captée par le satellite Landsat 7 de taille 10024×768 pixels [49]. Dans notre application, nous allons travailler avec deux images : une première image captée par les bandes spectrales 3 2 1 (figure FIG.3.9.a) et une deuxième avec les bandes spectrale 4 5 3 (figure FIG.3.9.b). L'image représente essentiellement trois zones différentes : une première zone qui représente des champs d'agricultures et des parcelles de terre, une deuxième zone d'habitation (milieu de l'image) représentée par une infrastructure composée de maison et de route, et enfin une troisième zone représentée par le canal Danube-Tisa-Danube qui traverse toute la région en formant par endroit des petits lacs et rivières.



(a) image captée par les bandes

(b) image captée par les bandes

spectrales 3 2 1

spectrales 4 5 3

FIG. 3.9 – Image test de Belgrade captée par le satellite Landsat 7.

Nous avons appliqué à chacune de ces images notre approche de segmentation statistique présentée dans le chapitre 2 définie par l'algorithme EMBB et la segmentation par la loi Gaussienne avec un nombre de classes $n = 9$ et pour 20 itérations. Les résultats de la segmentation sont présentés dans la figure FIG.3.10.

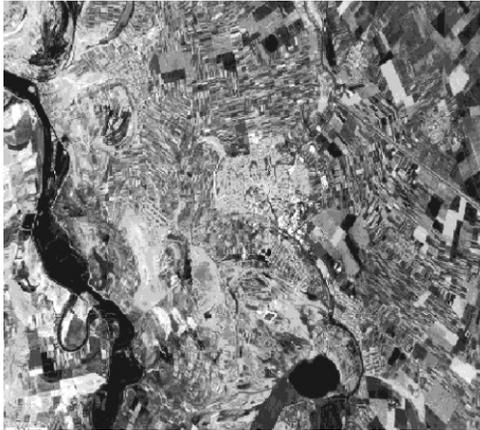


Image (Fig 3.9.a) segmentée en
utilisant les B-splines

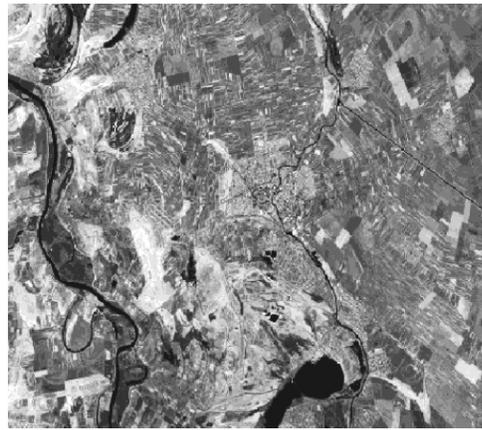


Image (Fig 3.9.b) segmentée en
utilisant les B-splines

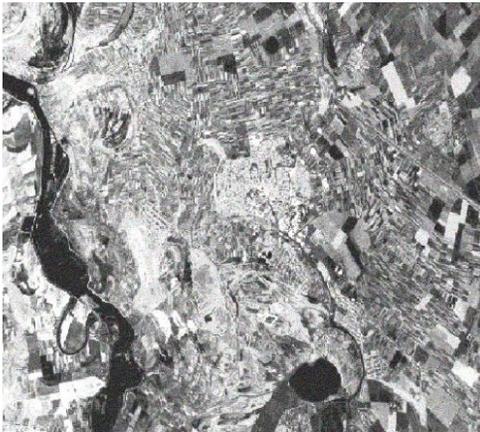


Image (Fig 3.9.a) segmentée avec la
loi Gaussienne

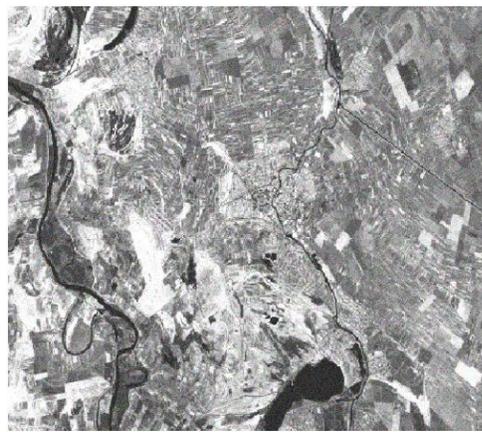
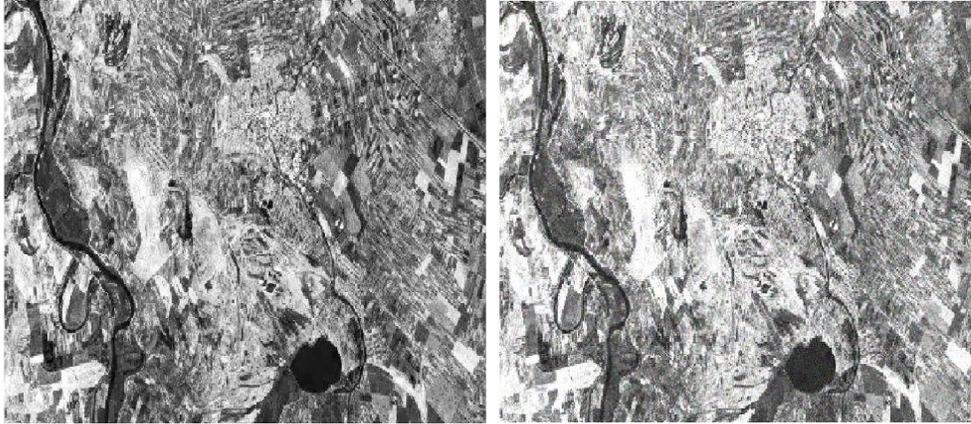


Image (Fig 3.9.b) segmentée avec
la loi Gaussienne

FIG. 3.10 – Résultats de la segmentation de l'image test de Belgrade.

Les quatre images segmentées vont être ensuite fusionnées deux à deux en utilisant l'approche bayésienne paramétrique et non paramétrique. Nous présentons dans la figure FIG.3.11 le résultat de fusion d'images segmentées.



(a) Avec les B-splines

(b) Avec la loi Gaussienne

FIG. 3.11 – Résultats de la fusion.

L'image fusionnée par notre approche montre bien une meilleure représentation des différentes régions de notre image de départ. En effet, la complémentarité de nos deux images a assuré une meilleure définition des classes de l'image. En effet, nous remarquons que dans la figure FIG.3.11 (a) une meilleure illustration de la zone représentée par le canal Danube-Tisa-Danube et des différentes rivières et lacs qui le composent et nous constatons aussi une meilleure identification des différents champs d'agricultures et de parcelles.

Image multi-temporaire

Notre deuxième image représente la ville de Pensacola (Floride, États-Unis) captée par la NASA avant et après le passage de l'ouragan Ivan en Septembre 2004. Cette image est de taille 1300×2500 pixels et a été captée à deux dates différentes. La première image présentée dans la Figure FIG.3.12.a a été captée le 4 janvier 2003, et la deuxième image

présentée dans la figure FIG.3.12.b a été captée le 18 septembre 2004. L'image représente les changements survenus après le passage de l'ouragan Ivan sur la ville de Pensacola et notamment les dégâts subits par son port maritime.



(a)



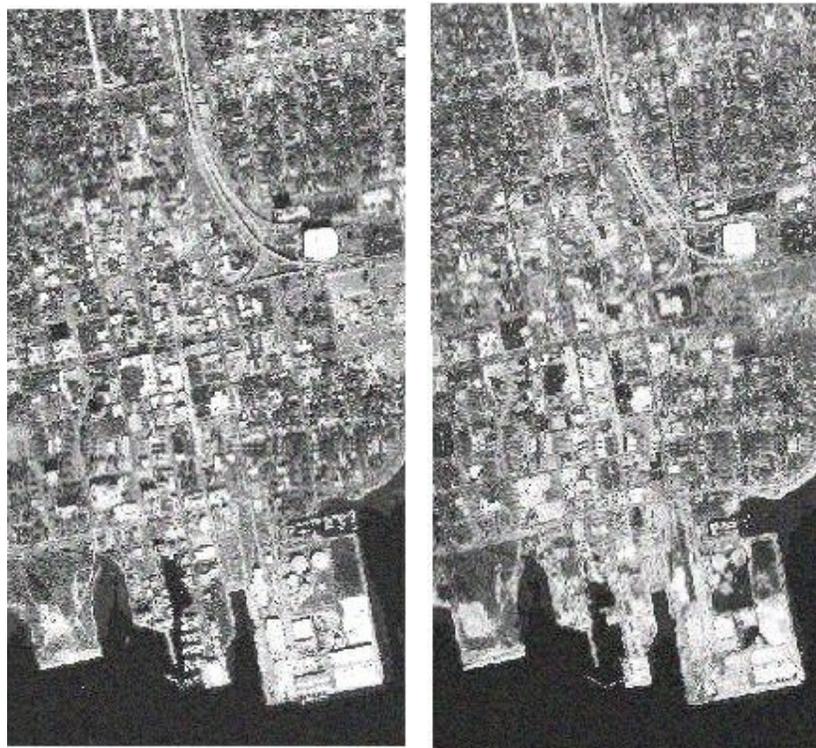
(b)

FIG. 3.12 – Image test représentant la ville de Pensacola : (a) Janvier 2003, (b) Septembre 2004.

Nous avons appliqué à chacune de ces images notre approche de segmentation présentée dans le chapitre 2 définie par l'algorithme EMBB et la segmentation dans un cadre paramétrique (mélange de lois Gaussiennes) avec un nombre de classes $n = 3$ et pour 20 itérations. Les résultats sont présentés dans la figure FIG.3.13.



Segmentation non paramétrique des images (a) et (b)



Segmentation paramétrique des images (a) et (b)

FIG. 3.13 – Résultats de la segmentation de l'image test de Pensacola.

Les quatre images segmentées vont être ensuite fusionnées deux à deux en utilisant l'approche bayésienne paramétrique et non paramétrique. Nous présentons dans la figure FIG.3.14 le résultat de fusion d'images segmentées.



Avec les B-splines

Avec la loi Gaussienne

FIG. 3.14 – Résultats de la fusion.

3.6 Conclusion

Face à la complexité des images réelles, il est devenu de plus en plus évident qu'aucun opérateur ne peut parvenir à lui seul à résoudre la question de la segmentation, mais la solution se trouve probablement dans la coopération ou la fusion de plusieurs méthodes. L'intérêt de la fusion dans le domaine de traitement des images n'est plus à justifier. En effet face à la diversité et

à l'augmentation du volume d'image issue de différents types de capteurs, plusieurs approches de fusion ont été développées dans le but de garantir un résultat de segmentation (classification) optimal. Parmi ces approches, la théorie Bayésienne qui a fait l'objet de notre chapitre. Nous avons présenté dans ce chapitre, une approche de fusion fondée sur cette théorie, basée sur la méthode de maximum de vraisemblance, à travers l'estimation non paramétrique des probabilités, pour chaque pixel de l'image. Nous avons ensuite appliqué cette approche à des images multi-spectrales et multi-temporelles segmentées par notre algorithme EMBB non paramétrique et non supervisé développé dans le chapitre 2. Les résultats trouvés sont satisfaisants, cependant le temps de calcul demeure l'inconvénient majeur de cette approche de fusion.

Conclusion et perspectives

Dans cette thèse nous avons traité le problème de l'estimation non paramétrique des lois de probabilités. Dans un premier temps, nous avons rappelé quelques méthodes d'estimation non paramétrique telles que la méthode d'histogramme, la méthode des noyaux, la méthode des fonctions orthogonales. Nous avons rappelé les fonctions B-splines ainsi la description des B-splines quadratiques. D'abord, nous avons supposé que la densité inconnue f a été approchée par un mélange de la base B-spline quadratique. Puis, nous avons proposé un nouvel estimateur de la densité inconnue f basé sur les fonctions B-splines quadratiques, avec deux méthodes d'estimation. La première est basée sur la méthode du maximum de vraisemblance et la deuxième est basée sur la méthode d'estimation Bayésienne MAP. Ensuite, nous avons généralisé notre étude d'estimation dans le cadre du mélange et nous avons proposé un nouvel estimateur du mélange de lois inconnues basé sur les deux méthodes d'estimation adaptées.

Dans un deuxième temps, nous avons traité le problème de la segmentation statistique non supervisée des images en se basant sur le modèle de Markov caché et les fonctions B-splines. Nous avons présenté la modélisation probabiliste en classification, nous avons rappelé le modèle Markovien en analyse d'images et nous avons décrit le modèle de Markov caché dans un cadre non paramétrique en utilisant les fonctions B-splines. Nous avons montré aussi l'apport de l'hybridation du modèle de Markov caché et les fonctions B-splines en segmentation statistique bayésienne non supervisée des images.

Dans un troisième temps, nous avons présenté une approche de fusion basée sur la méthode de maximum de vraisemblance, à travers l'estimation non paramétrique des probabilités, pour chaque pixel de l'image. Nous avons ensuite appliqué cette approche sur des images multi-spectrales et multi-temporelles segmentées par notre algorithme non paramétrique et non supervisé développé dans le chapitre 2. Les résultats obtenus sont satisfaisants, cependant le temps de calcul demeure l'inconvénient majeur de cette approche de fusion.

Différentes perspectives se dégagent de mes travaux. Premièrement, étendre notre étude d'estimation non paramétrique de la densité de probabilités par les fonctions B-splines dans un cadre multivarié. Deuxièmement, appliquer les fonctions B-splines multivariées pour le problème de la segmentation statistique et de la fusion des images. Troisièmement, réduire le nombre des paramètres du modèle mélange de fonctions B-splines et étudier le problème de l'estimation de la position des noeuds, et ce pour avoir une meilleur estimation. Enfin, étudier le problème de l'identification du modèle mélange de fonctions B-splines.

Annexe A

3.6.1 Mélange de lois

Dans ce paragraphe, nous décrivons la méthode classique du maximum de vraisemblance pour estimer les paramètres du modèle mélange. Cette méthode consiste à rechercher les valeurs des paramètres qui maximisent la vraisemblance (ou plutôt le logarithme de la vraisemblance) des données observées.

Données complètes et vraisemblance complète

Nous considérerons que les données observées x_1, \dots, x_n ne correspondent qu'à une connaissance partielle des données complètes h_1, \dots, h_n qui sont supposées être de la forme $h_j = (x_j, z_j), j = 1, \dots, n$, ou z_j est le numéro de la composante à laquelle appartient l'observation x_j . Nous verrons que dans le cas de la classification non supervisée, cette dernière information sera cachée et z_j sera appelé une donnée manquante. On appellera vraisemblance complète, la vraisemblance calculée à partir des données complètes $(x_1, z_1), \dots, (x_n, z_n)$ et elle sera notée $L((x, z); \Phi)$, ou plus simplement $L(\Phi)$.

Estimation par maximum de vraisemblance

La méthode du maximum de vraisemblance (MMV) permet d'estimer le paramètre Φ du modèle par $\hat{\Phi}_{MMV}$:

$$\hat{\Phi}_{MMV} = \arg \max_{\Phi} L(\Phi).$$

Les n données $(x_j, z_j), j = 1, \dots, n$, sont supposées indépendantes, on peut exprimer la vraisemblance du paramètre Φ par le produit de toutes les densités marginales. Pour simplifier le calcul on utilise le logarithme népérien de la fonction vraisemblance. Dans le cadre du modèle de mélange, la fonction log-vraisemblance s'écrit alors de la façon suivante :

$$l(\Phi) = \log(L(\Phi)) = \sum_{j=1}^n \log\left(\sum_{k=1}^K \pi_k f(x_j, \theta_k)\right).$$

Si de plus, les labels des observations sont connus, (*i.e.* dans le cas supervisé), cette fonction $l(\Phi)$ peut s'écrire sous la forme suivante :

$$l(\Phi) = \sum_{j=1}^n \sum_{k=1}^K \tau_{kj} \log(\pi_k f(x_j, \theta_k)),$$

où $\tau_{kj} = 1$ si l'observation x_j appartient à la composante k et $\tau_{kj} = 0$ sinon.

La maximisation de la fonction log-vraisemblance d'un modèle de mélange dans le cas non-supervisé conduit en général à des équations de vraisemblance qui ne possèdent pas de solutions analytiques. Il existe différents algorithmes permettant de maximiser la fonction log-vraisemblance quand les labels sont inconnus. L'algorithme le plus utilisé d'entre eux est l'algorithme itératif EM (Espérance-Maximisation).

Algorithme EM

L'algorithme EM repose sur l'idée qu'il est plus facile de maximiser la vraisemblance complète $L((x, z); \Phi)$ que la vraisemblance $L(x; \Phi)$ et se base sur la relation suivante entre les deux vrai-

semblances :

$$L(x; \theta) = L((x, z); \Phi) - \log f((x, z) | x, \Phi).$$

Cependant, la vraisemblance complète n'est pas calculable du fait que z n'est pas totalement connue. Dempster, Laird et Rubin [39] ont proposé pour maximiser cette vraisemblance, une procédure itérative qui se base sur la maximisation de l'espérance conditionnelle de la vraisemblance pour une valeur du paramètre courant Φ . L'algorithme consiste donc simplement à construire une suite $(\Phi^{(q)})$ qui vérifie :

$$\Phi^{(q+1)} = \arg \max_{\Phi} Q(\Phi, \Phi^{(q)}),$$

où $Q(\Phi, \Phi^{(q)}) = E[L((x, z); \Phi) | x, \Phi^{(q)}]$. Si l'on note $\tau_{kj} = E[Z_k | X = x_j, \Phi^{(q)}]$ l'espérance du $k^{\text{ème}}$ élément du vecteur aléatoire Z conditionnellement à $X = x_j$ et à $\Phi^{(q)}$, on peut alors écrire :

$$Q(\Phi, \Phi^{(q)}) = \sum_{j=1}^n \sum_{k=1}^K \tau_{kj}^{(q)} \log(\pi_k f(x_j, \theta_k))$$

De plus, il a été montré que la suite ainsi générée $\Phi^{(q)}$ fait croître la vraisemblance $L(x; \theta)$ et converge vers un maximum local de celle-ci sous certaines conditions de régularité. On pourra consulter [39] pour un plus simple développement du principe de cet algorithme d'estimation.

Les étapes de l'algorithme EM

Partant d'une solution initiale $\Phi^{(0)}$, l'algorithme EM alterne entre deux étapes appelées respectivement **E**, pour espérance et **M**, pour maximisation :

Étape E : cette étape consiste à estimer, à l'itération q , l'espérance de Z_k sachant la valeur du paramètre estimé à l'étape précédente et $X = x_j$:

$$\begin{aligned} \tau_{kj}^{(q)} &= E[Z_k | X = x_j, \hat{\Phi}^{(q)}] \\ &= P[Z = k | X = x_j, \hat{\Phi}^{(q)}] \end{aligned}$$

La formule de Bayes permet finalement d'exprimer les probabilités τ_{kj} d'appartenance des x_j aux composantes k , de la façon suivante :

$$\tau_{kj}^{(q)} = \frac{\pi_k^{(q)} f(x_j, \hat{\theta}_k^{(q)})}{\sum_{k=1}^K \pi_k^{(q)} f(x_j, \hat{\theta}_k^{(q)})}$$

Étape M : Cette étape consiste à estimer à chaque itération ($q + 1$) les proportions du mélange par :

$$\hat{\pi}_k^{(q+1)} = \frac{\sum_{j=1}^n \tau_{kj}^{(q)}}{n}$$

et les paramètres $\theta_1, \dots, \theta_K$ seront obtenus en résolvant les équations de vraisemblance correspondantes au modèle de mélange retenu.

Application de l'algorithme EM à un mélange de lois Gaussiennes

Dans cet exemple, nous nous sommes intéressés à l'estimation des vecteurs des paramètres du mélange de lois Gaussiennes univariées, du vecteur des paramètres $\theta_k = (\mu_k, \sigma_k^2)$ où μ_k représente la moyenne et σ_k^2 est la variance de la groupe k , en utilisant l'algorithme EM. La densité de ce mélange est proportionnelle à

$$f(x, \Phi) \propto \sum_{k=1}^K \frac{\pi_k}{\sigma_k} \exp\left(-\frac{(x_k - \mu_k)^2}{2\sigma_k^2}\right)$$

avec $x_k \in \mathbb{R}, \pi_k > 0, \sum_{k=1}^K \pi_k = 1$ et $\Phi = (\pi_1, \dots, \pi_K, \mu_1, \dots, \mu_K, \sigma_1^2, \dots, \sigma_K^2)$.

À l'étape **E** de l'itération (q), la probabilité a posteriori $\tau_{kj}^{(q)}$ est définie par

$$\tau_{k,i}^{(q)} = \frac{\hat{\pi}_k^{(q)} f(X_i / \hat{\theta}_k^{(q)})}{\sum_{k=1}^K \hat{\pi}_k^{(q)} f(X_i / \hat{\theta}_k^{(q)})}$$

À l'étape **M** de l'itération ($q + 1$), nous obtenons :

$$\hat{\pi}_k^{(q+1)} = \frac{1}{n} \sum_{i=1}^n \tau_{k,i}^{(q)}$$

$$\hat{\mu}_k^{(q+1)} = \frac{\sum_{i=1}^n X_i \tau_{k,i}^{(q)}}{\sum_{i=1}^n \tau_{k,i}^{(q)}}$$
$$[\hat{\sigma}_k^2]^{(q+1)} = \frac{\sum_{i=1}^n \tau_{k,i}^{(q)} (X_i - \hat{\mu}_k^{(q+1)})^2}{\sum_{i=1}^n \tau_{k,i}^{(q)}}.$$

Annexe B

Echantillonnage de MRF

La définition d'un champ de Markov passe par la définition de sa fonction d'énergie U . Celle-ci nécessite la définition d'un système de voisinage, qui définit alors le système de cliques, et de fonctions de potentiel associées aux cliques. Ces fonctions de potentiel permettent d'accéder à la probabilité globale d'une configuration, et aux probabilités conditionnelles locales. Le problème qui se pose alors est, étant défini un champ de Markov, comment pouvons nous réaliser le tirage d'une configuration (une image ici) en suivant la loi de probabilités de Gibbs caractéristique de ce champ ? Deux algorithmes ont été proposés pour synthétiser des réalisations d'un champ de Markov qui sont l'échantillonneur de Gibbs et l'algorithme de Métropolis que nous allons d'écrire dans le paragraphe suivant.

3.6.2 L'échantillonneur de Gibbs

Cet algorithme, proposé par Geman et Geman [59], repose sur la construction itérative d'une suite d'images. A la convergence, i.e après un nombre d'itérations suffisant, les images construites sont des réalisations tirées selon la loi de Gibbs globale. La méthode de construction de l'image à l'itération n , partant de l'image à l'itération $n - 1$ se fait par mise à jour successive des sites de l'image. A l'étape n :

- choix d'un site s ;
- au site s , selon la configuration des voisins V_s pour l'image $x^{(n-1)}$, calcul de la probabilité conditionnelle locale :

$$p(X_s = x_s/V_s) = \frac{\exp(-U_s(x_s/V_s))}{\sum_{x_s \in \Lambda} \exp(-U_s(\xi/V_s))}$$

- mise à jour du site s par tirage aléatoire selon la loi $p(X_s = x_s/V_s)$.

On considère que l'algorithme a convergé après un grand nombre d'itérations ou lorsque le nombre de changements est faible. Le choix du site s considéré à l'étape n peut se faire de n'importe quelle façon à condition de balayer tous les sites un très grand nombre de fois (théoriquement un nombre infini de fois). Les méthodes usuelles consistent à tirer un site selon une loi uniforme, ou exécuter un balayage classique, ligne par ligne, de l'image. Cet algorithme construit en réalité une suite d'images $x^{(n)}$ qui sont les observations d'une suite $X^{(n)}$ de champs aléatoires constituant une chaîne de Markov pour un certain noyau de transition. On peut montrer le théorème suivant, lorsque la séquence balaye chaque site une infinité de fois :

$$\forall x^{(0)} \forall x \in \Omega \quad \lim_{n \rightarrow +\infty} p(X^{(n)} = x / X^{(0)} = x^{(0)}) = p(x),$$

où p est la mesure de Gibbs associée au champ de Markov considéré. Ainsi, après un grand nombre d'itérations, les images $x^{(n)}$ générales sont des réalisations de la loi globale $p(x)$, et ceci indépendamment de la configuration initiale $x^{(0)}$.

On parle de l'échantillonneur de Gibbs comme d'un algorithme de "relaxation", car il procède par mise à jour successive des sites, et "probabiliste" car celle-ci est fondée sur un tirage aléatoire.

3.6.3 L'algorithme de Métropolis

L'échantillonneur de Gibbs est un algorithme très utilisé en traitement d'images pour la synthèse de champs de Markov. Néanmoins, un algorithme antérieur et issu de la physique statistique

avait été mis au point dans les années 50 par Métropolis [89]. Cet algorithme repose sur un principe similaire à l'échantillonneur de Gibbs, et il s'agit également d'un algorithme de relaxation probabiliste. Le principe est là encore de construire une suite d'images qui seront des tirages selon la loi du champ de Markov après un nombre suffisamment grand d'itérations. Mais la mise à jour en un site s'effectue de façon différente. Ainsi à l'étape n :

- choix d'un site s
- tirage aléatoire d'un descripteur λ dans E selon une loi uniforme ;
- calcul de la variation d'énergie pour $x_s^{(n-1)} \longrightarrow \lambda$:

$$\Delta U = U_s(\lambda/V_s^{(n-1)}) - U_s(x_s^{(n-1)}/V_s^{(n-1)})$$

- deux cas sont alors possibles :
 1. $\Delta U < 0$, le changement est accepté : $x_s^{(n)} = \lambda$;
 2. $\Delta U \geq 0$, le changement est accepté ou refusé par tirage selon la probabilité $p = \exp(-\Delta U)$ et $1 - p$.

Le système de balayage des sites et le critère d'arrêt sont similaires à ceux de l'échantillonneur de Gibbs. La différence avec l'échantillonneur de Gibbs réside dans le tirage au sort du nouveau niveau de gris (ou descripteur), au lieu de considérer la loi définie par tous les descripteurs. Comme on ne considère que la variation énergétique entre les deux configurations, l'algorithme de Métropolis est plus rapide à chaque étape que l'échantillonneur de Gibbs, qui lui nécessite le calcul de la fonction de partition locale. Mais la convergence peut être plus lente car le taux d'acceptation est strictement inférieur à 1 (les transitions ne sont pas toujours acceptées, contrairement au cas de l'échantillonneur de Gibbs).

Là encore, le principe est de construire une chaîne de Markov selon un certain noyau de transition (différent de celui intervenant dans l'échantillonneur de Gibbs).

3.6.4 Quelques MRF fondamentaux

Nous présentons ici quelques uns des champs de Markov les plus utilisés. Comme indiqué précédemment, ces champs sont définis par leur voisinage et leurs fonctions de potentiel. Ils sont illustrés par le tirage de réalisations selon l'échantillonneur de Gibbs.

Modèle d'Ising :

Ce modèle est le plus ancien (1925 [71]) et a été développé lors de l'étude du ferromagnétisme en physique statistique. L'espace des descripteurs est celui des états des spins, i.e $E = \{-1, 1\}$ (espace binaire), et le voisinage est constitué par les 4 ou 8 plus proches voisins dans un espace bidimensionnel. Les potentiels sont des potentiels en tout ou rien :

$$\begin{aligned} U_{c=(s,t)}(x_s, x_t) &= -\beta \text{ si } x_s = x_t \\ &= +\beta \text{ si } x_s \neq x_t. \end{aligned}$$

Ce qui s'écrit également :

$$U_{c=(s,t)}(x_s, x_t) = -\beta x_s x_t.$$

Les potentiels des cliques d'ordre 1 (clique constituée par un seul spin) sont de la forme $-Bx_s$.

L'énergie totale s'écrit :

$$U(x) = - \sum_{c=(s,t) \in C} \beta x_s x_t - \sum_{s \in S} B x_s$$

β est la constante de couplage entre sites voisins et B représente un champ magnétique externe.

Lorsque β est positif, les configurations les plus probables (i.e d'énergies plus faibles) sont celles pour lesquelles les spins sont de même signes (ferro-magnétisme), alors que dans le cas de β négatif, au contraire, on favorisera l'alternance de spins de signes opposés (antiferromagnétisme).

La valeur (signe et valeur absolue) de β conditionne donc la régularité du modèle d'Ising. Quant au champ magnétique externe relatif au potentiel d'ordre 1, il favorise a priori par son signe un spin ou un autre.

Modèle de Potts :

Il s'agit d'une extension du modèle d'Ising [110] pour un espace m -aire, i.e $E = \{0, m - 1\}$. Il peut s'agir de plusieurs niveaux de gris, mais plus souvent pour ce modèle, d'étiquettes (labels) pouvant représenter une classification de l'image (par exemple les classes eau, forêt, champ, ville). Le voisinage considéré est 4-ou 8-connexe et les potentiels sont comme précédemment en tout ou rien mais définis seulement pour les cliques d'ordre 2 :

$$\begin{aligned} U_{c=(s,t)}(x_s, x_t) &= -\beta \text{ si } x_s = x_t \\ &= +\beta \text{ si } x_s \neq x_t. \end{aligned}$$

Lorsque β est positif, les configurations les plus probables correspondent à des sites voisins de même niveau de gris ou descripteur, ce qui donne des réalisations constituées par des larges zones homogènes. La taille de ces régions est gouvernée par la valeur de β .

Modèle Markovien Gaussien :

Ce modèle est réservé aux images en niveaux de gris $E = \{0, \dots, 255\}$ et ne convient pas bien aux images d'étiquettes. Le voisinage est 4 ou 8-connexe et l'énergie est de la forme :

$$U(x) = \beta \sum_{c=(s,t)} (x_s - x_t)^2 + \alpha \sum_{s \in S} B(x_s - \mu_s)^2.$$

Le premier terme correspondant aux cliques d'ordre 2 est un terme de régularisation, qui favorise les faibles différences de niveaux de gris entre sites voisins pour $\beta > 0$. Le second terme peut correspondre à un terme d'attache aux données dans le cas où on possède une image de données extérieures.

Le modèle Gaussien favorise des niveaux de gris proches pour des pixels voisins dans tous les cas. Or si on considère une image naturelle cet aspect est néfaste à proximité des contours car il favorisera la présence d'un dégradé. Aussi, de nombreuses fonctions ϕ ont été proposées pour modéliser les potentiels des cliques d'ordre 2 : $U_{c=(s,t)} = \phi(x_s - x_t)$. En particulier, la fonction suivante permet de respecter les contours de l'image [60] :

$$\phi(x) = \frac{1}{1 + (\frac{x}{\delta})^2}$$

et est donc très utilisée en restauration. Ces modèles permettent de synthétiser des textures très variées.

Bibliographie

- [1] Abend, K., Harley, T. J., and Kanal, L. N. (1965). Classification of binary random pattern. IEEE Transactions on Information Theory, 11, p. 538-544.
- [2] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, Second International Symposium on Information Theory, p. 267-281.
- [3] Amini, M. R. (2001). Apprentissage automatique et recherche de l'information : application à l'extraction d'information de surface et au résumé de texte. PhD thesis, Université Paris 6.
- [4] Anderson GL, de Figueiredo RJP. (1980). An adaptive orthogonal series estimator for probability density functions. Ann Stat, 8, p.347-376.
- [5] Appriou, A. (1993). Formulation et traitement de l'incertain en analyse multi-senseurs. In Quatorzième Colloque GRETSI, p. 951-954, Juan les Pins.
- [6] Atefeh, F., Asghar, K., Morteza G., Mahdi, M. and Shahrokh N. (2012). 3D Reconstruction Using Cubic Bezier Spline Curves and Active Contours. Iranian journal of Medical Physics, volume 9, Issue 3, p. 169-176.
- [7] Aupetit, S., Slimane, M., and Monmarché, N. (2002). Les modèles de Markov cachés à substitutions de symboles. Technical Report 262, Laboratoire d'Informatique de l'Université François-Rabelais de Tours, Tours, France.

-
- [8] Aupetit, S. (2005). Contributions aux modèles de Markov cachés : métaheuristiques d'apprentissage, nouveaux modèles et visualisation de dissimilarité. Thèse de doctorat, Laboratoire d'Informatique de l'Université François-Rabelais de Tours, Tours, France.
- [9] Balabanovic, M., Shoham, Y. (1997). Fab : Content-Based, Collaborative Recommendation, *Communications of the ACM*, vol. 40, no. 3, p. 66-72.
- [10] Bartels, R., Beutty, J., Barsky, B. (1987). *An Introduction to Splines for use in Computer Graphics and Geometric Modeling*. Morgan Kaufmann Publishers, Lost Altos, CA. p. 347-376.
- [11] Ben Arab, T. (2014). Contribution des familles exponentielle en traitement des images. Thèse soutenue à l'université de Sfax, 26 avril, 2014.
- [12] Ben Arab, T., Zribi, M., Masmoudi, A. (2013). Unsupervised Satellite Image Segmentation using a Bivariate Beta type-II Mixture Model. *IEEE 9 th International Conference on Computational Cybernetics*, July 8-10, Tihany, Hungary, p. 95-100.
- [13] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Royal Statist. Soc. (series B)*, 36, p. 192-326.
- [14] Bernardo J., Smith A.F.M. (1994). *Bayesian Theory*. John Wiley, New York.
- [15] Beucher, S., Meyer, F. (1993). The morphological approach to segmentation : the watershed transformation. *Mathematical Morphology in Image Processing*, Ed. E.R. Dougherty, p. 433-481.
- [16] Bishop, M., Thompson, E. (1986). Maximum likelihood alignment of DNA sequences, *J Mol Biol*, vol. 190, no 2, p. 159-65.
- [17] Bloch, I., Maître, H. (1994). Fusion de données en traitement d'images : modèles d'information et décisions, *Traitement du Signal*, Vol. 11 N. 6.
-

-
- [18] Bloch, I., Maître, H. (2002). Fusion en traitement d'images : spécificités et approches numériques, *Technique de l'ingénieur*, TE 5 230, p. 1-26.
- [19] Bloch, I. (2003). Fusion d'information en traitement du signal et d'images, Hermès, Paris.
- [20] Bosq, D. (1998). Nonparametric statistics for stochastic processes, volume 110 of *Lecture Notes in Statistics*. Springer-Verlag, New York, second edition. Estimation and prediction.
- [21] Bouveyron, C. (2006). Modélisation et classification des données de grande dimension : Application à l'analyse d'image. Thèse soutenue à l'université Joseph Fourier, INRIA Rhône-Alpes, 28 septembre, 2006.
- [22] Bres, S., Jolion, J.M., Lebourgeois, F. (2003). *Traitement et analyse des images numériques*, Hermès, 412 pages.
- [23] Brouard, T. (1999). Algorithmes hybrides d'apprentissage de chaînes de Markov cachées : conception et applications à la reconnaissance des formes. PhD thesis, Laboratoire d'Informatique de l'Université François-Rabelais de Tours.
- [24] Carbon, M., Francq, C. (1999). Estimation non paramétrique de la densité et de la régression-Prévision non paramétrique. *Revista de Matemática : Teoría y Aplicaciones*, Vol 6(1), p.1-26.
- [25] Celeux, G and Diebolt, J (1985). The sem algorithm : a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly*, Vol 2(1), p.73-92.
- [26] Celeux, G and Govaert, G (1992). A classification EM algorithm for clustering and two stochastic versions. *Computational Statistics and Data Analysis*, Vol 14(3), p.315-332.
- [27] Chahboun, M. (1984). Estimation des densités de la loi et les probabilités de transition par les méthodes du noyau, des δ -suites et des fonctions orthogonales pour un processus de
-

- Markov stationnaire et mélangeant. PhD thesis, Université de Rouen.
- [28] Chauvin, S. (1995). Evaluation des performances du modèle bayésien de fusion appliqué à l'imagerie satellitaire, 15^e colloque GRETSI, Juan les Pins, Septembre 1995, pp 949-952.
- [29] Cheng, H.D., Jiang, X.H., Sun, Y., Wang, J. (2001). Color image segmentation : advances and projects. *Pattern Recognition*, 34(12), p. 2259-2281.
- [30] Chow C.K. (1965). Statistical Independence and Threshold Functions, *IEEE Transactions of Electrical Computer*, p. 66-68.
- [31] Clark, J.J., Yuille, A.L., (1990). *Data Fusion for Sensory Information Processing Systems*, Kluwer Academic Publishers, Boston, 1990.
- [32] Cook, R. (1988). Uncertainty in Risk Assessment : A Probabilist's Manifesto, *Reliability in Engineering and Systems Safety*, 23, p. 277-283.
- [33] Cook, R. (1991). *Experts in uncertainty*, Oxford university Press, GB.
- [34] Curry, H.B., Schoenberg, I.J. (1966). On Polya frequency functions. IV. The fundamental spline functions and their limits. *J. Analyse Math.* Vol. 17, p. 71-107.
- [35] Daga, I., Sakab, D. and Irk, D. (2004). Application of cubic B-splines for numerical solution of the RLW equation. *Applied Mathematics and Computation*, Vol.159, p. 373-389.
- [36] Daub, C., Steuer, R., Selbig, J. and Kloska, S. (2004). Estimating mutual information using B-spline functions : An improved similarity measure for analysing gene expression data, *Bioinformatics* 5(1), 118.
- [37] Dempster, A. Laird, N. and Rubin, D (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B*, 39 p.1-38.
- [38] Dempster, A. P. (1968). A generalization of Bayesian inference. *Journal of the Royal StatisticaZ Society, Series B*, 30, p. 205-247.
-

-
- [39] Dempster, A.P., Laird N.M., Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the E.M. Algorithm, *Journal of the Royal Statistical Society B* 977, Vol.39, New York.
- [40] Deheuvels, P. (1977). Estimation non paramétrique de la densité par histogrammes généralisés. *Revue de Statistique Appliquée*, Vol. 25, p. 5-42.
- [41] Delecroix, M. (1975). Sur l'estimation des densités marginales et de transition d'un processus stationnaire et mélangeant. PhD thesis, Université de Lille.
- [42] Diggle PJ, Hall P. (1986). The selection of terms in an orthogonal series density estimator. *J. Am Stat Assoc*, 81, p. 230-233.
- [43] Donald P, Connelly Keith E. Willard. (1992). Nonparametric probability density estimation : Improvements to the histogram for laboratory data. *Computers and Biomedical Research*, Vol. 25, p. 17-28.
- [44] Dromigny-Badin, A. (1998). Fusion d'images par la théorie de l'évidence en vue d'applications médicales et industrielles, thèse de doctorat de l'institut national des sciences appliquées de Lyon.
- [45] Dubois, D., Prade, H. (1988). Representation and combination of uncertainty with belief functions and possibility measures, *Computational Intelligence*, Vol.4, p. 244-264.
- [46] El-Yacoubi, A., Gilloux, M., Sabourin, R., and Suen, C. Y. (1999). An HMM-based approach for off-line unconstrained handwritten word modelling and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 21(8), p. 752-760.
- [47] Eskicioglu, A.M, Fisher, P.S. (1995). Image Quality Measures and their Performance, *IEEE Transactions On comm...*, Vol.43, No.12, p.2959-2965.
-

-
- [48] Estévez, G., Vieu, P., (2003). Nonparametric estimation under long memory dependence. *J. Nonparametr. Stat.*, 15(4-5), p. 535-551.
- [49] Euroimage, Image gallery, Landasat. [http : //www.eurimage.com/gallery/webfiles/ls_intro.html](http://www.eurimage.com/gallery/webfiles/ls_intro.html).
- [50] Fan, Z. and Cohen, F.S. (1986). Unsupervised texture image segmentation. In *proc . Int. Congress Intelligent Autonomous Systeme*, Amsterdam, Holland, Dec. 1986.
- [51] Fayon D. (2006). Clés pour Internet, chapitre 1, historique et principes du réseau. *Economica*.
- [52] Feller W. (1971). *An introduction to probability theory and its applications*. Wiley and Sons, Vol. 2, 2nd edition, Inc, New York.
- [53] Ferguson, T. (1973). A bayesian analysis of some nonparametric problems. *The annals of statistics* 1, 209-230.
- [54] Ferguson, T. (1974). Prior distributions on spaces of probability measures. *The annals of statistics* 2, p. 615-629.
- [55] Fine, S., Singer, Y., and Tishby, N. (1998). The hierarchical hidden markov model : Analysis and applications. *Machine Learning*, Vol.32(1), p. 41-62.
- [56] French, S. (1985). Group Consensus Probability Distributions : A Critical Survey. In J. Bernardo et al., editor, *Bayesian Statistics*, pp. 183-201, Elsevier, The Netherlands.
- [57] Friedland, N., Adam, D. (1989). Automatic ventricular cavity boundary detection from sequential ultrasound images using simulated annealing, *IEEE Trans. on Medical Imaging*, vol. 8, n° 4, p. 344-353.
- [58] Gebhard, M., Mattes, J., Eils, R. (2001). An Active Contour Model for Segmentation Based on Cubic B-splines and Gradient Vector Flow. *4th International Conference of Medical*
-

Image Computing and Computer-Assisted Intervention - MICCAI 2001, Utrecht, The Netherlands, October 14-17.

- [59] Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distribution, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6), p. 721-741.
 - [60] Geman, S. and McClure, D. E. (1985). Bayesian image analysis : an application to single photon emission tomography. *Proc. Statist. Comput. sect. (Amer. Statist. Assoc. Washington DC)*, p. 12-18.
 - [61] Grandin, J.-F, Marques, M. (2000). Robust Data Fusion, Fusion 2000, p. MoC3-1-9, Paris.
 - [62] Hadrich, A., Zribi, M., Masmoudi, A. (2012). An unsupervised image segmentation using B-splines function, *IEEE Spring World Congress on Engineering and Technology*, p. 1-5.
 - [63] Hadrich, A., Zribi, M., Masmoudi, A. (2012). A proposed normalized B-spline density estimator and it application in unsupervised statistical image segmentation, *2012 IEEE International Conference on Systems, Man, and Cybernetics*, Seoul, Korea, p. 480-483.
 - [64] Hadrich, A., Zribi, M., Masmoudi, A. (2013). Using B-splines functions and EM algorithm for Hidden Markov Model-based Unsupervised Image Segmentation, *ICAAM Special Issue on Advances in Applied Mathematics*, Hammamet, 16-19 December, In press.
 - [65] Hadrich, A., Zribi, M., Masmoudi, A. (2014). A proposed nonparametric mixture density estimation using B-splines functions, *Journal of Neurocomputing (soumis)*.
 - [66] Hadrich, A., Zribi, M., Masmoudi, A. (February 2016). Bayesian expectation maximization algorithm by using B-splines functions : Application in image segmentation, *journal of Mathematics and Computers in Simulation*, Vol. 120, p. 50-63.
-

-
- [67] Hagge, J. K, Wagner, R.J. (1992). High-Yield Assembly of Multi-Chip Modules through Known-Good IDs and Effective Test Strategies, Proceedings of the IEEE, Vol. 80, N. 12, p. 1965-1994.
- [68] Hall P. (1983). Orthogonal series distribution function estimation, with application,. J, R, Statist ; Soc, New York.
- [69] Hall P. (1986). On the rate of convergence of orthogonal series density estimators. J, R, Statist ; Soc, New York.
- [70] Hillion, A., Boucher, J. M., Roux, C. (1989). Le Traitement des images de télédétection : Aperçu et Perspectives, Télédétection en Francophonie, ED, AUPELF-UREF, JohnLibby, Eurotext, Paris, p. 97-110.
- [71] Ising, E. (1925). Beitrag zur theorie des ferromagnetisms. Zeitschrift fur Physik, 31, 253-258.
- [72] Janez, F., Appriou, A. (1998). Theory of evidence and non-exhaustive frames of discernment : Plausibilities correction methods, International Journal of Approximate Reasoning, Vol.18, p. 1-19.
- [73] Kanal, L. N. (1980). Markov mesh models. Image Modeling, New York, Academic Press.
- [74] Kelly P.A. et al. (1988). Adaptative segmentation of speckled images using a hierarchical random field model, IEEE transactions on Acoustic, Speech and Signal processing, p. 1628-1641.
- [75] Kervrann, C., Heitz, F. (1994). A hierarchical statistical framework for the segmentation of deformable objects in image sequences , IEEE Conf. Computer Vision Pattern Recognition, Seattle, p. 724-728.
- [76] Kumar M.P. and Koller.D (2009). Learning a small mixture of trees. In Neural Information Processing Systems (NIPS 2009).
-

-
- [77] Kundu, A. and Bahl, P. (1988). Recognition of handwritten script : a hidden Markov model based approach. In International Conference on Acoustics, Speech, Signal Processing (ICASSP'88), p. 928-931.
- [78] Lawrence R. Rabiner. (1989). A tutorial on Hidden Markov Models and selected applications in speech recognition, Proceedings of the IEEE, vol. 77, no 2, p. 257-286.
- [79] Leonard E. Baum and Ted Petrie. (1966). Statistical Inference for Probabilistic Functions of Finite State Markov Chains. Ann. Math. Statist. Volume 37, Number 6, p. 1554-1563.
- [80] Leonard E. Baum and Eagon J. A. (1967). An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. Bull. Amer. Math. Soc. 73, p. 360-363.
- [81] Likforman L., Sigelle M. (2005). Reconnaissance des Formes, Edition TSI-ENST Paris, 45 pages.
- [82] Massey, J.L. (1990). Causality, feedback and directed information.
- [83] Maritz J.S., Lwin T. (1989). Empirical Bayes Methods. Chapman and Hall, 2nd ed.
- [84] Martin, A. (2005). La fusion d'informations, Polycopié de cours ENSIETA, Réf 1484.
- [85] Marsault, X., Aubry, M., Fortin, J., Azizi, L. (2009). Classification markovienne automatique d'images aériennes de haute résolution. *www.aria.archi.fr/.../2009 – Classification – markovienne – automatique.pdf*
- [86] Marzouki A., Segmentation statistique d'images radar, Thèse de l'Université des Sciences et Technologies de Lille. Novembre 1996, 156 pages.
- [87] Manjunath and Chellappa R. (1991). Unsupervised texture segmentation using markov random fields. IEEE Trans . PAMI, Vol . 13, pp . 478-482,1991 .
-

-
- [88] Masri R, Redner R.A. (2005). Convergence rates for uniform B-splines density estimators on bounded and semi-infinite domains. *Nonparametric Statistics*, Vol. 17, pp.555-582.
- [89] Metropolis, N., Rosenbluth, A. W., Rosenbluth, N. M., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chemical Physics*, 21, p. 1087-1091.
- [90] McLachlan G.J., Peel D.(2000). *Finite Mixture Models*. John Wiley and Sons, New York.
- [91] Olivier, C., Avila, M., Courtellemont, P., Paquet, T., and Lecourtier, Y. (1993). Handwritten word recognition by image segmentation and hidden Markov model. In *Proceedings of IEEE IES 3*, pages 2093-2097.
- [92] Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics* 33 (3) : 1065.
- [93] Pearl, J. (1986). Fusion, Propagation and Structuring in Belief Networks, *Artificial Intelligence*, 29, p. 241-288.
- [94] Pearl, J. (1986). On Evidential Reasoning in a Hierarchy of Hypotheses, *Artificial Intelligence*, 28, p. 9-15.
- [95] Pieczynski W. (2003). Modèles de markov en traitement d'images, *Revue Traitement du Signal*, 20, p. 255-278.
- [96] Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, volume 77, pages 257-286.
- [97] Rekik, A., Zribi, M., Ben Hamida, A., Benjelloun, M. (2008). An Optimal unsupervised satellite image segmentation approach based on Pearson system and k-means clustering algorithm initialization. *IJSP, International Journal of Signal Processing*, Vol. 5 N° 1, p. 20-27.
-

-
- [98] Robbins, H. (1956). An empirical bayes approach to statistics, Proc. Third Berkley Symposium on Mathematical Statistics, 1, 157-163.
- [99] Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function". The Annals of Mathematical Statistics 27 (3) : p. 832-837.
- [100] Rouquet, C., Bonton, P. (1997). Segmentation en régions non supervisée par relaxation markovienne, une étude comparative, in Proc. Actes 16 ème colloque GRETSI, Grenoble, Sept. 15-19, p. 31-34.
- [101] Samaria, F. and Harter, A. (1994). Parameterisation of a stochastic model for human face identification. In IEEE workshop on Applications of Computer Vision, Florida.
- [102] Schwartz S. (1967). Estimation of probability density by an orthogonal series. Ann Math Stat, 38 : p. 1262-1265.
- [103] Soukhal, A., Kelarestaghi, M., Slimane, M., and Martineau, P. (2001). Hidden Markov Models and scheduling problem with transportation consideration. In 15th Annual European Simulation Multiconference (ESM'01), p. 836-840, Prague (Tchéquie).
- [104] Soukhal, A., Kelarestaghi, M., Slimane, M., and Martineau, P. (2001). Application des chaînes de Markov cachées au problème d'ordonnancement dans une cellule robotisée. In Conférence Internationale sur la productique (CIP'01), p. 151-156, Algérie.
- [105] Taskin Atilgan. (1992). Convergence properties of MLE's and asymptotic simultaneous confidence intervals in fitting cardinal B-splines for density estimation. Statistics and Probabilities letters, Vol. 13, p. 89-98, North-Holland.
- [106] Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2004). Hierarchical dirichlet processes. Journal of the American Statistical Association 101.
-

-
- [107] Wald, L. (1999). Some terms of reference in data fusion, *IEEE Trans. Geosci. Remote Sens.*, vol. 37, p. 1190-1193.
- [108] Watson GS. (1969). Density estimation by orthogonal series. *Ann Math Stat*, 38, p. 1262-1265.
- [109] Won C.S., Drein, H. (1992). Unsupervised segmentation of noisy and textured images using markov random fields, *CVG1P*, Vol. 54, N°4, p. 308- 328.
- [110] Wu, F. Y. (1982). The potts model. *Reviews of Modern Physics*, 54(1), p. 235-267.
- [111] Ypma T.J. (1995). Historical developement of the Newton-Raphson method. *SIAM Review*, 37, p. 531-551.
- [112] Zadeh, L.A. (1978). Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems*, vol. 1, p. 3-28.
- [113] Zaragoza, H. and Gallinari, P. (1998). Modèle hiérarchique et extraction de l'information et d'extraction de l'information textuelle de surface. In *Journée Francophones d'apprentissage (JFA'98)*, page 12.
- [114] Zribi, M., Ghorbel F. (2003). An Unsupervised and Non-parametric Bayesian Classifier. *Pattern Recognition. Letters*, vol. 24, pp. 97-112.
- [115] Zribi, M. (2007). Unsupervised Bayesian Image Segmentation Using Orthogonal Series, *Journal of Visual Communication and Image Representation*, 18, pp. 496-503.
-

Résumé : Dans cette thèse nous avons traité le problème de l'estimation non paramétrique des lois de probabilités. Dans un premier temps, nous avons supposé que la densité inconnue f a été approchée par un mélange de base B-spline quadratique. Puis, nous avons proposé un nouvel estimateur de la densité inconnue f basé sur les fonctions B-splines quadratiques, avec deux méthodes d'estimation. La première est basée sur la méthode du maximum de vraisemblance et la deuxième est basée sur la méthode d'estimation Bayésienne MAP. Ensuite, nous avons généralisé notre étude d'estimation dans le cadre du mélange et nous avons proposé un nouvel estimateur du mélange de lois inconnues basé sur les deux méthodes d'estimation adaptées. Dans un deuxième temps, nous avons traité le problème de la segmentation statistique semi supervisée des images en se basant sur le modèle de Markov caché et les fonctions B-splines. Nous avons montré l'apport de l'hybridation du modèle de Markov caché et les fonctions B-splines en segmentation statistique bayésienne semi supervisée des images. Dans un troisième temps, nous avons présenté une approche de fusion basée sur la méthode de maximum de vraisemblance, à travers l'estimation non paramétrique des probabilités, pour chaque pixel de l'image. Nous avons ensuite appliqué cette approche sur des images multi-spectrales et multi-temporelles segmentées par notre algorithme non paramétrique et non supervisé.

Mots clés : Estimation non paramétrique, Fonction B-spline, Fusion Bayésienne, Mélange des lois de probabilités, Méthode Bayésienne, Modèle de Markov caché, Segmentation statistique.

Abstract: In this thesis we are treated the problem of nonparametric estimation probability distributions. At first, we assumed that the unknown density f was approximated by a basic mixture quadratic B-spline. Then, we proposed a new estimate of the unknown density function f based on quadratic B-splines, with two methods estimation. The first is based on the maximum likelihood method and the second is based on the Bayesian MAP estimation method. Then we have generalized our estimation study as part of the mixture and we have proposed a new estimator mixture of unknown distributions based on the adapted estimation of two methods. In a second time, we treated the problem of semi supervised statistical segmentation of images based on the hidden Markov model and the B-spline functions. We have shown the contribution of hybridization of the hidden Markov model and B-spline functions in unsupervised Bayesian statistical image segmentation. Thirdly, we presented a fusion approach based on the maximum likelihood method, through the nonparametric estimation of probabilities, for each pixel of the image. We then applied this approach to multi-spectral and multi-temporal images segmented by our nonparametric and unsupervised algorithm.

Keywords: Nonparametric estimation, B-spline function, Bayesian Fusion, Mixing distributions of probability, Bayesian method, Hidden Markov model, Statistical segmentation.
