

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS 6

Spécialité : Mathématiques
Option : Statistiques

Présentée par
Emmanuel MONFRINI

Sujet de la thèse
**IDENTIFIABILITÉ ET MÉTHODE DES MOMENTS
DANS LES MÉLANGES GÉNÉRALISÉS
DE DISTRIBUTIONS DU SYSTÈME DE PEARSON**

TOME I

Thèse soutenue le 4 janvier 2002, devant le jury composé de :

M.	Marc YOR	Université Paris 6	<i>Président</i>
M.	Alain HILLION	ENST Bretagne	<i>Rapporteur</i>
M.	Christian ROBERT	Université Paris Dauphine	<i>Rapporteur</i>
M.	Michel BRONIATOWSKI	Université de Reims	<i>Examineur</i>
M.	Paul DEHEUEVELS	Université Paris 6	<i>Examineur</i>
M.	Wojciech PIECZYNSKI	INT	<i>Directeur</i>
M.	Daniel PIERRE-LOTI-VIAUD	Université Paris 6	<i>Directeur</i>
Mme	Annick VALIBOUZE	Université Paris 6	<i>Invitée</i>

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS 6

Spécialité : Mathématiques
Option : Statistiques

Présentée par
Emmanuel MONFRINI

Sujet de la thèse
IDENTIFIABILITÉ ET MÉTHODE DES MOMENTS
DANS LES MÉLANGES GÉNÉRALISÉS
DE DISTRIBUTIONS DU SYSTÈME DE PEARSON

TOME I

Thèse soutenue le 4 janvier 2002, devant le jury composé de :

M.	Marc YOR	Université Paris 6	<i>Président</i>
M.	Alain HILLION	ENST Bretagne	<i>Rapporteur</i>
M.	Christian ROBERT	Université Paris Dauphine	<i>Rapporteur</i>
M.	Michel BRONIATOWSKI	Université de Reims	<i>Examineur</i>
M.	Paul DEHEUEVELS	Université Paris 6	<i>Examineur</i>
M.	Wojciech PIECZYNSKI	INT	<i>Directeur</i>
M.	Daniel PIERRE-LOTI-VIAUD	Université Paris 6	<i>Directeur</i>
Mme	Annick VALIBOUZE	Université Paris 6	<i>Invitée</i>

REMERCIEMENTS.

En premier lieu, je souhaite remercier messieurs Deheuvels et Bosq qui m'ont accueilli au LSTA en DEA, qui m'ont donné envie et offert la possibilité d'entreprendre ce travail de thèse, et dont le soutien et la disponibilité m'ont permis de mener, en confiance, les travaux de recherche exposés dans la suite de ce document. Que messieurs Hillion et Robert soient assurés de ma plus sincère gratitude pour avoir accepté de s'investir dans leur rôle de rapporteur. Les critiques et les conseils constructifs qu'ils ont pu émettre à cette occasion m'ont permis de prendre du recul par rapport au sujet et d'effectuer une relecture efficace du manuscrit.

Je remercie, bien sûr, messieurs Broniatowski et Yor qui ont accepté de prendre sur leurs congés pour me faire l'honneur de participer au jury à l'occasion de ma soutenance.

Je remercie aussi Mme Valibouze qui a accepté notre invitation à participer au jury, et dont les encouragements pour que nous persévérions dans l'aspect "Calcul Formel" du sujet, ont été déterminants dans l'aboutissement de ce travail.

Ce travail a été réalisé dans le cadre d'une collaboration étroite entre mes deux codirecteurs, M. Pierre-Lotivaud du LSTA et M. Pieczynski de l'INT. Cette alliance, complexe sur le plan des objectifs statistiques, s'est révélée très enrichissante et m'aura permis, non seulement d'avancer sereinement dans mes recherches, mais surtout de progresser dans mon approche globale de la statistique. Je remercie mes directeurs pour l'énergie qu'ils ont mise dans cette collaboration, pour leur disponibilité et leur patience à mon égard, et surtout pour m'avoir permis de mener ces travaux à leur terme.

Il serait impensable, pour moi, de ne pas remercier chaleureusement madame Lamart, dont la gentillesse n'a d'égale que la bonne humeur dont elle sait faire preuve au quotidien. Elle a su proposer une solution à tous ces petits problèmes matériels ou administratifs qu'on ne soupçonne pas et auxquels j'ai été confronté durant ma présence au LSTA. Merci pour tout !!!

J'adresse mes remerciements complices à Jean-Baptiste et Florence, mes "assurances tous risques", toujours prêts à recentrer mes égarements mathématiques, ainsi qu'à Claude et Sébastien, qui m'ont associé à leur passion pour la musique mécanique en m'offrant la possibilité de développer une application originale de mes recherches. Je remercie aussi Michel, pour qui tout problème a sa solution ...

Enfin, et ce n'est pas un hasard si j'utilise ce terme, je souhaite embrasser comme ils le méritent ceux qui ont été mes plus proches soutiens, qui ont pu avoir l'impression d'être mes souffre-douleur, ceux qui ont, quoiqu'il arrive, autant de plaisir que moi à voir se profiler la fin du travail. Merci Francky, merci Hughette, merci Magali !!!

TABLE DES MATIÈRES.

COUVERTURE

PREMIÈRE PARTIE.

MÉLANGES DE LOIS DE PROBABILITÉ, IDENTIFIABILITÉ.	25
Chapitre 1 Introduction aux mélanges.	27
1.1 Définition.	27
1.2 Mélanges finis.	27
1.3 Définition probabiliste des mélanges.	28
1.3.1 Rappel sur les fonctions de répartition.	28
1.3.2 Théorème fondamental.	28
1.4 Distribution d'un échantillon.	29
Chapitre 2 Identifiabilité d'une classe de mélanges.	31
2.1 Définition.	31
2.2 Identifiabilité des mélanges de distributions à paramètres additifs et multiplicatifs.	32
2.2.1 Familles additivement fermées.	32
2.2.2 Théorème d'identifiabilité.	32

2.2.3	Exemples d'applications.	33
2.3	Identifiabilité des classes de mélanges de la famille exponentielle.	33
2.4	Identifiabilité des classes de mélanges de distributions gaussiennes.	35
2.4.1	Distribution normale à une dimension.	35
2.4.2	Théorème d'identifiabilité.	35
2.4.3	Extension au cas multidimensionnel.	36
2.5	Identifiabilité des classes de mélanges de distributions continues à un paramètre sur \mathbb{R}	36
2.6	Identifiabilité des classes de mélanges de mesures produits.	37
2.7	Indépendance de l'identifiabilité par rapport à la paramétrisation.	38
Chapitre 3	Identifiabilité des mélanges finis.	39
3.1	Condition suffisante pour l'identifiabilité des classes de mélanges finis.	40
3.1.1	Théorème d'identifiabilité de Teicher.	40
3.1.2	Identifiabilité de la classe des mélanges finis de distributions normales.	40
3.1.3	Identifiabilité des classes de mélanges de distributions gammas dont les densités ont le même support.	40
3.1.4	Identifiabilité de la classe des mélanges finis de distributions inverse-gamma dont le support des densités est le même.	42
3.2	Une CNS d'identifiabilité de la classe des mélanges finis.	42
3.2.1	Théorème de Yakowitz et Spragins.	42
3.2.2	Applications du théorème de Yakowitz et Spragins.	43
3.2.3	Identifiabilité de la classe des mélanges finis de distributions de Cauchy.	43
3.3	Mélanges finis de mesures produits.	44
3.4	D'autres mélanges finis particuliers.	45
Chapitre 4	Introduction des mélanges généralisés.	47

4.1	Définition générale	47
4.2	Mélanges finis généralisés.	48
Chapitre 5	Système des distributions de Pearson.	49
5.1	Définition.	49
5.2	Densités appartenant au système de Pearson.	50
5.3	Graphe de Pearson.	51
5.4	Comparaison entre la paramétrisation classique et celle du système de Pearson.	52
Chapitre 6	Identifiabilité des mélanges finis de distributions du système de Pearson.	57
6.1	Un cas particulier.	57
6.2	Décomposition des mélanges finis généralisés.	58
6.3	Support des densités et identifiabilité.	59
6.4	Lien entre identifiabilité et négligeabilité.	61
6.5	Application à l'étude des mélanges généralisés.	64
6.6	Support et comportement asymptotique des lois du système de Pearson.	66
DEUXIÈME PARTIE.		
	ESTIMATION DES PARAMÈTRES DE DISTRIBUTIONS DE MÉLANGES PAR LA MÉTHODE DES MOMENTS.	69
Chapitre 7	Estimation des paramètres des distributions de mélanges finis.	73
7.1	Estimation par le maximum de vraisemblance.	74
7.2	La méthode des moments de Pearson.	75
7.2.1	Mélange de deux distributions normales de moyennes différentes.	75
7.2.2	Mélange de deux distributions normales de moyennes identiques.	76
7.2.3	Choix de la méthode.	77
Chapitre 8	Etude du système des cinq premières équations de moments centrés.	79

8.1	Existence et unicité des solutions de \mathfrak{S}_1	79
8.1.1	Existence.	79
8.1.2	Unicité.	79
8.1.3	Existence et unicité locale.	80
8.2	Inversion numérique du système \mathfrak{S}_1 par l'algorithme de Newton-Raphson.	81
8.3	Inversion algébrique du système.	82
8.3.1	Cas général où $x_1 \neq x_2$	83
8.3.2	Cas où les variances sont les mêmes.	89
8.3.3	Discussion dans le cas : $ x_1 - x_2 = (6\sqrt{6} - 9)^{1/4} \sqrt{ \sigma_1 - \sigma_2 }$	90
8.3.3.1	Restriction de la problématique.	90
8.3.3.2	Racine double.	91
8.3.3.3	Représentation graphique.	92
8.3.4	Conclusion.	93
8.4	Exemples d'inversion du système.	93
8.4.1	Cas général (éloigné des pôles du jacobien).	94
8.4.2	Cas où $\sigma_1 = \sigma_2$ (éloigné des pôles du jacobien).	94
8.4.2.1	Si $p \neq 0.5$	94
8.4.2.2	Si $p = 0.5$	94
8.4.3	Cas où les variables appartiennent à la variété V_2	95
8.4.4	Cas où x_1 et x_2 sont proches (voisinage de V_1).	95
8.4.4.1	Moyennes proches variances éloignées.	95
8.4.4.2	Moyennes et variances proches.	95
8.4.5	Conclusion.	96

Chapitre 9	Etude du système formé des quatre premières et de la sixième équations de moments centrés.	99
9.1	Existence et unicité des solutions.	99
9.2	Inversion algébrique du système.	103
9.2.1	Si $x_1 = x_2$.	103
9.2.2	Si $x_1 \neq x_2$.	103
9.3	Exemples d'inversion du système.	109
9.3.1	Cas général (éloigné des pôles du jacobien).	110
9.3.2	Cas où $\sigma_1 = \sigma_2$ (éloigné des pôles du jacobien).	110
9.3.2.1	Si $p \neq 0.5$.	110
9.3.2.2	Si $p = 0.5$.	110
9.3.3	Cas où les variables appartiennent à la variété V_2 .	110
9.3.4	Cas où x_1 et x_2 sont proches (voisinage de V_1).	110
9.3.4.1	Moyennes proches variance éloignées.	110
9.3.4.2	Moyennes et variances proches.	110
Chapitre 10	Le problème des moments dans le cas des moments empiriques.	113
10.1	Cas général (éloigné des pôles du jacobien).	113
10.2	Cas intermédiaire.	114
10.3	Cas où $\sigma_1 = \sigma_2$ (éloigné de V_2 ou V_4).	114
10.4	Cas où x_1 et x_2 sont proches (voisinage de V_1).	119
10.5	Cas où les variables appartiennent à un voisinage de la variété V_2 .	119
10.6	Cas où les variables appartiennent à un voisinage de la variété V_4 .	124
10.7	Conclusion.	124
Chapitre 11	Méthode d'estimation des paramètres d'une distribution de mélange de deux distributions normales inconnues.	127

11.1	Exposé de la méthode proposée.	127
11.1.1	Estimation des paramètres avec P_σ	127
11.1.2	Cas où les variances sont égales.	128
11.1.3	Cas où les moyennes sont les mêmes.	129
11.1.4	Estimation des paramètres avec P_{12}	129
11.1.5	Méthode générale d'estimation des paramètres d'un mélange de deux distributions normales.	130
Chapitre 12	Estimateurs des moments d'un mélange exponentielle-exponentielle.	133
12.1	Calcul du jacobien de \mathfrak{S}_3	133
12.2	Résolution algébrique du système \mathfrak{S}_3	134
12.3	Exemples d'inversion du système \mathfrak{S}_3	134
Chapitre 13	Estimateurs des moments d'un mélange normale-exponentielle.	137
13.1	Calcul du jacobien.	137
13.2	Inversion algébrique du système.	140
13.3	Méthode d'inversion du système \mathfrak{S}_4	152
13.4	Exemples d'inversions du système \mathfrak{S}_4	153
13.4.1	Exemples d'inversion théorique.	153
13.4.2	Exemples d'estimation des paramètres d'un échantillon.	153
13.4.3	Conclusion.	153
Chapitre 14	Estimation des paramètres d'un mélange généralisé simple.	157
14.1	Enoncé de la méthode.	157
14.2	Exemples d'applications.	158
14.2.1	Exemples dans le cas théorique.	158
14.2.2	Exemples dans le cas empirique.	159

TROISIÈME PARTIE.

SEGMENTATION STATISTIQUE D'IMAGES.	167
Chapitre 15 Classification bayésienne d'images - Modèles de Markov cachés.	169
15.1 Démarche bayésienne.	169
15.1.1 Modélisation du bruit - lois de Y conditionnelles à X	170
15.1.2 Aspect bayésien.	170
15.1.3 La classification bayésienne.	170
15.2 Méthodes de classification bayésienne.	171
15.2.1 Position du problème	171
15.2.2 Méthodes locales	171
15.2.3 Méthodes globales	172
15.3 Les champs de Markov cachés.	172
15.3.1 Définition.	172
15.3.2 Théorème de Hammersley-Clifford.	172
15.4 Utilisation des champs de Markov cachés en imagerie.	173
15.4.1 Loi de (X, Y)	173
15.4.2 Mode des marginales a posteriori (<i>MPM</i>)	174
15.4.2.1 L'échantillonneur de Gibbs.	174
15.4.2.2 L'algorithme de Maroquin et al.	175
15.4.3 Maximum a posteriori (<i>MAP</i>)	175
15.4.3.1 ICM.	175
15.4.3.2 Algorithme du recuit simulé.	176
15.5 Utilisation des chaînes de Markov cachées	177
15.5.1 Introduction	177

15.5.2	Loi de X	177
15.5.3	Loi de (X,Y) .	179
15.5.4	Méthode du <i>MPM</i> .	179
15.5.4.1	Calcul du <i>MPM</i>	179
15.5.4.2	Probabilités forward et backward - Algorithme de Baum-Welch.	179
15.5.5	Méthode du <i>MAP</i> .	180
15.5.5.1	Calcul du <i>MAP</i> .	180
15.5.5.2	Algorithme de Viterbi.	181
Chapitre 16	Modélisations sur arbres.	183
16.1	Graphes et probabilités.	183
16.1.1	Graphes de dépendance	183
16.1.2	Règle de retrait d'un sommet	184
16.2	Exemples de graphes de dépendances.	185
16.3	La structure d'arbre.	186
16.3.1	Définitions.	186
16.3.2	Exemples.	186
16.3.3	Interprétation de l'arbre	186
16.3.4	Passage d'un champs aléatoire à un arbre homogène	188
16.3.4.1	Principe de la modélisation	188
16.3.4.2	Des arbres particuliers.	188
16.3.4.3	Comparaison de cette modélisation avec les champs de Markov.	190
Chapitre 17	Application à l'imagerie.	193
17.1	Simulations de champs par les arbres.	193
17.1.1	Simulations par le quadarbre.	194

17.1.1.1	Le quadarbre à transitions constantes.	194
17.1.1.2	Le quadarbre évolutif.	194
17.1.2	Simulations par diarbre.	194
17.1.2.1	Simulations par le diarbre droit.	195
17.1.2.2	Simulations par le diarbre alterné.	195
17.1.2.3	Simulations par le diarbre diagonal.	195
17.2	Modèle markovien caché sur arbre.	196
17.3	Algorithmes de classification sur les arbres de Markov cachés.	197
17.3.1	Algorithme déterministe du MAP sur les arbres de Markov cachés.	197
17.3.2	Algorithme non-itératif du <i>MPM</i> sur les arbres de Markov cachés.	198
17.3.3	Algorithme simplifié sur les arbres de Markov cachés.	201
17.3.4	Segmentation supervisée sur les arbres.	202
17.3.4.1	Segmentations par quadarbres.	202
17.3.4.2	Segmentations par diarbres.	202
Chapitre 18	Estimation des paramètres dans les modèles cachés.	207
18.1	Cas des champs de Markov.	207
18.1.1	L'algorithme EM Gibbsien :	207
18.1.2	L'algorithme du gradient stochastique <i>GS</i>	209
18.1.3	L'algorithme Iterative Conditional Estimation (<i>ICE</i>).	209
18.2	Cas des chaînes de Markov.	211
18.2.1	L'algorithme Expectation Maximisation (EM classique).	211
18.2.2	Algorithmes <i>ICE</i> et <i>MICE</i> de Benmiloud-Pieczynski.	212
18.2.3	L'algorithme <i>SEM</i>	214
18.3	Cas des arbres de Markov cachés.	214

18.3.1	Algorithmes <i>EM</i> sur le diarbre.	215
18.3.2	Algorithmes <i>ICE</i> et <i>MICE</i>	216
18.3.3	Algorithme <i>SEM</i> sur le diarbre.	217
Chapitre 19	Segmentation non supervisée par arbres de Markov cachés.	219
19.1	Choix de l'algorithme d'estimation des paramètres.	219
19.2	Comparaison des performances des modèles sur arbres de Markov et champs de Markov.	227
19.2.1	Images simulées par les différents modèles.	227
19.2.2	Images de synthèse quelconques.	230
19.3	Limites du modèle par arbres.	232
19.4	Images réelles.	234
Chapitre 20	Le modèle des mélanges généralisés.	239
20.1	Présentation de la méthode.	239
20.2	Etalonnage de la méthode proposée.	240
20.3	Lecture de cartons d'orgue de barbarie.	246
20.3.1	Description d'un carton.	246
20.3.2	Protocole d'acquisition de l'image.	246
20.3.3	Segmentation de l'image du carton 1.	248
20.3.4	Segmentation de l'image du carton sale.	253
20.3.5	Une méthode de lecture.	256
Appendice A.:	Théorème fondamental de Robbins.	259
Appendice B.:	Identifiabilité de la classe de mesures produits.	261
Appendice C.:	Théorème d'identifiabilité de Teicher.	263
Appendice D.:	CNS d'identifiabilité de Yakowitz et Spragins.	265

Appendice E.: Les moments.	267
Appendice F.: Les moments empiriques.	271
Appendice G.: La méthode des moments.	273
Appendice H.: Les moments de mélanges.	277
Appendice I. : Le Théorème des fonctions implicites.	279
Appendice J. : La méthode de Newton de résolution des systèmes algébriques non linéaires.	281

INTRODUCTION.

Afin de mieux situer les domaines de l'étude et les circonstances ayant motivé nos travaux, précisons le contexte général dans lequel s'inscrivent nos recherches. Nous avons travaillé dans le cadre d'une collaboration entre le LSTA de l'Université Paris 6 et le département CITI de l'INT d'Evry, autour du thème de la segmentation statistique d'images. Rappelons qu'une image, en général de 256×256 pixels, est constituée de plusieurs classes, dont nous modélisons le bruitage par différentes distributions de probabilité. Le problème de la segmentation statistique d'images se décompose en deux phases complémentaires. La première consiste à estimer les paramètres des distributions modélisant les bruitages des classes ainsi que les paramètres caractérisant le modèle, la seconde consistant en une classification des pixels. Ainsi, la phase d'estimation consiste à estimer les paramètres d'un mélange de distributions bruitantes, sur de larges échantillons, de couramment 256×256 individus. Observons, cependant, qu'en segmentation d'images on cherche à privilégier l'estimation des distributions composantes du mélange, à l'estimation du mélange en tant que distribution, puisque les algorithmes de classification, qui permettent la reconstruction des images, utilisent les distributions composantes pour identifier les différentes classes. L'objectif d'une partie de notre travail a été d'analyser, à travers l'estimation de ses paramètres, la complexité de la distribution de mélange, trop souvent considérée comme une somme pondérée de distributions courantes. Le fait d'inscrire nos travaux dans le cadre général de la segmentation statistique d'images, nous permet de comparer l'approche probabiliste du mélange en tant que distribution, à laquelle sera consacrée la majorité de ce document, avec celle utilisée dans un domaine de pointe de la statistique appliquée. La segmentation statistique des images fait, en effet, partie des grands thèmes de recherche du GET (Telecom Paris, Telecom Bretagne et INT Evry). Un certain nombre de modèles et de traitements originaux ont été proposés depuis une dizaine d'années, desquels il ressort que les méthodes statistiques parfois très lourdes et souvent très complexes, sont aussi très puissantes (voir en particulier [Benmiloud95], [Giordana97], [Laferté00], [Monfrini99], [Peng95] et [Pieczynski99]). En particulier, l'estimation des paramètres, dans le cadre de l'imagerie, peut être effectuée de façon très fiable, grâce à des algorithmes de type *EM* (voir [Dempster77] et [Redner84]), *SEM* (voir [Broniatowski83], [Celeux86], [Celeux92] et [Celeux95]) ou *ICE* (voir [Pieczynski94] et [Delmas97]), adaptés à des modèles markoviens cachés (champs, chaînes ou arbres), modélisant les structures bidimensionnelles sous-jacentes de l'image. Outre l'introduction de nouvelles modélisations markoviennes permettant d'améliorer la qualité des estimations et de diminuer le temps des calculs, la modélisation du bruit, elle aussi, se voit étendue, avec l'introduction, en particulier, du bruit corrélé (voir [Pieczynski00]) ou, et c'est ce qui nous intéresse ici, des mélanges généralisés (voir [Bendjebour00], [Bendjebour01], [Delignon97] et [Fouque99]), qui permettent de considérer un élargissement du bruitage à des distributions non gaussiennes et non issues d'une même famille paramétrique.

Toutes ces avancées pratiques nécessitent de s'intéresser à leurs justifications théoriques. Ainsi, l'absence de preuve théorique de la convergence des algorithmes mis en oeuvre, ou d'étude de l'identifiabilité des mélanges généralisés utilisés, nous a amené à préciser le cadre théorique des mélanges de distributions, et à la présentation d'une méthode d'estimation théoriquement fondée.

Présentation du cadre probabiliste.

Pour des raisons autant historiques que pratiques, nous choisissons de circonscrire l'ensemble des distributions dont on étudiera la classe des mélanges, à la famille des distributions du système de Pearson. Dans sa série d'articles intitulée "Contribution to the Mathematical Theory of Evolution" ([Pearson1894] et [Pearson1895]), dont l'article de 1894 est considéré comme le point de départ de l'étude des mélanges, Karl Pearson définit au moyen d'une équation différentielle, un système regroupant les principales distributions de probabilité. Ce système propose une paramétrisation générale des distributions de cette famille dans laquelle on distinguera, entre autres, les distributions normales, gammas et betas.

Nous traitons, donc, de l'identifiabilité et de l'estimation des paramètres dans les mélanges finis de distributions de la famille de Pearson, dans le but de contribuer, en particulier, à la segmentation statistique d'images. Afin de mettre en évidence les trois axes principaux que nous avons dégagé dans ce sujet, nous séparons notre travail en trois parties. La première, dont une part non négligeable revisite la bibliographie, ancienne sur le sujet, est consacrée à une refondation de la problématique des mélanges et de l'identifiabilité, avec un élargissement de la théorie aux nécessités dictées par l'étude des mélanges généralisés. La seconde consiste, principalement à travers l'étude détaillée de l'exemple du mélange de deux distributions normales, à montrer les difficultés de mise en oeuvre de la méthode des moments pour l'estimation des paramètres des mélanges finis. Nous présentons, aussi, dans cette partie, une extension de la méthode des moments à l'estimation des paramètres de mélanges généralisés. La troisième partie, enfin, plus particulièrement consacrée à l'imagerie, met en évidence la grande variété des possibilités offertes par l'utilisation des algorithmes d'estimation de type *EM* ou *ICE*, dans leurs applications sous des hypothèses de mélanges généralisés.

Nous détaillons maintenant le contenu de ces trois parties.

PREMIÈRE PARTIE : Mélanges de lois de probabilité, identifiabilité.

Outre la nécessité de resituer la problématique, qui nous intéresse ici, dans un cadre probabiliste, la diversité des écritures et formalismes des publications, assez anciennes, sur le sujet nous a conduit à effectuer un important travail de synthèse et de reformulation, qui alimente les trois premiers chapitres de cette partie. Le premier chapitre rappelle le fondement probabiliste de la notion de mélange (voir [Robbins48]), et confirme que la distribution d'un échantillon de variables aléatoires iid issues d'un mélange, est une distribution de mélange. Le deuxième chapitre consacré à l'identifiabilité des mélanges continus, détaille les différentes publications établissant les cas d'identifiabilité et de non-identifiabilité de certaines classes de mélanges. Nous concluons ce chapitre par une définition non paramétrique des mélanges et de l'identifiabilité, qui se rapproche du cadre des mélanges généralisés, mais dont la mise en oeuvre est très complexe. Les résultats d'identifiabilité énoncés pour les mélanges continus peuvent, évidemment, être réinvestis dans le cas particulier des mélanges finis, développé dans le chapitre suivant. Les classes de mélanges finis étant nettement moins complexes que les précédentes, les résultats d'identifiabilité sont plus nombreux. Nous signalerons ici les deux conditions nécessaires et suffisantes d'identifiabilité de Teicher (voir [Teicher63]) et de Yakowitz et Spragins (voir [Yakowitz68]), qui permettent de conclure, entre autres, à l'identifiabilité de la classe des mélanges finis de distributions normales ainsi qu'à celle des distributions gamma de support \mathbb{R}^{+*} . Notons, que dans le cadre de l'identifiabilité des mélanges finis nous montrons l'identifiabilité des classes de tous les mélanges finis des distributions gamma à support fixé ainsi que son équivalent pour les distributions inverse-gamma à support fixé.

Les trois chapitres suivants sont consacrés à la définition des mélanges généralisés, ainsi qu'à l'identifiabilité de la classe de tous les mélanges finis de la famille des distributions de Pearson. Nous établissons, dans les chapitres 3, 5 et 6, une succession de résultats permettant de montrer finalement, que sous certaines restrictions, la classe des mélanges finis des distributions du système de Pearson est identifiable. Par ailleurs, les diverses propositions établies dans ce chapitre n'étant pas liées au seul système de distributions de Pearson, elles seront présentées de façon à pouvoir être réinvesties en dehors de cette famille. Le résultat final d'identifiabilité, ne peut malheureusement pas être considéré comme complet, bien que nous sachions la nécessité de restreindre la famille de Pearson pour espérer l'identifiabilité de la classe de ses mélanges finis. Il nous permet cependant d'envisager légitimement l'estimation des paramètres des mélanges qui nous intéresseront par la suite.

DEUXIÈME PARTIE : **Estimation des paramètres de distributions de mélanges par la méthode des moments.**

Nous avons souhaité développer, dans cette partie du travail, une méthode d'estimation des paramètres d'un mélange fini (voir [Everitt81], [Lindsay95], [McLachlan88] et [Titterington85]) en tant que distribution et non une méthode, comme *EM* ou *ICE*, donnant les paramètres du mélange, par l'estimation des paramètres des distributions composantes. La vraisemblance n'est, le plus souvent, pas bornée pour les distributions de mélange. Sous certaines conditions de régularité, Redner et Walker (voir [Redner84]) montrent que l'on peut trouver des solutions aux équations de la vraisemblance, ce qui autorise le développement d'algorithmes itératifs d'estimation du maximum de vraisemblance, paramétriques ou non paramétriques (voir [Dempster77] et [Laird78]). La convergence de ces algorithmes est assez lente, et très fortement liée à l'initialisation. Il est, de plus, impossible d'assurer la non convergence éventuelle vers un extremum local (voir [Fiorin01]). Compte tenu de la grande taille, notée ici n , des échantillons auxquels nous nous référons, le choix de l'estimation des paramètres par une méthode des moments, dont on connaît la convergence en \sqrt{n} , semble indiqué. Cependant cette méthode considérée, jusqu'ici, comme inférieure à la précédente, en raison, principalement, de manifestations d'instabilité inexplicables constatées lors de son utilisation (voir [Fryer72], [McLachlan88] et [Tan72]). C'est aux causes de cette instabilité que nous nous intéressons ici, grâce à une étude approfondie de la question de l'unicité des solutions aux divers systèmes de moments, qui fait apparaître l'intérêt d'un ré-examen de la méthode des moments.

La principale étape de la méthode des moments consiste à inverser un système algébrique non linéaire, avec des contraintes associées aux domaines de définition des paramètres. D'une façon générale, la littérature sur le sujet délaisse les conditions d'existence et surtout d'unicité des solutions de tels systèmes. Dans son article de 1894, Pearson s'intéresse au mélange de deux distributions normales dont il ramène, par une méthode de substitution, l'inversion à la recherche des racines négatives d'un polynôme de degré 9 (voir [Pearson1894] et [Cohen67]). Pourtant, les questions de l'existence et de l'unicité des solutions sont à peine effleurées (voir, par exemple [Bowman73]). Nous avons donc choisi de reprendre intégralement la méthode des moments sur ce cas particulier, notamment en étendant le principe initialement proposé par Pearson, d'utiliser la sixième équation de moments comme critère de décision dans les cas de non unicité.

Ainsi, les quatre premiers chapitres de cette partie sont consacrés à la mise au point d'une méthode fiable d'estimation des paramètres du mélange de deux distributions normales, basée sur la méthode générale des moments. Les chapitres suivants sont consacrés à la généralisation de la méthode des moments pour l'estimation des paramètres d'un mélange fini de deux distributions plus générales.

Précisons le cas du mélange de deux distributions normales $N(\mu_1, \sigma_1)$ et $N(\mu_2, \sigma_2)$. C'est une distribution paramétrique à cinq paramètres, puisqu'aux deux fois deux paramètres des distributions composantes, il convient d'ajouter le paramètre, noté p , de mélangeance. L'estimation de ces paramètres nécessite, par conséquent, l'aide d'au moins cinq équations. La qualité des estimateurs des moments diminuant avec l'augmentation de l'ordre de ceux-ci, nous choisissons, comme le fit Pearson, d'estimer ces paramètres grâce aux cinq premières équations de moments de la distribution de mélange. En prenant la précaution de calculer le jacobien, noté J_1 , de ce système, nous montrons l'existence de deux variétés sur lesquelles ce jacobien s'annule et qui correspondent à une non inversibilité du système des moments. Si le cas où les deux moyennes sont égales, signalé par Pearson, se retrouve effectivement dans notre variété V , nous mettons en évidence le cas nouveau de la sous-variété V_2 de $\mathbb{R}^2 \times \mathbb{R}^{+2} \times [0.5; 1[$ définie par :

$$V_2 = \left\{ (\mu_1, \mu_2, \sigma_1, \sigma_2, p), |\mu_1 - \mu_2| = \left(6\sqrt{6} - 9\right)^{1/4} \sqrt{|\sigma_1 - \sigma_2|} \right\}.$$

Remarquons que dans le cas où les moyennes sont égales, le système de Pearson se réduit à trois équations pour quatre inconnues. Par contre, la non-unicité des solutions sur V_2 ne se manifeste qu'au moment de rechercher les racines négatives du polynôme de Pearson. Nous montrons, en effet, que ce cas d'annulation se manifeste par la présence d'une racine négative double pour le polynôme de Pearson, qui permet d'inverser le système. Ce cas n'est mentionné ni par Pearson, ni après.

Lorsque nous nous plaçons dans le cas théorique où les moments sont effectivement les moments du mélange de deux distributions normales, l'existence d'une solution, que nous savons unique par le résultat d'identifiabilité des mélanges finis de distributions normales, nous assure, quitte à faire entrer en jeu suffisamment d'équations, de l'unicité de la solution au problème des moments. Le cas où les paramètres appartiennent à V_2 ne pose, de ce fait, pas de réel problème dans l'étude théorique de l'inversibilité, puisqu'il suffit de tester l'adéquation des candidats solution sur un nombre suffisant d'équations de moments. Nous pouvons cependant supposer, et cela se confirme lorsque l'on remplace les moments théoriques par les moments empiriques, que l'estimation de paramètres appartenant à un voisinage de V_2 sera source d'un double problème. D'une part l'absence de racine permettant l'inversion statistique du système, d'autre part, l'obtention de deux racines donnant deux ensembles de paramètres solutions plus ou moins proches, que l'incertitude sur les moments empiriques ne permettra pas de supprimer de façon fiable.

Une idée naturelle, mentionnée par Pearson, est de se servir de la sixième équation de moment pour construire un test permettant d'éliminer les solutions parasites susceptibles d'apparaître lors de l'inversion du système des moments empiriques. Nous proposons d'aller plus loin, en considérant deux systèmes des moments, le premier étant celui de Pearson et le second étant obtenu en remplaçant dans le système de Pearson, la cinquième équation de moment du mélange, par la sixième. Nous procédons alors, avec le second système, comme avec le premier, en calculant son jacobien, noté J_2 . Nous établissons, dans un premier temps que ce jacobien s'annule sur la sous-variété V_4 de $\mathbb{R}^2 \times \mathbb{R}^{+2} \times [0.5; 1[$ définie par:

$$V_4 = \left\{ \begin{array}{l} (\mu_1, \mu_2, \sigma_1, \sigma_2, p), p = \left(\frac{1}{2} + \frac{3(\sigma_1 - \sigma_2)(15(\sigma_1 - \sigma_2)^4 - 10(\mu_1 - \mu_2)^4(\sigma_1 - \sigma_2)^2 - (\mu_1 - \mu_2)^8)}{2(\mu_1 - \mu_2)^2(135(\sigma_1 - \sigma_2)^4 - 18(\mu_1 - \mu_2)^4(\sigma_1 - \sigma_2)^2 - (\mu_1 - \mu_2)^8)} \right) \\ \text{et } (\mu_1 - \mu_2)^2 \left(135(\sigma_1 - \sigma_2)^4 - 18(\mu_1 - \mu_2)^4(\sigma_1 - \sigma_2)^2 - (\mu_1 - \mu_2)^8 \right) \neq 0 \end{array} \right\},$$

qui couvre, entre autres, le cas symétrique de la variété V_3 où les variances sont égales avec un coefficient de mélangeance de 0.5. Nous montrons aussi, et c'est ce qui nous permettra, par la suite, de définir une nouvelle méthode d'estimation, que ce jacobien ne s'annule pas en même temps que J_1 . Ce résultat nous assure de la complémentarité entre les deux systèmes des moments. Une utilisation combinée, que nous discutons dans le chapitre 10, de ces deux systèmes, tenant compte de l'importance de privilégier les moments empiriques de bas degré, permettra une estimation fiable et stable des paramètres par la méthode des moments. Notons que l'inversion algébrique du second système des moments permet de réinvestir pour une large partie les calculs effectués dans le cadre de l'inversion du premier système, et conduit à rechercher les racines positives d'un polynôme de degré 12. En effet, la méthode que nous avons adoptée pour l'inversion du premier système des moments, permet d'obtenir aisément l'inversion du second, puisqu'il ne reste à effectuer que la dernière étape de substitution des résultats dans cette la sixième équation de moment. Grâce à l'étude des deux jacobiens, nous savons, aussi, que nous pouvons limiter la méthode à l'inversion de deux systèmes.

La mise au point d'un protocole d'estimation des paramètres basé sur ces deux systèmes, passe alors, aussi, par une évaluation de "l'épaisseur" de la zone d'incertitude autour des variétés sur lesquelles les jacobiens sont nuls. Nous avons choisi d'estimer heuristiquement cette épaisseur en procédant à des tests d'inversion par les deux systèmes, de la façon la plus complète possible, au voisinage des différentes variétés susceptibles de provoquer des dégradations, voir des impossibilités d'estimation. L'ensemble des résultats de ces estimations, nécessaires pour comprendre le mécanisme de la méthode d'estimation que nous proposons, sont présentés en annexe, et constituent, par ailleurs, une base de données susceptibles d'inspirer au lecteur une approche plus théorique de la détermination des "épaisseurs" à risque. Nous concluons ainsi, cette partie du travail, en proposant, dans le chapitre 12, une méthode pratique originale d'estimation des paramètres du mélange de deux distributions normales, basée sur la méthode des moments.

Les deux chapitres suivants permettent d'étendre la méthode des moments à d'autres distributions du système de Pearson. Le premier, qui est consacré au mélange de deux distributions exponentielles (voir [Rider61] et [Rider62]) nous permet d'aboutir directement à des expressions littérales des paramètres en fonction des moments non centrés. Le jacobien, J_3 , ne s'annulant que dans le cas du mélange de deux distributions identiques, il n'y a pas de discussion dans ce cas. Le second, approfondissant le cas du mélange d'une distribution normale et d'une distribution exponentielle, $pN(\mu_1, \sigma_1) + (1-p)E(m_2)$, illustre les difficultés supplémentaires que l'on peut

rencontrer lorsque le système ne peut être simplifié en utilisant la symétrie des équations. Il nous a été impossible de donner une factorisation générale du jacobien J_4 :

$$J_4 = 9p^2(1-p) \left(\begin{array}{c} X^3 + \frac{-2}{3} (\mu_1^4 - 8 \mu_1^3 m_2 + 24 \mu_1^2 m_2^2 - 40 \mu_1 m_2^3 + 26 m_2^4) X \\ + \frac{8}{27} (\mu_1^6 - 12 \mu_1^5 m_2 + 60 \mu_1^4 m_2^2 - 160 \mu_1^3 m_2^3 + 222 \mu_1^2 m_2^4 - 120 \mu_1 m_2^5 + 16 m_2^6) \end{array} \right),$$

avec :

$$X = \sigma_1 - \frac{(\mu_1^2 - 4\mu_1 m_2 - 2m_2^2)}{3}.$$

Nous parvenons, cependant au prix de manipulations algébriques complexes et de calculs inenvisageables sans le soutien d'un logiciel de calcul formel, à donner une méthode rigoureuse d'inversion, basée principalement sur la recherche des racines appartenant à $]0; 1[$ d'un polynôme de degré 12 en p . L'absence de factorisation du jacobien, ne nous permet pas de présenter une étude aussi complète que pour le mélange de deux distributions normales, et notamment de cerner les variétés à risque. Nous aboutissons ainsi, à une méthode des moments pour un mélange normale-exponentielle, à manipuler avec la même précaution que la méthode initiale de Pearson.

Grâce à l'ensemble des méthodes vues précédemment, nous concluons cette partie du travail par un chapitre présentant une méthode originale d'estimation des paramètres (ainsi que des distributions composantes) d'un mélange généralisé de deux distributions issues de la famille réunissant les distributions normales et exponentielles (sans imposer a priori le type des distributions). Cette étude fait apparaître les nouvelles difficultés liées au choix du type des distributions composant le mélange. Compte tenu du mode de décision choisi pour cette étape de l'estimation, basée sur les équations de moments d'ordre 5 ou 6, nous mettons en évidence l'importance de la taille des échantillons considérés pour le choix du type de mélange étudié.

TROISIÈME PARTIE : **Segmentation statistique d'images.**

Cette partie du travail se distingue fondamentalement des deux premières, mais n'en est pas moins leur complément naturel. Nous avons, jusqu'ici, traité le sujet sur un plan théorique, ce qui a montré la nécessité de circonscrire le sujet au maximum, pour pouvoir l'étudier dans des conditions raisonnables. L'esprit dans lequel il faut aborder ce qui va suivre est radicalement différent. Notre ambition, ici, est de construire des procédures de segmentation d'images, les plus précises et les plus rapides possible, que nous validerons par le biais de l'expérimentation informatique. Ces procédures nous permettent, au prix de concessions sur la connaissance théorique de leur comportement, de dépasser largement le cadre des mélanges que nous nous sommes fixés dans la DEUXIÈME PARTIE. Nous estimons, couramment, les paramètres de mélanges de plus de deux composantes, lesquelles pourront être cherchées dans une palette de familles de distributions bruitantes quasi illimitée. A ce propos, le Théorème d'identifiabilité, que nous avons proposé dans la PREMIÈRE PARTIE, permet de justifier, sous réserve des restrictions que nous y avons signalées, du choix de distributions bruitantes appartenant à la famille de Pearson.

Bien que la base du travail, en particulier, la détermination du modèle et le développement des algorithmes, demande un fondement statistique théorique très complexe, la partie de mise en oeuvre informatique offre une large place à l'expression de l'ingéniosité du programmeur. L'ensemble des choix de programmation, et les diverses astuces informatiques, sont, en effet, déterminants pour l'optimisation de la rapidité d'un algorithme défini théoriquement, et les "plus" ajoutés au cours du développement, permettent d'améliorer empiriquement la qualité des résultats. Nous chercherons, donc, dans cette troisième partie du travail, à faire valoir l'efficacité du modèle et des algorithmes que nous proposons, et à montrer à travers les exemples que nous y traitons, la distance existant entre les cas étudiables théoriquement et ceux traités en Statistique Appliquée.

Nous commençons par présenter les modèles des champs de Markov cachés et des chaînes de Markov cachées utilisés pour la segmentation d'image, ainsi que les méthodes de classification bayésienne **Mode des Marginales a posteriori** (*MPM*) et **Maximum A Posteriori** (*MAP*) correspondantes (voir [Besag74], [Geman84] et [Marroquin87]). Après quelques rappels de théorie des graphes, nous donnons les équivalents des algorithmes

précédents pour les arbres de Markov cachés. La structure des arbres de Markov étant proche de celle des chaînes de Markov, les algorithmes du *MAP* et du *MPM* y sont entièrement déterministes et donc très rapides, ce qui justifie qu'on s'y intéresse (voir [Baum70] et [Chardin00]). Le chapitre suivant, dans lequel nous rappelons les différents algorithmes d'estimation des paramètres de type *EM* et *ICE*, associés à ces différents modèles, permet d'achever de nous familiariser avec les modèles markoviens.

L'étalonnage du modèle de diarbre que nous présentons, ensuite, se déroule alors en trois étapes. Dans un premier temps, il s'agit d'associer au modèle, l'algorithme d'estimation des paramètres permettant d'optimiser les résultats de la classification par le *MPM* sur le diarbre. La comparaison, dans un second temps, sur des exemples, des résultats de segmentation non supervisée, par le diarbre, avec ceux issus d'un modèle de champs de Markov simple, nous permet de mettre en évidence la compétitivité du modèle que nous proposons, par rapport à un modèle ayant fait ses preuves. Enfin, une étude des limites de notre modèle, dans le cas de distributions bruitantes très proches, nous assure de la stabilité du modèle du diarbre de Markov caché, en matière de segmentation d'images.

Nous introduisons le chapitre suivant en proposant une nouvelle méthode d'estimation des paramètres, sur le diarbre, dans le cadre des mélanges généralisés. Bien que la méthode soit valable pour une famille plus grande, nous ne considérons que la famille des distributions normales et exponentielles, déjà étudiée dans la DEUXIEME PARTIE. La comparaison, sur des images, des performances de la méthode des moments pour les mélanges généralisés, avec les algorithmes d'estimation que nous avons proposés, montre, compte tenu de la taille (128×128) des images de synthèse considérées, la plus grande efficacité des méthodes *SEM* généralisées pour cette taille d'échantillon.

Nous détaillons, enfin, une application originale de la méthode des mélanges généralisés, à la segmentation d'images réelles pour la mise au point d'un logiciel de lecture d'un carton d'orgue de barbarie. Cette application illustre, outre l'intérêt d'avoir étendu la méthode gaussienne classique aux mélanges généralisés, la rapidité des algorithmes sur les arbres de Markov cachés. En effet, l'algorithme mis en oeuvre autorise une conversion, en temps réel de l'image lue par la caméra, en un fichier utilisable par un ordinateur, que ce soit pour reproduire le carton, ou générer un fichier musical de type midi.

PREMIÈRE PARTIE.

**MÉLANGES DE LOIS DE
PROBABILITÉ, IDENTIFIABILITÉ.**

- CHAPITRE 1 -

Introduction aux mélanges.

1.1 Définition.

Soit deux entiers non nuls m et n . Considérons $F = \{F(x, \theta), x \in \mathbb{R}^n, \theta \in \Omega^m\}$ une famille paramétrique de fonctions de répartition sur \mathbb{R}^n , paramétrée par $\theta = (\theta_1, \dots, \theta_m) \in \Omega^m$, $\Omega^m \subset \mathbb{R}^m$, telle que $F(x, \theta)$ est mesurable sur $\mathbb{R}^n \times \Omega^m$. Considérons aussi \mathcal{G} , une classe de fonctions de répartition m -dimensionnelles. Un élément de \mathcal{G} est noté G et la mesure associée à G est notée μ_G . On définit, alors, la classe \mathcal{H} , des fonctions de répartition H sur \mathbb{R}^n , par la correspondance :

$$\begin{aligned} \mathcal{M} &: \mathcal{G} \rightarrow \mathcal{H} \\ G &\mapsto H, \text{ telle que } H(x) = \int_{\Omega^m} F(x, \theta) dG(\theta), x \in \mathbb{R}^n. \end{aligned} \quad [1.1]$$

L'élément de \mathcal{H} associé à G est appelé G -mélange de F ou plus simplement mélange.

1.2 Mélanges finis.

Si μ_G est une probabilité discrète qui ne charge, avec des poids positifs, qu'un nombre fini N de points, notés $\theta^1, \dots, \theta^N$, de Ω^m , les éléments de \mathcal{H} sont des mélanges dits finis. Dans ce cas l'écriture des mélanges se simplifie considérablement et prend la forme :

$$H(x) = \sum_{i=1}^N c_i F(x, \theta^i), \text{ avec } \sum_{i=1}^N c_i = 1 \text{ et } \forall i, c_i > 0.$$

1.3 Définition probabiliste des mélanges.

Nous reprenons, dans ce paragraphe, les développements de H. Robbins dans [Robbins48]. Il s'agit d'établir l'existence des mélanges continus des distributions d'une famille de lois. Le cas des mélanges finis est un cas particulier qui en découle.

1.3.1 Rappel sur les fonctions de répartition.

Soit $x = (x_1, \dots, x_n)$ un vecteur de l'espace euclidien \mathbb{R}^n et soit \mathcal{B}_n la tribu engendrée par les Boréliens de \mathbb{R}^n . On notera S_x le pavé semi-ouvert de \mathbb{R}^n défini par :

$$S_x = \{(w_1, \dots, w_n) \in \mathbb{R}^n; w_1 \leq x_1, \dots, w_n \leq x_n\}.$$

On rappelle que, si on se donne la probabilité μ sur \mathcal{B}_n , la fonction de répartition F associée à μ est définie par:

$$F(x) = \mu(S_x), \text{ pour tout } x \text{ de } \mathbb{R}^n. \quad [1.2]$$

μ est alors l'unique probabilité de \mathcal{B}_n vérifiant [1.2].

1.3.2 Théorème fondamental.

Soit, maintenant G une fonction de répartition une fonction de répartition définie sur $\Omega^m \subset \mathbb{R}^n$. Soit aussi une fonction F définie sur $\Omega^m \times \mathbb{R}^n$ vérifiant pour μ_G -presque tout y , $F(\cdot, y)$ est une fonction de répartition, et pour tout x de \mathbb{R}^n , $F(x, \cdot)$ est une fonction borélienne. La probabilité sur \mathcal{B}_n associée à la fonction de répartition $F(\cdot, y)$, sera notée μ_y .

THÉORÈME 1.1: *La fonction*

$$H(x) = \int_{\Omega^m} F(x, y) dG(y) \quad [1.3]$$

est une fonction de répartition sur \mathbb{R}^n . De plus, si μ est la probabilité sur \mathcal{B}_n associée à H , alors pour tout $S \in \mathcal{B}_n$, $\mu_y(S)$ est une fonction borélienne de y et

$$\mu(S) = \int_{\Omega^m} \mu_y(S) dG(y). \quad [1.4]$$

Démonstration : voir Appendice A.♦

Ceci établit l'existence des mélanges en tant qu'objet mathématique et à titre d'exemple on peut montrer que la convolution sur \mathbb{R} est un mélange. En effet, G et F étant des fonctions de répartition sur \mathbb{R} , $F(x - y)$ est une fonction de $(x, y) \in \mathbb{R}^2$ vérifiant les hypothèses du théorème 1.1. Par conséquent :

$$H(x) = \int_{\mathbf{R}} F(x - y) dG(y)$$

est la fonction de répartition d'un mélange sur \mathbb{R} que l'on note habituellement $H(x) = F(x) \otimes G(x)$, où \otimes est l'opérateur de convolution.

1.4 Distribution d'un échantillon.

Si l'on se place maintenant sur \mathbb{R} , que l'on y considère le mélange

$$H(u) = \int_{\mathbb{R}} F(u, v) dG(v) \text{ pour } u \in \mathbb{R} \quad [1.5]$$

et que l'on définit, pour $x = (x_1, \dots, x_n)$,

$$\begin{aligned} \overline{H}(x) &= H(x_1) \cdots H(x_n) \\ \text{et } \overline{G}(x) &= G(x_1) \cdots G(x_n), \end{aligned}$$

alors \overline{H} et \overline{G} sont des fonctions de répartition sur \mathbb{R}^n . En particulier $\overline{H}(x)$ est la fonction de répartition d'un échantillon de n variables indépendantes ayant chacune pour fonction de répartition H donnée en [1.5].

Si l'on pose

$$\overline{F}(x, y) = F(x_1, y_1) \cdots F(x_n, y_n) \text{ pour } x = (x_1, \dots, x_n) \in \mathbb{R}^n \text{ et } y = (y_1, \dots, y_n) \in \mathbb{R}^n,$$

alors, pour $\mu_{\overline{G}}$ -presque tout y , $\overline{F}(\cdot, y)$ est une fonction de répartition et $F(x, \cdot)$ est une fonction borélienne, pour tout $x \in \mathbb{R}^n$. Par le théorème de Fubini nous avons alors

$$\begin{aligned} \overline{H}(x) &= \int_{\mathbb{R}} F(x_1, y_1) dG(y_1) \cdots \int_{\mathbb{R}} F(x_n, y_n) dG(y_n), \\ \overline{H}(x) &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} F(x_1, y_1) \cdots F(x_n, y_n) dG(y_1) \cdots dG(y_n), \\ \overline{H}(x) &= \int_{\mathbb{R}^n} \overline{F}(x, y) d\overline{G}(y), \end{aligned}$$

ce qui montre que $\overline{H}(x)$ est une fonction de répartition de mélange.

- CHAPITRE 2 -

Identifiabilité d'une classe de mélanges.

L'identifiabilité de la classe de tous les mélanges d'une famille donnée de distributions, assure l'unicité de la décomposition d'un mélange sur cette classe de distributions. La question de l'identifiabilité d'une classe de mélanges de distributions est un préalable nécessaire à toute étude statistique faisant intervenir cette classe. Il est, en effet, primordial, dans les problèmes statistiques afférents aux mélanges de distributions, qu'un échantillon suivant une telle loi ne puisse pas être décomposé de plusieurs façons différentes.

La complexité des résultats présentés dans cette partie montre la difficulté d'obtenir des règles générales sur l'identifiabilité des classes de mélanges.

2.1 Définition.

Reprenons les notations du chapitre 1 : $F = \{F(x, \theta), x \in \mathbb{R}^n, \theta \in \Omega^m\}$ étant une famille paramétrique de fonctions de répartition sur \mathbb{R}^n , paramétrées par $\theta \in \Omega^m$, \mathcal{G} étant une classe de fonctions de répartition m -dimensionnelles, la classe \mathcal{H} des mélanges est définie par la correspondance :

$$\begin{aligned} \mathcal{M} &: \mathcal{G} \rightarrow \mathcal{H} \\ G &\mapsto H, \text{ telle que } H(x) = \int_{\Omega^m} F(x, \theta) dG(\theta) \text{ pour } x \in \mathbb{R}^n. \end{aligned} \quad [2.6]$$

on dira que \mathcal{H} est identifiable si \mathcal{M} est une correspondance bijective de \mathcal{G} sur \mathcal{H} .

De même toute sous-classe \mathcal{H}_0 de \mathcal{H} est dite identifiable si \mathcal{M} réalise une bijection de $\mathcal{G}_0 = \mathcal{M}^{-1}(\mathcal{H}_0)$ sur \mathcal{H}_0 .

REMARQUE 2.1: Une conséquence immédiate de ces définitions est que si $\mathcal{H}_0 \subset \mathcal{H}_1 \subset \mathcal{H}$, l'identifiabilité de \mathcal{H}_1 entraîne clairement l'identifiabilité de \mathcal{H}_0 .

Les paragraphes suivants présentent des résultats sur l'identifiabilité de certaines classes de mélanges.

Identifiabilité d'une classe de mélanges.

2.2 Identifiabilité des mélanges de distributions à paramètres additifs et multiplicatifs.

2.2.1 Familles additivement fermées.

En 1960, Henry Teicher entame une série d'articles sur l'identifiabilité des mélanges. Le premier article, [Teicher60], est principalement consacré à l'étude de la classe des mélanges finis de distributions normales. Nous détaillerons l'identifiabilité des classes de mélanges finis dans le chapitre suivant. Dans le second [Teicher61], qui nous intéresse ici, il montre l'identifiabilité des classes de mélanges des familles additivement fermées de distributions à un paramètre. Rappelons qu'une famille de fonctions de répartition $F_D = \{F(\cdot, \theta), \theta \in D\}$, est dite additivement fermée si on a :

$$F(\cdot, \theta_1) \otimes F(\cdot, \theta_2) = F(\cdot, \theta_1 + \theta_2), \forall \theta_1, \theta_2 \in D,$$

où D est un semi-groupe abélien additif.

2.2.2 Théorème d'identifiabilité.

On sait, depuis [Teicher54], que les distributions d'une famille paramétrique en dimension 1, additivement fermée sur $D = \mathbb{N}$, $D = \mathbb{Q}^+$ ou $D = \mathbb{R}^+$, possèdent une fonction caractéristique, $\Psi_F(\cdot, \theta)$, de la forme :

$$\Psi_F(\cdot, \theta) = (\Psi_F(\cdot, 1))^\theta, \quad t \in \mathbb{R},$$

où $\Psi_F(\cdot, 1)$ est une fonction caractéristique indépendante de θ . Dans ce cas, tout élément de \mathcal{H}_D a une fonction caractéristique $\Psi_H(t)$, de la forme :

$$\Psi_H(t) = \int_D (\Psi_F(t, 1))^\theta dG(\theta), \quad t \in \mathbb{R}.$$

D'après l'étude de l'intégrale $I(z, G) = \int_D z^\theta dG(\theta)$ sur la boule ouverte de rayon 1, privée de l'origine, et grâce au théorème d'identité des transformées de Fourier, on établit que, si deux fonctions de répartition G_1 et G_2 engendrent le même mélange de \mathcal{H}_D , elles sont nécessairement égales.

C'est le théorème :

THÉORÈME 2.1: *La classe \mathcal{H}_D des mélanges d'une famille paramétrique sur \mathbb{R} additivement fermée lorsque D est l'un des trois \mathbb{N} , \mathbb{Q}^+ ou \mathbb{R}^+ , est identifiable.*

Ce théorème trouve l'une de ses applications dans le cas où une fonction de répartition F engendre la famille F_D , par exemple pour $F_1 = \{F_1(\cdot, \theta) = F(\cdot, \theta), \theta \in D\}$ ou $F_2 = \{F_2(\cdot, \theta) = F(\cdot - \theta), \theta \in D\}$. On peut ainsi établir la proposition 2.1 :

PROPOSITION 2.1: *Si la transformée de Fourier de la fonction de répartition $\overline{F}(\cdot) = F(e^\cdot)$ n'est pas identiquement nulle sur un intervalle non dégénéré de \mathbb{R} , les classes \mathcal{H}_1 , respectivement \mathcal{H}_2 , des mélanges de distributions des F_1 , respectivement F_2 , sont identifiables.*

Identifiabilité d'une classe de mélanges.

2.2.3 Exemples d'applications.

Le théorème 2.1 et la proposition 2.1 s'appliquent au cas de la distribution gamma à deux paramètres. Soit, en effet, la famille de distributions :

$$F_{\Gamma} = \left\{ \Gamma(\cdot, p, q) = \int_0^x \frac{1}{\Gamma(q)} \frac{u^{q-1}}{p^q} \exp\left(-\frac{u}{p}\right) du, p \in \mathbb{R}^{*+}, q \in \mathbb{R}^{*+} \right\}.$$

Notons pour $p \in \mathbb{R}^{*+}$ et $q \in \mathbb{R}^{*+}$:

$$\Phi(t, p, q) = (1 - ipt)^{-q},$$

la fonction caractéristique de la loi Gamma de fonction de répartition $\Gamma(\cdot, p, q)$. A p fixé, F_{Γ} est une famille additivement fermée et le théorème assure donc l'identifiabilité des $G(q)$ -mélanges de F_{Γ} . Lorsque q est fixé, F_{Γ} est engendrée par la fonction $\Gamma(\cdot, 1, q)$, de plus, la fonction caractéristique de \overline{F} est $\frac{\Gamma(q+it)}{\Gamma(q)}$, ce qui permet, grâce à la proposition 2.1, de conclure à l'identifiabilité des $G(p)$ -mélanges de F_{Γ} .

Remarquons que la non identifiabilité, en général, de la classe \mathcal{H} des $G(p, q)$ -mélanges qui a été montrée dans [Robbins49], n'empêche pas l'identifiabilité à p ou à q fixé.

On montre, de la même façon, que la classe des $G(\sigma)$ -mélanges, à θ fixé, et celle des $G(\theta)$ -mélanges, à σ fixé, de la famille des distributions uniformes :

$$F_U = \left\{ U(\cdot, \theta, \sigma) = \int_{-\infty}^{\cdot} \frac{1}{2\sigma} \mathbf{1}_{\{\theta-\sigma, \theta+\sigma\}}(u) du, \sigma > 0, \theta \in \mathbb{R} \right\},$$

sont identifiables. Il n'en n'est pas de même, bien sûr, pour la classe \mathcal{H} de tous les mélanges de F_U , une distribution uniforme étant elle-même un mélange de distributions uniformes :

$$U\left(\cdot, \frac{1}{2}, \frac{1}{2}\right) = \frac{1}{2}U\left(\cdot, \frac{1}{4}, \frac{1}{4}\right) + \frac{1}{2}U\left(\cdot, \frac{3}{4}, \frac{1}{4}\right).$$

Pour ce qui est de la classe des distributions binomiales :

$$F_B = \left\{ B(\cdot, n, p) = \sum_{j < \cdot} \binom{n}{j} p^j (1-p)^{n-j} \right\},$$

on obtient, par un raisonnement analogue, l'identifiabilité, à p fixé, de la classe des $G(n)$ -mélanges de F_B . Par contre, à n fixé, la classe des $G(p)$ -mélanges de F_B n'est pas identifiable.

2.3 Identifiabilité des classes de mélanges de la famille exponentielle.

En 1965, Barndorff-Nielsen tente d'aborder le problème sous un angle plus général, en discutant l'identifiabilité de la classe de mélanges de la famille exponentielle, (voir [Barndorff65]).

Identifiabilité d'une classe de mélanges.

Soit $F = \{F(x, \theta), x \in \mathbb{R}^n, \theta \in \Omega^m\}$, où, pour une mesure σ -finie μ sur \mathcal{B}_n on a :

$$dF(x, \theta) = a(\theta) b(x) \exp\left(\sum_{j=1}^m \theta_j h_j(x)\right) d\mu(x),$$

avec $a(\theta) > 0$, $b(x) \geq 0$ et $a, b, h_i, 1 \leq i \leq m$, mesurables.

Si $G_1, G_2 \in \mathcal{G}$, on pose :

$$f_k(x) = \frac{dH}{d\mu}(x) = b(x) \int_{\Omega^m} a(\theta) \exp\left(\sum_{j=1}^m \theta_j h_j(x)\right) dG_k(\theta), \quad k = 1, 2.$$

Soit alors :

$$\begin{aligned} \xi &= \{x : f_1(x) = f_2(x) \neq 0\} \\ \text{et } \eta &= \{y = (h_1(x), \dots, h_m(x)) : x \in \xi\}, \end{aligned}$$

et enfin, $\langle \theta, y \rangle$ étant le produit scalaire de θ et de y ,

$$f_k^*(y) = \int_{\Omega^m} a(\theta) \exp[\langle \theta, y \rangle] dG_k(\theta), \quad k = 1, 2.$$

Sous ces notations, on a $f_1^*(y) = f_2^*(y)$ si $y \in \eta$. Il est alors possible, si l'on restreint η , que cela implique $G_1 = G_2$. Pour cela, notons $c(\eta)$ l'enveloppe convexe¹ de η et distinguons les quatre cas suivants :

(i) η est fini.

Le cas (i) regroupe les mélanges ne prenant qu'un nombre fini de valeurs non nulles et ceux des distributions pour lesquelles les fonctions h_i sont à valeurs dans un ensemble fini. C'est typiquement le cas de la distribution Binomiale qui ne prend qu'un nombre fini de valeurs non nulles. La discussion de Teicher dans [Teicher63], pour cette distribution, donne une idée des problèmes soulevés dans ce cas.

(ii) η est infini, $c(\eta)$ est bornée et η n'a pas de point d'accumulation dans l'intérieur² de $c(\eta)$.

Le cas (ii) est le moins intéressant d'un point de vue statistique.

(iii) comme (ii) sauf que $c(\eta)$ est supposé non borné.

Le cas (iii) est celui des distributions discrètes (η n'a pas de point d'accumulation), pour lesquelles on ne peut pas borner l'une, au moins, des fonctions h_i . C'est le cas de la distribution de Poisson.

(iv) η est infini et η a un point d'accumulation à l'intérieur de $c(\eta)$.

Le cas (iv) est essentiellement celui de distributions continues.

Grâce à quelques précisions d'ordre techniques, on obtient, dans les cas (iii) et (iv) légèrement modifiés, l'égalité de G_1 et G_2 . Par application de ces résultats, Barndorff-Nielsen prouve l'identifiabilité de la classe des mélanges de distributions de Poisson et de celle des mélanges des produits de $n > 1$ distributions normales identiques. Teicher, dans [Teicher60], a d'ailleurs montré que ce n'est pas le cas si $n = 1$ et dans [Teicher67] que ce n'est pas non plus le cas si les distributions ne sont pas identiques.

¹L'enveloppe convexe d'une partie A d'un espace affine E est la plus petite partie convexe de E contenant A .

2.4 Identifiabilité des classes de mélanges de distributions gaussiennes.

Dans un article, très complet, sur l'identifiabilité des mélanges, Bruni et Koch (voir [Bruni85]), énoncent deux théorèmes d'identifiabilité de la classe des mélanges de distributions normales. Nous savons que la classe des mélanges de distributions normales $N(x, \lambda(y))$, $x \in \mathbb{R}^p$,

$$N(x, \lambda(y)) = \frac{1}{(2\pi)^{p/2} |\Sigma(y)|^{1/2}} \exp \left\{ -\frac{1}{2} (x - m(y))^T \Sigma^{-1}(y) (x - m(y)) \right\}$$

n'est pas identifiable en général, c'est pourquoi il est nécessaire de réduire l'ensemble des paramètres $\lambda = (m, \Sigma)$.

2.4.1 Distribution normale à une dimension.

Soit D un compact de \mathbb{R} , on définit le sous-ensemble Λ_1 de $C^1(D)^3$:

$$\Lambda_1 = \left\{ \lambda = (\lambda_1, \lambda_2) \in C^1(D) : \begin{array}{l} |\lambda_1(y)| \leq K_1 < +\infty, 0 < s_1 \leq \lambda_2(y) \leq s_2 < +\infty, \\ |\lambda_1'(y)| + |\lambda_2'(y)| \leq K_2 < +\infty, \forall y \in D \end{array} \right\},$$

puis l'on précise cet ensemble :

$$\Lambda_2 = \left\{ \begin{array}{l} \text{sous ensemble de } \Lambda_1 \text{ avec, de plus,} \\ \forall \lambda, \lambda' \in \Lambda_2, \forall y, y' \in D, \lambda(y) = \lambda'(y') \Rightarrow y = y' \end{array} \right\},$$

où aussi :

$$\Lambda_3 = \left\{ \begin{array}{l} \lambda \in \Lambda_1 : \lambda(y) = \lambda(y') \Rightarrow y = y', \forall y, y' \in D, \text{ et} \\ \text{pour } \lambda_1, \lambda_2 \text{ monotones, } |\lambda_1(y)| + |\lambda_2(y)| > 0, \forall y \in D \end{array} \right\}.$$

La topologie sur ces ensembles est la topologie naturellement induite par la norme sup. Il est important de remarquer qu'un ensemble Λ_3 ne contient pas nécessairement d'ensemble Λ_2 . Il n'est pas non plus nécessairement contenu dans un ensemble Λ_2 .

2.4.2 Théorème d'identifiabilité.

Si l'on entend par $\mathbf{M}(D)$ l'ensemble des mesures signées de support D , à variations bornées, muni de la topologie vague [Billingsley68], on notera $\mathcal{P}(D)$ le sous ensemble des probabilités de $\mathbf{M}(D)$. On définit l'opérateur

$T : \Lambda_1 \times \mathbf{M}(D) \rightarrow L_\infty(\mathbf{R})^4$ par :

$$f = T(\lambda, \mu) : f(x) = \int_D N(x, \lambda(y)) \mu(dy).$$

On montre aisément que T est uniformément continue séparément par rapport à λ et μ et est donc continue sur $\Lambda_1 \times \mathbf{M}(D)$.

Bruni et Koch établissent alors les deux théorèmes suivants :

³Ensemble des fonctions à valeurs réelles, continues et dérivables de D .

⁴ $L_\infty(\mathbf{R})$ est l'espace des fonctions de \mathbf{R} dans \mathbf{R} intégrables pour la norme sup.

THÉORÈME 2.2: Pour $\lambda, \lambda' \in \Lambda_2$, $\mu, \mu' \in \mathcal{P}(D)$, si $T(\lambda, \mu) = T(\lambda', \mu')$, alors

$$\mu = \mu' \text{ et } \lambda = \lambda'(\mu - pp.),$$

qui montre l'unicité pour $\lambda, \lambda' \in \Lambda_2$.

THÉORÈME 2.3: Pour $\lambda, \lambda' \in \Lambda_3$, $\mu, \mu' \in \mathcal{P}(D)$, si $T(\lambda, \mu) = T(\lambda', \mu')$, il existe, alors, une fonction C^1 , $h : D \rightarrow D$, strictement monotone et définie de manière unique sur le support de μ tel que :

$$\mu'(dh(y)) = \mu(dy) \text{ et } \lambda'(h(y)) = \lambda(y)(\mu - pp.),$$

qui établit un résultat plus fort que le précédent, sur un ensemble différent.

Ces résultats confirment ceux, que nous énoncerons dans le chapitre suivant, sur l'identifiabilité des classes de mélanges finis pour les distributions normales. Dans ce cas, μ est une probabilité discrète, il n'y a plus de problème sur la continuité de λ et les hypothèses de monotonie de Λ_4 sont inutiles.

2.4.3 Extension au cas multidimensionnel.

Les développements effectués dans le cas des lois sur \mathbb{R} peuvent être étendus aux lois normales multidimensionnelles. Afin de comprendre le principe de l'extension, on peut regarder le cas $n = 2$.

On considère pour cela :

$$\begin{aligned} m(y) &= (\lambda_1(y), \lambda_2(y)) \\ \text{et } \Sigma(y) &= \begin{pmatrix} \lambda_3(y) & \lambda_4(y) \\ \lambda_4(y) & \lambda_5(y) \end{pmatrix} \end{aligned}$$

et l'on définit Λ_1^2 par :

$$\Lambda_1^2 = \left\{ \begin{array}{l} \lambda \in C^1(D) : |\lambda_i(y)| \leq k_1 < +\infty, i = 1, \dots, 5, 0 \leq \lambda_i(y), i = 3, 5; \\ |\lambda_3(y)\lambda_5(y) - \lambda_4^2(y)| \geq s_1 > 0; \sum_{i=1}^5 |\lambda'_i(y)| \leq k_2 < +\infty, \forall y \in D \end{array} \right\}.$$

Si l'on définit Λ_2^2 et Λ_3^2 comme précédemment, mais en fonction de Λ_1^2 , on peut étendre les théorèmes 2.2 et 2.3.

2.5 Identifiabilité des classes de mélanges de distributions continues à un paramètre sur \mathbb{R} .

Le théorème qui va suivre, présenté par Tallis dans [Tallis69], donne une condition nécessaire et suffisante d'identifiabilité de la classe des mélanges continus de distributions $F(x, \theta)$ à un paramètre unidimensionnel. C'est un théorème très peu pratique à mettre en oeuvre, que l'on peut cependant adapter dans certains cas particuliers (voir [Tallis69]).

Quand $F(x, \theta)$ est continue en θ , on peut, sans perte de généralité, supposer que $\theta \in [-1; +1]$.

Considérons, donc le mélange

$$H(x) = \int_{-1}^{+1} F(x, \theta) dG(\theta). \quad [2.7]$$

Identifiabilité d'une classe de mélanges.

Soit $T(x, \theta) = \frac{\partial F(x, \theta)}{\partial \theta}$, que l'on supposera continue en θ et de carré intégrable sur $[-1; +1] \times [-1; +1]$. Le mélange [2.7] peut, alors, après une intégration par parties, être mis sous la forme :

$$L(x) = \int_{-1}^{+1} T(x, \theta) G(\theta) d\theta, \quad [2.8]$$

où $L(x) = F(x, 1) - H(x)$. Tallis établit que la classe des mélanges [2.7] est identifiable si et seulement s'il existe une unique solution de carré intégrable à [2.8]. En posant $K(x, y) = \int_{-1}^{+1} T(x, z) T(y, z) dz$ on a :

THÉORÈME 2.4: *La classe des mélanges [2.7] est identifiable si et seulement si l'ensemble des fonctions propres associées à des valeurs propres non nulles de K est complet.*

Ce résultat est la conséquence d'une application du théorème de Hilbert-Schmidt pour les noyaux non-symétriques [Tricomi57].

2.6 Identifiabilité des classes de mélanges de mesures produits.

En 1967 Teicher ajoute un volet à sa série d'articles sur l'identifiabilité des mélanges en donnant des conditions d'identifiabilité pour les mélanges de mesures produits. Si l'on note, en se référant aux notations de la définition des classes de mélanges, :

$$F_{n, mn}^* = \left\{ F^*(x, \tilde{\theta}) : F^*(x, \tilde{\theta}) = \prod_{i=1}^n F(x, \theta^i), \text{ où } F(x, \theta^i) \in F_{1, m}, 1 \leq i \leq n \right\},$$

pour $\tilde{\theta} = (\theta^1, \dots, \theta^n) \in \Omega^{nm} = \underbrace{\Omega^m \times \Omega^m \times \dots \times \Omega^m}_{n \text{ fois}}$,

on définit une classe $F_{n, m \cdot n}^*$ particulière, sur Ω^{nm} , dont la distribution générique est la mesure produit $F^*(x, \tilde{\theta})$. En fait, $F_{n, mn}^*$ est la famille des distributions jointes des échantillons de n variables indépendantes ayant chacune une distribution dans $F_{1, m}$. La justification et la définition des mélanges de distributions de $F_{n, nm}^*$ est donc assurée par le développement sur le problème des échantillons du paragraphe I.1.4. Nous avons alors le résultat suivant :

THÉORÈME 2.5: *Si la classe de tous les mélanges de $F_{1, m}$ est identifiable, alors pour tout $n > 1$, la classe des mélanges de $F_{n, mn}^*$ est aussi identifiable. Réciproquement, si pour au moins un $n > 1$, la classe de tous les mélanges de $F_{n, mn}^*$ est identifiable, il en est de même pour $F_{1, m}$.*

Démonstration : voir Appendice B. ♦

Grâce à ce théorème, on peut, par exemple, établir que la classe de tous les mélanges de lois normales monodimensionnelles, $F = F_{1, 2}$ n'étant pas identifiable, dans le cas général, il en sera de même pour la classe de tous les mélanges de distributions de $F_{n, 2n}^*$ des produits de n distributions normales. Ce résultat est bien sûr modulé par le paragraphe I.2.3 qui assurait l'identifiabilité des mélanges des distributions produit de n normales identiques.

Identifiabilité d'une classe de mélanges.

2.7 Indépendance de l'identifiabilité par rapport à la paramétrisation.

Tallis et Chesson proposent, en 1981, de préciser la modélisation de mélange et de l'identifiabilité, afin de mettre en évidence l'indépendance de l'identifiabilité, de la classe, \mathcal{H} , des mélanges, par rapport à la paramétrisation des distributions de la famille F . Il est, en effet, souhaitable que le choix des paramètres à étudier, ne soit pas limité par la possibilité de perdre l'identifiabilité. Si l'on généralise les discussions de Robbins, on se trouve en présence d'un espace mesurable (X, \mathcal{S}) , ainsi que d'une famille paramétrique de probabilités $\{P_\theta, \theta \in \Omega^m\}$. Soit, aussi, \mathcal{B} , une tribu sur Ω^m , telle que $P_\theta(S)$ soit \mathcal{B} -mesurable pour tout $S \in \mathcal{S}$. Un mélange de P_θ est alors une mesure,

$$\mu_H(S) = \int_{\Omega^m} P_\theta(S) d\mu_G(\theta), \quad S \in \mathcal{S},$$

où μ_G est une mesure sur (Ω, \mathcal{B}) . Les mélanges de P_θ sont alors identifiables, d'après la définition classique, si la correspondance :

$$\mu_G \longmapsto \int_{\Omega^m} P_\theta d\mu_G(\theta), \quad \mu_G \text{ étant une probabilité sur } (\Omega^m, \mathcal{B})$$

est bijective.

Pour avoir une discussion plus globale, c'est à dire indépendamment de θ , sur l'identifiabilité des mélanges de P_θ , il faut paramétrer P_θ , par lui-même. L'ensemble des paramètres est alors un ensemble Θ de probabilités sur (X, \mathcal{S}) , sur lequel on définit la tribu \mathcal{P} . Un mélange est alors :

$$\int_{\Theta} P d\mu(P), \quad \text{où } \mu \text{ est une probabilité sur } (\Theta, \mathcal{P}).$$

et l'identifiabilité de tels mélanges est équivalente à la bijectivité de la correspondance

$$\mu \longmapsto \int_{\Theta} P d\mu(P),$$

où μ est une probabilité sur (Θ, \mathcal{P}) . Grâce à cette définition on peut s'affranchir de la dépendance en θ .

REMARQUE 2.2: Ces deux définitions des mélanges sont équivalentes si l'on considère à la fois $\Theta = \{P_\theta, \theta \in \Omega^m\}$ et les deux tribus \mathcal{S} et \mathcal{B} les plus petites pour que les correspondances respectives, $\theta \mapsto P_\theta(S)$ et $P \mapsto P(S)$, soient mesurables pour tout $S \in \mathcal{S}$. Cette propriété assure, en effet, la bijectivité du "changement de variable" pour s'affranchir de la paramétrisation.

La complexité de la mise en oeuvre de cette définition et principalement de l'établissement de l'expression de la mesure μ , n'a pas permis, jusqu'ici, d'étendre tous les résultats obtenus dans le cas classique de paramétrisation. Il permet cependant aux auteurs d'obtenir un résultat théorique intéressant sur l'identifiabilité des mélanges dont l'inverse est bornée, ainsi qu'une CNS d'identifiabilité des classes de mélanges sur des familles finies de distributions.

- CHAPITRE 3 -

Identifiabilité des mélanges finis.

La simplification du modèle, lorsque l'on passe des mélanges continus aux mélanges finis, n'est pas que formelle. Il est, en effet, beaucoup plus facile d'interpréter statistiquement ces derniers; Si une population X , de fonction de répartition F , peut être divisée en n sous populations X^i , de fonction de répartition F_i , dont la proportion dans la population X est, respectivement, c_i , on imagine aisément que X suit une loi de mélange : $F = \sum_{i=1}^n c_i F_i$. De nombreux exemples d'études pratiques des mélanges finis sont donnés dans les deux ouvrages de référence sur le sujet que sont [Everitt81] et [Titterington85].

La notion d'identifiabilité, elle aussi, prend un sens plus accessible dans le cas des mélanges finis. L'identifiabilité de la classe des mélanges finis sur la classe $\mathcal{F} = \{F(x, \theta), x \in \mathbb{R}^k, \theta \in \Omega^m\}$

$$H(\cdot) = \sum_{i=1}^n c_i F(\cdot, \theta_i) \text{ avec } \sum_{i=1}^n c_i = 1 \text{ et } \forall i, c_i > 0,$$

se traduit, si l'on suppose les θ_i différents deux à deux⁵, par l'unicité du n ainsi que des couples (c_i, θ_i) , $i = 1, \dots, n$, réalisant la décomposition de H . Par extension, on dit que le mélange H est identifiable sur \mathcal{F} , ou plus simplement identifiable, s'il n'y a pas de doute.

Dans le cadre de l'étude de l'identifiabilité des classes de mélanges finis, l'idéal est de montrer l'identifiabilité de \mathcal{H}' qui est la classe de tous les mélanges finis et qui contient tous les \mathcal{H}_k , $\forall k \in \mathbb{N}^*$ (\mathcal{H}_k et \mathcal{H}' sont définis dans la partie I.1.2). Même si des résultats ont été obtenus dans des cas simplifiés, notamment lorsque l'ensemble F est fini, nous ne développerons que le cas général des ensembles de paramètres continus, qui est celui que nous rencontrons dans le cadre de notre étude pratique.

⁵Ce que l'on peut toujours faire quitte à regrouper tous les termes contenant le même θ_i .

3.1 Condition suffisante pour l'identifiabilité des classes de mélanges finis.

Les premiers théorèmes sur l'identifiabilité des mélanges finis ont été établis par Teicher qui montre dès 1963 que les classes de tous les mélanges finis de lois normales et de loi gammas sur \mathbb{R}^{*+} sont identifiables sans hypothèses particulières. Toujours en 1963, Teicher montre que la classe des mélanges finis de distributions binomiales n'est pas identifiable.

3.1.1 Théorème d'identifiabilité de Teicher.

Les résultats énoncés en introduction sont obtenus grâce au théorème suivant.

THÉORÈME 3.1: *Soit $F = \{F_i(x) = F(x, \theta^i), x \in \mathbb{R}^n, \theta^i \in \Omega^m\}$ une famille de fonction de répartition pour lesquelles il existe une transformation Φ , de domaine de définition S_Φ , telle que $M : F \mapsto \Phi$, soit linéaire et bijective. Supposons de plus qu'il existe, sur F , un ordre total, noté \trianglelefteq , tel que $F_1 \triangleleft F_2$ implique*

- $S_{\Phi_1} \subseteq S_{\Phi_2}$,
- $\exists t_1 \in \overline{S_{\Phi_1}}$ ⁶, (t_1 indépendant de S_{Φ_2}) tel que $\lim_{t \rightarrow t_1} \frac{\Phi_2(t)}{\Phi_1(t)} = 0$.

Alors la classe \mathcal{H}' de tous les mélanges finis de F est identifiable.

Démonstration : voir Appendice C. ♦

3.1.2 Identifiabilité de la classe des mélanges finis de distributions normales.

La proposition qui suit découle du théorème d'identifiabilité de Teicher.

PROPOSITION 3.1: *La classe de tous les mélanges finis de lois normales est identifiable.*

Démonstration : Cette démonstration est donnée à titre d'exemple. On considère la classe

$$F = \left\{ N(x, \mu, \sigma^2) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) du, x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^{*+} \right\},$$

des fonctions de répartition des lois normales de moyenne μ et de variance σ^2 . Prenons pour $\Phi(t, \mu, \sigma^2) = \exp\left(\frac{\sigma^2 t^2}{2} - \mu t\right)$, $t \in \mathbb{R}$, la transformée de Laplace associée. On peut ordonner F au moyen d'un ordre lexicographique : si $N_1 = N(x, \mu_1, \sigma_1^2)$ et $N_2 = N(x, \mu_2, \sigma_2^2)$ sont deux éléments différents de F , on a $N_1 \triangleleft N_2$ si $\sigma_1^2 < \sigma_2^2$, ou, si $\sigma_1^2 = \sigma_2^2$ et $\mu_1 < \mu_2$. Il suffit, alors, pour conclure, d'appliquer le théorème 3.1 avec $S_{\Phi_i} =]-\infty; +\infty[$ et $t_1 = +\infty$. ♦

3.1.3 Identifiabilité des classes de mélanges de distributions gammas dont les densités ont le même support.

Teicher établit aussi la proposition suivante concernant les lois gammas.

⁶Ceci désigne le plus petit ensemble fermé contenant S_{Φ_1} .

PROPOSITION 3.2: La classe de tous les mélanges finis de lois gammas de la famille :

$$F = \left\{ \Gamma(x, p, q) = \int_0^x \frac{1}{\Gamma(q)} \frac{u^{q-1}}{p^q} \exp\left(-\frac{u}{p}\right) du, x \in \mathbb{R}^{*+}, p \in \mathbb{R}^{*+}, q \in \mathbb{R}^{*+} \right\},$$

des fonctions de répartition des lois gammas à deux paramètres p et q , est identifiable.

Il est important de constater que cette proposition se limite aux distributions définies sur \mathbb{R}^{*+} . Pour la suite de l'exposé, il est nécessaire pour nous d'élargir cette proposition. Nous présentons le résultat suivant :

PROPOSITION 3.3: La classe de tous les mélanges finis de lois gammas, dont les densités ont le même support, $]r_0; +\infty[$, est identifiable.

La preuve est une adaptation de celle de Teicher.

Démonstration : On considère la classe :

$$F_{r_0} = \left\{ \Gamma(x, p, q, r_0) = \int_{-\infty}^x \frac{1}{\Gamma(q)} \frac{(u - r_0)^{q-1}}{p^q} \exp\left(-\frac{(u - r_0)}{p}\right) \mathbb{I}_{[r_0; +\infty[}(u) du, \left| \begin{array}{l} x \in \mathbb{R}, r_0 \in \mathbb{R}, \\ p \in \mathbb{R}^{*+}, q \in \mathbb{R}^{*+} \end{array} \right. \right\},$$

des fonctions de répartition des lois gammas à trois paramètres p, q et r_0 , dont le dernier est fixé. Considérons la transformation linéaire qui à $\Gamma(\cdot, p, q, r_0)$ associe $\Phi_{r_0}(\cdot, p, q)$ telle que :

$$\begin{aligned} \Phi_{r_0}(t, p, q) &= \mathbb{E} \left[e^{-(X-r_0)t} \right] = \int_{-\infty}^{+\infty} \frac{1}{\Gamma(q)} \frac{(u - r_0)^{q-1}}{p^q} \exp\left(-\frac{u - r_0}{p} - (u - r_0)t\right) \mathbb{I}_{[r_0; +\infty[}(u) du, \\ &= \int_{-\infty}^{+\infty} \frac{1}{\Gamma(q)} \frac{u^{q-1}}{p^q} \exp\left(-\frac{u}{p} - ut\right) \mathbb{I}_{[r_0; +\infty[}(u + r_0) du, \text{ après changement de variable,} \\ &= \int_{-\infty}^{+\infty} \frac{1}{\Gamma(q)} \frac{(ut)^{q-1}}{p^q} \exp\left(-\frac{u(1+pt)}{p}\right) \mathbb{I}_{[0; +\infty[}(u) du, \\ &= (1+pt)^{-q} \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\Gamma(q)} (ut)^{q-1} \left(\frac{(1+pt)}{p}\right)^q \exp\left(-\frac{u(1+pt)}{p}\right) \mathbb{I}_{[0; +\infty[}(u) du}_{=1}, \\ &= (1+pt)^{-q}, t > -\frac{1}{p}. \end{aligned}$$

On vérifie que, r_0 étant fixé, la correspondance $\Gamma(\cdot, p, q, r_0) \mapsto \Phi_{r_0}(\cdot, p, q)$ est bien une bijection.

On peut toujours ordonner F_{r_0} au moyen d'un ordre lexicographique : si $\Gamma_1 = \Gamma(x, p_1, q_1, r_0)$ et $\Gamma_2 = \Gamma(x, p_2, q_2, r_0)$ sont deux éléments différents de F_{r_0} , on a $\Gamma_1 < \Gamma_2$ si $p_1 > p_2$, ou, si $p_1 = p_2$ et $q_1 > q_2$. Il suffit, alors, pour conclure, d'appliquer le théorème 3.1 avec

$$S_{\Phi_{r_0,1}} = \left] -\frac{1}{p_1}; +\infty \left[\subseteq \right] -\frac{1}{p_2}; +\infty \left[= S_{\Phi_{r_0,2}} \text{ et } t_1 = -\frac{1}{p_1}. \blacklozenge$$

Identifiabilité des mélanges finis.

3.1.4 Identifiabilité de la classe des mélanges finis de distributions inverse-gamma dont le support des densités est le même.

Afin d'élargir la collection de familles identifiables de Teicher nous montrons la proposition qui suit :

PROPOSITION 3.4: : La classe de tous les mélanges finis de lois inverse-gammas, dont les densités sont de même support, est identifiable.

Démonstration : On considère la classe

$$F'_{r_0} = \left\{ \text{Inv}\Gamma(x, p, q, r_0) = \int_{-\infty}^x \frac{1}{\Gamma(q)} \frac{(u - r_0)^{-q-1}}{p^q} \exp\left(-\frac{2}{p(u - r_0)}\right) \mathbb{I}_{[r_0; +\infty[}(u) du, \left| \begin{array}{l} x \in \mathbb{R}, r_0 \in \mathbb{R}, \\ p \in \mathbb{R}^{*+}, q \in \mathbb{R}^{*+} \end{array} \right. \right\},$$

des fonctions de répartition des lois gammas à trois paramètres p, q et r_0 , dont le dernier est fixé. Considérons la transformation linéaire qui à $\text{Inv}\Gamma(\cdot, p, q, r_0)$ associe $\Phi_{r_0}(\cdot, p, q)$ telle que :

$$\begin{aligned} \Phi_{r_0}(t, p, q) &= \mathbb{E} \left[\exp\left(-\frac{2t}{X - r_0}\right) \right] = \int_{-\infty}^{+\infty} \frac{1}{\Gamma(q)} \frac{(u - r_0)^{-q-1}}{p^q} \exp\left(-\frac{2}{p(u - r_0)} - \frac{2t}{u - r_0}\right) \mathbb{I}_{[r_0; +\infty[}(u) du, \\ &= \int_{-\infty}^{+\infty} \frac{1}{\Gamma(q)} \frac{u^{-q-1}}{p^q} \exp\left(-\frac{2}{pu} - \frac{2t}{u}\right) \mathbb{I}_{[r_0; +\infty[}(u + r_0) du, \text{ après changement de variable,} \\ &= (1 + pt)^{-q} \int_{-\infty}^{+\infty} \frac{1}{\Gamma(q)} u^{-q-1} \left(\frac{1 + pt}{p}\right)^q \exp\left(-\frac{2}{u} \left(\frac{1 + pt}{p}\right)\right) \mathbb{I}_{[0; +\infty[}(u) du, \\ &= (1 + pt)^{-q}, t > -\frac{1}{p}, \end{aligned}$$

comme précédemment. r_0 étant fixé, la correspondance $\Gamma(\cdot, p, q, r_0) \mapsto \Phi_{r_0}(\cdot, p, q)$ est bien une bijection.

On peut aussi ordonner F'_{r_0} au moyen du même ordre lexicographique que ci-dessus : si $\text{Inv}\Gamma_1 = \text{Inv}\Gamma(x, p_1, q_1, r_0)$ et $\text{Inv}\Gamma_2 = \text{Inv}\Gamma(x, p_2, q_2, r_0)$ sont deux éléments différents de F'_{r_0} , on a $\text{Inv}\Gamma_1 < \text{Inv}\Gamma_2$ si $p_1 > p_2$, ou, si $p_1 = p_2$ et $q_1 > q_2$. Il suffit, alors, pour conclure, d'appliquer le théorème 3.1 avec $S_{\Phi_{r_0,1}} = \left] -\frac{1}{p_1}; +\infty \right[\subseteq \left] -\frac{1}{p_2}; +\infty \right[= S_{\Phi_{r_0,2}}$ et $t_1 = -\frac{1}{p_1}$. ♦

3.2 Une CNS d'identifiabilité de la classe des mélanges finis.

C'est en 1968 que Yakowitz et Spragins énoncent une CNS d'identifiabilité de la classe \mathcal{H}' de tous les mélanges finis d'une famille $F = \{F_i = F(x, \theta^i), \theta^i \in \Omega^m, x \in \mathbb{R}^n\}$, de distribution donnée.

3.2.1 Théorème de Yakowitz et Spragins.

Identifiabilité des mélanges finis.

THÉOREME 3.2: *La classe de tous les mélanges finis de F est identifiable si et seulement si F est une famille linéairement indépendante sur \mathbb{R}^7 .*

Démonstration : Appendice D. ♦

Déoulant du fait que F est linéairement indépendante si et seulement si son image par un isomorphisme est linéairement indépendante sur l'espace image, on a le corollaire :

COROLLAIRE 3.1: *La classe de tous les mélanges finis de F est identifiable si et seulement si l'image de F par un isomorphisme vectoriel est linéairement indépendante sur l'espace image.*

Ce résultat permet de montrer l'identifiabilités des classes de mélanges finis de plusieurs familles paramétriques.

3.2.2 Applications du théorème de Yakowitz et Spragins.

Grâce au théorème 3.2 et à son corollaire, il est possible d'établir les 5 propositions suivantes, dont nous ne donnerons pas les démonstrations.

PROPOSITION 3.5: *Si k est un entier positif et F_1 la famille des produits de k fonctions de répartition de la loi exponentielle, la classe \mathcal{H}_1 de tous les mélanges finis de F_1 est identifiable.*

PROPOSITION 3.6: *La famille F_2 des fonctions de répartition gaussiennes k -dimensionnelles génère une classe \mathcal{H}_2 de mélanges finis identifiable.*

PROPOSITION 3.7: *Si $F_3 = F_1 \cup F_2$, la classe \mathcal{H}_3 de tous les mélanges finis d'éléments de F_3 est identifiable.*

PROPOSITION 3.8: *La famille F_4 des distributions binomiales négatives, non dégénérées, engendre une classe \mathcal{H}_4 de mélanges finis identifiable.*

PROPOSITION 3.9: *Soit F_5 une famille, quelconque, de fonctions de répartition engendrée par translation par la distribution univariée F . La classe \mathcal{H}_5 de tous les mélanges finis de F_5 est, alors, identifiable.*

Cette proposition est complétée en 1972 par Rennie (voir[Rennie72]), qui l'étend au cas multivarié.

3.2.3 Identifiabilité de la classe des mélanges finis de distributions de Cauchy.

La proposition qui suit est une autre conséquence du théorème de Yakowitz et Spragins.

PROPOSITION 3.10: *La classe \mathcal{H}_6 de tous les mélanges finis de la famille F_6 des densités de Cauchy, est identifiable.*

Démonstration : Cette démonstration est donnée compte tenu de l'appartenance de cette distribution au système de Pearson. Un élément de F_6 est de la forme :

$$f(x, \mu, \sigma) = \frac{\sigma}{\pi \left(\sigma^2 + (x - \mu)^2 \right)}, x \in \mathbb{R}, \sigma \in \mathbb{R}^{*+}, \mu \in \mathbb{R},$$

⁷ F est une famille linéairement indépendante sur R , si $\forall k \in N; \forall x \in R^n; (a_1, \dots, a_k) \in R^k; \forall i \neq j, F_j \neq F_i, \sum_{i=1}^k a_i F_i(x) = 0 \Rightarrow (a_1, \dots, a_k) = (0, \dots, 0)$.

et sa fonction caractéristique est :

$$\Psi(t, \mu, \sigma) = \exp(i\mu t - \sigma |t|), t \in \mathbb{R}, \sigma \in \mathbb{R}^{*+}, \mu \in \mathbb{R}.$$

La correspondance $\mathcal{C}_1 : f \mapsto \Psi$ est linéaire. La correspondance $\mathcal{C}_2 : \Psi \mapsto \Psi'$, avec

$$\begin{aligned} \Psi'(t, \mu, \sigma) &= \Psi(t, \mu, \sigma), \text{ si } t \geq 0, \\ &= 0, \text{ sinon,} \end{aligned}$$

est elle aussi linéaire. Si \mathcal{C}_3 associe à Ψ' sa transformée de Laplace, \mathcal{C}_3 est aussi une application linéaire. On a donc, $\forall s \in D_\sigma^8$, si $b = \sigma - i\mu$,

$$\mathcal{C}_3 \circ \mathcal{C}_2 \circ \mathcal{C}_1 (f(x, \mu, \sigma)) = \mathcal{C}(f(x, \mu, \sigma)) = \sigma (s + b)^{-1}, \sigma \in \mathbb{R}^{*+}, \mu \in \mathbb{R},$$

\mathcal{C} peut être étendue à un isomorphisme sur F_6 .

De plus, si l'on a

$$\sum_{j=1}^n c_j (s + b_j)^{-1}, s \in D = \bigcap_{j=1}^n D_{\sigma_j}, \quad [3.9]$$

on peut supposer, quitte à réindexer la somme que $\sigma_1 \leq \sigma_j$, si $j > 1$. Si on multiplie par $(s + b_1)$ les deux membres de [3.9], et que l'on fait tendre s vers $-b_1$ (en restant dans D), on a :

$$|c_1| \leq \lim_{s \rightarrow -b_1} |(s + b_1)| \sum_{j=2}^n |c_j (s + b_j)^{-1}| = 0.$$

On montre de même, que tous les autres c_j sont nuls. L'image de F_6 par un isomorphisme étant linéairement indépendante, le corollaire 3.1 nous assure l'identifiabilité de \mathcal{H}_6 . ♦

3.3 Mélanges finis de mesures produits.

Dans son article de 1967, Teicher développe un théorème, sur les mélanges finis de mesures produits, analogue à celui sur les mélanges continus. Sa démonstration est une adaptation de celle du théorème 2.5. L'énoncé du théorème, dans le cas des mélanges finis, est alors :

THÉORÈME 3.3: *Si la classe de tous les mélanges finis de $F_{1,m}$ est identifiable, alors pour tout $n > 1$, la classe des mélanges finis de $F_{n,mn}^*$ est aussi identifiable. Réciproquement, si pour au moins un $n > 1$, la classe de tous les mélanges finis de $F_{n,mn}^*$ est identifiable, il en est de même pour $F_{1,m}$.*

⁸ D_σ est l'ensemble des complexes dont la partie réelle est supérieure à $-\sigma$.

3.4 D'autres mélanges finis particuliers.

D'autres théorèmes d'identifiabilité sont développés dans des cas très particuliers, que nous ne ferons que citer ici. Dans [Fraser81], on montre l'identifiabilité de la classe des mélanges finis de distributions de Von-Mises et dans [Kent83] il est montré que dans une grande majorité des cas, la classe des mélanges finis des distribution de données directionnelles est identifiable.

- CHAPITRE 4 -

Introduction des mélanges généralisés.

Nous avons jusqu'ici évoqué les mélanges de distributions toutes issues de la même famille paramétrique. Les mélanges faisant intervenir des familles paramétriques différentes sont actuellement à l'étude en statistique appliquée, notamment dans le traitement statistique des images issues de la télédétection. Nous tentons ici de donner un cadre théorique à ce type de mélange.

4.1 Définition générale

Soit (X, \mathfrak{X}) , un espace métrique, et soit \mathcal{P} , l'ensemble des probabilités, P , sur cet espace. Notons \mathfrak{B} , une tribu définie sur \mathcal{P} . Soit, alors, $\mathcal{G}(\mathcal{P})$, la classe des probabilités, \mathbb{P} , sur $(\mathcal{P}, \mathfrak{B})$ et soit \mathcal{P}_0 , un sous ensemble de \mathcal{P} . On définit, alors, la classe $\mathcal{H}(\mathcal{P}_0)$, des probabilités \mathbb{H} sur (X, \mathfrak{X}) , par la correspondance :

$$\begin{aligned} \mathbb{M} & : \mathcal{G}(\mathcal{P}) \rightarrow \mathcal{H}(\mathcal{P}_0) \\ \mathbb{P} & \mapsto \mathbb{H}, \text{ telle que } \mathbb{H}(A) = \int_{\mathcal{P}_0} P(A) d\mathbb{P}(P), A \in \mathfrak{X}. \end{aligned} \quad [4.10]$$

Un élément de \mathcal{H} est appelé \mathbb{P} -mélange de \mathcal{P} ou plus simplement mélange généralisé.

Cette définition des mélanges est à rapprocher de celle de Tallis et Chesson exposée dans [Tallis82]. Nous ne l'utilisons, pourtant pas dans le même but, et n'exploitons pas les théorèmes d'identifiabilité présentés par Tallis et Chesson, trop complexes à mettre en oeuvre.

Si \mathcal{P}_0 est une réunion finie ou dénombrable de familles paramétriques disjointes, $\mathcal{P}_0 = \bigcup_{i=1}^{+\infty} \mathcal{P}_i$, l'écriture de [4.10] se simplifie en :

$$\mathbb{H}(A) = \sum_{i=1}^N \mathbb{P}(\mathcal{P}_i) \int_{\mathcal{P}_i} P(A) d\mathbb{P}(P | \mathcal{P}_i),$$

$$\text{ou } \mathbb{H}(A) = \sum_{i=1}^{+\infty} \mathbb{P}(\mathcal{P}_i) \int_{\mathcal{P}_i} P(A) d\mathbb{P}(P | \mathcal{P}_i), A \in \mathfrak{X}$$

et si l'on note $\mathbb{P}_i = \mathbb{P}(\cdot | \mathcal{P}_i)$,

$$\mathbb{H}(A) = \sum_{i=1}^N \mathbb{P}(\mathcal{P}_i) \int_{\mathcal{P}_i} P(A) d\mathbb{P}_i(P), \quad [4.11]$$

$$\text{ou } \mathbb{H}(A) = \sum_{i=1}^{+\infty} \mathbb{P}(\mathcal{P}_i) \int_{\mathcal{P}_i} P(A) d\mathbb{P}_i(P), A \in \mathfrak{X}, \quad [4.12]$$

ce qui nous donne une somme dénombrable pondérée de mélanges “classiques”, ou encore un mélange dénombrable de mélanges.

4.2 Mélanges finis généralisés.

Si, pour un mélange du type [4.11] les \mathbb{P}_i ne chargent, avec des poids positifs, qu'un nombre fini d'éléments de \mathcal{P}_i , nous sommes en présence d'un mélange fini généralisé.

Soit, pour k_0 et n , entiers positifs, et pour tout i , $1 \leq i \leq k_0$, des entiers non nuls m_i , ainsi que des familles de distributions paramétriques disjointes de fonctions de répartition F^i sur \mathbb{R}^n ,

$$F_n^i = \{F^i(x, \alpha^i), x \in \mathbb{R}^n, \alpha^i \in \Omega^{m_i}\},$$

indexées respectivement par $\alpha^i \in \Omega^{m_i}$, sous ensemble borélien du m_i -espace euclidien \mathbb{R}^{m_i} . $F^i(x, \alpha^i)$ est mesurable sur $\mathbb{R}^n \times \Omega^{m_i}$. On peut alors définir la famille

$$\mathcal{F}_{n, k_0} = \bigcup_{i=1}^{k_0} F_n^i,$$

qui est une famille de distributions de \mathbb{R}^n .

Pour N , un entier positif et tout j , $1 \leq j \leq N$, on définit une suite $(k_j)_j$ d'entiers positifs, inférieurs ou égaux à k_0 et on se donne $(c_j, \alpha_j^{k_j}) \in \mathbb{R}^{*+} \times \Omega^{m_{k_j}}$. La fonction de répartition H sur \mathbb{R}^n définie par :

$$H(x) = \sum_{j=1}^N c_j F^{k_j}(x, \alpha_j^{k_j}), \text{ avec } \sum_{j=1}^N c_j = 1. \quad [4.13]$$

est un mélange fini généralisé. On vérifie aisément que tout mélange de la forme [4.13] peut se mettre sous la forme [4.11], et [4.13] décrit donc la classe de tous les mélanges finis généralisés de distributions de \mathcal{F}_{n, k_0} .

- CHAPITRE 5 -

Systeme des distributions de Pearson.

Afin de donner un cadre précis à notre étude, nous avons choisi de nous consacrer à l'étude des mélanges finis pris dans la famille des distributions de Pearson. Ce système, défini par une équation différentielle à quatre paramètres, contient les principales distributions statistiques et se prête parfaitement au cadre appliqué de notre étude.

5.1 Définition.

Une densité f , sur \mathbb{R} , appartient au système des densités de Pearson, si elle est solution de l'équation différentielle à quatre paramètres :

$$\frac{1}{f(x)} \frac{df(x)}{dx} = -\frac{x+a}{c_0+c_1x+c_2x^2} \quad [5.14]$$

où a, c_0, c_1, c_2 sont les paramètres.

C'est une équation différentielle dont la forme des solutions est fortement liée à l'existence et au type des solutions du polynôme $P(x) = c_0 + c_1x + c_2x^2$. On ne peut, par conséquent, pas donner une forme générique des solutions. On ne s'intéresse, de plus, qu'aux solutions qui sont des densités. La discussion qui suit, permet d'ordonner toutes ces densités.

5.2 Densités appartenant au système de Pearson.

On peut, alors, répartir les densités solutions $f_{i,1 \leq i \leq 8}$, en huit familles distinctes $\{F_1, \dots, F_8\}$. Les notations, qui reprennent celles de Pearson, ne sont pas celles utilisées couramment, mais ont été introduites sous cette forme par l'auteur.

Nous détaillons ces familles en faisant le lien avec la discussion sur les racines de $P(X)$ permettant la résolution de l'équation différentielle.

La famille F_4 correspond au cas où $P(x)$ n'a pas de racines réelles. Ses éléments ont des distributions dont la densité vérifie :

$$f_4(x) = K_4 \left[C_0 + c_2 (x + C_1)^2 \right]^{-\left(\frac{1}{2c_2}\right)} \exp \left(-\frac{a-C_1}{\sqrt{C_0 c_2}} \arctan \left(\sqrt{\frac{c_2}{C_0}} (x + C_1) \right) \right),$$

avec $C_0 = c_0 - \frac{c_1^2}{4c_2}$, $C_1 = \frac{c_1}{2c_2}$,

Les éléments de la famille F_7 sont des distributions de densités:

$$f_7(x) = K_7 \left[c_0 + c_2 x^2 \right]^{-\left(\frac{1}{2c_2}\right)} \exp \left(-\frac{a}{\sqrt{c_0 c_2}} \arctan \left(\sqrt{\frac{c_2}{c_0}} x \right) \right),$$

Elles sont obtenues dans le sous cas de la famille F_4 pour lequel $c_1 = 0$.

K_4 et K_7 sont les constantes de normalisation $K_i = \int_{\mathbb{R}} f_i(x) dx$, $i = 4, 7$.

Ces deux cas sont les moins intéressants en pratique, si l'on excepte la loi $K_{\sigma, \mu}$ [5.15] de Cauchy et la loi centrale de Student $t_{\mu, \nu}$, qui sont des lois de type *IV*.

Les autres distributions sont très utilisées en statistique, et ont un réel intérêt pratique. Nous présentons ces distributions sous leur forme habituelle, en exprimant les paramètres "classiques" en fonction des paramètres de l'équation différentielle.

La famille $F_1 U F_2$ est la famille des lois **betas**, de densité:

$$f_{1,2}(x) = \frac{1}{B(p,q)} \frac{(x-b_1)^{p-1} (b_2-x)^{q-1}}{(b_2-b_1)^{p+q-1}} \mathbf{I}_{[b_1; b_2]}(x),$$

avec $b_1 = -\frac{1}{2c_2} (c_1 + \sqrt{\Delta})$, $b_2 = -\frac{1}{2c_2} (c_1 - \sqrt{\Delta})$, $p = \frac{a+b_1}{c_2(b_2-b_1)} + 1$, $q = -\frac{a+b_2}{c_2(b_2-b_1)} + 1$,

où $\Delta = c_1^2 - 4c_0 c_2$.

Ces deux types correspondent au cas où $P(x)$ a deux racines réelles de signes opposés, F_2 étant le cas où $p = q$.

La famille F_6 , pour laquelle $P(x)$ a deux racines réelles, distinctes et de même signe, est la famille des lois **betas du second type**. Leur densité est de la forme :

$$f_6(x) = \frac{s^q}{B(p,q)} \frac{(x-r)^{p-1}}{(x-(r-s))^{p+q}} \mathbf{I}_{[r; +\infty]}(x),$$

avec $s = \frac{\sqrt{\Delta}}{c_2}$, $r = -\frac{1}{2c_2} (c_1 - \sqrt{\Delta})$, $p = \frac{a+r}{c_2 s} + 1$, $q = \frac{1}{c_2} - 1$.

La famille F_5 est la famille des lois **inverse-gammas**, obtenue dans le cas où $P(x)$ est un carré parfait, propose des distributions de densité :

$$f_5(x) = \frac{1}{\Gamma(q)} \frac{(x-r)^{-q-1}}{p^q} \exp \left(-\frac{2}{p(x-r)} \right) \mathbf{I}_{[r; +\infty]}(x),$$

Système des distributions de Pearson.

avec $p = \frac{c_2}{a - \frac{c_1}{2c_2}}$, $q = \frac{1}{c_2} - 1$, $r = -\frac{c_1}{2c_2}$.

Il reste, ensuite, les cas où $P(x)$ n'est pas du second degré.

Si $c_2 = 0$ mais que $c_1 \neq 0$ on obtient la famille F_3 est la famille des lois **gammas**, de densité

$$f_3(x) = \frac{1}{\Gamma(q)} \frac{(x-r)^{q-1}}{p^q} \exp\left(-\frac{(x-r)}{p}\right) \mathbf{I}_{[r;+\infty[}(x),$$

avec $p = c_1$, $q = \frac{1}{c_1} \left(\frac{c_0}{c_1} - a\right) + 1$, $r = -\frac{c_0}{c_1}$.

Et pour finir, si $c_2 = c_1 = 0$ on aura la famille F_8 est la famille des lois **normales**, de densité

$$f_8(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

avec $\mu = -a$, $\sigma^2 = c_0$.

La définition des distributions du système de Pearson par la même équation différentielle permet de déterminer ces distributions uniquement à partir de leur quatre premiers moments. Cette propriété, que nous développons dans le paragraphe suivant, donne tout son intérêt à cette famille, pour la problématique qui nous intéresse.

5.3 Graphe de Pearson.

Soit X une variable aléatoire dont la densité appartient au système de Pearson et considérons son moment d'ordre 1,

$$\mu_1 = E[X]$$

et pour $q = 2, 3, 4$, les moments centrés :

$$\mu_q = E[(X - E[X])^q],$$

grâce auxquels on définit les coefficients,

$$\gamma_1 = \frac{(\mu_3)^2}{(\mu_2)^3} \text{ et } \gamma_2 = \frac{\mu_4}{(\mu_2)^2},$$

où $\sqrt{\gamma_1}$ est appelée "skewness" et γ_2 "kurtosis".

Les quatre coefficients de l'équation différentielle [5.14] sont alors reliés à μ_1 et μ_2 par les relations:

$$\begin{aligned} a &= \frac{(\gamma_2 + 3)\sqrt{\gamma_1\mu_2}}{10\gamma_2 - 12\gamma_1 - 18} - \mu_1, \\ c_0 &= \frac{\mu_2(4\gamma_2 - 3\gamma_1) - \mu_1(\gamma_2 + 3)\sqrt{\gamma_1\mu_2} + (\mu_1)^2(2\gamma_2 - 3\gamma_1 - 6)}{10\gamma_2 - 12\gamma_1 - 18}, \end{aligned}$$

$$c_1 = \frac{(\gamma_2 + 3)\sqrt{\gamma_1\mu_2} - 2\mu_1(2\gamma_2 - 3\gamma_1 - 6)}{10\gamma_2 - 12\gamma_1 - 18}$$

et $c_2 = \frac{(2\gamma_2 - 3\gamma_1 - 6)}{10\gamma_2 - 12\gamma_1 - 18}$,

ce qui permet de discuter les solutions de [5.14] suivant les valeurs de γ_1 , γ_2 et du paramètre λ , permettant de discuter les différentes solutions de $c_0 + c_1x + c_2x^2$,

$$\lambda = \frac{c_1^2}{4c_0c_2} = \frac{\gamma_1(\gamma_2 + 3)^2}{4(4\gamma_2 - 3\gamma_1)(2\gamma_2 - 3\gamma_1)(2\gamma_2 - 3\gamma_1 - 6)}.$$

On retrouve, alors, les densités solutions $f_{i,1 \leq i \leq 8}$, et les huit familles $\{F_1, \dots, F_8\}$, définies précédemment :

$$\begin{aligned} \{f_1 \in F_1\} &\iff \{\lambda < 0\}, \\ \{f_2 \in F_2\} &\iff \{\lambda = 0 \text{ avec } \{\gamma_1 = 0 \text{ et } \gamma_2 < 3\}\}, \\ \{f_3 \in F_3\} &\iff \{\lambda \text{ est } \infty \text{ car } \{2\gamma_2 - 3\gamma_1 - 6 = 0\}\}, \\ \{f_4 \in F_4\} &\iff \{0 < \lambda < 1\}, \\ \{f_5 \in F_5\} &\iff \{\lambda = 1\}, \\ \{f_6 \in F_6\} &\iff \{\lambda > 1\}, \\ \{f_7 \in F_7\} &\iff \{\lambda = 0 \text{ avec } \{\gamma_1 = 0 \text{ et } \gamma_2 > 3\}\}, \\ \{f_8 \in F_8\} &\iff \{\lambda = 0 \text{ avec } \{\gamma_1 = 0 \text{ et } \gamma_2 = 3\}\}. \end{aligned}$$

La connaissance des moments μ_1, μ_2, μ_3 et μ_4 permet donc de déterminer, la sous famille F_i à laquelle appartient la distribution, ainsi que les paramètres de cette distribution. Notons, de plus, qu'on ne peut jamais avoir $\gamma_2 - \gamma_1 - 1 < 0$.

Le graphe de Pearson, voir FIG. I.1, exprimant γ_2 en fonction de γ_1 montre la répartition des différentes familles de distributions dans le plan (γ_1, γ_2) .

5.4 Comparaison entre la paramétrisation classique et celle du système de Pearson.

Dans la mesure où nous souhaitons étudier l'identifiabilité de mélanges de distributions du système de Pearson, et que cette identifiabilité dépend de la paramétrisation choisie, il est nécessaire, si nous voulons utiliser les résultats connus dans le cas de paramétrisation classique des distributions, de comparer les deux paramétrisations. Nous développons ici cette comparaison.

Nous retenons, pour F_4 les distributions de cauchy qui sont de la forme :

$$K_{\sigma,\mu}(x) = \frac{\sigma}{\pi(\sigma^2 + (x - \mu)^2)}, x \in \mathbb{R}, \sigma \in \mathbb{R}^{*+}, \mu \in \mathbb{R}, \quad [5.15]$$

qui correspond au cas où $c_2 = 0.5$, $c_1 = -\mu$, $a = c_1$ et $c_0 = \sigma^2 + \frac{\mu^2}{2}$. Toutes les distributions de Cauchy sont donc contenues une et une seule fois dans le système de Pearson sous la paramétrisation :

$$a \in \mathbb{R}, c_2 = 0.5, c_1 = a \text{ et } c_0 \geq \frac{a^2}{2}.$$

Système des distributions de Pearson.

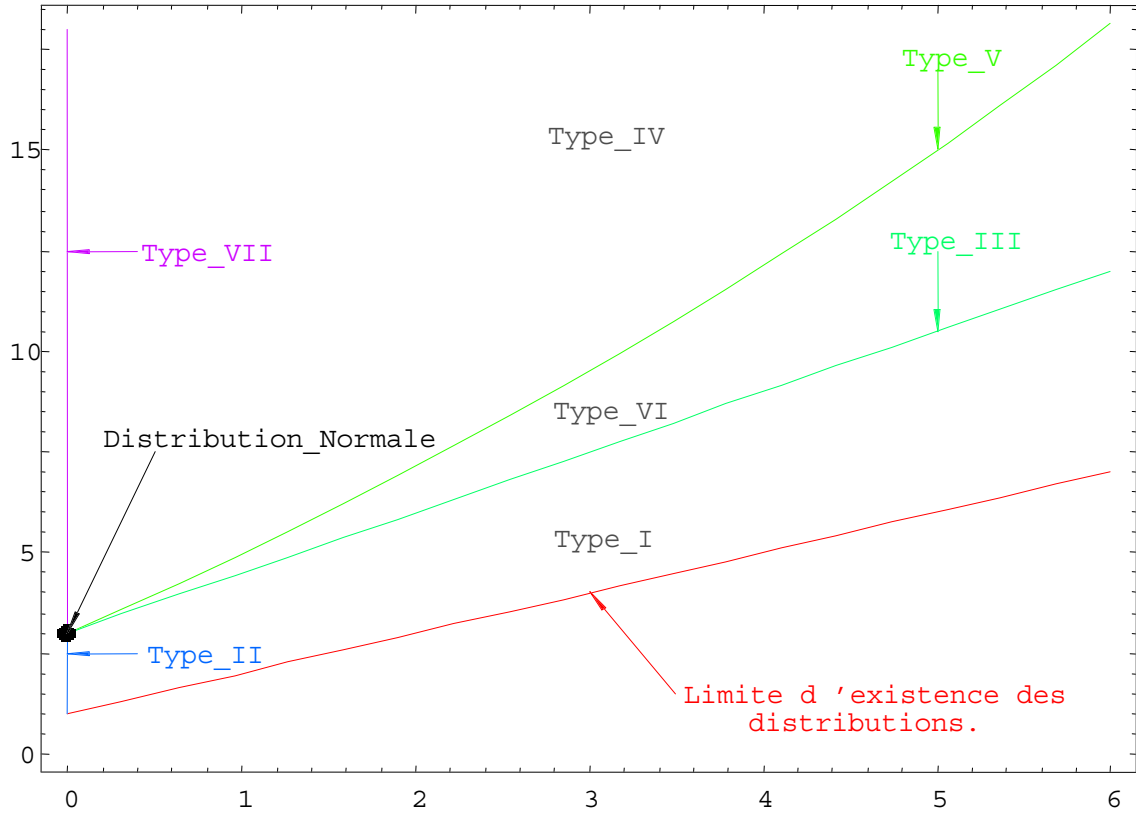


FIG. I.1. - GRAPHE DE PEARSON.

Pour les distributions de Student, qui appartiennent aussi à F_4 , nous avons :

$$t_{\mu,\nu}(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{(x-\mu)^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

avec $a = -\mu$, $c_2 = \frac{1}{\nu+1}$, $c_1 = 2ac_2$ et $c_0 = 1 + c_2(a^2 - 1)$. Comme pour la famille de distributions précédente, toute distribution de Student est contenue une et une seule fois dans la famille des distributions de Pearson. Notons que certaines distributions appartiennent à la fois à la famille des distributions de Cauchy et à celle des distributions de Student, qui ont une intersection non vide, bien qu'aucune ne contienne l'autre.

Pour les distributions betas de F_1UF_2 on a :

$$f_{1,2}(x) = \frac{1}{B(p,q)} \frac{(x-b_1)^{p-1}(b_2-x)^{q-1}}{(b_2-b_1)^{p+q-1}} \mathbf{I}_{[b_1;b_2]}(x),$$

où $b_1 = -\frac{1}{2c_2}(c_1 + \sqrt{\Delta})$, $b_2 = -\frac{1}{2c_2}(c_1 - \sqrt{\Delta})$, $p = \frac{a+b_1}{c_2(b_2-b_1)} + 1$, $q = -\frac{a+b_2}{c_2(b_2-b_1)} + 1$ et $\Delta = c_1^2 - 4c_0c_2$.

Signalons que, pour que les racines b_1 et b_2 de $P(X)$ soient de signes opposés, il est nécessaire que c_0 et c_2 soient de signes opposés.

Système des distributions de Pearson.

Par inversion des formules de paramétrisation ci-dessus on obtient :

$$c_2 = \frac{1}{2 - (q + p)}, \quad c_1 = -\frac{b_1 + b_2}{2 - (q + p)}, \quad c_0 = \frac{b_1 b_2}{2 - (q + p)} \quad \text{et} \quad a = \frac{b_1(q - 1) + b_2(p - 1)}{2 - (q + p)},$$

ce qui assure une correspondance bijective de l'ensemble des distributions beta de type 1 du système de Pearson sur l'ensemble des distributions beta "classiques" dont les bornes du support sont de signes opposés et dont la somme des paramètres p et q n'est pas égale à 2.

Pour les distributions beta de la famille F_6 , la densité est de la forme :

$$f_6(x) = \frac{s^q}{B(p, q)} \frac{(x - r)^{p-1}}{(x - (r - s))^{p+q}} \mathbf{I}_{[r; +\infty[}(x),$$

avec $s = \frac{\sqrt{\Delta}}{c_2}$, $r = -\frac{1}{2c_2} (c_1 - \sqrt{\Delta})$, $p = \frac{a+r}{c_2 s} + 1$, $q = \frac{1}{c_2} - 1$, où $\Delta = c_1^2 - 4c_0 c_2$.

Contrairement au cas précédent, il faut que les racines r et $r - s$ de $P(X)$ soient de même signe, (mais pas égales), ce qui est le cas si et seulement si c_2 et c_0 sont de même signe.

Par inversion des formules on obtient :

$$c_2 = \frac{1}{q + 1}, \quad c_1 = -\frac{2r - s}{q + 1}, \quad c_0 = \frac{r(r - s)}{q + 1} \quad \text{et} \quad a = \frac{ps}{q + 1} - r,$$

ce qui assure une correspondance bijective de l'ensemble des distributions beta de type 2 du système de Pearson sur l'ensemble des distributions beta dégénérées dont les bornes du support sont de même signe et dont le paramètre q n'est pas égale à -1 .

Le dernier cas où le polynôme P a des racines est celui où les deux racines sont identiques et où l'on est en présence de distributions inverse gamma de densité :

$$f_5(x) = \frac{1}{\Gamma(q)} \frac{(x - r)^{-q-1}}{p^q} \exp\left(-\frac{2}{p(x - r)}\right) \mathbf{I}_{[r; +\infty[}(x),$$

avec $p = \frac{c_2}{a+r}$, $q = \frac{1}{c_2} - 1$, $r = -\frac{c_1}{2c_2}$.

Notons, pour commencer, que c_0 n'intervient pas dans la reparamétrisation. Comme nous sommes dans le cas où le discriminant du binôme est nul on a :

$$c_1^2 - 4c_0 c_2 = 0 \Leftrightarrow c_0 = \frac{c_1^2}{4c_2},$$

ce qui donne, pour la reparamétrisation :

$$c_2 = \frac{1}{q + 1}, \quad c_1 = -\frac{2r}{q + 1}, \quad a = \frac{1}{p(q + 1)} - r \quad \text{et} \quad c_0 = \frac{r^2}{q + 1}.$$

La correspondance est bijective car $p > 0$ et $q > 0$.

Passons maintenant aux distributions gammas du système de Pearson, dont la densité est donnée par :

$$f_3(x) = \frac{1}{\Gamma(q)} \frac{(x - r)^{q-1}}{p^q} \exp\left(-\frac{(x - r)}{p}\right) \mathbf{I}_{[r; +\infty[}(x),$$

Système des distributions de Pearson.

avec $p = c_1$, $q = \frac{1}{c_1} \left(\frac{c_0}{c_1} - a \right) + 1$, $r = -\frac{c_0}{c_1}$.

Il n'y a ici que trois paramètres dans les deux paramétrisations (puisque $c_2 = 0$). On établit, comme précédemment les correspondances :

$$c_1 = p, c_0 = -pr \text{ et } a = -p(q - 1) - r,$$

ce qui montre, à nouveau l'équivalence entre la classe F_3 des distributions de Pearson et la famille des distributions gammas (à trois paramètres).

Enfin, pour la famille des lois normales, de densité:

$$f_8(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right),$$

avec $\mu = -a$, $\sigma^2 = c_0$. La correspondance est clairement bijective (puisque c_1 et c_2 sont nulles), entre la famille des distributions normales "classiques" et celles du système de Pearson.

- CHAPITRE 6 -

Identifiabilité des mélanges finis de distributions du système de Pearson.

Compte tenu de l'équivalence des paramétrisations que nous venons de montrer, nous pouvons indifféremment étudier les problèmes posés par l'identifiabilité sur la paramétrisation classique ou sur celle de Pearson. Nous choisissons de le faire avec les variables classiques.

L'identifiabilité des classes des mélanges finis généralisés se traduit par l'unicité du nombre N de composantes du mélange, ainsi que l'unicité de l'ensemble des paramètres du mélange :

$$\left\{ \left(c_j, \theta_j^{k_j} \right) \in]0; 1[\times \Omega_{k_j}, 1 \leq j \leq N \right\}.$$

Dans le cas des mélanges généralisés, la question de l'identifiabilité est le plus souvent réglée par l'étude du comportement aux bornes des supports des différentes familles paramétriques en présence.

6.1 Un cas particulier.

L'identifiabilité de tous les mélanges finis de distributions du système de Pearson, n'est pas vérifiée. C'est le cas, comme nous l'établissons ici, pour les distributions du type F_1UF_2 . En effet, on a :

$$(1-x)^2 = 1 - 2x + x^2,$$

d'où

$$x^{p-1}(1-x)^{q-1} = x^{p-1}(1-x)^{q-3}(1-x)^2,$$

et

$$x^{p-1}(1-x)^{q-1} = x^{p-1}(1-x)^{q-3} - 2x^p(1-x)^{q-3} + x^{p+1}(1-x),$$

soit

$$x^{p-1}(1-x)^{q-1} + 2x^p(1-x)^{q-3} = x^{p-1}(1-x)^{q-3} + x^{p+1}(1-x)^{q-3}.$$

On a donc :

$$\begin{aligned} B(p, q) \frac{x^{p-1}(1-x)^{q-1}}{B(p, q)} + 2B(p+1, q-2) \frac{x^p(1-x)^{q-3}}{B(p+1, q-2)} &= B(p, q-2) \frac{x^{p-1}(1-x)^{q-3}}{B(p, q-2)} \\ &+ B(p+2, q-2) \frac{x^{p+1}(1-x)^{q-3}}{B(p+2, q-2)}, \end{aligned} \quad [6.16]$$

et en intégrant [6.16] sur $[0; 1]$,

$$B(p, q) + 2B(p+1, q-2) = B(p, q-2) + B(p+2, q-2) = S.$$

L'égalité [6.16] nous donne l'égalité, pour tout $q > 2$, de deux mélanges différents de deux lois betas :

$$p_1 \frac{x^{p-1}(1-x)^{q-1}}{B(p, q)} + (1-p_1) \frac{x^p(1-x)^{q-3}}{B(p+1, q-2)} = p_2 \frac{x^{p-1}(1-x)^{q-3}}{B(p, q-2)} + (1-p_2) \frac{x^{p+1}(1-x)^{q-3}}{B(p+2, q-2)},$$

où $p_1 = \frac{1}{S}B(p, q)$, $1-p_1 = \frac{1}{S}2B(p+1, q-2)$, $p_2 = \frac{1}{S}B(p, q-2)$ et $1-p_2 = \frac{1}{S}B(p+2, q-2)$.

Ceci contredit la propriété d'identifiabilité de la classe des mélanges de lois betas du premier type. Signalons qu'une démonstration identique peut être faite pour les distributions betas du deuxième type.

PROPOSITION 6.1: *Les classes des mélanges finis des distributions betas du premier et du deuxième type ne sont pas identifiables.*

6.2 Décomposition des mélanges finis généralisés.

Pour montrer des propriétés d'identifiabilité des classes de mélanges finis généralisés, il faut séparer les différents types de distributions en présence, afin de mettre en évidence les spécificités de ces types qui permettront, éventuellement, de conclure à l'identifiabilité.

Pour la clarté du propos, considérons un mélange de deux types de distributions. Ce cas s'étend naturellement à un mélange quelconque. Supposons que les n premières distributions sont toutes du même type et que les m suivantes sont du second type. On aura alors si $\sum_{i=1}^{n+m} p_i = 1$:

$$\sum_{i=1}^{n+m} p_i F_i(\cdot) = \sum_{i=1}^n p_i F_i(\cdot) + \sum_{i=n+1}^{n+m} p_i F_i(\cdot) \quad [6.17]$$

Identifiabilité des mélanges finis de distributions du système de Pearson.

et en notant :

$$\sum_{i=1}^n p_i = S \text{ et } \sum_{i=1}^n p_i = S', \text{ on a } S + S' = 1.$$

[6.17] devient donc :

$$\sum_{i=1}^{n+m} p_i F_i(\cdot) = S \sum_{i=1}^n \frac{p_i}{S} F_i(\cdot) + S' \sum_{i=n+1}^{n+m} \frac{p_i}{S'} F_i(\cdot).$$

En posant,

$$H(\cdot) = \sum_{i=1}^n \frac{p_i}{S} F_i(\cdot) \text{ et } H'(\cdot) = \sum_{i=1}^n \frac{p_i}{S'} F_i(\cdot),$$

nous mettons en évidence un mélange des mélanges H et H' :

$$\sum_{i=1}^{n+m} p_i F_i(\cdot) = SH(\cdot) + S'H'(\cdot), \quad S + S' = 1.$$

La réciproque est immédiate. La démonstration du cas général est identique.

PROPOSITION 6.2: *La classe de tous les mélanges finis d'une famille de distributions réunissant k types différents de distributions, est la classe de tous les mélanges finis de la famille réunissant les k classes des mélanges finis sur les familles de distributions de chacun des types séparément.*

On peut étendre cette proposition au cas où la famille de distributions est composée d'une infinité de types différents de distributions.

PROPOSITION 6.3: *Soit $(F_a)_{a \in A}$ une famille de familles de distributions indicée par un ensemble A . Soit \varkappa_{a_0} la classe de tous les mélanges finis de la famille F_{a_0} et \varkappa_A la classe de tous les mélanges finis de la famille $\bigcup_{a \in A} F_a$.*

Alors la classe \varkappa_A est la classe, \varkappa , de tous les mélanges finis de la famille $\bigcup_{a \in A} \varkappa_a$.

Démonstration : On a clairement $\varkappa \subseteq \varkappa_A$.

Montrons, alors que tout élément H_{A_0} de \varkappa_A est un élément de \varkappa . H_{A_0} étant un mélange fini, H_{A_0} est un élément de la classe \varkappa_{A_0} de tous les mélanges finis de la famille $\bigcup_{a \in A_0} F_a$ où A_0 peut être, sans perte de généralité, considéré comme un ensemble fini, puisque le mélange est fini. D'après la proposition 6.2, \varkappa_{A_0} est la classe de tous les mélanges finis de la famille $\bigcup_{a \in A_0} \varkappa_a \subseteq \bigcup_{a \in A} \varkappa_a$. On a donc $H_{A_0} \in \varkappa_{A_0} \subseteq \varkappa$. ♦

6.3 Support des densités et identifiabilité.

On se place dans le cadre d'une famille F de distributions totalement ordonnée au sens de l'inclusion. Si on note " \sqsubset " la relation d'ordre, on aura, pour tout couple (f_1, f_2) de distributions de F :

$$f_1 \sqsubset f_2 \Leftrightarrow S_{f_1} \subset S_{f_2} \text{ au sens strict,}$$

Identifiabilité des mélanges finis de distributions du système de Pearson.

où S_{f_i} est le support de f_i .

Dans le cadre de ce type de familles de distributions, on peut établir la proposition suivante :

PROPOSITION 6.4: *Soit F une famille de fonctions, positives sur leur support, F étant totalement ordonnée, au sens strict, au moyen de l'ordre défini ci-dessus. F est une famille libre.*

Démonstration : Pour la démonstration, nous appellerons S_n le support de $\sum_{i=1}^n a_i f_i(x)$ et si S_{f_i} est le support de f_i , on a $S_n = \bigcup_{i=1}^n S_{f_i}$. Montrons, alors, par récurrence, que :

$$\left[\begin{array}{l} \forall n \in \mathbb{N}, \forall x \in S_n, \sum_{i=1}^n a_i f_i(x) = 0, \\ \text{avec } \forall i, 1 \leq i \leq n, a_i \in \mathbb{R} \text{ et } f_i \in F \end{array} \right] \implies [\forall i, 1 \leq i \leq n, \text{ on a } a_i = 0].$$

Au rang 1, on a $\forall x \in S_{f_1}$,

$$a_1 f_1(x) = 0, \text{ avec } a_1 \in \mathbb{R} \text{ et } f_1 \in F, \quad [6.18]$$

f_1 étant une fonction positive sur son support, on a nécessairement $a_1 = 0$.

Supposons que la propriété est vraie au rang n , et montrons qu'elle l'est toujours au rang $n+1$.

Considérons pour cela : $\forall x \in S_{n+1}$:

$$\sum_{i=1}^{n+1} a_i f_i(x) = 0, \text{ avec } \forall i, 1 \leq i \leq n+1, a_i \in \mathbb{R} \text{ et } f_i \in F. \quad [6.19]$$

F étant totalement ordonnée, il existe un entier k_0 , $1 \leq k_0 \leq n$, tel que

$$\forall i, 1 \leq i \leq n, i \neq k_0 \Rightarrow f_i \sqsubset f_{k_0},$$

c'est à dire :

$$S_{n+1} = S_{f_{k_0}}.$$

Par conséquent, $B = S_{n+1} \setminus \bigcup_{\substack{i=1 \\ i \neq k_0}}^{n+1} S_{f_i}$ est un ensemble de mesure non nulle sur lequel on a :

$$a_{k_0} f_{k_0}(x) = 0, \forall x \in B, B \subseteq S_{f_{k_0}},$$

ce qui implique, f_{k_0} étant positive sur son support, que $a_{k_0} = 0$. On tire alors de [6.19] :

$$\sum_{i=1}^n a_i f_i(x) = 0, \forall x \in B$$

et on peut faire valoir l'hypothèse de récurrence. ♦

D'où l'on tire le corollaire immédiat,

COROLLAIRE 6.1: *La classe de tous les mélanges finis d'une famille de distributions totalement ordonnée au sens strict de la relation d'ordre \sqsubset , est identifiable.*

Identifiabilité des mélanges finis de distributions du système de Pearson.

On peut aussi étendre les propositions 3.3 et 3.4. Nous démontrons grâce à la proposition 6.4 la proposition générale d'identifiabilité des mélanges finis de la famille des distributions gammas.

PROPOSITION 6.5: *La classe \varkappa de tous les mélanges finis de la famille :*

$$F = \left\{ \Gamma(x, p, q, r) = \int_{-\infty}^x \frac{(u-r)^{q-1}}{\Gamma(q)p^q} \exp\left(-\frac{(u-r)}{p}\right) \mathbb{I}_{[r;+\infty[}(u) du, \left| \begin{array}{l} x \in \mathbb{R}, r \in \mathbb{R}, \\ p \in \mathbb{R}^{*+}, q \in \mathbb{R}^{*+} \end{array} \right. \right\},$$

des distributions gamma est identifiable.

Démonstration : Un élément de \varkappa se décompose, d'après la proposition 6.2, en un mélange de mélanges finis, H_1, \dots, H_k , des classes $\varkappa_{r_1}, \dots, \varkappa_{r_k}$ de mélanges des familles F_{r_1}, \dots, F_{r_k} , de la proposition 3.3, les r_i étant différents deux à deux. Chaque distribution H_i est une distribution paramétrique positive sur son support et la classe \varkappa_{r_i} est identifiable. La famille des distributions H_i est ordonnée au sens défini dans la proposition 6.4. D'après le corollaire 6.1, \varkappa est donc identifiable. \blacklozenge

La proposition assurant de l'identifiabilité de la classe de tous les mélanges finis de la famille inverse-gamma se démontre de la même façon :

PROPOSITION 6.6: *La classe de \varkappa' tous les mélanges finis de la famille :*

$$F' = \left\{ \text{Inv}\Gamma(x, p, q, r) = \int_{-\infty}^x \frac{(u-r)^{-q-1}}{\Gamma(q)p^q} \exp\left(-\frac{2}{p(u-r)}\right) \mathbb{I}_{[r;+\infty[}(u) du, \left| \begin{array}{l} x \in \mathbb{R}, r \in \mathbb{R}, \\ p \in \mathbb{R}^{*+}, q \in \mathbb{R}^{*+} \end{array} \right. \right\},$$

des distributions inverse-gamma est identifiable.

Démonstration : La démonstration est identique à la précédente. \blacklozenge

6.4 Lien entre identifiabilité et négligeabilité.

Certaines conditions sur le comportement limite des densités des distributions de la famille considérée, assurent l'identifiabilité de la classe de ses mélanges finis.

PROPOSITION 6.7: *Considérons une famille F_a , $a \in \mathbb{R}$, de distributions, dont les supports des densités ont tous une borne a (finie ou infinie), au moins, en commun. Supposons que les comportements limites (par exemple la décroissance vers 0) des densités en cette borne sont totalement ordonnés au sens de la négligeabilité et deux à deux non équivalents.*

La classe de tous les mélanges finis de distributions de cette famille est identifiable.

Démonstration : Montrons, par récurrence que la famille F_a est libre. Considérons la combinaison linéaire de n densités dont les vitesses de décroissance sont échelonnées. Supposons, sans perte de généralité, que

$$f_i \ll_a f_{i+1}, 1 \leq i \leq n-1$$

et montrons que

$$\forall x \in D, \sum_{i=1}^n p_i f_i(x) = 0 \iff p_i = 0, 1 \leq i \leq n. \quad [6.20]$$

Identifiabilité des mélanges finis de distributions du système de Pearson.

Si $n = 1$ la propriété est immédiate. Supposons [6.20] vraie au rang $n - 1$. On a :

$$\begin{aligned} \forall x \in D, \sum_{i=1}^n p_i f_i(x) &= 0 \\ \iff \forall x \in D, f_n(x) \left(\sum_{i=1}^{n-1} p_i \frac{f_i(x)}{f_n(x)} + p_n \right) &= 0, \\ \iff \forall x \in D, \sum_{i=1}^{n-1} p_i \frac{f_i(x)}{f_n(x)} + p_n &= 0. \end{aligned} \quad [6.21]$$

En passant à la limite, en a , dans [6.21] on a $p_n = 0$ et on se ramène au rang $n - 1$ ce qui permet de conclure. ♦

Cette proposition permet d'avancer sur l'identifiabilité des classes de mélanges finis de distributions de Student. En effet, si l'on considère une classe de telles distributions, à moyenne μ_0 fixée, on a pour équivalent en $+\infty$ et $-\infty$:

$$t_{\mu_0, \nu}(x) \underset{\infty}{\sim} \frac{\Gamma\left(\frac{\nu+1}{2}\right) \nu^{\frac{(\nu+1)}{2}}}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} x^{-(\nu+1)}.$$

Ces équivalents sont totalement ordonnés au sens de la proposition 6.7 et l'on en tire la proposition :

PROPOSITION 6.8: *La classe de tous les mélanges finis de la famille de Student :*

$$\mathcal{F}_\mu = \left\{ t_{\mu, \nu}, t_{\mu, \nu}(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(x-\mu)^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \nu \in \mathbb{N}^* \right\}$$

est identifiable.

La famille des distributions betas de support fixé, $[0; a]$, $a > 0$, entre aussi dans le cadre des applications de cette proposition. En effet de telles distributions ont pour densité :

$$f_{a,p,q}(x) = \frac{1}{B(p,q)} \frac{x^{p-1} (a-x)^{q-1}}{a^{p+q-1}} \mathbf{I}_{[0;a]}(x),$$

et ont donc un équivalent en a :

$$f_{a,p,q}(x) \underset{a}{\sim} \frac{1}{B(p,q)} \frac{(a-x)^{q-1}}{a^q},$$

et un équivalent en 0 :

$$f_{a,p,q}(x) \underset{0}{\sim} \frac{1}{B(p,q)} \frac{x^{p-1}}{a^p}.$$

Ces équivalents sont totalement ordonnés au sens de la négligeabilité, et l'on peut faire valoir la proposition 6.7 pour établir que :

PROPOSITION 6.9: *La classe de tous les mélanges finis de la famille $\mathcal{F}_{a,p_0} = \{f_{a,p_0,q}(x), q > 0\}$ et celle de tous les mélanges finis de la famille $\mathcal{F}_{a,q_0} = \{f_{a,p,q_0}(x), p > 0\}$ des distributions betas I de support $[0; a]$, $a > 0$, sont identifiables.*

Démonstration : cette proposition est une conséquence directe de la proposition 6.7. ♦

Identifiabilité des mélanges finis de distributions du système de Pearson.

Cette dernière proposition nous amène à nous intéresser à l'identifiabilité des classes de tous les mélanges finis de la famille des familles de distributions $(\mathcal{F}_{a,p_0})_{a \in \mathbb{R}^{**+}}$ et $(\mathcal{F}_{a,q_0})_{a \in \mathbb{R}^{**+}}$. En effet ces deux familles entrent dans la catégorie des familles dont les supports des densités sont totalement ordonnés au sens de l'inclusion. Par une démonstration analogue à celle conduite dans le cadre des distributions gammas et inverse-gammas on démontre, grâce à la proposition 6.4 que :

PROPOSITION 6.10: *La classe de tous les mélanges finis de la famille $(\mathcal{F}_{a,p_0})_{a \in \mathbb{R}^{**+}}$ est identifiable.*

PROPOSITION 6.11: *La classe de tous les mélanges finis de la famille $(\mathcal{F}_{a,q_0})_{a \in \mathbb{R}^{**+}}$ est identifiable.*

Et, pour les mêmes raisons :

PROPOSITION 6.12: *La classe de tous les mélanges finis de la famille des distributions beta I, dont le paramètre p est fixé à p_0 et dont la borne supérieure du support est 0, est identifiable.*

PROPOSITION 6.13: *La classe de tous les mélanges finis de la famille des distributions beta I, dont le paramètre q est fixé à q_0 et dont la borne supérieure du support est 0, est identifiable.*

Ces propositions ont pour corollaire, qui nous intéresse directement :

COROLLAIRE 6.2: *La classe de tous les mélanges finis de la famille des distributions de type $FI \cup FII$ du système de Pearson, restreinte au cas $c_0 = 0$ et $c_1 < 0$, avec p ou q fixé, est identifiable. (cas correspondant aux propositions 6.10 et 6.11).*

et de même :

COROLLAIRE 6.3: *La classe de tous les mélanges finis de la famille des distributions de type $FI \cup FII$ du système de Pearson, restreinte au cas $c_0 = 0$ et $c_1 > 0$, avec p ou q fixé, est identifiable. (cas correspondant aux propositions 6.12 et 6.13).*

Dans le cadre de l'étude d'images réelles (radar, satellites ...), les échantillons sont fournis sous la forme de niveaux de gris, que l'on peut toujours considérer comme positifs. Il est donc raisonnable de s'intéresser aux distributions des propositions 6.10 ou 6.11 dans un tel cadre.

Afin de compléter l'identifiabilité des mélanges finis des distributions betas, intéressons nous aux distributions de la famille F_6 . C'est la famille des distributions de type :

$$f_6(x) = \frac{s^q}{B(p,q)} \frac{(x-r)^{p-1}}{(x-(r-s))^{p+q}} \mathbf{I}_{[r;+\infty[}(x).$$

Si l'on travaille sur de telles distributions à support fixé (s_0 et r_0 fixés), et que l'on prend la précaution de fixer aussi p_0 ou q_0 , on montre comme précédemment que

PROPOSITION 6.14: *La classe de tous les mélanges finis de la famille de distributions :*

$$\mathcal{F}_{6,r_0,s_0,p_0} = \left\{ f_6(x) = \frac{s_0^{q_0}}{B(p_0,q_0)} \frac{(x-r_0)^{p_0-1}}{(x-(r_0-s_0))^{p_0+q_0}} \mathbf{I}_{[r_0;+\infty[}(x) \right\}$$

et celle de tous les mélanges finis de la famille

$$\mathcal{F}_{6,r_0,s_0,q_0} = \left\{ f_6(x) = \frac{s_0^{q_0}}{B(p,q_0)} \frac{(x-r_0)^{p-1}}{(x-(r_0-s_0))^{p+q_0}} \mathbf{I}_{[r_0;+\infty[}(x) \right\}$$

sont identifiables.

Identifiabilité des mélanges finis de distributions du système de Pearson.

Démonstration : Cette proposition est une application directe de la proposition 6.7. ♦

Si l'on considère maintenant les familles des distributions :

$$\mathcal{F}_{6,s_0,p_0} = \left\{ f_6(x) = \frac{s_0^q}{B(p_0, q)} \frac{(x-r)^{p_0-1}}{(x-(r-s_0))^{p_0+q}} \mathbf{I}_{[r;+\infty[}(x) \right\}, s_0 \text{ et } p_0 \text{ sont fixés,}$$

et

$$\mathcal{F}_{6,s_0,q_0} = \left\{ f_6(x) = \frac{s_0^{q_0}}{B(p, q_0)} \frac{(x-r)^{p-1}}{(x-(r-s_0))^{p+q_0}} \mathbf{I}_{[r;+\infty[}(x) \right\}, s_0 \text{ et } q_0 \text{ sont fixés,}$$

on s'aperçoit que ces familles sont deux familles de familles de distributions dont les supports sont totalement ordonnés au sens de la proposition 6.4. En recyclant, à nouveau, le principe des démonstrations de l'identifiabilité des classes de tous les mélanges des familles de distributions gamma et inverse-gammas, on conclut que :

PROPOSITION 6.15: *La classe de tous les mélanges finis de la famille \mathcal{F}_{6,s_0,p_0} et celle de tous les mélanges finis de la famille \mathcal{F}_{6,s_0,q_0} sont identifiables.*

Si l'on souhaite compléter la famille de distributions beta I envisagée plus haut, on établit le corollaire suivant:

COROLLAIRE 6.4: *La classe de tous les mélanges finis de la famille*

$$\mathcal{F}_{6,p_0} = \left\{ f_6(x) = \frac{r^q}{B(p_0, q)} \frac{(x-r)^{p_0-1}}{x^{p_0+q}} \mathbf{I}_{[r;+\infty[}(x) \right\}, r > 0$$

et celle de tous les mélanges finis de la famille de distributions

$$\mathcal{F}_{6,q_0} = \left\{ f_6(x) = \frac{r^{q_0}}{B(p, q_0)} \frac{(x-r)^{p-1}}{x^{p+q_0}} \mathbf{I}_{[r;+\infty[}(x) \right\}, r > 0$$

sont identifiables.

Démonstration : Ce corollaire ne se déduit pas, à proprement parler, de la proposition précédente, mais sa démonstration est identique. ♦

6.5 Application à l'étude des mélanges généralisés.

On peut étendre les deux propositions 6.4 et 6.7, afin de les appliquer aux mélanges généralisés :

PROPOSITION 6.16: *Soit $(F_A)_{A \subseteq \mathcal{A}}$ une famille de familles de distributions dont le support respectif des densités est $A \subseteq \mathcal{A} \subset \mathbb{R}^n$. Supposons, de plus que cette famille est totalement ordonnée au sens strict de l'inclusion. Soit $\varkappa_{\mathcal{A}}$ la classe de tous les mélanges finis de la famille F_A .*

La classe, \varkappa , de tous les mélanges finis de la famille $F = \bigcup_{A \subseteq \mathcal{A}} F_A$ est identifiable si et seulement si $\forall A \subseteq \mathcal{A}$, \varkappa_A est identifiable.

Démonstration : D'après la proposition 6.3, l'élément H de \varkappa s'écrit sous la forme d'un mélange de k distributions, H_{A_1}, \dots, H_{A_k} provenant respectivement des classes des mélanges finis de F_{A_1}, \dots, F_{A_k} . On peut supposer sans perte de généralité que les A_i sont différents deux à deux. Les $(H_{A_i})_{1 \leq i \leq k}$ sont des distributions dont les supports des densités sont totalement ordonnés. Par application de la proposition 6.4, la classe de tels mélanges est identi-

Identifiabilité des mélanges finis de distributions du système de Pearson.

fiable si et seulement si, $\forall i, 1 \leq i \leq k, H_{A_i}$ est uniquement déterminée dans \varkappa_A . L'identifiabilité, $\forall A \subseteq \mathcal{A}$, des \varkappa_A est une condition suffisante pour cela.

La réciproque est immédiate. \blacklozenge

On traite de la même façon, la généralisation de la proposition 6.7 :

PROPOSITION 6.17: *Considérons une famille, $F = \bigcup_{i \in A} F_{a_i}$, réunion de familles de distributions de $(F_{a_i})_{i \in A}$, $A \subseteq \mathbb{R}^k$. Les éléments de F ont des densités dont les supports ont tous une extrémité a , finie ou infinie, au moins, en commun. On supposera, de plus, que*

$$\forall i, j \in A, \forall f \in F_{a_i}, \forall g \in F_{a_j}, f \ll_a g . \quad [6.22]$$

Alors, la classe \varkappa_A de tous les mélanges finis de F est identifiable si et seulement si les classes \varkappa_{a_i} des mélanges finis de chacune des F_{a_i} , $i \in A$ le sont.

Démonstration : D'après la proposition 6.3, un élément de la classe de tous les mélanges finis de la famille de distributions F s'écrit sous la forme d'un mélange de k distributions, H_{a_1}, \dots, H_{a_k} provenant respectivement des classes des mélanges finis de F_{a_1}, \dots, F_{a_k} , que l'on peut supposer, sans perte de généralité, différentes deux à deux. D'après [6.22] on a, quitte à réindicer l'ensemble,

$$\forall i, j \in A, H_{a_i} \ll_a H_{a_j} .$$

Par application de la proposition 6.7, la classe de tous les mélanges finis de F est identifiable si et seulement si chacun des H_{a_1}, \dots, H_{a_k} est uniquement déterminé dans \varkappa_A . L'identifiabilité de chacune des \varkappa_{a_i} est une condition suffisante pour cela.

La réciproque est immédiate. \blacklozenge

Enfin, pour pouvoir appliquer au système de Pearson les différents théorèmes que nous venons d'établir, nous aurons besoin de la proposition suivante :

PROPOSITION 6.18: *Considérons trois familles de distributions, s_1, s_2 et s_3 . Les distributions de s_1 sont des distributions dont le support des densités est de la forme $\bigcup_{i=1}^k [a_i; b_i]$, $a_i < b_i$, $(a_i; b_i) \in \mathbb{R}^2$. Les distributions de s_2 sont des distributions dont le support des densités est de la forme $\bigcup_{i=1}^k [a_i; +\infty[$, $a_i \in \mathbb{R}$. Les distributions de s_3 sont des distributions dont le support des densités est \mathbb{R} .*

La classe de tous les mélanges finis de distributions de la famille $s_1 \cup s_2 \cup s_3$ est identifiable si et seulement si les classes de tous les mélanges finis de chacune des familles s_1, s_2 et s_3 le sont.

Démonstration : D'après la proposition 6.2, un élément de la classe \varkappa de tous les mélanges finis de la famille de distributions $s_1 \cup s_2 \cup s_3$ s'écrit sous la forme d'un mélange de trois distributions, H_1, H_2 et H_3 provenant respectivement des classes des mélanges finis \varkappa_1, \varkappa_2 et \varkappa_3 de s_1, s_2 et s_3 . Montrons qu'une telle famille est libre. Si on a :

$$\forall x \in \mathbb{R}, c_1 H_1(x) + c_2 H_2(x) + c_3 H_3(x) = 0,$$

on a nécessairement $c_3 = 0$, puisque $c_1 H_1(x) + c_2 H_2(x)$ est identiquement nul sur une partie de mesure non nul du support de H_3 . Ceci implique que $c_2 = 0$, $c_1 H_1(x)$ étant identiquement nul sur une partie de mesure non nulle du support de H_2 . On a alors $c_1 = 0$, et la famille (H_1, H_2, H_3) est libre.

\varkappa est donc identifiable si et seulement si H_1, H_2 et H_3 sont uniquement déterminées, ce qui est le cas si et seulement si \varkappa_1, \varkappa_2 et \varkappa_3 sont identifiables. \blacklozenge

Identifiabilité des mélanges finis de distributions du système de Pearson.

6.6 Support et comportement asymptotique des lois du système de Pearson.

Conformément à ce que nous avons établi ci dessus, il convient d'analyser les comportements des différents types de loi du système de Pearson.

Type I et II : Les lois des types I et II (betas) ont des densités dont le support est borné et ce sont les seules.

Type III : Les lois de type III (gammas) ont des densités dont le support est minoré, ont un équivalent en $+\infty$ qui est:

$$\frac{1}{\Gamma(q)} \frac{x^{q-1}}{p^q} \exp\left(-\frac{x}{p}\right)$$

et un équivalent en r , la borne inférieure du support, qui est

$$\frac{1}{\Gamma(q)} \frac{(x-r)^{q-1}}{p^q}.$$

Type IV et VII : Les distributions de Cauchy, définies sur \mathbb{R} ont pour équivalent en $+\infty$ et en $-\infty$:

$$\frac{\sigma}{\pi x^2}$$

et celles de Student définies sur le même support ont pour équivalent en $+\infty$ et en $-\infty$:

$$\frac{\Gamma\left(\frac{\nu+1}{2}\right) \nu^{\frac{(\nu+1)}{2}}}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} x^{-(\nu+1)}.$$

Type V : Les lois du type V (inverse gammas) ont des densités dont le support est minoré et ont un équivalent en $+\infty$ qui est :

$$\frac{1}{\Gamma(q)} \frac{x^{-q-1}}{p^q},$$

l'équivalent en r est :

$$\frac{1}{\Gamma(q)} \frac{(x-r)^{-q-1}}{p^q} \exp\left(-\frac{2}{p(x-r)}\right).$$

Type VI : Les lois du type VI (betas du second type) ont des densités dont le support est minoré, ont un équivalent en $+\infty$ qui est :

$$\frac{s^q}{B(p, q)} x^{-q-1}$$

et un équivalent en r qui est :

$$\frac{1}{B(p, q) s^p} (x-r)^{p-1}.$$

Identifiabilité des mélanges finis de distributions du système de Pearson.

Type VIII : Les lois du type VIII (normales) ont des densités définies sur \mathbb{R} et un équivalent en $+\infty$ et en $-\infty$ qui est :

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Le cas litigieux entre les distributions de Cauchy et de Student, qui ont des comportements en $+\infty$ équivalents, survient sur l'intersection entre les deux familles (appartenant toutes les deux au type IV), il n'y a donc, pas de problème d'identifiabilité dans ce cas. Les lois des types V et VI sont, par conséquent, les seules à avoir des comportements limites en $+\infty$ qui peuvent être équivalents. Cependant, les équivalents des densités en la borne inférieure sont bien différents. En effet, si deux densités, $f_5 \in F_5$ et $f_6 \in F_6$ ont le même support, $[r; +\infty]$, on aura $f_6 \gg_r f_5$.

Si l'on se réfère aux notations de la proposition 6.18, $F_1 \cup F_2$ est une famille du type s_1 , $F_3 \cup F_5 \cup F_6$ est de la forme s_2 et $F_4 \cup F_7 \cup F_8$ est de la forme s_3 . Regardons, alors, l'identifiabilité des classes des mélanges finis de ces trois familles.

Nous avons précisé plus haut que l'identifiabilité de la classe de tous les mélanges finis de distributions de la famille de Pearson n'était pas vraie à cause de la non identifiabilité de la classe des mélanges finis des distributions betas. Nous avons vu, aussi, que si nous restreignons $F_1 \cup F_2$ à une famille $F'_1 \cup F'_2$ de distributions dont les densités ont des supports dont la borne inférieure est nulle et dont l'un des paramètres p ou q est fixé, la classe des mélanges sur cette famille est identifiable.

La famille $F'_4 \cup F_8$ est composée de distributions normales, de Student et de Cauchy. Nous savons les classes de tous les mélanges finis de distributions normales et de Cauchy identifiables, et nous avons montré celle de la classe des mélanges finis de la famille des distributions de Student pour une moyenne fixée. D'après la proposition 6.17 et ce qui précède au sujet du comportement des densités de ces familles de distributions en l'infini, la classe de tous les mélanges finis de la famille $F'_4 \cup F_8$ est identifiable.

Le problème de l'identifiabilité de la classe des mélanges finis de la famille des distributions de $F_3 \cup F_5 \cup F_6$ est le plus compliqué. D'après la proposition 6.17 et le comportement des densités en $+\infty$, la classe des mélanges finis de cette famille est identifiable si et seulement si les classes des mélanges finis de F_3 , F_5 et F_6 le sont. C'est acquis pour la famille F_3 des distributions gammas et la famille F_5 des distributions inverse-gammas. Les éléments de la classe des mélanges finis de la famille F_6 sont, d'après la proposition 6.2, des mélanges de mélanges H_6 , provenant de F_6 . On en conclut que la classe des mélanges finis de $F_3 \cup F_5 \cup F_6$ est identifiable si et seulement si la classes des mélanges finis de F_6 l'est. Cette identifiabilité n'est pas acquise mais nous avons défini une sous-famille F'_6 , restreinte à \mathcal{F}_{6,p_0} ou \mathcal{F}_{6,q_0} , telle que la classe de ses mélanges finis soit identifiable.

Tout ceci conduit au théorème général suivant :

THÉORÈME 6.1: *La classe de tous les mélanges finis de la famille $F'_1 \cup F'_2 \cup F_3 \cup F'_4 \cup F_5 \cup F'_6 \cup F_8$ est identifiable. Sans restriction de la famille de distributions de Pearson, la classe de tous ses mélanges finis n'est pas identifiable.*

DEUXIÈME PARTIE.

ESTIMATION DES PARAMÈTRES DE DISTRIBUTIONS DE MÉLANGES PAR LA MÉTHODE DES MOMENTS.

INTRODUCTION.

Cette partie du travail est consacrée à l'estimation des paramètres des distributions de mélanges, dans le cadre général des mélanges de deux distributions du système de distributions de Pearson. Nous choisissons de n'étudier que des familles de mélanges que nous savons identifiables, ce qui assure l'existence et l'unicité théorique des paramètres recherchés. Même dans le cas simple du mélange de deux distributions gaussiennes, de nombreuses questions restent actuellement sans réponse. Il faut, en effet, approfondir les problèmes liés à l'unicité de l'ensemble des paramètres, dans le cas théorique, et à l'existence et l'unicité, de cet ensemble, lorsque les moments théoriques sont remplacés par les moments empiriques. Ce n'est qu'une fois ce travail effectué qu'il est raisonnable d'envisager l'inversion du système des moments et la procédure d'estimation des paramètres. Quatre questions sont, donc, à élucider :

Question 1 : Les conditions d'existence et d'unicité n'étant pas abordées dans la littérature, quelles sont-elles ?

Question 2 : Le problème général d'existence et d'unicité étant très complexe, quelles solutions peut-on apporter dans le cas du mélange de deux distributions gaussiennes ?

Question 3 : Le problème théorique étant en suspend, peut-on espérer avoir recours à une méthode de type Newton pour une recherche numérique de l'ensemble des solutions du système ?

Question 4 : S'il n'est pas possible d'inverser numériquement le système, comment généraliser la méthode de résolution par substitution du type de celle de Pearson.

Afin d'apporter le maximum de réponses à ces interrogations, nous commencerons par faire un tour d'horizon des principales méthodes d'estimation utilisées dans les mélanges, puis, dans un second temps, nous revisiterons le cas gaussien en répondant au problème de l'unicité et en précisant ceux liés à l'existence avant de tenter de généraliser la méthode à l'étude des mélanges généralisés de la famille réunissant les familles des distributions normales et exponentielles.

- CHAPITRE 7 -

Estimation des paramètres des distributions de mélanges finis.

L'étude des distributions de mélanges finis, qui date de la fin du dix-neuvième siècle (voir [Pearson1894] et [Newcomb1886]) a révélé sa complexité, tout au long du siècle dernier, et a connu un tournant avec l'avènement de l'informatique. La méthode historique des moments de Pearson (voir [Pearson1894]), a longtemps été la seule envisagée dans la littérature (voir [Charlier24], [Rao48], [Cohen67] et [Fryer72]). Les comportements instables et inexpliqués liés à la méthode des moments, et sur lesquels nous reviendrons plus loin, ont conduit au développement d'autres types de méthodes d'estimation des paramètres de mélanges. La première piste suivie a été celle des méthodes graphiques (voir [Tanner62], [Bhattacharya67] et plus récemment [Fowlkes79]). Ces techniques, basées sur les propriétés de l'interprétation graphique de la distribution normale et l'analyse des fréquences ou des fréquences cumulées de l'échantillon considéré, donnent des résultats intéressants dans le cas de mélanges distinctement multimodaux. Signalons aussi les méthodes basées sur les estimateurs de distance minimum (voir [Wolfowitz57] et [Choi68]), ainsi que leurs extensions aux fonctions caractéristiques, (voir [Bryant83]), ou aux fonctions génératrices des moments, (voir [Quandt78]). Certaines méthodes numériques de décomposition ont aussi été utilisées (voir, par exemple [Medgyessy77]). On trouvera une synthèse de ces méthodes dans [Titterington85] ou [Redner84]. Les deux approches les plus couramment utilisées à l'heure actuelle, et depuis ces trente dernières années, sont basées sur l'estimation bayésienne (voir [Diebolt90] et [Robert92]), que nous présentons plus en détail dans la TROISIEME PARTIE, et surtout celles basées sur l'estimation du maximum de vraisemblance. En effet, malgré la non existence, en général, pour les mélanges finis, d'un tel estimateur (la vraisemblance n'est pas nécessairement bornée) et la difficulté d'une inversion algébrique du système des équations de vraisemblance, il est possible, moyennant quelques hypothèses de régularité sur les distributions composantes (voir [Redner84]), d'assurer l'existence de solutions aux équations de la vraisemblance. On peut alors, lorsque le calcul direct n'est pas possible, approcher l'estimateur du maximum de vraisemblance par un calcul algorithmique.

7.1 Estimation par le maximum de vraisemblance.

L'exemple connu pour le mélange de deux normales s'adapte par exemple au cas de la distribution du mélange d'une distribution normale et d'une distribution gamma :

$$L(X_1, \dots, X_n) = \prod_{X_i} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i - \mu)^2}{2\sigma^2}\right) + \frac{1}{\Gamma(q)} \frac{(X_i - a)^{q-1}}{p^q} \exp\left(-\frac{(X_i - a)}{p}\right) \mathbf{I}_{[a; +\infty[}(X_i) \right).$$

Si l'on initialise μ à X_1 la vraisemblance s'écrit :

$$L(X_1, \dots, X_n) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} + \frac{1}{\Gamma(q)} \frac{(X_1 - a)^{q-1}}{p^q} \exp\left(-\frac{(X_1 - a)}{p}\right) \mathbf{I}_{[a; +\infty[}(X_1) \right) \prod_{X_i \neq X_1} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i - X_1)^2}{2\sigma^2}\right) + \frac{(X_i - a)^{q-1}}{\Gamma(q)p^q} \exp\left(-\frac{(X_i - a)}{p}\right) \mathbf{I}_{[a; +\infty[}(X_i) \right).$$

Lorsque l'on fait tendre $\sigma \rightarrow 0$ la première partie tend vers $+\infty$ et la deuxième reste constante, ce qui fait diverger la vraisemblance. La vraisemblance n'est pas bornée. Il est possible, en pratique, d'éviter la divergence de la vraisemblance, par exemple en interdisant aux variances de tendre vers 0, ou en supposant l'égalité des variances (on pourra voir aussi [Cox74]).

L'algorithme *EM* dont la définition fondatrice a été proposée par Dempster, Laird et Rubin dans [Dempster77], est un algorithme itératif d'approximation de l'estimateur du maximum de vraisemblance. Ce type de méthode avait déjà été utilisé, au préalable, dans certains cas particuliers d'applications pratiques, pour rechercher le maximum de vraisemblance (voir, par exemple [Baum70]), mais le formalisme général est donné dans [Dempster77]. Le principe de cet algorithme, est de répéter, jusqu'à convergence (jusqu'à ce qu'un critère d'arrêt soit vérifié), une association de deux étapes dites *Expectation* et *Maximization*. *EM* permet de construire une suite $(\theta^{(k)})_k$ de paramètres telle que, si l'on note L la log-vraisemblance du mélange,

$$L(\theta^{(k+1)}) \geq L(\theta^{(k)}).$$

La littérature présente de très nombreux exemples d'applications de cet algorithme (voir [McLachlan97] pour une synthèse récente) ainsi que de ses versions stochastiques *SEM* et *SAEM* (voir [Broniatowski83], [Celeux86] et [Celeux92]). Nous présentons plusieurs exemples d'applications de ces différentes méthodes dans la TROISIEME PARTIE de ce document.

L'important travail de Redner et Walker (voir [Redner84]), qui montre l'existence, sous certaines conditions de régularité, de solutions convergentes aux équations de vraisemblance, permet d'assurer que, sous de bonnes conditions d'initialisation, l'utilisation des algorithmes de type *EM* permet une estimation par le maximum de vraisemblance. Le problème de l'initialisation des paramètres est important à double titre dans le cadre de ces algorithmes. D'une part, la convergence étant très lente, la qualité de l'initialisation a une réelle influence sur la rapidité des algorithmes. D'autre part, et nous touchons là à la principale difficulté dans l'utilisation des méthodes de type *EM*, une initialisation inadéquate conduit à bloquer la suite $L(\theta^{(k)})$ de maximisation de la vraisemblance dans un maximum local, et par conséquent à une estimation erronée. Le choix d'un $\theta^{(0)}$ pertinent constitue, donc, une étape cruciale dans la mise en oeuvre de telles procédures. Plusieurs des méthodes d'estimation, citées plus haut, ont été utilisées pour faire la "pré-estimation des paramètres" nécessaire à une bonne initialisation des méthodes itératives. Furman et Lindsay (voir [Furman94a], [Furman94b] et [Lindsay95]), montrent l'intérêt de recourir à la méthode des moments pour cette partie essentielle de *EM*, qui permet d'obtenir une estimation souvent aussi bonne que celle obtenue en initialisant l'algorithme avec les paramètres de simulation des différents exemples traités.

Il est nécessaire de signaler, enfin, les récents travaux de Fiorin (voir [Fiorin01]), qui montrent que l'on ne peut pas assurer, et particulièrement dans le cadre des mélanges, de la non convergence vers un maximum local.

7.2 La méthode des moments de Pearson.

La méthode des moments se résume principalement à l'inversion d'un système d'équations algébriques avec contraintes, (voir Appendice E à Appendice H). Ainsi en notant \vec{M} , resp. \vec{M}_n , le vecteur (éventuellement de dimension infinie) dont la $i^{\text{ème}}$ composante est le moment d'ordre i , resp. le moment empirique d'ordre i , d'une distribution F et θ le vecteur de ses paramètres, il existe une fonctionnelle, \mathfrak{F} telle que $\vec{M} = \mathfrak{F}(\theta)$. Ceci nous permet de définir un estimateur des moments $\hat{\theta}$, de θ , en résolvant l'équation $\vec{M}_n = \mathfrak{F}(\hat{\theta})$. Il s'agit pour nous d'inverser \mathfrak{F} en considérant la restriction de cette fonctionnelle à un nombre fini d'équations, dont la détermination mérite quelques précautions. En effet, l'estimation d'un nombre fini, noté k , de paramètres nécessite, en général, le recours à un même nombre k d'équations de moments, qu'il faut choisir en tenant compte de la perte de précision de l'estimation liée à l'augmentation de l'ordre des moments. Le système "optimal" que l'on espère ainsi obtenir est celui que nous avons alors à inverser. Le degré de complexité du problème, lorsque l'on passe de l'étude des estimateurs des moments d'une distribution simple à celle de distributions de mélanges est telle, que, même dans le cas des mélanges de deux lois, il est très difficile, et le plus souvent impossible, de donner une expression littérale de ces estimateurs.

Compte tenu de la grande taille, notée n , des échantillons auxquels nous nous référons, le choix de l'estimation des paramètres par une méthode des moments, dont on connaît la convergence en \sqrt{n} , semble indiqué. Pourtant cette méthode est considérée comme inférieure à celle du maximum de vraisemblance, son utilisation ayant révélé des cas d'instabilité jusqu'ici inexpliqués (voir sur ce point les deux articles de référence [Fryer70] et [Fryer72] ainsi que [McLachlan88] et [Tan72]).

Nous chercherons dans ce qui suit à préciser la problématique du choix des équations de moments afin de trouver une solution stable au problème des moments, l'efficacité prouvée de la méthode des moments pour l'initialisation de méthodes itératives d'estimation, suffisant à justifier que nous nous y intéressions.

7.2.1 Mélange de deux distributions normales de moyennes différentes.

Le système des équations de moments du mélange de deux distributions normales $N(\mu_1, \sigma_1^2)$ et $N(\mu_2, \sigma_2^2)$, en fonction des moments centrés s'écrit (voir Appendice H) :

$$\mathfrak{S}_1 = \begin{cases} 0 = px_1 + (1-p)x_2 \\ M_2^c = px_1^2 + (1-p)x_2^2 + (p\sigma_1 + (1-p)\sigma_2) \\ M_3^c = px_1^3 + (1-p)x_2^3 + 3(p\sigma_1x_1 + (1-p)\sigma_2x_2) \\ M_4^c = px_1^4 + (1-p)x_2^4 + 6(p\sigma_1x_1^2 + (1-p)\sigma_2x_2^2) + 3(p\sigma_1^2 + (1-p)\sigma_2^2) \\ M_5^c = px_1^5 + (1-p)x_2^5 + 10(p\sigma_1x_1^3 + (1-p)\sigma_2x_2^3) + 15(p\sigma_1^2x_1 + (1-p)\sigma_2^2x_2) \end{cases} \quad [7.23]$$

où M_1 est la moyenne du mélange et M_i^c est le $i^{\text{ème}}$ moment centré du mélange. On note, pour des raisons de symétrie, $x_1 = \mu_1 - M_1$ et $x_2 = \mu_2 - M_1$.

L'étude de ce cas particulier de mélange a été initiée par Pearson, en 1894, dans son fameux article qui est considéré comme le point de départ de la problématique des mélanges (voir [Pearson1894]). Il se ramène à l'étude des racines d'un polynôme de degré 9. Si l'on se donne un échantillon, \mathfrak{M} , \mathfrak{M}_2^c et \mathfrak{M}_3^c sont respectivement les estimateurs des moyenne, deuxième et troisième moments centrés. κ_4 et κ_5 sont les estimateurs des quatrième et cinquième cumulants. Le polynôme de Pearson est :

$$P_{\text{Pearson}}(x) = 24x^9 + 84\kappa_4x^7 + 36\mathfrak{M}_3^{c2}x^6 + (90\kappa_4^2 + 72\kappa_5\mathfrak{M}_3^c)x^5 + (444\kappa_4\mathfrak{M}_3^{c2} - 18\kappa_5^2)x^4 \\ + (288\mathfrak{M}_3^{c4} - 108\mathfrak{M}_3^c\kappa_4\kappa_5 + 27\kappa_4^3)x^3 - (63\mathfrak{M}_3^{c2}\kappa_4^2 + 72\mathfrak{M}_3^{c3}\kappa_5)x^2 - 96\mathfrak{M}_3^{c4}\kappa_4x - 24\mathfrak{M}_3^{c6},$$

Estimation des paramètres des distributions de mélanges finis.

alors, si ν^* est une racine négative de $P_{Pearson}$, on calcule :

$$r^* = \frac{-8\mathfrak{M}_3^c \nu^{*3} + 3\kappa_5 \nu^{*2} + 6\mathfrak{M}_3^c \kappa_4 \nu^* + 2\mathfrak{M}_3^c}{2\nu^{*4} + 3\kappa_4 \nu^{*2} + 4\mathfrak{M}_3^c \nu^*},$$

d'où l'on tire, si l'on suppose que $x_1^* \leq x_2^*$:

$$\begin{aligned} x_1^* &= \frac{1}{2} \left(r^* - \sqrt{r^{*2} - 4\nu^*} \right), \\ x_2^* &= \frac{1}{2} \left(r^* + \sqrt{r^{*2} - 4\nu^*} \right) \end{aligned}$$

et

$$\beta^* = \frac{1}{3} (2r^* - \mathfrak{M}_3^c / \nu^*),$$

qui permettent finalement d'obtenir :

$$\begin{aligned} \mu_1^* &= x_1^* + \mathfrak{M}, \\ \mu_2^* &= x_2^* + \mathfrak{M}, \\ p &= \frac{\mu_2^*}{\mu_2^* - \mu_1^*}, \\ \sigma_1^{*2} &= \mu_1^* \beta^* + \mathfrak{M}_2^c - \mu_1^{*2} \\ \text{et } \sigma_2^{*2} &= \mu_2^* \beta^* + \mathfrak{M}_2^c - \mu_2^{*2}. \end{aligned}$$

Si le polynôme fournit plusieurs racines donnant des jeux de paramètres plausibles on aura recours à la sixième équation de moments afin d'établir un test permettant de ne garder qu'un seul ensemble de paramètres.

7.2.2 Mélange de deux distributions normales de moyennes identiques.

Nous avons vu plus haut que $\mu_1 = \mu_2 = M_1$ où M_1 est la moyenne du mélange, entraîne que

$$x_1 = x_2 = 0.$$

Les estimateurs de μ_1 et μ_2 seront donc

$$\mu_1^* = \mu_2^* = \mathfrak{M},$$

où \mathfrak{M} est la moyenne empirique de l'échantillon étudié.

Il reste trois paramètres à déterminer et il faut avoir recours à la sixième équation de moments. Le système à résoudre est, alors :

$$\begin{cases} (p\sigma_1 + (1-p)\sigma_2) = M_2^c \\ 3(p\sigma_1^2 + (1-p)\sigma_2^2) = M_4^c \\ 15(p\sigma_1^3 + (1-p)\sigma_2^3) = M_6^c \end{cases} . \quad [7.24]$$

D'après Cohen (voir [Cohen67]) l'inversion de ce système donne :

$$\begin{aligned}\sigma_1 &= M_2^c + \frac{1}{2} \left[\frac{k_6}{5k_4} - \sqrt{\left(\frac{k_6}{5k_4}\right)^2 + \frac{4k_4}{3}} \right], \\ \sigma_2 &= M_2^c + \frac{1}{2} \left[\frac{k_6}{5k_4} + \sqrt{\left(\frac{k_6}{5k_4}\right)^2 + \frac{4k_4}{3}} \right] \\ \text{et } p &= \frac{\sigma_2}{\sigma_2 + \sigma_1},\end{aligned}$$

ou k_6 est le sixième cumulant du mélange $k_6 = M_6^c - 15M_4^cM_2^c - 10M_3^{c2} + 30M_2^{c3}$. L'estimation de paramètres en découle naturellement en remplaçant les moments et cumulants théoriques par les moments et cumulants empiriques.

7.2.3 Choix de la méthode.

Si l'on se donne un échantillon issu du mélange de deux distributions normales, il est nécessaire, dans l'optique de l'estimation des paramètres, de choisir la méthode d'estimation adéquate (moyennes égales ou différentes).

D'après Cohen, (voir [Cohen67]) la connaissance de M_3^c et k_4 permet de déterminer dans lequel des cas précédents on se trouve.

Si $M_3^c = 0$ et $k_4 > 0$, $x_1 = x_2$, il faut avoir recours au cas particulier où les moyennes sont égales.

Si $M_3^c = 0$ et $k_4 \leq 0$, $\sigma_1 = \sigma_2$ et $p = \frac{1}{2}$ on utilise le polynôme $P_{varégales}$, cas particulier où les variances sont égales et dont les calculs sont détaillés plus loin.

Si $M_3^c \neq 0$, on utilise le polynôme de degré 9, $P_{Pearson}$, et si σ_1 et σ_2 sont proches, on peut essayer d'affiner le résultat grâce à $P_{varégales}$.

Il est important de remarquer que la probabilité d'obtenir, à partir d'un échantillon, un troisième moment centré égal à zéro, est très petite. En pratique, le problème n'est donc pas censé arriver et une précision suffisante dans le calcul des racines de P_σ devrait permettre l'inversion du système et l'estimation des paramètres. Comme nous le signalions dans l'introduction, cette méthode n'est accompagnée d'aucune vérification d'existence ou d'unicité des solutions et la proposition de recourir à la sixième équation de moment pour éliminer d'éventuelles solutions du système ne correspondant pas aux paramètres recherchés n'est en aucun cas une assurance d'unicité. Nous choisissons donc de reprendre l'étude en commençant par ces questions.

- CHAPITRE 8 -

Etude du système des cinq premières équations de moments centrés.

8.1 Existence et unicité des solutions de \mathfrak{S}_1 .

8.1.1 Existence.

Nous choisissons, dans cette partie théorique, de travailler sur des moments que nous supposons être des moments de mélange de deux distributions gaussiennes. Ceci nous permet d'assurer l'existence d'une solution au moins pour le système que nous étudions.

8.1.2 Unicité.

Le fait de travailler sur un mélange fini de distributions normales nous assure l'identifiabilité c'est à dire l'unicité des paramètres des deux distributions. Ceci n'induit pas, cependant, l'unicité de l'ensemble solution du système qui n'est composé que de cinq équations de moments.

Notons pour la suite que la notion d'identifiabilité est définie à la symétrie de l'ensemble des paramètres près $(\mu_1 \leftrightarrow \mu_2, \sigma_1^2 \leftrightarrow \sigma_2^2 \text{ et } p_1 \leftrightarrow p_2)$. Conformément à ce que suggèrent Mc Lachlan et Peel dans [McLachlan00], nous introduirons, pour contrôler ce phénomène, un ordre total sur l'espace des paramètres du mélange. Cette opération s'effectuera d'ailleurs naturellement lors de l'inversion algébrique du système.

8.1.3 Existence et unicité locale.

Calculons le jacobien J_1 du système des moments \mathfrak{S}_1 , comme le préconise le théorème des fonctions implicites, afin d'étudier l'unicité locale de l'ensemble des paramètres du mélange. Signalons que dans le cas du mélange de deux distributions normales, la matrice jacobienne associée au système des cinq premiers moments centrés est la même que celle associée au système des cinq premiers moments non centrés. Le déterminant à calculer est :

$$J_1 = p^2 (1-p)^2 \begin{vmatrix} 1 & 1 & x_1 - x_2 & 0 & 0 \\ 2x_1 & 2x_2 & \begin{pmatrix} x_1^2 + \sigma_1 \\ -x_2^2 - \sigma_2 \end{pmatrix} & 1 & 1 \\ 3(x_1^2 + \sigma_1) & 3(x_2^2 + \sigma_2) & \begin{pmatrix} x_1^3 + 3\sigma_1 x_1 \\ -x_2^3 - 3\sigma_2 x_2 \end{pmatrix} & 3x_1 & 3x_2 \\ 4(x_1^3 + 3\sigma_1 x_1) & 4(x_2^3 + 3\sigma_2 x_2) & \begin{pmatrix} x_1^4 + 6\sigma_1 x_1^2 + 3\sigma_1^2 \\ -x_2^4 - 6\sigma_2 x_2^2 - 3\sigma_2^2 \end{pmatrix} & 6(x_1^2 + \sigma_1) & 6(x_2^2 + \sigma_2) \\ 5 \begin{pmatrix} x_1^4 + 6\sigma_1 x_1^2 \\ +3\sigma_1^2 \end{pmatrix} & 5 \begin{pmatrix} x_2^4 + 6\sigma_2 x_2^2 \\ +3\sigma_2^2 \end{pmatrix} & \begin{pmatrix} x_1^5 + 10\sigma_1 x_1^3 + 15\sigma_1^2 x_1 \\ -x_2^5 - 10\sigma_2 x_2^3 - 15\sigma_2^2 x_2 \end{pmatrix} & 10x_1(x_1^2 + 3\sigma_1) & 10x_2(x_2^2 + 3\sigma_2) \end{vmatrix},$$

que l'on développe en :

$$J_1 = p^2 (1-p)^2 \begin{pmatrix} -x_1^9 + 9 x_1^8 x_2 - 36 x_1^7 x_2^2 + 84 x_1^6 x_2^3 - 126 x_1^5 x_2^4 + 126 x_1^4 x_2^5 \\ -84 x_1^3 x_2^6 + 36 x_1^2 x_2^7 - 9 x_1 x_2^8 + x_2^9 - 18 x_1^5 \sigma_1^2 + 90 x_1^4 x_2 \sigma_1^2 \\ -180 x_1^3 x_2^2 \sigma_1^2 + 180 x_1^2 x_2^3 \sigma_1^2 - 90 x_1 x_2^4 \sigma_1^2 + 18 x_2^5 \sigma_1^2 \\ +135 x_1 \sigma_1^4 - 135 x_2 \sigma_1^4 + 36 x_1^5 \sigma_1 \sigma_2 - 180 x_1^4 x_2 \sigma_1 \sigma_2 \\ +360 x_1^3 x_2^2 \sigma_1 \sigma_2 - 360 x_1^2 x_2^3 \sigma_1 \sigma_2 + 180 x_1 x_2^4 \sigma_1 \sigma_2 \\ -36 x_2^5 \sigma_1 \sigma_2 - 540 x_1 \sigma_1^3 \sigma_2 + 540 x_2 \sigma_1^3 \sigma_2 - 18 x_1^5 \sigma_2^2 \\ +90 x_1^4 x_2 \sigma_2^2 - 180 x_1^3 x_2^2 \sigma_2^2 + 180 x_1^2 x_2^3 \sigma_2^2 - 90 x_1 x_2^4 \sigma_2^2 \\ +18 x_2^5 \sigma_2^2 + 810 x_1 \sigma_1^2 \sigma_2^2 - 810 x_2 \sigma_1^2 \sigma_2^2 - 540 x_1 \sigma_1 \sigma_2^3 \\ +540 x_2 \sigma_1 \sigma_2^3 + 135 x_1 \sigma_2^4 - 135 x_2 \sigma_2^4 \end{pmatrix},$$

que l'on peut simplifier en :

$$J_1 = -p^2 (1-p)^2 (x_1 - x_2) \left((x_1 - x_2)^8 + 18 (x_1 - x_2)^4 (\sigma_1 - \sigma_2)^2 - 135 (\sigma_1 - \sigma_2)^4 \right)$$

et que l'on factorise :

$$\begin{aligned} J_1 &= -p^2 (1-p)^2 (x_1 - x_2) \left((x_1 - x_2)^4 + 3 (3 + 2\sqrt{6}) (\sigma_1 - \sigma_2)^2 \right) \\ &\times \left((x_1 - x_2)^2 + \sqrt{3} \sqrt{2\sqrt{6} - 3} |\sigma_1 - \sigma_2| \right) \\ &\times \left((x_1 - x_2) + (6\sqrt{6} - 9)^{1/4} \sqrt{|\sigma_1 - \sigma_2|} \right) \left((x_1 - x_2) - (6\sqrt{6} - 9)^{1/4} \sqrt{|\sigma_1 - \sigma_2|} \right). \end{aligned}$$

Les cas de nullité de ce déterminant sont, par conséquent :

$$x_1 = x_2 \text{ ou } |x_1 - x_2| = (6\sqrt{6} - 9)^{1/4} \sqrt{|\sigma_1 - \sigma_2|}.$$

Dans le cas où $x_1 = x_2$ on est en présence d'un mélange de deux distributions de même moyenne. On a donc $x_1 = x_2 = 0$ et les troisième et cinquième équations du système des moments, toujours vérifiées, ne sont plus pertinentes pour notre problématique. Il faudrait, dans cette optique, utiliser une autre équation de moments.

Etude du système des cinq premières équations de moments centrés.

Le second cas, qui ne s'explique pas par une détérioration évidente d'une partie du système, n'apparaît qu'à la suite des différentes manipulations algébriques nécessaires à l'inversion de celui-ci.

Nous montrons donc qu'il n'y a pas unicité locale des solutions de \mathfrak{S}_1 dans le voisinage des sous-variétés V_1 et V_2 de $\mathbb{R}^2 \times \mathbb{R}^{+2} \times [0.5; 1[$:

$$V_1 = \{(x_1, x_2, \sigma_1, \sigma_2, p), x_1 = x_2 = 0\}$$

$$\text{et } V_2 = \left\{ (x_1, x_2, \sigma_1, \sigma_2, p), |x_1 - x_2| = (6\sqrt{6} - 9)^{1/4} \sqrt{|\sigma_1 - \sigma_2|} \right\},$$

et il faudra analyser précisément le comportement du système dans les voisinages de ces variétés afin d'approfondir la connaissance que nous avons de l'ensemble des solutions. Pour cela, nous devons inverser le système pour étudier la continuité des solutions sur les variétés V_1 et V_2 .

8.2 Inversion numérique du système \mathfrak{S}_1 par l'algorithme de Newton-Raphson.

Avant d'entrer dans de longs et fastidieux calculs, souvenons nous que la méthode de Newton-Raphson est une méthode numérique reconnue d'inversion de systèmes assurément inversibles. L'exemple suivant illustre les différents types de problèmes que l'on peut rencontrer avec cette méthode, dans le cas du système des moments. Pour cela nous nous placerons dans un voisinage de la variété problématique V_2 .

Calculons les moments à l'aide des valeurs initiales :

$$x_1 = \frac{1}{2} - \frac{1}{\sqrt{2}} (6\sqrt{6} - 9)^{1/4} \approx -0.592434 \text{ à } 10^{-6} \text{ près,}$$

$$x_2 = \frac{1}{\sqrt{2}} (6\sqrt{6} - 9)^{1/4} \approx 1.092434 \text{ à } 10^{-6} \text{ près,}$$

$$\text{donc } p = 0.648379 \text{ à } 10^{-6} \text{ près,}$$

$$\sigma_1 = 3$$

$$\text{et } \sigma_2 = 1.$$

Si les paramètres sont initialisés à :

$$x_1^{ini} = 2,$$

$$x_2^{ini} = -1,$$

$$p^{ini} = 0.5,$$

$$\sigma_1^{ini} = 2.9 \approx M_2$$

$$\text{et } \sigma_2^{ini} = 2.9,$$

il y a convergence vers :

$$x_1^* = 1.092434,$$

$$x_2^* = -0.592434,$$

$$p^* = 0.35162,$$

$$\sigma_1^* = 1$$

Etude du système des cinq premières équations de moments centrés.

$$\text{et } \sigma_2^* = 3,$$

qui est le bon ensemble de paramètres.

Si les paramètres sont initialisés à :

$$\begin{aligned} x_1^{ini} &= 1, \\ x_2^{ini} &= -1, \\ p^{ini} &= 0.5, \\ \sigma_1^{ini} &= 2.9 \\ \text{et } \sigma_2^{ini} &= 2.9, \end{aligned}$$

il y a convergence vers :

$$\begin{aligned} x_1^* &= 0.804433, \\ x_2^* &= -1.45849, \\ p^* &= 0.644516, \\ \sigma_1^* &= 1.4933 \\ \text{et } \sigma_2^* &= 2.27365 \end{aligned}$$

qui est une solution parasite.

Enfin, si les paramètres sont initialisés à :

$$\begin{aligned} x_1^{ini} &= 1, \\ x_2^{ini} &= -1, \\ p^{ini} &= 0.5, \\ \sigma_1^{ini} &= 3 \\ \text{et } \sigma_2^{ini} &= 2.9, \end{aligned}$$

il n'y a pas convergence ...

Si nous excluons les cas où le système n'a pas de solution, il y a, liés à la méthode de Newton-Raphson, trois types de cas problématiques, qui peuvent se produire :

- La procédure peut tout simplement ne pas converger, à cause d'une mauvaise initialisation.
- La Procédure peut ne pas converger, malgré une initialisation très proche des vrais paramètres. Dans ce cas, quelle que soit l'initialisation, on n'obtiendra pas de solution par cette méthode.
- La procédure peut converger vers un ensemble de paramètres acceptables, qui n'est pas celui que nous recherchons.

Il ya trop d'aléas dans l'utilisation d'une telle méthode qui empêchent de faire confiance aux résultats obtenus, lorsqu'on en obtient... Nous nous tournons donc vers une inversion algébrique de \mathfrak{S}_1 , afin d'en comprendre les mécanismes sous-jacents.

8.3 Inversion algébrique du système.

Suite au calcul du jacobien, nous sommes amenés à séparer le cas où les moyennes sont égales de celui où elles sont différentes.

Etude du système des cinq premières équations de moments centrés.

8.3.1 Cas général où $x_1 \neq x_2$.

On considère le mélange de deux distributions normales $N(\mu_1, \sigma_1)$ et $N(\mu_2, \sigma_2)$ dont les moyennes sont différentes. Nous rappelons que le système à inverser, pour un tel mélange, peut-être :

$$\mathfrak{S}_1 = \begin{cases} px_1 + (1-p)x_2 = 0 \\ px_1^2 + (1-p)x_2^2 = M_2^c - (p\sigma_1 + (1-p)\sigma_2) \\ px_1^3 + (1-p)x_2^3 = M_3^c - 3(p\sigma_1x_1 + (1-p)\sigma_2x_2) \\ px_1^4 + (1-p)x_2^4 = M_4^c - 6(p\sigma_1x_1^2 + (1-p)\sigma_2x_2^2) - 3(p\sigma_1^2 + (1-p)\sigma_2^2) \\ px_1^5 + (1-p)x_2^5 = M_5^c - 10(p\sigma_1x_1^3 + (1-p)\sigma_2x_2^3) - 15(p\sigma_1^2x_1 + (1-p)\sigma_2^2x_2) \end{cases} . \quad [8.25]$$

Commençons par poser :

$$\begin{aligned} \sigma &= p\sigma_1 + (1-p)\sigma_2 \\ \text{et } \sigma' &= p\sigma_1x_1 + (1-p)\sigma_2x_2, \end{aligned}$$

ce qui est équivalent, lorsque $x_1 - x_2 \neq 0$, à :

$$\begin{aligned} \sigma_1 &= \frac{\sigma' - \sigma x_2}{p(x_1 - x_2)} \\ \text{et } \sigma_2 &= \frac{\sigma' - \sigma x_1}{(1-p)(x_2 - x_1)}. \end{aligned}$$

Le système après transformation s'écrit :

$$\mathfrak{S}'_1 = \begin{cases} px_1 + (1-p)x_2 = 0 \\ px_1^2 + (1-p)x_2^2 = M_2^c - \sigma \\ px_1^3 + (1-p)x_2^3 = M_3^c - 3\sigma' \\ px_1^4 + (1-p)x_2^4 = M_4^c - 6(\sigma'(x_1 + x_2) - \sigma x_1x_2) - \frac{3((\sigma')^2 + \sigma^2(p x_1^2 + (1-p)x_2^2))}{p(1-p)(x_1 - x_2)^2} \\ px_1^5 + (1-p)x_2^5 = M_5^c - 10(\sigma'((x_1 + x_2)^2 - x_1x_2) - \sigma x_2x_1(x_1 + x_2)) + \frac{15}{x_1x_2}(\sigma'^2(x_1 + x_2) - 2\sigma\sigma'x_1x_2) \end{cases} . \quad [8.26]$$

Exprimons alors, au moyen des trois premières équations, p , x_1 , et x_2 en fonction de σ' et σ . Des deux premières équations on tire :

$$\begin{aligned} x_1 &= \pm \sqrt{\frac{(1-p)}{p}} (M_2^c - \sigma) \\ \text{et } x_2 &= -\text{sgn}(x_1) \cdot \sqrt{\frac{p}{(1-p)}} (M_2^c - \sigma), \end{aligned} \quad [8.27]$$

où $\text{sgn}(x)$ est la fonction qui à x associe son signe.

La troisième équation devient alors :

$$\frac{\pm(1-p)^2 \mp p^2}{(p(1-p)^{\frac{1}{2}})} = \frac{M_3^c - 3\sigma'}{(M_2^c - \sigma)^{\frac{3}{2}}} \quad [8.28]$$

et en élevant au carré les deux membres de l'égalité [8.28] :

Etude du système des cinq premières équations de moments centrés.

$$\frac{(1-2p)^2}{p(1-p)} = \frac{(M_3^c - 3\sigma')^2}{(M_2^c - \sigma)^3} = A,$$

ce qui nous ramène à la résolution de l'équation du second degré :

$$p^2 - p + \frac{1}{4+A} = 0. \quad [8.29]$$

Or A est strictement positif, puisque, d'après les deux premières équations du système,

$$M_2^c - \sigma = p(1-p)(\mu_1 - \mu_2)^2 \geq 0,$$

et que $\mu_1 - \mu_2 = x_1 - x_2 \neq 0$.

Le discriminant de [8.29]

$$\Delta = \frac{A}{4+A}$$

est donc positif aussi et par conséquent, le polynôme de la variable p , [8.29], a deux racines réelles positives p_1 et p_2 , données par, si l'on suppose $p_1 \geq p_2$:

$$\begin{aligned} p_1 &= \frac{\sqrt{4+A} + \sqrt{A}}{2\sqrt{4+A}} \\ \text{et } p_2 &= \frac{\sqrt{4+A} - \sqrt{A}}{2\sqrt{4+A}}. \end{aligned} \quad [8.30]$$

Comme $p_1 = 1 - p_2$, ces deux racines sont les deux coefficients du mélange. Nous pouvons supposer, compte tenu de l'objectif que nous poursuivons, que $p \geq (1-p)$. Ce choix implique alors celui de x_1 et de x_2 sur lesquels nous avons encore un doute [8.27]. En effet, on a dans ce cas :

$$\begin{aligned} p|x_1|^3 &= \left(\frac{\sqrt{4+A} - \sqrt{A}}{2\sqrt{4+A}} \right) \left(\frac{\sqrt{4+A} - \sqrt{A}}{\sqrt{4+A} + \sqrt{A}} \right)^{\frac{1}{2}} (M_2^c - \sigma)^{\frac{3}{2}} \\ \text{et } (1-p)|x_2|^3 &= \left(\frac{\sqrt{4+A} + \sqrt{A}}{2\sqrt{4+A}} \right) \left(\frac{\sqrt{4+A} + \sqrt{A}}{\sqrt{4+A} - \sqrt{A}} \right)^{\frac{1}{2}} (M_2^c - \sigma)^{\frac{3}{2}}, \end{aligned}$$

c'est à dire que $p|x_1|^3 \leq (1-p)|x_2|^3$, ou encore, x_1 et x_2 étant de signes contraires, $px_1^3 + (1-p)x_2^3$ est du signe de x_2 . Or le signe de cette dernière expression est donné par la troisième équation du système, ce qui fixe de manière unique x_1 et x_2 .

Le seul détail à régler est le cas particulier où $M_3^c - 3\sigma' = 0$, qui n'est pas couvert par le raisonnement ci-dessus. Dans cette hypothèse, on a nécessairement, d'après [8.30], $p = \frac{1}{2}$ et donc $x_1 = -x_2$. Nous choisirons, dans ce cas $x_1 < 0$ et $x_2 > 0$, étant donné le rôle symétrique joué par les deux paramètres.

Etude du système des cinq premières équations de moments centrés.

Ainsi, les trois premières équations de moments fixent de manière unique, en fonction des variances, les moyennes des distributions ainsi que le coefficient de mélange :

$$p = \frac{\sqrt{4(M_2^c - \sigma)^3 + (M_3^c - 3\sigma')^2} + |M_3^c - 3\sigma'|}{2\sqrt{4(M_2^c - \sigma)^3 + (M_3^c - 3\sigma')^2}},$$

$$\mu_1 = M_1 - \operatorname{sgn}(M_3^c - 3\sigma') \sqrt{\frac{\sqrt{4(M_2^c - \sigma)^3 + (M_3^c - 3\sigma')^2} - |M_3^c - 3\sigma'|}{\sqrt{4(M_2^c - \sigma)^3 + (M_3^c - 3\sigma')^2} + |M_3^c - 3\sigma'|}} (M_2^c - \sigma)$$

$$\text{et } \mu_2 = M_1 + \operatorname{sgn}(M_3^c - 3\sigma') \cdot \sqrt{\frac{\sqrt{4(M_2^c - \sigma)^3 + (M_3^c - 3\sigma')^2} + |M_3^c - 3\sigma'|}{\sqrt{4(M_2^c - \sigma)^3 + (M_3^c - 3\sigma')^2} - |M_3^c - 3\sigma'|}} (M_2^c - \sigma).$$

Nous pouvons désormais nous attaquer au calcul des variances, qui est plus compliqué. La quatrième équation de \mathfrak{S}'_1 devient :

$$(px_1^2 + (1-p)x_2^2) \left((x_1 + x_2)^2 - 2x_1x_2 \right) - (x_1x_2)^2 = M_4^c - 6(\sigma'(x_1 + x_2) - \sigma x_1x_2) \quad [8.31]$$

$$- \frac{3 \left((\sigma')^2 + \sigma^2 (px_1^2 + (1-p)x_2^2) \right)}{p(1-p)(x_1 - x_2)^2}.$$

D'après ce qui précède, on établit que :

$$x_1 + x_2 = \frac{M_3^c - 3\sigma'}{M_2^c - \sigma},$$

$$x_1 - x_2 = -\operatorname{sgn}(M_3^c - 3\sigma') \frac{\sqrt{4(M_2^c - \sigma)^3 + (M_3^c - 3\sigma')^2}}{M_2^c - \sigma}, \quad [8.32]$$

$$x_1x_2 = \sigma - M_2^c$$

$$\text{et } p(1-p) = \left(\frac{(M_2^c - \sigma)^3}{4(M_2^c - \sigma)^3 + (M_3^c - 3\sigma')^2} \right).$$

Notons que dans la suite du développement le signe de $M_3^c - 3\sigma'$ n'intervient plus et qu'il n'est pas nécessaire de discuter le cas où $M_3^c - 3\sigma' = 0$.

Par substitution, [8.31] devient :

$$M_3^{c2} = 6\sigma'^2 + (-M_2^{c2} - 4\sigma M_2^c + 2\sigma^2 + M_4^c)(M_2^c - \sigma), \quad [8.33]$$

que l'on peut mettre sous la forme :

$$M_3^{c2} = 6\sigma'^2 + \left(2(M_2^c - \sigma)^2 + M_4^c - 3M_2^{c2} \right) (M_2^c - \sigma)$$

et donc :

$$\sigma'^2 = \frac{M_3^{c2} - \left(2(M_2^c - \sigma)^2 + k_4 \right) (M_2^c - \sigma)}{6}, \quad [8.34]$$

Etude du système des cinq premières équations de moments centrés.

où $k_4 = M_4^c - 3M_2^{c2}$ est le quatrième cumulant du mélange.

La cinquième équation de \mathfrak{S}'_1 est :

$$px_1^5 + (1-p)x_2^5 = M_5^c - 10 \left(\sigma' \left((x_1 + x_2)^2 - x_1x_2 \right) - \sigma x_2 x_1 (x_1 + x_2) \right) + \frac{15}{x_1x_2} \left(\sigma'^2 (x_1 + x_2) - 2\sigma\sigma' x_1x_2 \right).$$

Posons :

$$S = M_5^c - 10 \left(\sigma' \left((x_1 + x_2)^2 - x_1x_2 \right) - \sigma x_2 x_1 (x_1 + x_2) \right) + \frac{15}{x_1x_2} \left(\sigma'^2 (x_1 + x_2) - 2\sigma\sigma' x_1x_2 \right)$$

et $S' = px_1^5 + (1-p)x_2^5$.

Si l'on substitue, dans S , les premiers résultats, on obtient :

$$S = M_5^c - 10 \left(\sigma' \left(\left(\frac{M_3^c - 3\sigma'}{M_2^c - \sigma} \right)^2 + M_2^c - \sigma \right) + \sigma (M_3^c - 3\sigma') \right) - 15\sigma'^2 \frac{M_3^c - 3\sigma'}{(M_2^c - \sigma)^2} - 30\sigma\sigma',$$

où encore :

$$\begin{aligned} S(M_2^c - \sigma)^2 &= M_5^c (M_2^c - \sigma)^2 - 10\sigma' (M_3^c - 3\sigma')^2 - 10\sigma' (M_2^c - \sigma)^3 - 10\sigma (M_3^c - 3\sigma') (M_2^c - \sigma)^2 \\ &\quad - 15\sigma'^2 (M_3^c - 3\sigma') - 30\sigma\sigma' (M_2^c - \sigma)^2, \\ &= M_5^c (M_2^c - \sigma)^2 - 10\sigma' M_3^c + 60\sigma'^2 M_3 - 90\sigma'^3 - 10\sigma' (M_2^c - \sigma)^3 - 10\sigma M_3^c (M_2^c - \sigma)^2 \\ &\quad - 15\sigma'^2 M_3 + 45\sigma'^3, \\ &= -10\sigma' M_3^{c2} + 45\sigma'^2 M_3^c - 45\sigma'^3 - 10\sigma' (M_2^c - \sigma)^3 + (M_5^c - 10\sigma M_3^c) (M_2^c - \sigma)^2, \\ &= 45\sigma'^2 M_3^c - \sigma' \left(10 \left((M_2^c - \sigma)^3 + M_3^{c2} \right) + 45\sigma'^2 \right) \\ &\quad + (M_5^c - 10M_3^c M_2^c + 10M_3^c (M_2^c - \sigma)) (M_2^c - \sigma)^2, \end{aligned}$$

ce qui donne, pour finir, si l'on pose $X = M_2^c - \sigma$ et $k_5 = M_5^c - 10M_3^c M_2^c$, le cinquième cumulant du mélange,

$$SX^2 = 45\sigma'^2 M_3^c - \sigma' (10(X^3 + M_3^{c2}) + 45\sigma'^2) + (k_5 + 10M_3^c X) X^2.$$

Calculons S' :

$$\begin{aligned} S' &= (px_1^3 + (1-p)x_2^3) (x_1^2 + x_2^2), \\ &= (px_1^3 + (1-p)x_2^3) \left((x_1 + x_2)^2 - 2x_1x_2 \right), \\ S' &= (M_3^c - 3\sigma') \left[\left(\frac{M_3^c - 3\sigma'}{M_2^c - \sigma} \right)^2 + 2(M_2^c - \sigma) \right], \\ &= \frac{(M_3^c - 3\sigma') \left[M_3^{c2} - 6\sigma' M_3^c + 9\sigma'^2 + 2(M_2^c - \sigma)^3 \right]}{(M_2^c - \sigma)^2}, \\ &= \frac{M_3^{c3} - 6\sigma' M_3^{c2} + 9\sigma'^2 M_3^c + 2(M_2^c - \sigma)^3 M_3^c - 3\sigma' M_3^{c2} + 18\sigma'^2 M_3^c - 27\sigma'^3 - 6\sigma' (M_2^c - \sigma)^3}{(M_2^c - \sigma)^2}, \end{aligned}$$

Etude du système des cinq premières équations de moments centrés.

$$= \frac{M_3^{c3} + 27M_3^c\sigma'^2 + 2M_3^c(M_2^c - \sigma)^3 - 3\sigma' \left(2(M_2^c - \sigma)^3 + 3M_3^{c2} + 9\sigma'^2 \right)}{(M_2^c - \sigma)^2},$$

ou encore :

$$X^2 S' = M_3^{c3} + 27M_3^c\sigma'^2 + 2M_3^c X^3 - 3\sigma' (2X^3 + 3M_3^{c2} + 9\sigma'^2),$$

et donc :

$$X^2 S' = 27M_3^c\sigma'^2 - 3\sigma' (2X^3 + 3M_3^{c2} + 9\sigma'^2) + M_3^{c3} + 2M_3^c X^3.$$

La cinquième équation du système devient :

$$\begin{aligned} 0 &= X^2 (S - S'), \\ 0 &= 18\sigma'^2 M_3^c - \sigma' (4X^3 + M_3^{c2} + 18\sigma'^2) + (k_5 + 8M_3^c X) X^2 - M_3^{c3}, \end{aligned}$$

ou encore :

$$\sigma' = \frac{18\sigma'^2 M_3^c + (k_5 + 8M_3^c X) X^2 - M_3^{c3}}{4X^3 + M_3^{c2} + 18\sigma'^2}.$$

D'après [8.34] on a :

$$\sigma' = \frac{3(M_3^{c2} - (2X^2 + k_4) X) M_3^c + (k_5 + 8M_3^c X) X^2 - M_3^{c3}}{4X^3 + M_3^{c2} + 3(M_3^{c2} - (2X^2 + k_4) X)},$$

c'est à dire :

$$\sigma' = \frac{2M_3^c X^3 + k_5 X^2 - 3k_4 X M_3^c + 2M_3^{c3}}{-2X^3 - 3k_4 X + 4M_3^{c2}}.$$

[8.34] devient alors :

$$6(2M_3^c X^3 + k_5 X^2 - 3k_4 X M_3^c + 2M_3^{c3})^2 = (M_3^{c2} - (2X^2 + k_4) X) (-2X^3 - 3k_4 X + 4M_3^{c2})^2, \quad [8.35]$$

où nous posons afin de rendre les calculs plus clairs :

$$\begin{aligned} S_1^2 &= (2M_3^c X^3 + k_5 X^2 - 3k_4 X M_3^c + 2M_3^{c3})^2 \\ \text{et } S_2^2 &= (M_3^{c2} - (2X^2 + k_4) X) (-2X^3 - 3k_4 X + 4M_3^{c2})^2. \end{aligned}$$

On a pour S_1^2 :

$$\begin{aligned} S_1^2 &= (2M_3^{c3} + k_5 X^2 + (2X^2 - 3k_4) M_3^c X)^2, \\ &= 4M_3^{c6} + k_5^2 X^4 + (2X^2 - 3k_4)^2 M_3^{c2} X^2 + 4M_3^{c3} k_5 X^2 + 4M_3^{c4} (2X^2 - 3k_4) X + 2(2X^2 - 3k_4) k_5 M_3^c X^3, \\ &= 4M_3^{c6} + k_5^2 X^4 + 4M_3^{c2} X^6 + 9k_4^2 M_3^{c2} X^2 - 12k_4 M_3^{c2} X^4 + 4M_3^{c3} k_5 X^2 + 8M_3^{c4} X^3 \end{aligned}$$

Etude du système des cinq premières équations de moments centrés.

$$\begin{aligned}
& -12M_3^{c^4}k_4X + 4k_5M_3^cX^5 - 6k_4k_5M_3^cX^3, \\
= & 4M_3^{c^2}X^6 + 4k_5M_3^cX^5 + (k_5^2 - 12k_4M_3^{c^2})X^4 + (8M_3^{c^3} - 6k_4k_5)M_3^cX^3 \\
& + (9k_4^2 + 4M_3^ck_5)M_3^{c^2}X^2 - 12M_3^{c^4}k_4X + 4M_3^{c^6},
\end{aligned}$$

de même, on calcule :

$$\begin{aligned}
S_2^2 &= (M_3^{c^2} - (2X^2 + k_4)X)(4M_3^{c^2} + (-2X^2 - 3k_4)X)^2, \\
&= (-2X^3 - k_4X + M_3^{c^2})(16M_3^{c^4} + (4X^4 + 9k_4^2 + 12k_4X^2)X^2 + (-16M_3^{c^2}X^3 - 24M_3^{c^2}k_4X)), \\
&= (-2X^3 - k_4X + M_3^{c^2})(4X^6 + 12k_4X^4 - 16M_3^{c^2}X^3 + 9k_4^2X^2 - 24M_3^{c^2}k_4X + 16M_3^{c^4}), \\
&= -8X^9 - 24k_4X^7 + 32M_3^{c^2}X^6 - 18k_4^2X^5 + 48M_3^{c^2}k_4X^4 - 32M_3^{c^4}X^3 \\
&\quad - 4k_4X^7 - 12k_4^2X^5 + 16M_3^{c^2}k_4X^4 - 9k_4^3X^3 + 24M_3^{c^2}k_4^2X^2 - 16k_4M_3^{c^4}X \\
&\quad + 4M_3^{c^2}X^6 + 12M_3^{c^2}k_4X^4 - 16M_3^{c^4}X^3 + 9M_3^{c^2}k_4^2X^2 - 24M_3^{c^4}k_4X + 16M_3^{c^6}
\end{aligned}$$

et donc

$$\begin{aligned}
S_2^2 &= -8X^9 - 28k_4X^7 + 36M_3^{c^2}X^6 - 30k_4^2X^5 + 76M_3^{c^2}k_4X^4 - (48M_3^{c^4} + 9k_4^3)X^3 \\
&\quad + 33M_3^{c^2}k_4^2X^2 - 40M_3^{c^4}k_4X + 16M_3^{c^6}.
\end{aligned}$$

D'après [8.35], on tire finalement :

$$\begin{aligned}
S_2^2 - 6S_1^2 &= -8X^9 - 28k_4X^7 + 36M_3^{c^2}X^6 - 30k_4^2X^5 + 76M_3^{c^2}k_4X^4 - (48M_3^{c^4} + 9k_4^3)X^3 \\
&\quad + 33M_3^{c^2}k_4^2X^2 - 40M_3^{c^4}k_4X + 16M_3^{c^6} - 24M_3^{c^2}X^6 - 24k_5M_3^cX^5 - (6k_5^2 - 72k_4M_3^{c^2})X^4 \\
&\quad - (48M_3^{c^3} - 36k_4k_5)M_3^cX^3 - (54k_4^2 + 24M_3^ck_5)M_3^{c^2}X^2 + 72M_3^{c^4}k_4X - 24M_3^{c^6},
\end{aligned}$$

c'est à dire :

$$\begin{aligned}
S_2^2 - 6S_1^2 &= -8X^9 - 28k_4X^7 + 12M_3^{c^2}X^6 - (30k_4^2 + 24k_5M_3^c)X^5 + (-6k_5^2 + 148k_4M_3^{c^2})X^4 \\
&\quad + (-96M_3^{c^4} + 36k_4k_5M_3^c - 9k_4^3)X^3 + (-21k_4^2 - 24M_3^ck_5)M_3^{c^2}X^2 + 32M_3^{c^4}k_4X - 8M_3^{c^6}.
\end{aligned}$$

La résolution du système se ramène donc à la recherche des racines strictement positives (car $X = M_2^c - \sigma \geq 0$ avec $X = 0$ si et seulement si $x_1 = x_2 = 0$) du polynôme de degré 9 :

$$\begin{aligned}
P_\sigma(X) &= 8X^9 + 28k_4X^7 - 12M_3^{c^2}X^6 + (30k_4^2 + 24k_5M_3^c)X^5 + (6k_5^2 - 148k_4M_3^{c^2})X^4 \\
&\quad - (-96M_3^{c^4} + 36k_4k_5M_3^c - 9k_4^3)X^3 + (21k_4^2 + 24M_3^ck_5)M_3^{c^2}X^2 - 32M_3^{c^4}k_4X + 8M_3^{c^6}.
\end{aligned}$$

Comme $P_\sigma(0) = 8M_3^{c^6} > 0$, si $M_3^c \neq 0$, et que $\lim_{X \rightarrow +\infty} P_\sigma(X) = +\infty$, si le polynôme admet une racine positive, il en admettra, d'après le théorème des valeurs intermédiaires, au moins une deuxième (sauf dans les cas où ce sera une racine double). Il y a donc, en général, plusieurs racines positives, mais, le plus souvent, une seule d'entre elle permettra d'obtenir des paramètres ayant une existence statistique ($p \in]0; 1[$ et σ_1 et σ_2 positifs). Dans le cas où il y aurait plusieurs ensembles de solutions ayant une existence statistique, il est intéressant de vérifier le degré de conformité des paramètres avec la sixième équation de moments, au moyen d'un test à déterminer.

Etude du système des cinq premières équations de moments centrés.

Signalons, aussi, que dans le cas où $M_3^c = 0$,

$$P_\sigma(X) = 8X^9 + 28k_4X^7 + 30k_4^2X^5 + 6k_5^2X^4 + 9k_4^3X^3$$

et P_σ n'a pas de racine strictement positive si $k_4 \geq 0$, ce qui correspond, d'après [Cohen67], au cas où les moyennes sont égales.

8.3.2 Cas où les variances sont les mêmes.

Il n'est pas nécessaire, à en juger par le calcul du jacobien, de distinguer ce cas. Cependant, les calculs sont beaucoup plus simples et les variances seront réellement égales dans les applications, alors qu'elles varient légèrement si l'on reste dans le cas général. Enfin, on s'affranchit, de l'interrogation sur le signe de σ' .

Si l'on travaille, donc, sur un mélange de deux normales dont les variances sont les mêmes, $\sigma' = 0$ et le système [8.25] devient :

$$\mathfrak{S}_{1 \cap 2} = \begin{cases} px_1 + (1-p)x_2 = 0 \\ px_1^2 + (1-p)x_2^2 = M_2^c - \sigma \\ px_1^3 + (1-p)x_2^3 = M_3^c \\ px_1^4 + (1-p)x_2^4 = M_4^c - 6\sigma(px_1^2 + (1-p)x_2^2) - 3\sigma^2 \end{cases} \quad [8.36]$$

ou encore

$$\mathfrak{S}_{1 \cap 2} = \begin{cases} px_1 + (1-p)x_2 = 0 \\ px_1^2 + (1-p)x_2^2 = M_2^c - \sigma \\ px_1^3 + (1-p)x_2^3 = M_3^c \\ px_1^4 + (1-p)x_2^4 = M_4^c - 6\sigma(M_2^c - \sigma) - 3\sigma^2 \end{cases} \quad [8.37]$$

Des trois premières équations on tire, comme précédemment :

$$p = \frac{\sqrt{4(M_2^c - \sigma)^3 + M_3^{c2}} + |M_3^c|}{2\sqrt{4(M_2^c - \sigma)^3 + M_3^{c2}}},$$

$$\mu_1 = M_1 - \operatorname{sgn}(M_3^c) \sqrt{\frac{\sqrt{4(M_2^c - \sigma)^3 + M_3^{c2}} - |M_3^c|}{\sqrt{4(M_2^c - \sigma)^3 + M_3^{c2}} + |M_3^c|}} (M_2^c - \sigma)$$

$$\text{et } \mu_2 = M_1 + \operatorname{sgn}(M_3^c) \cdot \sqrt{\frac{\sqrt{4(M_2^c - \sigma)^3 + M_3^{c2}} + |M_3^c|}{\sqrt{4(M_2^c - \sigma)^3 + M_3^{c2}} - |M_3^c|}} (M_2^c - \sigma).$$

Si $M_3^c = 0$, on a $p = \frac{1}{2}$ les paramètres μ_1 et μ_2 ont des rôles symétriques et l'on choisira, par exemple, $\mu_1 < \mu_2$. La quatrième équation devient :

$$M_3^{c2} = 2(M_2^c - \sigma)^3 + k_4(M_2^c - \sigma).$$

L'inversion du système, dans ce cas particulier, se ramène à la recherche des racines positives du polynôme :

$$P_{\text{varégales}}(X) = 2X^3 + k_4X - M_3^{c2}.$$

Etude du système des cinq premières équations de moments centrés.

L'étude de la dérivée du polynôme et le fait que $P_{varégales}(0) \leq 0$ montrent qu'il n'y a qu'une seule racine positive (ou nulle) pour ce polynôme.

8.3.3 Discussion dans le cas : $|x_1 - x_2| = (6\sqrt{6} - 9)^{1/4} \sqrt{|\sigma_1 - \sigma_2|}$.

Il est intéressant de remarquer qu'aucun problème n'apparaît dans ce cas, lors de la résolution du système, ce qui explique pourquoi Pearson et al. ne le mentionnent pas. En effet, si le fait de prendre $x_1 \neq x_2$ autorise la division par $(x_1 - x_2)$, qui est nécessaire à l'inversion du système par une méthode de substitution, il n'y a pas de nécessité similaire faisant intervenir la variété V_2 . La solution doit donc être donnée par l'une des racines du polynome P_σ .

Restriction de la problématique.

Ce cas regroupe aussi bien des mélanges unimodaux que bimodaux et l'explication du phénomène n'est pas à chercher dans cette voie. La figure FIG. II.2 montre les deux cas.

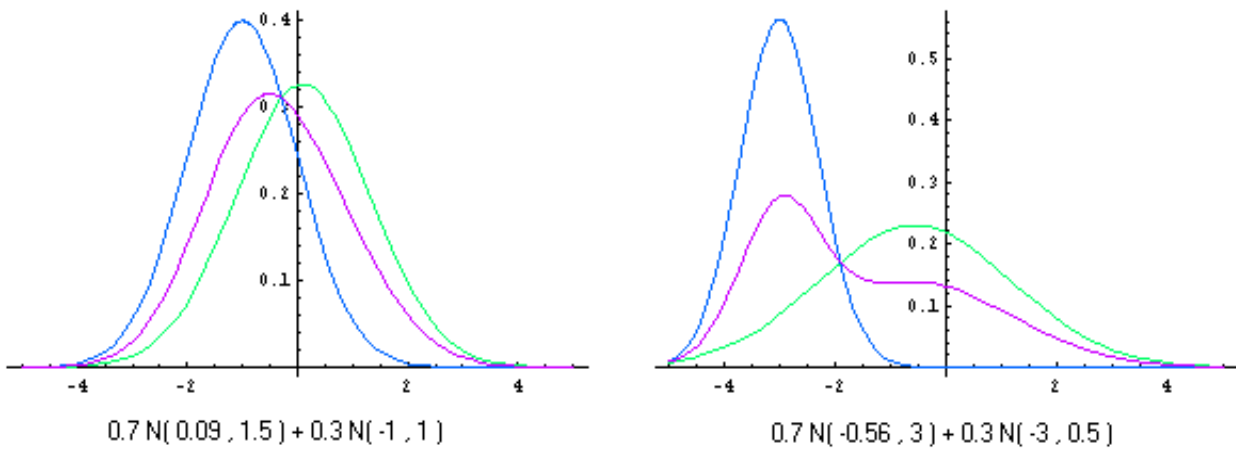


FIG. II.2. - MÉLANGE UNIMODAL ET MÉLANGE BIMODAL POUR V_2 (EN MAUVE).

Commençons par remarquer que les quadruplets $(x_1, x_2, \sigma_1, \sigma_2)$ vérifiant

$$|x_1 - x_2| = (6\sqrt{6} - 9)^{1/4} \sqrt{|\sigma_1 - \sigma_2|},$$

ne sont pas tous des quadruplets de paramètres. En effet, x_1 et x_2 étant de signes contraires,

$$|x_1 - x_2| = |x_1| + |x_2|.$$

Comme $|x_1| > 0$ et $|x_2| > 0$, on aura nécessairement :

$$|x_1| < (6\sqrt{6} - 9)^{1/4} \sqrt{|\sigma_1 - \sigma_2|} \text{ et } |x_2| < (6\sqrt{6} - 9)^{1/4} \sqrt{|\sigma_1 - \sigma_2|}.$$

Etude du système des cinq premières équations de moments centrés.

Racine double.

Il s'avère que si l'on se place sur la variété V_2 , le polynôme P_σ admet une racine double positive, qui est celle permettant d'estimer les paramètres. Commençons par exprimer, grâce aux équations du système, k_4 , k_5 et M_3^c en fonction de x_1 , x_2 , σ_1 et σ_2 . En fait, P_σ s'exprime en fonction de $x_1 + x_2$, $x_1 - x_2$ et de la différence $\sigma_1 - \sigma_2$. Posons donc :

$$\begin{aligned} S &= x_1 + x_2, \\ D &= x_1 - x_2 \\ \text{et } Z &= \sigma_1 - \sigma_2. \end{aligned}$$

Il vient alors :

$$\begin{aligned} M_3^c &= \frac{(D-S)(D+S)(DS+3Z)}{4D}, \\ k_4 &= -\frac{(D-S)(D+S)(D^4-3(D^2S^2+4DSZ+2Z^2))}{8D^2} \\ \text{et } k_5 &= -\frac{(D-S)(D+S)(2D^4S-3D^2S^3+5D^3Z-15DS^2Z-15SZ^2)}{4D^2} \end{aligned}$$

et

$$P_\sigma(X) = \frac{D^2 - S^2 - 4X}{1024D^6} \left[\begin{array}{l} -2048 D^6 X^8 \\ -512 D^6 (D^2 - S^2) X^7 \\ +256 D^4 (D^2 - S^2) (3 D^4 - 10 D^2 S^2 - 42 D S Z - 21 Z^2) X^6 \\ +64 D^4 (D^2 - S^2)^2 (3 D^4 - 7 D^2 S^2 - 24 D S Z + 6 Z^2) X^5 \\ -8D^2 (D^2 - S^2)^2 \left[\begin{array}{l} D^4 (9D^4 - 166D^2S^2 + 265S^4) - 168D^3 (5D^2S - 13S^3) Z \\ -24D^2 (38D^2 - 233S^2) Z^2 + 4320DSZ^3 + 540Z^4 \end{array} \right] X^4 \\ -2D^2 (D^2 - S^2)^2 \left[\begin{array}{l} D^4 (9D^6 + 165D^4S^2 - 737D^2S^4 + 611S^6) \\ +48D^3S (21D^4 - 138D^2S^2 + 137S^4) Z \\ +12D^2 (135D^4 - 1704D^2S^2 + 2089S^4) Z^2 \\ -576DS (42D^2 - 67S^2) Z^3 + 108 (-69D^2 + 169S^2) Z^4 \end{array} \right] X^3 \\ +(D^2 - S^2)^3 (D S + 3 Z)^2 \left[\begin{array}{l} D^4 (21D^4 - 254D^2S^2 + 317S^4) \\ +24D^3S (-49D^2 + 95 S^2) Z \\ +12D^2 (-93 D^2 + 379S^2) Z^2 + 1584DSZ^3 - 108Z^4 \end{array} \right] X^2 \\ +8 (D^2 - S^2)^4 (DS + 3Z)^4 (2D^4 - 5D^2S^2 - 18DSZ - 3Z^2) X \\ +2(D^2 - S^2)^5 (DS + 3Z)^6 \end{array} \right].$$

On sait alors, d'après [8.32], que

$$-x_1x_2 = \frac{D^2 - S^2}{4}$$

doit être racine du polynôme P_σ , ce qui est confirmé par la factorisation du terme

$$D^2 - S^2 - 4X$$

dans l'écriture du polynôme.

Pour savoir quand le polynôme admet une racine double en $-x_1x_2$, il faut examiner le reste de la division polynomiale de $P_\sigma(X)$ par

$$(X + x_1x_2)^2 = \frac{1}{16} (4X - D^2 + S^2)^2,$$

et voir quand il s'annule. Ce reste est :

$$r = -\frac{(D - S)^5 (D + S)^5 (D^2 - S^2 - 4 X)}{4096 D^6} \underbrace{(D^4 + 6 D S Z + 27 Z^2)}_{F_1} \underbrace{(D^8 + 18 D^4 Z^2 - 135 Z^4)}_{F_2}.$$

Le premier facteur (F_1) de ce produit ne s'annule jamais. Le deuxième facteur (F_2) apparaît aussi lors de la factorisation du jacobien. Il correspond au cas que nous étudions et nous permet de confirmer le fait que, si l'on se place sur la variété V_2 , le reste r s'annule et la racine de P_σ , permettant d'estimer les paramètres, est double. Il apparaît aussi que c'est le seul endroit où cela peut se produire.

Représentation graphique.

Pour un Z fixé, il est possible de représenter graphiquement l'intersection entre P_σ et la variété annulant le jacobien. La représentation graphique résumera les différents développements que nous venons de faire. Supposons pour cela, sans perte de généralité, que $x_1 > 0$ et donc, sur V_2 ,

$$x_1 = (6\sqrt{6} - 9)^{1/4} \sqrt{|Z|} + x_2.$$

On injecte ce résultat dans l'expression de P_σ utilisée ci-dessus, et pour un Z_0 fixé, on fait varier X sur les valeurs positives et x_2 entre 0 et $-(6\sqrt{6} - 9)^{1/4} \sqrt{|Z_0|}$. La figure FIG. II.3 montre l'allure de la surface pour $Z = 1$ (les polynômes sont parallèles aux lignes jaunes), et la figure FIG. II.4 souligne l'endroit des racines doubles (en rouge).

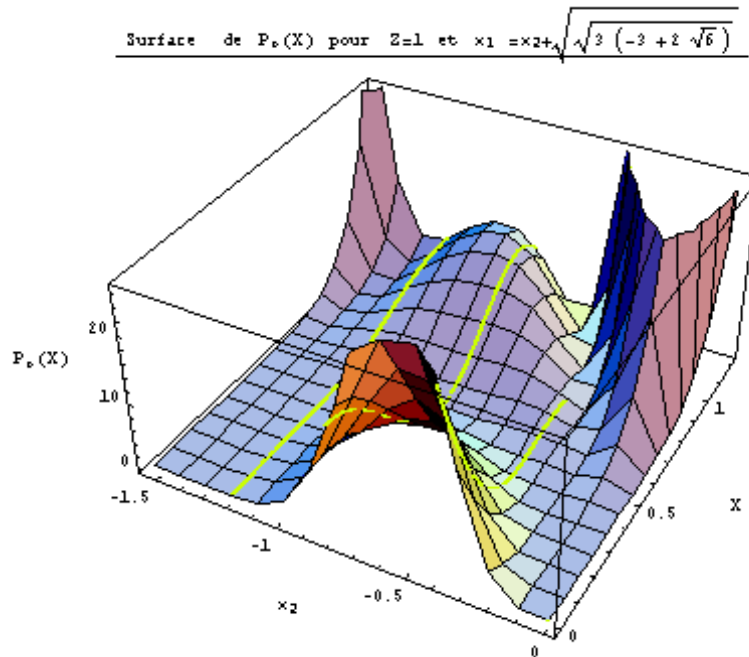


FIG. II.3. -

Comme on peut le constater, dans les deux exemples qui suivent, si l'on se place sur la variété où le jacobien s'annule, l'estimation, par le polynôme P_σ , est adaptée.

Etude du système des cinq premières équations de moments centrés.

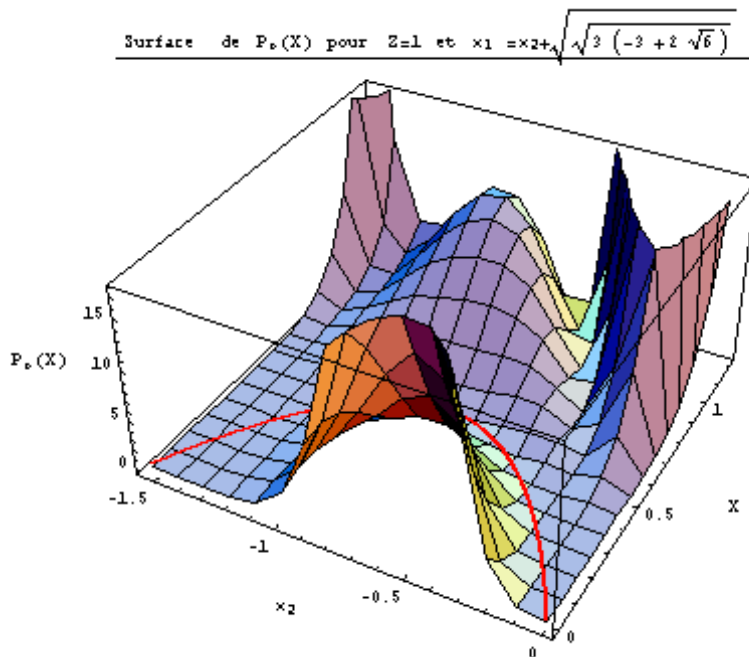


FIG. II.4. -

8.3.4 Conclusion.

Grâce au théorème des fonctions implicites nous avons montré l'existence et l'unicité locales des solutions, en dehors des variétés V_1 et V_2 .

Nous avons établi d'une part que le problème n'a pas de solution pour la variété V_1 , puisqu'il n'y a pas assez d'équations pour le nombre d'inconnues. Notons qu'une étude numérique du comportement du système pour x_1 et x_2 très proches, permettra d'observer d'éventuelles dégénérescences des résultats.

D'autre part, nous avons identifié les effets des paramètres, lorsqu'ils appartiennent à la variété V_2 , sur le comportement du polynôme, dont la racine permettant d'inverser le système est double. Il n'y a donc, pas d'impossibilité à inverser le système sur V_2 . Notons cependant, qu'un ensemble de moments empiriques entraînant un "décalage" du polynôme, peut conduire à deux racines très proches générant deux jeux de paramètres presque identiques ou même à la disparition de la racine, ce qui interdirait l'inversion du système

Les exemples suivants illustrent la fiabilité de la méthode dans l'inversion théorique des cas favorables ainsi que dans les cas problématiques.

8.4 Exemples d'inversion du système.

Dans les exemples suivants, les calculs sont effectués avec une précision de 1000 chiffres significatifs et les racines des différents polynômes sont calculées avec une précision de 50000 chiffres significatifs.

Etude du système des cinq premières équations de moments centrés.

Le test que nous présenterons avec la sixième équation est le rapport :

$$T_6 = \left| \frac{p(15\sigma_1^3 + 45\sigma_1^2 x_1^2 + 15\sigma_1 x_1^4 + x_1^6) + (1-p)(15\sigma_2 x_2^4 + 45\sigma_2^2 x_2^2 + 15\sigma_2^3 + x_2^6) - M_6^c}{2M_6^c} \right|,$$

où le numérateur est calculé avec les paramètres obtenus par inversion du système et le dénominateur est calculé avec les paramètres connus initialement. T_6 doit, donc, être "proche" de 0 (égal à 0 dans les études théoriques), et s'il y a plusieurs ensembles de solutions acceptables, issus de l'inversion du système, nous choisirons celui qui donnera le rapport le plus proche de 0. De même, nous pourrions utiliser, pour évaluer les éventuels candidats solution obtenus avec $P_{varégales}$, l'équivalent T_5 de T_6 avec la cinquième équation,

$$T_5 = \left| \frac{p(x_1^5 + 10\sigma_1 x_1^3 + 15\sigma_1^2 x_1) + (1-p)(x_2^5 + 10\sigma_2 x_2^3 + 15\sigma_2^2 x_2) - M_5^c}{2M_5^c} \right|.$$

Les commentaires que nous faisons dans les paragraphes suivants se rapportent aux différents exemples proposés en annexe, et dont nous présentons les tableaux récapitulatifs.

8.4.1 Cas général (éloigné des pôles du jacobien).

Nous faisons référence à l'annexe 1. Le polynôme admet deux racines positives R_1 et R_2 , dont seule la seconde donne un ensemble de paramètres qui soit une solution appartenant au domaine de définition des distributions normales. Le test T_6 est nul pour cet ensemble de solutions, qui est effectivement celui que nous espérons.

8.4.2 Cas où $\sigma_1 = \sigma_2$ (éloigné des pôles du jacobien).

Si $p \neq 0.5$.

Nous faisons référence à l'annexe 2. Le polynôme P_σ a deux racines positives, R_1 et R_2 , ce qui offre deux possibilités d'ensembles solutions au problème des moments. Le système \mathfrak{S}_1 inversé par R_1 donne un ensemble n'ayant pas d'existence dans l'ensemble des paramètres. L'inversion par R_2 donne un ensemble qui est celui que nous espérons en inversant notre système par le polynôme P_σ .

En inversant le système $\mathfrak{S}_{1 \cap 2}$ par l'unique racine positive de $P_{varégales}$ on obtient, aussi, l'inversion exacte du système. Les deux méthodes conviennent à l'inversion du système.

Si $p = 0.5$.

Nous faisons référence à l'annexe 3. Nous nous plaçons, ici, dans le cas d'une distribution de mélange symétrique. C'est le cas particulier où M_3^c et σ' sont simultanément nuls. Moyennant quelques précautions dans l'utilisation de la méthode par P_σ , les résultats sont ceux auxquels nous pouvons nous attendre. En effet, le polynôme a deux racines positives, dont une est une racine double, qui donnent toutes les deux un ensemble de paramètres acceptables pour un mélange de distributions normales. Le test T_6 permet toutefois de trancher en faveur de l'ensemble des paramètres théoriques. La différence entre les deux ensembles de paramètres ne peut se faire que sur le moment d'ordre 6. Il faut, de plus, remarquer que le test donne, pour l'ensemble issu de la racine double du polynôme, un résultat suffisamment proche de zéro, pour qu'il puisse générer des sources d'erreur dans les études empiriques.

La racine positive de $P_{varégales}$ engendre l'ensemble de paramètres attendu qui est, évidemment, en accord avec les deux tests.

Etude du système des cinq premières équations de moments centrés.

Les deux méthodes conviennent à la recherche des solutions du problème des moments et il n'est pas nécessaire, théoriquement, de distinguer les méthodes. Il sera, cependant, utile de vérifier que l'inversion par les racines de P_σ n'est pas source d'erreur lors de la phase "appliquée" d'estimation des paramètres d'un échantillon de mélange.

8.4.3 Cas où les variables appartiennent à la variété V_2 .

La seule précaution est de choisir x_1 et x_2 de signes opposés et appartenant au domaine de restriction décrit plus haut. Comme le montrait l'étude théorique, la racine permettant d'inverser le système, dans le cas présenté en annexe 4, est double. L'inversion donne deux ensembles de paramètres identiques, qui sont ceux recherchés.

Si l'on se place au voisinage de cette variété, et nous faisons, alors, référence à l'annexe 5, on obtient deux racines positives, certes distinctes, mais toutefois très proches, et qui engendrent deux ensembles de paramètres acceptables et quasiment équivalents. L'inversion théorique de \mathfrak{S}_1 permet d'aboutir à une différence au niveau du test, mais la valeur du test de l'ensemble rejeté est pourtant de l'ordre de 7×10^{-7} . Nous confirmons, ainsi, la possibilité de défaillance de l'inversion du système dans le cadre d'un échantillon.

Le problème apparu lors du calcul du jacobien n'empêche pas l'inversion du système pour des paramètres appartenant à la variété V_2 , mais laisse présager des difficultés lors du passage à une utilisation de la méthode pour l'estimation des paramètres.

8.4.4 Cas où x_1 et x_2 sont proches (voisinage de V_1).

Le cas où les moyennes sont égales est résolu par la méthode de Pearson. Le comportement au voisinage de V_1 , qui n'est pas couvert par ce sous exemple particulier, mérite d'être abordé ici.

Moyennes proches variances éloignées.

Nous faisons référence à l'annexe 6. Même si les racines de P_σ sont très rapprochées, une seule génère une solution statistiquement acceptable au problème des moments.

Moyennes et variances proches.

Dans le cas présenté en annexe 7, les deux distributions composant le mélange sont très proches l'une de l'autre. Le fait de travailler avec les moments théoriques permet d'isoler clairement l'une des deux solutions statistiquement acceptables après inversion. Cependant, comme nous l'avons remarqué dans d'autres cas, l'utilisation de cette méthode risque d'être problématique dans une démarche d'estimation.

8.4.5 Conclusion.

Dans le cadre d'un mélange de deux distributions normales, l'inversion du système des cinq premières équations de moments a des lacunes. Il est impossible d'avoir recours à une méthode de type Newton-Raphson, à cause des effets indésirables liés à l'initialisation. L'utilisation des polynômes P_σ (ou $P_{Pearson}$) est à faire avec précaution, et nécessite l'appui de la sixième équation de moments pour éliminer certaines solutions du système.

Comme nous sommes amenés à avoir recours à la sixième équation de moments, que ce soit dans le cas où les moyennes sont les mêmes (exemple de Pearson), ou pour choisir l'une des solutions quand plusieurs sont valables, nous proposons de remplacer l'une des équations du système [8.25] par cette sixième équation. Etant donnée la méthode utilisée pour inverser le système \mathfrak{S}_1 , si nous remplaçons la cinquième équation, les quatre premières étapes de la résolution seront communes, et il suffira de substituer les résultats obtenus, lors de ces étapes, dans la sixième équation. Ce choix permettra notamment de supprimer le problème lié à l'égalité des moyennes (variété V_1).

Cette idée est développée dans le chapitre suivant.

INVERSION DE S1 POUR DES MOMENTS THEORIQUES

	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2		racines >0	acceptable	Test T6	Commentaires
Annexe 1	5.0	3.0	0.6	1.0	1.0		0.54761436	non		
	5.0	3.0	0.6	1.0	1.0		3.84	oui	0.0	
Annexe 2	1.0	1.0	0.7	3.0	1.0		0.26803419	non		
	1.0	1.0	0.7	3.0	1.0		0.84	oui	0.0	
	1.0	1.0	0.7	3.0	1.0		0.84	oui	0.0	On a aussi T5=0.0
Annexe 3	1.0	1.0	0.5	2.0	1.0		0.25	oui	0.0	
	1.0	1.0	0.5	2.0	1.0		0.43301270	oui	0.066876	Cette racine double n'est pas celle qui permet l'inversion.
	1.0	1.0	0.5	2.0	1.0		0.43301270	oui	0.066876	T5 n'est pas calculable
Annexe 4	2.092434442	3.0	0.5	-0.092434442	1.0		1.19341301	oui	0.0	Cette racine double caractérise V2.
	2.092434442	3.0	0.5	-0.092434442	1.0		1.19341301	oui	0.0	
Annexe 5	2.092444442	3.0	0.5	-0.092444442	1.0		1.19341522	oui	7.36E-7	Les deux racines sont proches
	2.092444442	3.0	0.5	-0.092444442	1.0		1.19343486	oui	0.0	Les deux tests aussi.
Annexe 6	1.001	5.0	0.6	1.0	1.0		2.4E-7	oui	0.0	
	1.001	5.0	0.6	1.0	1.0		5.57893E-7	non		
Annexe 7	1.0001	1.0001	0.6	1.0	1.0		2.4E-9	oui	0.0	
	1.0001	1.0001	0.6	1.0	1.0		5.57898E-9	oui	6.12E-10	

Légende :


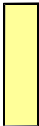


Paramètres théoriques :		Racine choisie pour l'inversion :	
Paramètres calculés :		Paramètres issus de Pyrégalés :	

FIG. II.5. -

Etude du système formé des quatre premières et de la sixième équations de moments centrés.

- CHAPITRE 9 -

Etude du système formé des quatre premières et de la sixième equations de moments centrés.

Nous nous proposons dans cette partie de remédier aux problèmes de l'utilisation des polynômes $P_{Pearson}$ et P_σ dans le voisinage des variétés annulant le jacobien J_1 , voisinage dans lequel nous sommes confrontés à des problèmes d'unicité et même d'existence des solutions. Pour cela, considérons le système :

$$\mathfrak{S}_2 = \begin{cases} 0 = px_1 + (1-p)x_2 \\ M_2^c = px_1^2 + (1-p)x_2^2 + (p\sigma_1 + (1-p)\sigma_2) \\ M_3^c = px_1^3 + (1-p)x_2^3 + 3(p\sigma_1x_1 + (1-p)\sigma_2x_2) \\ M_4^c = px_1^4 + (1-p)x_2^4 + 6(p\sigma_1x_1^2 + (1-p)\sigma_2x_2^2) + 3(p\sigma_1^2 + (1-p)\sigma_2^2) \\ M_6^c = p x_1^6 + (1-p) x_2^6 + 15 (p \sigma_1 x_1^4 + (1-p) \sigma_2 x_2^4) + 45 (p \sigma_1^2 x_1^2 + (1-p) \sigma_2^2 x_2^2) \\ + 15 (p\sigma_1^3 + (1-p)\sigma_2^3) \end{cases}, \quad [9.38]$$

où la cinquième équation de \mathfrak{S}_1 a été remplacée par la sixième équation des moments centrés.

Afin de montrer que ce système permet de pallier aux problèmes soulevés par [8.25] étudions son jacobien.

9.1 Existence et unicité des solutions.

Le jacobien J_2 associé au système \mathfrak{S}_2 est :

$$J_2 = \begin{bmatrix} p & 1-p & x_1-x_2 & 0 & 0 \\ 2px_1 & 2(1-p)x_2 & x_1^2+\sigma_1-x_2^2-\sigma_2 & p & 1-p \\ 3p(x_1^2+\sigma_1) & 3(1-p)(x_2^2+\sigma_2) & x_1^3+3\sigma_1x_1-x_2^3-3\sigma_2x_2 & 3px_1 & 3(1-p)x_2 \\ 4p(x_1^3+3\sigma_1x_1) & 4(1-p)(x_2^3+3\sigma_2x_2) & x_1^4+6\sigma_1x_1^2+3\sigma_1^2-x_2^4-6\sigma_2x_2^2-3\sigma_2^2 & 6p(x_1^2+\sigma_1) & 6(1-p)(x_2^2+\sigma_2) \\ 6p \begin{pmatrix} x_1^5+10\sigma_1x_1^3 \\ +15\sigma_1^2x_1 \end{pmatrix} & 6(1-p) \begin{pmatrix} x_2^5+10\sigma_2x_2^3 \\ +15\sigma_2^2x_2 \end{pmatrix} & x_1^6+15\sigma_1x_1^4+45\sigma_1^2x_1^2+\sigma_1^3-x_2^6-15\sigma_2x_2^4-45\sigma_2^2x_2^2-\sigma_2^3 & 15p(x_1^4+6\sigma_1x_1^2+3\sigma_1^2) & 15(1-p)(x_2^4+6\sigma_2x_2^2+3\sigma_2^2) \end{bmatrix}$$

Etude du système formé des quatre premières et de la sixième equations de moments centrés.

après avoir factorisé p dans les première et quatrième colonnes, et $(1-p)$ dans les deuxième et cinquième colonnes, le calcul du jacobien donne :

$$J_2 = 3(-1+p)^2 p^2 \begin{bmatrix} x_1^{10} - 8x_1^9 x_2 + 27x_1^8 x_2^2 - 48x_1^7 x_2^3 + 42x_1^6 x_2^4 - 42x_1^4 x_2^6 + 48x_1^3 x_2^7 - 27x_1^2 x_2^8 + 8x_1 x_2^9 - x_2^{10} \\ + 3x_1^8 \sigma_1 - 24x_1^7 x_2 \sigma_1 + 84x_1^6 x_2^2 \sigma_1 - 168x_1^5 x_2^3 \sigma_1 + 210x_1^4 x_2^4 \sigma_1 - 168x_1^3 x_2^5 \sigma_1 + 84x_1^2 x_2^6 \sigma_1 \\ - 24x_1 x_2^7 \sigma_1 + 3x_2^8 \sigma_1 + 18x_1^6 \sigma_1^2 - 72x_1^5 x_2 \sigma_1^2 + 90x_1^4 x_2^2 \sigma_1^2 - 90x_1^3 x_2^3 \sigma_1^2 + 72x_1^2 x_2^4 \sigma_1^2 \\ - 18x_1 x_2^5 \sigma_1^2 + 30x_1^4 \sigma_1^3 - 120x_1^3 x_2 \sigma_1^3 + 180x_1^2 x_2^2 \sigma_1^3 - 120x_1 x_2^3 \sigma_1^3 + 30x_2^4 \sigma_1^3 - 135x_1^2 \sigma_1^4 \\ + 135x_2^2 \sigma_1^4 - 45\sigma_1^5 - 3x_1^8 \sigma_2 + 24x_1^7 x_2 \sigma_2 - 84x_1^6 x_2^2 \sigma_2 + 168x_1^5 x_2^3 \sigma_2 - 210x_1^4 x_2^4 \sigma_2 \\ + 168x_1^3 x_2^5 \sigma_2 - 84x_1^2 x_2^6 \sigma_2 + 24x_1 x_2^7 \sigma_2 - 3x_2^8 \sigma_2 - 36x_1^6 \sigma_1 \sigma_2 + 144x_1^5 x_2 \sigma_1 \sigma_2 \\ - 180x_1^4 x_2^2 \sigma_1 \sigma_2 + 180x_1^3 x_2^3 \sigma_1 \sigma_2 - 144x_1^2 x_2^4 \sigma_1 \sigma_2 + 36x_2^6 \sigma_1 \sigma_2 - 90x_1^4 \sigma_1^2 \sigma_2 + 360x_1^3 x_2 \sigma_1^2 \sigma_2 \\ - 540x_1^2 x_2^2 \sigma_1^2 \sigma_2 + 360x_1 x_2^3 \sigma_1^2 \sigma_2 - 90x_2^4 \sigma_1^2 \sigma_2 + 540x_1^2 \sigma_1^3 \sigma_2 - 540x_2^3 \sigma_1^3 \sigma_2 + 225\sigma_1^4 \sigma_2 \\ + 18x_1^6 \sigma_2^2 - 72x_1^5 x_2 \sigma_2^2 + 90x_1^4 x_2^2 \sigma_2^2 - 90x_1^3 x_2^3 \sigma_2^2 + 72x_1^2 x_2^4 \sigma_2^2 - 18x_1 x_2^5 \sigma_2^2 + 90x_1^4 \sigma_1 \sigma_2^2 \\ - 360x_1^3 x_2 \sigma_1 \sigma_2^2 + 540x_1^2 x_2^2 \sigma_1 \sigma_2^2 - 360x_1 x_2^3 \sigma_1 \sigma_2^2 + 90x_2^4 \sigma_1 \sigma_2^2 - 810x_1^2 \sigma_1^2 \sigma_2^2 + 810x_2^3 \sigma_1^2 \sigma_2^2 \\ - 450\sigma_1^3 \sigma_2^2 - 30x_1^4 \sigma_2^3 + 120x_1^3 x_2 \sigma_2^3 - 180x_1^2 x_2^2 \sigma_2^3 + 120x_1 x_2^3 \sigma_2^3 - 30x_2^4 \sigma_2^3 + 540x_1^2 \sigma_1 \sigma_2^3 \\ - 540x_2^3 \sigma_1 \sigma_2^3 + 450\sigma_1^2 \sigma_2^3 - 135x_1^2 \sigma_2^4 + 135x_2^3 \sigma_2^4 - 225\sigma_1 \sigma_2^4 + 45\sigma_2^5 \end{bmatrix},$$

que l'on peut factoriser en :

$$J_2 = -3(1-p)^2 p^2 \left[\begin{array}{l} (x_1 - x_2)^9 (x_1 + x_2) + 3(x_1 - x_2)^8 (\sigma_1 - \sigma_2) + 18(x_1 - x_2)^5 (x_1 + x_2) (\sigma_1 - \sigma_2)^2 \\ + 30(x_1 - x_2)^4 (\sigma_1 - \sigma_2)^3 - 135(x_1 + x_2) (x_1 - x_2) (\sigma_1 - \sigma_2)^4 - 45(\sigma_1 - \sigma_2)^5 \end{array} \right],$$

ou encore, en posant $Z = \sigma_1 - \sigma_2$:

$$J_2 = 3(1-p)^2 p^2 \left[\begin{array}{l} 45Z^5 + 135(x_1 + x_2)(x_1 - x_2)Z^4 - 30(x_1 - x_2)^4 Z^3 - 18(x_1 - x_2)^5 (x_1 + x_2)Z^2 \\ - 3(x_1 - x_2)^8 Z - (x_1 - x_2)^9 (x_1 + x_2) \end{array} \right].$$

Remarquons dès maintenant que ce déterminant admet au moins un pôle, puisqu'il est sous forme d'un polynôme de degré impair en Z .

Il est possible d'ordonner le jacobien sous une autre forme plus instructive:

$$J_2 = 3(1-p)^2 p^2 \left[\begin{array}{l} \overbrace{3Z [(x_1 - x_2)^8 + 10(x_1 - x_2)^4 Z^2 - 15Z^4]}^{F_3} \\ + \underbrace{(x_1 + x_2)(x_1 - x_2) [(x_1 - x_2)^8 + 18(x_1 - x_2)^4 Z^2 - 135Z^4]}_{F_4} \end{array} \right]. \quad [9.39]$$

Le second terme, F_4 , s'annule pour $|x_1| = |x_2|$ et pour $|x_1 - x_2| = (6\sqrt{6} - 9)^{1/4} \sqrt{|Z|}$. Les variétés sur lesquelles le jacobien J_1 s'annule étant caractérisées par $x_1 = x_2$ et $|x_1 - x_2| = (6\sqrt{6} - 9)^{1/4} \sqrt{|Z|}$, le terme F_4 de [9.39] s'annule sur les variétés annulant J_1 . Regardons si le terme F_3 s'annule pour ces mêmes jeux de paramètres.

Le cas particulier où $x_1 = x_2 = 0$, qui correspond au cas du système inversé pour la méthode de Pearson avec les moyennes égales, s'écrit sous la forme :

$$J_2 = -135(1-p)^2 p^2 Z^5.$$

Il ne s'annule que pour $Z = \sigma_1 - \sigma_2 = 0$, correspondant au cas où les deux distributions normales sont les mêmes, qui n'entre plus dans l'étude des paramètres des distributions de mélange.

Si l'on se place sur V_2

$$|x_1 - x_2| = (6\sqrt{6} - 9)^{1/4} \sqrt{|Z|},$$

Etude du système formé des quatre premières et de la sixième equations de moments centrés.

[9.39] devient

$$J_2 = 9(1-p)^2 p^2 Z^5 \left[(6\sqrt{6} - 9)^2 + 10(6\sqrt{6} - 9) - 15 \right],$$

qui ne s'annule que lorsque $Z = 0$, c'est à dire lorsque les deux distributions normales composant le mélange sont identiques. Ce déterminant ne s'annule donc pas pour les mêmes valeurs que J_1 . Le système devrait permettre d'obtenir des solutions dans les cas posant, précédemment, un problème.

Le terme F_4 de [9.39] s'annule encore si $x_1 = -x_2 \neq 0$, qui n'est pas un cas de nullité recensé pour J_1 . Dans ce cas on aura:

$$J_2 = 9(1-p)^2 p^2 Z [256x_1^8 + 160x_1^4 Z^2 - 15Z^4]. \quad [9.40]$$

Le discriminant de [9.40] considéré comme un polynôme en Z^2 étant :

$$\Delta_{J_2} = 25600x_1^8 + 15360x_1^8 = 40960x_1^8 > 0,$$

on peut factoriser [9.40] en:

$$\begin{aligned} J_2 &= -135(1-p)^2 p^2 Z \left[\left(Z^2 + \frac{(-160 - \sqrt{40960}) x_1^4}{30} \right) \left(Z^2 + \frac{(-160 + \sqrt{40960}) x_1^4}{30} \right) \right], \\ &= -135(1-p)^2 p^2 Z \left[\left(Z^2 - \frac{16(5 + 2\sqrt{10}) x_1^4}{15} \right) \left(Z^2 + \frac{16(2\sqrt{10} - 5) x_1^4}{15} \right) \right], \\ &= -135(1-p)^2 p^2 Z \left(|Z| - 4\sqrt{\frac{(5 + 2\sqrt{10})}{15}} x_1^2 \right) \left(|Z| + 4\sqrt{\frac{(5 + 2\sqrt{10})}{15}} x_1^2 \right) \times \\ &\quad \left(Z^2 + \frac{16(2\sqrt{10} - 5) x_1^4}{15} \right). \end{aligned} \quad [9.41]$$

Si $Z = 0$, comme nous étudions le cas où $x_1 = -x_2 \neq 0$ (ce qui est équivalent à $p = 0.5$), nous nous plaçons dans le cas symétrique où

$$\begin{aligned} \sigma_1 &= \sigma_2 \\ \text{et } p &= 0.5. \end{aligned}$$

ce qui définit une sous-variété V_3 , de $\mathbb{R}^2 \times \mathbb{R}^{+2} \times [0.5; 1[$:

$$V_3 = \{(x_1, x_2, \sigma_1, \sigma_2, p), \sigma_1 = \sigma_2, p = 0.5, x_1 = -x_2 \neq 0\},$$

sur laquelle J_2 est nul. Ce cas de nullité de J_2 est, a priori, couvert par le sous cas du premier système des moments où les variances sont égales.

[9.41] définit une autres variété sur laquelle J_2 s'annule. Il s'agit de la sous-variété V'_3 , de $\mathbb{R}^2 \times \mathbb{R}^{+2} \times [0.5; 1[$ définies par :

$$V'_3 = \left\{ (x_1, x_2, \sigma_1, \sigma_2, p), |\sigma_1 - \sigma_2| = 4\sqrt{\frac{(5 + 2\sqrt{10})}{15}} x_1^2 \text{ et } p = 0.5 \right\}.$$

Etude du système formé des quatre premières et de la sixième equations de moments centrés.

Puisque l'on sait que les deux déterminants ne sont pas simultanément nuls, on peut se placer, désormais, hors des variétés annulant J_1 (ce qui assure que $135X^2Z^4 - 18X^6Z^2 - X^{10} \neq 0$). Le jacobien J_2 peut se factoriser.

En effet, si l'on s'aperçoit que

$$x_1 + x_2 = (1 - 2p)(x_1 - x_2)$$

et que l'on pose

$$X = (x_1 - x_2),$$

l'équation

$$J_2 = 3(1-p)^2 p^2 \left[\begin{array}{c} 45Z^5 + 135(x_1 + x_2)(x_1 - x_2)Z^4 - 30(x_1 - x_2)^4 Z^3 - 18(x_1 - x_2)^5 (x_1 + x_2)Z^2 \\ - 3(x_1 - x_2)^8 Z - (x_1 - x_2)^9 (x_1 + x_2) \end{array} \right]$$

devient

$$J_2 = 3(1-p)^2 p^2 [45Z^5 - 3X^8 Z - 30X^4 Z^3 + (1-2p)X(135XZ^4 - 18X^5 Z^2 - X^9)],$$

que l'on peut factoriser sous la forme d'un polynôme en p , puisque $135X^2Z^4 - 18X^6Z^2 - X^{10} \neq 0$:

$$\begin{aligned} J_2 &= -6X^2(135Z^4 - 18X^4Z^2 - X^8)(1-p)^2 p^2 \times \\ &\quad \left[p - \frac{45Z^5 - 3X^8 Z - 30X^4 Z^3 + 135X^2Z^4 - 18X^6Z^2 - X^{10}}{2(135X^2Z^4 - 18X^6Z^2 - X^{10})} \right], \\ &= -6X^2(135Z^4 - 18X^4Z^2 - X^8)(1-p)^2 p^2 \left[p - \left(\frac{1}{2} + \frac{3Z(15Z^4 - 10X^4Z^2 - X^8)}{2X^2(135Z^4 - 18X^4Z^2 - X^8)} \right) \right]. \end{aligned}$$

Comme

$$\begin{aligned} 15Z^4 - 10X^4Z^2 - X^8 &= \left(|Z| - \sqrt{\frac{(5+2\sqrt{10})}{15}} X^2 \right) \left(|Z| + \sqrt{\frac{(5+2\sqrt{10})}{15}} X^2 \right) \times \\ &\quad \left(Z^2 + \frac{16(2\sqrt{10}-5)x_1^4}{15} \right), \end{aligned}$$

la sous-variété V_4 , de $\mathbb{R}^2 \times \mathbb{R}^{+2} \times [0.5; 1[$:

$$V_4 = \left\{ \begin{array}{l} (x_1, x_2, \sigma_1, \sigma_2, p), p = \left(\frac{1}{2} + \frac{3(\sigma_1 - \sigma_2)(15(\sigma_1 - \sigma_2)^4 - 10(x_1 - x_2)^4(\sigma_1 - \sigma_2)^2 - (x_1 - x_2)^8)}{2(x_1 - x_2)^2(135(\sigma_1 - \sigma_2)^4 - 18(x_1 - x_2)^4(\sigma_1 - \sigma_2)^2 - (x_1 - x_2)^8)} \right) \\ \text{et } (x_1 - x_2)^2 (135(\sigma_1 - \sigma_2)^4 - 18(x_1 - x_2)^4(\sigma_1 - \sigma_2)^2 - (x_1 - x_2)^8) \neq 0 \end{array} \right\}$$

contient la variété V_3 issue du sous-cas où $\sigma_1 = \sigma_2$ et $p = 0.5$, ainsi que la sous-variété V_3' , puisque cela correspond à la sous-variété sur laquelle $X = 2x_1 \neq 0$ et $p = 0.5$ qui appartient bien à V_4 .

THÉORÈME 9.1: *Le jacobien J_2 est nul si et seulement si $(x_1, x_2, \sigma_1, \sigma_2, p) \in V_4$ dont l'intersection avec $V_1 \cup V_2$ est nulle, par définition même de V_4 .*

Il ne nous reste qu'à inverser le système [9.38] afin de pallier aux lacunes du précédent système des moments.

Etude du système formé des quatre premières et de la sixième equations de moments centrés.

9.2 Inversion algébrique du système.

Considérons le système des équations des moments du mélange de deux distributions normales $N(\mu_1, \sigma_1^2)$ et $N(\mu_2, \sigma_2^2)$, en fonction des moments centrés M_2^c, M_3^c, M_4^c et M_6^c . Ce système s'écrit, comme nous l'avons signalé plus haut :

$$\mathfrak{S}_2 = \begin{cases} 0 = px_1 + (1-p)x_2 \\ M_2^c = px_1^2 + (1-p)x_2^2 + (p\sigma_1 + (1-p)\sigma_2) \\ M_3^c = px_1^3 + (1-p)x_2^3 + 3(p\sigma_1x_1 + (1-p)\sigma_2x_2) \\ M_4^c = px_1^4 + (1-p)x_2^4 + 6(p\sigma_1x_1^2 + (1-p)\sigma_2x_2^2) + 3(p\sigma_1^2 + (1-p)\sigma_2^2) \\ M_6^c = px_1^6 + (1-p)x_2^6 + 15(p\sigma_1x_1^4 + (1-p)\sigma_2x_2^4) + 45(p\sigma_1^2x_1^2 + (1-p)\sigma_2^2x_2^2) \\ + 15(p\sigma_1^3 + (1-p)\sigma_2^3) \end{cases} \quad [9.42]$$

9.2.1 Si $x_1 = x_2$.

Nous nous ramenons au système donnant, dans l'étude de Pearson la solution au problème du mélange de deux distributions de moyennes égales. Il s'agit du cas où $M_3^c = 0$ et $k_4 > 0$

9.2.2 Si $x_1 \neq x_2$.

Nous pouvons, comme précédemment, effectuer le changement de variable, lorsque $x_1 - x_2 \neq 0$, :

$$\begin{aligned} \sigma_1 &= \frac{\sigma' - \sigma x_2}{p(x_1 - x_2)} \\ \text{et } \sigma_2 &= \frac{\sigma' - \sigma x_1}{(1-p)(x_2 - x_1)}, \end{aligned}$$

pour obtenir le système :

$$\mathfrak{S}'_2 = \begin{cases} px_1 + (1-p)x_2 = 0 \\ px_1^2 + (1-p)x_2^2 = M_2^c - \sigma \\ px_1^3 + (1-p)x_2^3 = M_3^c - 3\sigma' \\ px_1^4 + (1-p)x_2^4 = M_4^c - 6(\sigma'(x_1 + x_2) - \sigma x_1 x_2) - \frac{3((\sigma')^2 + \sigma^2(px_1^2 + (1-p)x_2^2))}{p(1-p)(x_1 - x_2)^2} \\ (px_1^6 + (1-p)x_2^6)(x_1 x_2)^2 = -15(x_1 x_2)^2 \left[\sigma'(x_1 + x_2) \left((x_1 + x_2)^2 - 2x_2 x_1 \right) \right] \\ -15(x_1 x_2)^2 \left[\sigma x_1 x_2 \left((x_1 + x_2)^2 - x_1 x_2 \right) \right] + (x_1 x_2)^2 M_6 \\ + 45x_1 x_2 \left[\sigma'^2 \left((x_1 + x_2)^2 - x_2 x_1 \right) - 2\sigma' \sigma x_1 x_2 (x_1 + x_2) + (\sigma x_1 x_2)^2 \right] \\ -15(x_1 - x_2) \left[(1-2p)\sigma'^3 + (3\sigma'^2 \sigma - \sigma^3 x_1 x_2) \left[\frac{-(1-2p)x_1 x_2}{(x_1 + x_2)} \right] \right] \end{cases} \quad [9.43]$$

Les calculs pour obtenir la dernière équation sont détaillés ici :

$$\begin{aligned} px_1^6 + (1-p)x_2^6 &= M_6^c - 15 \left(\left(\frac{\sigma' - \sigma x_2}{(x_1 - x_2)} \right) x_1^4 - \left(\frac{\sigma' - \sigma x_1}{(x_1 - x_2)} \right) x_2^4 \right) \\ &\quad - 45 \left(\frac{(\sigma' - \sigma x_2)^2}{p(x_1 - x_2)^2} x_1^2 + \frac{(\sigma' - \sigma x_1)^2}{(1-p)(x_2 - x_1)^2} x_2^2 \right) \end{aligned}$$

Etude du système formé des quatre premières et de la sixième équations de moments centrés.

$$-15 \left(\frac{(\sigma' - \sigma x_2)^3}{p^2 (x_1 - x_2)^3} - \frac{(\sigma' - \sigma x_1)^3}{(1-p)^2 (x_1 - x_2)^3} \right),$$

puis :

$$\begin{aligned} px_1^6 + (1-p)x_2^6 &= M_6^c - 15 \left(\frac{\sigma' (x_1^4 - x_2^4) - \sigma x_1 x_2 (x_1^3 - x_2^3)}{(x_1 - x_2)} \right) \\ &\quad - 45 \left(\frac{(1-p)(\sigma' x_1 - \sigma x_1 x_2)^2 + p(\sigma' x_2 - \sigma x_1 x_2)^2}{p(1-p)(x_1 - x_2)^2} \right) \\ &\quad - 15 \left(\frac{(1-p)^2 (\sigma' - \sigma x_2)^3 - p^2 (\sigma' - \sigma x_1)^3}{p^2 (1-p)^2 (x_1 - x_2)^3} \right), \end{aligned}$$

donc :

$$\begin{aligned} px_1^6 + (1-p)x_2^6 &= M_6^c - 15 (\sigma' (x_2^3 + x_2^2 x_1 + x_2 x_1^2 + x_1^3) - \sigma x_1 x_2 (x_1^2 + x_1 x_2 + x_2^2)) \\ &\quad + \frac{45}{x_1 x_2} \left(\sigma'^2 ((1-p)x_1^2 + px_2^2) - 2\sigma' \sigma x_1 x_2 (x_1 + x_2) + (\sigma x_1 x_2)^2 \right) \\ &\quad - \frac{15(x_1 - x_2)}{(x_1 x_2)^2} \left[(1-p)^2 [\sigma'^3 - 3\sigma'^2 \sigma x_2 + 3\sigma' (\sigma x_2)^2 - (\sigma x_2)^3] \right] \\ &\quad - \frac{15(x_1 - x_2)}{(x_1 x_2)^2} \left[-p^2 [\sigma'^3 - 3\sigma'^2 \sigma x_1 + 3\sigma' (\sigma x_1)^2 - (\sigma x_1)^3] \right], \end{aligned}$$

c'est à dire :

$$\begin{aligned} (px_1^6 + (1-p)x_2^6) (x_1 x_2)^2 &= -15 (x_1 x_2)^2 \left(\sigma' (x_1 + x_2) \left((x_1 + x_2)^2 - 2x_2 x_1 \right) \right) \\ &\quad - 15 (x_1 x_2)^2 \left(-\sigma x_1 x_2 \left((x_1 + x_2)^2 - x_1 x_2 \right) \right) \\ &\quad + 45 x_1 x_2 \left(\sigma'^2 ((1-p)x_1^2 + px_2^2) - 2\sigma' \sigma x_1 x_2 (x_1 + x_2) + (\sigma x_1 x_2)^2 \right) \\ &\quad - 15 (x_1 - x_2) \left[(1-2p) \sigma'^3 + 3\sigma'^2 \sigma \left[p^2 x_1 - (1-p)^2 x_2 \right] \right] \\ &\quad - 15 (x_1 - x_2) \left[-3\sigma' \sigma^2 \left[p^2 x_1^2 - (1-p)^2 x_2^2 \right] + \sigma^3 \left[p^2 x_1^3 - (1-p)^2 x_2^3 \right] \right] \\ &\quad + (x_1 x_2)^2 M_6^c. \end{aligned}$$

Comme :

$$\begin{aligned} p^2 x_1^3 - (1-p)^2 x_2^3 &= -x_1 x_2 \left(p^2 x_1 - (1-p)^2 x_2 \right), \\ p^2 x_1^2 - (1-p)^2 x_2^2 &= 0 \\ \text{et } p^2 x_1 - (1-p)^2 x_2 &= \frac{-(1-2p)x_1 x_2}{(x_1 + x_2)}, \end{aligned}$$

on a enfin :

$$(px_1^6 + (1-p)x_2^6) (x_1 x_2)^2 = +45 x_1 x_2 \left[\sigma'^2 \left((x_1 + x_2)^2 - x_2 x_1 \right) - 2\sigma' \sigma x_1 x_2 (x_1 + x_2) + (\sigma x_1 x_2)^2 \right]$$

Etude du système formé des quatre premières et de la sixième equations de moments centrés.

$$\begin{aligned}
& -15\sigma' (x_1x_2)^2 (x_1 + x_2) \left((x_1 + x_2)^2 - 2x_2x_1 \right) \\
& +15 (x_1x_2)^3 \sigma \left((x_1 + x_2)^2 - x_1x_2 \right) + (x_1x_2)^2 M_6^c \\
& -15 (x_1 - x_2) \left[(1 - 2p) \sigma'^3 + (3\sigma'^2\sigma - \sigma^3x_1x_2) \left[\frac{-(1 - 2p) x_1x_2}{(x_1 + x_2)} \right] \right].
\end{aligned}$$

Les quatre premières équations de \mathfrak{S}'_2 étant les mêmes que celles de \mathfrak{S}'_1 , on tire, comme plus haut :

$$\begin{aligned}
p &= \frac{\sqrt{4(M_2^c - \sigma)^3 + (M_3^c - 3\sigma')^2} + |M_3^c - 3\sigma'|}{2\sqrt{4(M_2^c - \sigma)^3 + (M_3^c - 3\sigma')^2}}, \\
\mu_1 &= M_1 - \operatorname{sgn}(M_3^c - 3\sigma') \sqrt{\frac{\sqrt{4(M_2^c - \sigma)^3 + (M_3^c - 3\sigma')^2} - |M_3^c - 3\sigma'|}{\sqrt{4(M_2^c - \sigma)^3 + (M_3^c - 3\sigma')^2} + |M_3^c - 3\sigma'|} (M_2^c - \sigma)} \\
\text{et } \mu_2 &= M_1 + \operatorname{sgn}(M_3^c - 3\sigma') \cdot \sqrt{\frac{\sqrt{4(M_2^c - \sigma)^3 + (M_3^c - 3\sigma')^2} + |M_3^c - 3\sigma'|}{\sqrt{4(M_2^c - \sigma)^3 + (M_3^c - 3\sigma')^2} - |M_3^c - 3\sigma'|} (M_2^c - \sigma)},
\end{aligned}$$

ainsi que :

$$\sigma'^2 = \frac{M_3^{c2} - (2(M_2^c - \sigma)^2 + k_4)(M_2^c - \sigma)}{6}. \quad [9.44]$$

Il suffit, désormais, de substituer ces résultats dans la dernière équation. Pour éclaircir le calcul, posons :

$$\begin{aligned}
S_3 &= -15 (x_1x_2)^2 \left(\sigma' (x_1 + x_2) \left((x_1 + x_2)^2 - 2x_2x_1 \right) + \sigma x_1x_2 \left((x_1 + x_2)^2 - x_1x_2 \right) \right), \\
S_4 &= 45x_1x_2 \left(\sigma'^2 \left((x_1 + x_2)^2 - x_2x_1 \right) - 2\sigma' \sigma x_1x_2 (x_1 + x_2) + (\sigma x_1x_2)^2 \right), \\
S_5 &= -15 (x_1 - x_2) \left[(1 - 2p) \sigma'^3 + (3\sigma'^2\sigma - \sigma^3x_1x_2) \left[\frac{-(1 - 2p) x_1x_2}{(x_1 + x_2)} \right] \right], \\
S_6 &= (px_1^6 + (1 - p)x_2^6 - M_6) (x_1x_2)^2,
\end{aligned}$$

et commençons les calculs:

$$\begin{aligned}
S_3 &= -15 (x_1x_2)^2 \left(\sigma' (x_1 + x_2) \left((x_1 + x_2)^2 - 2x_2x_1 \right) - \sigma x_1x_2 \left((x_1 + x_2)^2 - x_1x_2 \right) \right), \\
&= -15 \sigma' (M_2^c - \sigma)^2 \left(\frac{M_3^c - 3\sigma'}{M_2^c - \sigma} \right) \left(\left(\frac{M_3^c - 3\sigma'}{M_2^c - \sigma} \right)^2 + 2(M_2^c - \sigma) \right) \\
&\quad -15 \sigma (M_2^c - \sigma)^3 \left(\left(\frac{M_3^c - 3\sigma'}{M_2^c - \sigma} \right)^2 + (M_2^c - \sigma) \right),
\end{aligned}$$

$$\begin{aligned}
(M_2^c - \sigma) S_3 &= -15 \sigma' (M_3^c - 3\sigma') \left((M_3^c - 3\sigma')^2 + 2(M_2^c - \sigma)^3 \right) \\
&\quad -15\sigma (M_2^c - \sigma)^2 \left((M_3^c - 3\sigma')^2 + (M_2^c - \sigma)^3 \right),
\end{aligned}$$

Etude du système formé des quatre premières et de la sixième équations de moments centrés.

$$\begin{aligned}
&= -15\sigma' (M_3^c - 3\sigma') \left((M_3^c - 3\sigma')^2 + 2(M_2^c - \sigma)^3 \right) \\
&\quad + 15(M_2^c - \sigma)^3 (M_3^c - 3\sigma')^2 - 15(M_2^c - \sigma)^3 \\
&\quad - 15M_2^c (M_2^c - \sigma)^2 \left((M_3^c - 3\sigma')^2 + (M_2^c - \sigma)^3 \right),
\end{aligned}$$

d'où :

$$\begin{aligned}
XS_3 &= -15 \left(\sigma' (M_3^c - 3\sigma') \left((M_3^c - 3\sigma')^2 + 2X^3 \right) - X^3 \left((M_3^c - 3\sigma')^2 + X^3 \right) \right) \\
&\quad - 15 M_2^c X^2 \left((M_3^c - 3\sigma')^2 + X^3 \right), \\
&= -15 \left((M_3^c \sigma' - 3\sigma'^2) (M_3^{c2} - 6M_3^c \sigma' + 9\sigma'^2 + 2X^3) - X^3 (M_3^{c2} - 6M_3^c \sigma' + 9\sigma'^2 + X^3) \right) \\
&\quad - 15 M_2^c X^2 (M_3^{c2} - 6M_3^c \sigma' + 9\sigma'^2 + X^3), \\
XS_3 &= -15 (M_3^c \sigma' - 6M_3^c \sigma'^2 + 9M_3^c \sigma'^3 + 2M_3^c X^3 \sigma' - 3M_3^{c2} \sigma'^2 + 18M_3^c \sigma'^3 - 27\sigma'^4 - 6X^3 \sigma'^2) \\
&\quad - 15 (-M_3^{c2} X^3 + 6X^3 M_3^c \sigma' - 9X^3 \sigma'^2 - X^6 + M_2^c M_3^{c2} X^2) \\
&\quad - 15 (-6M_2^c M_3^c X^2 \sigma' + 9M_2^c X^2 \sigma'^2 + M_2^c X^5), \\
&= -15 (-27\sigma'^4 + 27M_3^c \sigma'^3 + 3(-3M_3^{c2} + -5X^3 + 3M_2^c X^2) \sigma'^2) \\
&\quad - 15 ((M_3^c \sigma' - 6M_2^c M_3^c X^2 + 8X^3 M_3^c) \sigma' + (-X^4 + M_2^c X^3 - M_3^{c2} X + M_2^c M_3^{c2}) X^2), \\
&= -45 (-3M_3^{c2} - 5X^3 + 3M_2^c X^2 - 9\sigma'^2) \sigma'^2 - 15 (-X^4 + M_2^c X^3 - M_3^{c2} X + M_2^c M_3^{c2}) X^2 \\
&\quad - 15 (M_3^c \sigma' - 6M_2^c M_3^c X^2 + 8X^3 M_3^c + 27M_3^c \sigma'^2) \sigma'.
\end{aligned}$$

De même pour S_4 :

$$\begin{aligned}
S_4 &= 45x_1x_2 \left(\sigma'^2 \left((x_1 + x_2)^2 - x_2x_1 \right) - 2\sigma' \sigma x_1x_2(x_1 + x_2) + (\sigma x_1x_2)^2 \right) \\
&= -45 (M_2^c - \sigma) \left(\sigma'^2 \left(\frac{(M_3^c - 3\sigma')^2 + (M_2^c - \sigma)^3}{(M_2^c - \sigma)^2} \right) + 2\sigma' \sigma (M_3^c - 3\sigma') + (\sigma (M_2^c - \sigma))^2 \right), \\
(M_2^c - \sigma) S_4 &= -45 \left(\sigma'^2 \left((M_3^c - 3\sigma')^2 + (M_2^c - \sigma)^3 \right) - 2\sigma' (M_2^c - \sigma)^3 (M_3^c - 3\sigma') \right) \\
&\quad - 45 \left(2\sigma' M_2^c (M_2^c - \sigma)^2 (M_3^c - 3\sigma') + (M_2^c - \sigma)^2 \left(-(M_2^c - \sigma)^2 + M_2^c (M_2^c - \sigma) \right)^2 \right),
\end{aligned}$$

donc :

$$\begin{aligned}
XS_4 &= -45 \left(\sigma'^2 \left((M_3^c - 3\sigma')^2 + X^3 \right) - 2\sigma' X^3 (M_3^c - 3\sigma') \right) \\
&\quad - 45 \left(2\sigma' M_2^c X^2 (M_3^c - 3\sigma') + (X^3 - M_2^c X^2)^2 \right),
\end{aligned}$$

$$\begin{aligned}
XS_4 &= -45 (M_3^{c2} \sigma'^2 - 6M_3^c \sigma'^3 + 9\sigma'^4 + X^3 \sigma'^2 - 2\sigma' M_3^c X^3 + 6\sigma'^2 X^3) \\
&\quad - 45 \left(2\sigma' M_3^c M_2^c X^2 - 6\sigma'^2 M_2^c X^2 + (X^3 - M_2^c X^2)^2 \right), \\
&= -45 \left((7X^3 - 6M_2^c X^2 + M_3^{c2}) \sigma'^2 - 6M_3^c \sigma'^3 + 9\sigma'^4 \right) \\
&\quad - 45 \left(-2 (M_3^c X^3 - M_3^c M_2^c X^2) \sigma' + (X^3 - M_2^c X^2)^2 \right), \\
&= -45 \left((7X^3 - 6M_2^c X^2 + M_3^{c2} + 9\sigma'^2) \sigma'^2 - 2 (M_3^c X^3 - M_3^c M_2^c X^2 + 3M_3^c \sigma'^2) \sigma' \right)
\end{aligned}$$

Etude du système formé des quatre premières et de la sixième equations de moments centrés.

$$-45 (X^3 - M_2^c X^2)^2,$$

Passons à S_5 :

$$S_5 = -15 (x_1 - x_2) \left[(1 - 2p) \sigma'^3 + (3\sigma'^2 \sigma - \sigma^3 x_1 x_2) \left[\frac{-(1 - 2p) x_1 x_2}{(x_1 + x_2)} \right] \right],$$

$$S_5 = -15 \left(-\operatorname{sgn}(M_3^c - 3\sigma') \frac{\sqrt{4(M_2^c - \sigma)^3 + (M_3^c - 3\sigma')^2}}{M_2^c - \sigma} \right) \left[\left(\frac{-|M_3^c - 3\sigma'|}{\sqrt{4(M_2^c - \sigma)^3 + (M_3^c - 3\sigma')^2}} \right) \sigma'^3 - \frac{\operatorname{sgn}(M_3^c - 3\sigma') (3\sigma'^2 \sigma + \sigma^3 (M_2^c - \sigma)) (M_2^c - \sigma)^2}{\sqrt{4(M_2^c - \sigma)^3 + (M_3^c - 3\sigma')^2}} \right],$$

$$S_5 = 15 \left(\operatorname{sgn}(M_3^c - 3\sigma') \frac{1}{M_2^c - \sigma} \right) \left[(-|M_3^c - 3\sigma'|) \sigma'^3 - \operatorname{sgn}(M_3^c - 3\sigma') (3\sigma'^2 \sigma + \sigma^3 (M_2^c - \sigma)) (M_2^c - \sigma)^2 \right],$$

donc :

$$S_5 = 15 \left[-\frac{(M_3^c - 3\sigma') \sigma'^3}{M_2^c - \sigma} - (3\sigma'^2 \sigma + \sigma^3 (M_2^c - \sigma)) (M_2^c - \sigma) \right],$$

$$(M_2^c - \sigma) S_5 = 15 \left[-(M_3^c - 3\sigma') \sigma'^3 - \left(3\sigma'^2 (- (M_2^c - \sigma) + M_2^c) + (- (M_2^c - \sigma) + M_2^c)^3 (M_2^c - \sigma) \right) (M_2^c - \sigma) \right],$$

c'est à dire :

$$\begin{aligned} X S_5 &= 15 \left[-(M_3^c - 3\sigma') \sigma'^3 - \left(3\sigma'^2 (-X + M_2^c) + (-X + M_2^c)^3 X \right) X^2 \right], \\ &= 15 \left[-(M_3^c - 3\sigma') \sigma'^3 - 3\sigma'^2 (-X + M_2^c) X^2 - (-X^3 + 3M_2^c X^2 - 3M_2^{c2} X + M_2^{c3}) X^3 \right], \end{aligned}$$

et donc :

$$X S_5 = 15 \left[-M_3^c \sigma'^3 - 3\sigma'^2 (-X^3 + M_2^c X^2 - \sigma'^2) + X^6 - 3M_2^c X^5 + 3M_2^{c2} X^4 - M_2^{c3} X^3 \right].$$

Enfin :

$$\begin{aligned} S_6 &= (p x_1^6 + (1 - p) x_2^6 - M_6^c) (x_1 x_2)^2, \\ &= \left[(p x_1^4 + (1 - p) x_2^4) (x_1^2 + x_2^2) + (x_1 x_2)^3 \right] (x_1 x_2)^2 - M_6^c (x_1 x_2)^2, \\ &= \left[\left[(p x_1^2 + (1 - p) x_2^2) \left((x_1 + x_2)^2 - 2x_1 x_2 \right) - (x_1 x_2)^2 \right] \left((x_1 + x_2)^2 - 2x_1 x_2 \right) + (x_1 x_2)^3 \right] (x_1 x_2)^2 \\ &\quad - M_6^c (x_1 x_2)^2, \\ &= \left[\left[(-x_1 x_2) (x_1 + x_2)^2 + 2(x_1 x_2)^2 - (x_1 x_2)^2 \right] \left((x_1 + x_2)^2 - 2x_1 x_2 \right) + (x_1 x_2)^3 \right] (x_1 x_2)^2 \\ &\quad - M_6^c (x_1 x_2)^2, \\ &= \left[(-x_1 x_2) (x_1 + x_2)^2 + (x_1 x_2)^2 \right] \left[(x_1 + x_2)^2 (x_1 x_2)^2 - 2(x_1 x_2)^3 \right] + (x_1 x_2)^5 - M_6^c (x_1 x_2)^2, \end{aligned}$$

Etude du système formé des quatre premières et de la sixième equations de moments centrés.

ou encore :

$$\begin{aligned} S_6 &= \left(\frac{(M_3^c - 3\sigma')^2}{M_2^c - \sigma} + (M_2^c - \sigma)^2 \right) \left((M_3^c - 3\sigma')^2 + 2(M_2^c - \sigma)^3 \right) - (M_2^c - \sigma)^5 - M_6^c (M_2^c - \sigma)^2, \\ &= \left(\frac{(M_3^{c2} - 6\sigma' M_3^c + 9\sigma'^2 + (M_2^c - \sigma)^3) (M_3^{c2} - 6\sigma' M_3^c + 9\sigma'^2 + 2(M_2^c - \sigma)^3)}{M_2^c - \sigma} \right) \\ &\quad - \frac{(M_2^c - \sigma)^6}{M_2^c - \sigma} - M_6^c \frac{(M_2^c - \sigma)^3}{M_2^c - \sigma}, \end{aligned}$$

donc :

$$\begin{aligned} (M_2^c - \sigma) S_6 &= \left(54M_3^{c2} + 27(M_2^c - \sigma)^3 + 81\sigma'^2 \right) \sigma'^2 - \left(18M_3^c (M_2^c - \sigma)^3 + 12M_3^{c3} + 108\sigma'^2 M_3^c \right) \sigma' \\ &\quad + (M_2^c - \sigma)^6 + (3M_3^{c2} - M_6^c) (M_2^c - \sigma)^3 + M_3^{c4}, \end{aligned}$$

c'est à dire :

$$X S_6 = (54M_3^{c2} + 27X^3 + 81\sigma'^2) \sigma'^2 - (18M_3^c X^3 + 12M_3^{c3} + 108\sigma'^2 M_3^c) \sigma' + X^6 + (3M_3^{c2} - M_6^c) X^3 + M_3^{c4}.$$

Rassemblons les quatre parties du calcul :

$$X (S_3 + S_4 + S_5 - S_6) = 0$$

et donc :

$$\begin{aligned} 0 &= \sigma'^2 (72X^3 - 90X^2 M_2^c - 36M_3^{c2} + 36\sigma'^2) + \sigma' (12X^3 M_3^c + 3M_3^{c3} + 42\sigma'^2 M_3^c) \\ &\quad + 16X^6 - 30X^5 M_2^c + (15M_2^{c3} - 12M_3^{c2} - M_6^c) X^3 + 15X^2 M_2^c M_3^{c2} + M_3^{c4}. \end{aligned}$$

En posant $k_6 = M_6^c - 15M_4^c M_2^c - 10M_3^{c2} + 30M_2^{c3}$, le sixième cumulante du mélange :

$$\begin{aligned} 0 &= \sigma'^2 (72X^3 - 90X^2 M_2^c - 36M_3^{c2} + 36\sigma'^2) + \sigma' (12X^3 M_3^c + 3M_3^{c3} + 42\sigma'^2 M_3^c) \\ &\quad + 16X^6 - 30X^5 M_2^c + (-15k_4 M_2^c - 22M_3^{c2} - k_6) X^3 + 15X^2 M_2^c M_3^{c2} + M_3^{c4}. \end{aligned}$$

Il est, alors, nécessaire de séparer deux cas :

$$\begin{aligned} 12X^3 M_3^c + 3M_3^{c3} + 42\sigma'^2 M_3^c &= 0 \\ \text{ou } 12X^3 M_3^c + 3M_3^{c3} + 42\sigma'^2 M_3^c &\neq 0, \end{aligned}$$

or

$$\begin{aligned} 12X^3 M_3^c + 3M_3^{c3} + 42\sigma'^2 M_3^c &= 0 \Leftrightarrow \\ M_3^c (12X^3 + 3M_3^{c2} + 42\sigma'^2) &= 0 \Leftrightarrow \\ M_3^c &= 0 \text{ car } X > 0 \text{ par définition,} \end{aligned}$$

Etude du système formé des quatre premières et de la sixième equations de moments centrés.

Or, si $M_3^c = 0$, deux cas sont possible : Si $M_3^c = 0$ et $k_4 > 0$, les moyennes sont égales et la résolution du système peut être menée différemment. Si $M_3^c = 0$ et $k_4 \leq 0$, on est dans le cas où les variances sont égales et le système se restreint aux quatre premières équations, système déjà inversé dans le cas précédemment étudié.

Si $M_3^c \neq 0$, on a :

$$\sigma' = \frac{\sigma'^2(-72X^3 + 90X^2M_2^c + 36M_3^{c2} - 36\sigma'^2) - 16X^6 + 30X^5M_2^c + (15k_4M_2^c + 22M_3^{c2} + k_6)X^3 - 15X^2M_2^cM_3^{c2} - M_3^{c4}}{12X^3M_3^c + 3M_3^{c3} + 42\sigma'^2M_3^c},$$

et, d'après [9.44] :

$$\sigma' = \frac{4X^6 + 8X^4k_4 + (k_6 + 2M_3^{c2})X^3 - k_4^2X^2 - 4k_4M_3^{c2}X + 4M_3^{c4}}{M_3^c(-2X^3 - 7Xk_4 + 10M_3^{c2})}.$$

Il suffit alors de substituer ce résultat dans [9.44] :

$$6(4X^6 + 8k_4X^4 + (2M_3^{c2} + k_6)X^3 - k_4^2X^2 - 4M_3^{c2}k_4X + 4M_3^{c4})^2 = (M_3^{c2} - (2X^2 + k_4)X)(-2M_3^cX^3 - 7M_3^ck_4X + 10M_3^{c3})^2,$$

c'est à dire, après développement :

$$\begin{aligned} &96X^{12} + 384k_4X^{10} + 8(6k_6 + 13M_3^{c2})X^9 + 336k_4^2X^8 + 12k_4(8k_6 + 5M_3^{c2})X^7 \\ &+ 6(-16k_4^3 + k_6^2 + 4k_6M_3^{c2} + 22M_3^{c4})X^6 - 6k_4^2(2k_6 + 47M_3^{c2})X^5 + 6k_4(k_4^3 - 2(4k_6M_3^{c2} + 5M_3^{c4}))X^4 \\ &+ (97k_4^3M_3^{c2} + 48M_3^{c4}(k_6 + 7M_3^{c2}))X^3 - 141k_4^2M_3^{c4}X^2 + 48k_4M_3^{c6}X - 4M_3^{c8} = 0. \end{aligned}$$

La résolution du système \mathfrak{S}_2 , si $M_3^c \neq 0$, se ramène donc à la recherche des racines strictement positives du polynôme de degré 12 :

$$\begin{aligned} P_{12}(X) = &96X^{12} + 384k_4X^{10} + 8(6k_6 + 13M_3^{c2})X^9 + 336k_4^2X^8 + 12k_4(8k_6 + 5M_3^{c2})X^7 \\ &+ 6(-16k_4^3 + k_6^2 + 4k_6M_3^{c2} + 22M_3^{c4})X^6 - 6k_4^2(2k_6 + 47M_3^{c2})X^5 \\ &+ 6k_4(k_4^3 - 2(4k_6M_3^{c2} + 5M_3^{c4}))X^4 + (97k_4^3M_3^{c2} + 48M_3^{c4}(k_6 + 7M_3^{c2}))X^3 \\ &- 141k_4^2M_3^{c4}X^2 + 48k_4M_3^{c6}X - 4M_3^{c8}. \end{aligned}$$

Comme $P_{12}(0) = -4M_3^{c8} < 0$ et que $P_{12}(X) \xrightarrow{X \rightarrow +\infty} +\infty$, P_{12} a toujours au moins une racine strictement positive.

9.3 Exemples d'inversion du système.

Dans les exemples suivants, les calculs sont effectués avec une précision de 1000 chiffres significatifs et les racines des différents polynômes sont calculées avec une précision de 50000 chiffres significatifs.

Le test que nous présenterons avec la cinquième équation est le rapport T_5 défini plus haut :

$$T_5 = \left| \frac{p(x_1^5 + 10\sigma_1x_1^3 + 15\sigma_1^2x_1) + (1-p)(x_2^5 + 10\sigma_2x_2^3 + 15\sigma_2^2x_2) - M_5^c}{2M_5^c} \right|,$$

où le numérateur est calculé avec les paramètres obtenus par inversion du système et le dénominateur est calculé avec les paramètres connus initialement. T_5 doit être "proche" de 0 (égal à 0 dans les études théoriques), et comme

Etude du système formé des quatre premières et de la sixième equations de moments centrés.

précédemment, s'il y a plusieurs ensembles de solutions acceptables, nous choisirons celui qui donnera le rapport le plus proche de 0.

9.3.1 Cas général (éloigné des pôles du jacobien).

Nous faisons référence à l'annexe 8. Le polynôme P_{12} admet une seule racine positive r_1 qui donne un ensemble de paramètres solution appartenant au domaine de définition des distributions normales. Le test T_5 est nul pour cet ensemble de solutions, qui est effectivement celui que nous espérons.

9.3.2 Cas où $\sigma_1 = \sigma_2$ (éloigné des pôles du jacobien).

Si $p \neq 0.5$.

Nous faisons référence à l'annexe 9. Le polynôme P_{12} a une seule racine positive, r_1 , ce qui permet d'éviter de choisir entre plusieurs possibilités d'ensemble solution au problème des moments.

Si $p = 0.5$.

Nous faisons référence à l'annexe 10. C'est un cas particulier de la variété V_4 posant problème pour ce second système. Nous constatons que le cas théorique est couvert par la méthode.

9.3.3 Cas où les variables appartiennent à la variété V_2 .

Comme nous pouvions le prévoir après les calculs de jacobien, les calculs d'inversion, présentés en annexe 11, sont menés sans problème dans ce cas. Le polynôme P_{12} n'a qu'une racine positive, ce qui montre que l'utilisation de ce polynôme peut permettre de remédier à d'éventuelles défaillances du problème classique des moments.

9.3.4 Cas où x_1 et x_2 sont proches (voisinage de V_1).

Le cas où les moyennes sont égales est résolu par la méthode de Pearson. Le comportement au voisinage de V_1 , qui n'est pas couvert par ce sous exemple particulier, mérite d'être abordé ici.

Moyennes proches variance éloignées.

Il y a, dans l'exemple cité en annexe 12, trois racines positives à P_{12} qui ne donnent qu'une solution statistique au problème des moments. La différence entre les variances suffit à l'unicité statistique de la solution, mais comme nous l'avons déjà constaté pour P_σ , il n'y a pas unicité globale des solutions.

Moyennes et variances proches.

Nous faisons référence à l'annexe 13. Les trois racines de P_{12} donnent cette fois deux solutions potentielles au problème des moments. Le cas des distributions de paramètres très proches est moins aisé à inverser. Il peut, dans ce cas, comme nous le signalions déjà pour P_σ , y avoir des difficultés dans les inversions de système empiriques des moments

INVERSION DE S2 POUR DES MOMENTS THEORIQUES

	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2	racines >0	acceptable	Test T5	Commentaires
Annexe 8	5.0	3.0	0.6	1.0	1.0	3.84	oui	0.0	
	5.0	3.0	0.6	1.0	1.0				
Annexe 9	1.0	1.0	0.7	3.0	1.0	0.84	oui	0.0	
	1.0	1.0	0.7	3.0	1.0				
Annexe 10	1.0	1.0	0.5	2.0	1.0	0.84	oui	0.0	
	1.0	1.0	0.5	2.0	1.0				L'inversion se fait sur V4
Annexe 11	2.092434442	3.0	0.5	-0.092434442	1.0	0.25	oui	incalculable	
	2.092434442	3.0	0.5	-0.092434442	1.0	1.19341301	oui	0.0	Il n'y a qu'une seule racine sur V4
Annexe 12	1.001	5.0	0.6	1.0	1.0	9.05089E-8	non		
	1.001	5.0	0.6	1.0	1.0	2.4E-7	oui	0.0	
Annexe 13	1.0001	1.0001	0.6	1.0	1.0	6.53347E-7	non		
	1.0001	1.0001	0.6	1.0	1.0	9.05117E-10	non		
						2.4E-9	oui	0.0	
						6.53359E-9	oui	0.03422	

Légende :

Paramètres théoriques : Racine choisie pour l'inversion :

Paramètres calculés : Paramètres issus de Pvarégales :

FIG. II.6. -

Etude du système formé des quatre premières et de la sixième equations de moments centrés.

- CHAPITRE 10 -

Le problème des moments dans le cas des moments empiriques.

Afin d'illustrer les problèmes que nous laissions présager les inversions des systèmes avec moments théoriques, nous étendons ces inversions au cas de moments empiriques issus d'un échantillon de 100000 individus. Les racines des polynômes sont calculées avec une précision de 50000 chiffres, pour assurer, comme nous en avons montré la nécessité plus haut, une précision suffisante.

Précisons, enfin, que ce chapitre n'est pas un chapitre d'estimation, à proprement parler, mais plutôt une collection d'exemples destinés à mettre au point une méthode originale d'estimation cumulant les deux systèmes des moments, et que nous présenterons dans le chapitre suivant. Nous cherchons, en fait, à nous faire une idée empirique sur "l'épaisseur" du voisinage des variétés V_1 et V_2 pour lequel l'inversion du premier système des moments ne donne pas une estimation fiable pour les moments empiriques, et doit être suppléée par l'une des deux autres méthodes. C'est pourquoi nous présentons dans tous les cas, les résultats pour les polynômes P_σ et P_{12} , et si nécessaire, ceux du polynôme $P_{varégales}$. Les exemples suivants illustrent de façon presque exhaustive les différents cas de mélanges susceptibles d'être étudiés, et nécessaire à l'optimisation du protocole d'estimation que nous souhaitons développer.

10.1 Cas général (éloigné des pôles du jacobien).

Ce cas de figure présenté en annexe 14, montre l'efficacité des deux méthodes pour un mélange ne devant théoriquement pas poser problème. Nous avons, par ailleurs, la confirmation que l'estimation par les cinq premières équations est un peu plus précise que celle utilisant le remplacement de la cinquième par la sixième. Signalons, enfin, que le test T_6 est vérifié à deux millièmes près.

10.2 Cas intermédiaire.

Le cas d'un tel mélange ne semble pas particulièrement proche de l'une des variétés que nous avons identifiées comme pouvant être source d'erreur. Pourtant, l'utilisation de P_σ met en évidence une instabilité de la méthode à cet endroit. Dans le premier des trois exemples présentés en annexe 15, 16 et 17, les ensembles de paramètres obtenus par les deux méthodes sont équivalentes et proches de celles que nous attendons. Le test T_6 remplit son rôle dans la sélection du bon ensemble de paramètres. Dans le second exemple, aucune des deux solutions issues de P_σ ne s'approche de celle issue de P_{12} . De plus, la solution désignée par le test T_6 n'est pas celle qui semble la plus raisonnablement proche des paramètres de simulation. Enfin, le dernier exemple, dans lequel P_σ n'a pas de racine positive, confirme que l'on est apparemment trop proche d'une variété problématique (probablement V_2), pour pouvoir faire confiance à cette méthode.

10.3 Cas où $\sigma_1 = \sigma_2$ (éloigné de V_2 ou V_4).

L'objectif de ce paragraphe est de montrer l'évolution de la qualité des estimations, lorsque, pour deux moyennes théoriques fixées, on augmente les variances théoriques (en les maintenant égales). La distribution de mélange se rapproche ainsi du cas symétrique, qui peut être source d'erreur. De plus, la qualité de l'estimation des paramètres est altérée, dans les estimations par EM ou ICE , lorsque les variances théoriques sont grandes par rapport à la différence des moyennes théoriques.

Nous fixerons, dans les exemples présentés dans les annexes 18 à 29, les deux moyennes théoriques à 1 et 5, et les variances théoriques s'étaleront progressivement de 4 à 16, ce qui mettra en évidence une limite de la méthode.

Pour des variances théoriques égales à 4, 5 et 6, (annexes 18 à 20), l'inversion par P_σ reste sensiblement de qualité constante, alors que P_{12} donne des solutions qui se dégradent. Les inversions par $P_{varégales}$ évoluent globalement comme celles par P_σ . Notons que pour les variances théoriques 5 et 6, les tests T_5 et T_6 préconisent de choisir les solutions issues de $P_{varégales}$.

A partir de variances théoriques égales à 7, (annexe 21 et 22), la qualité des résultats par P_σ se dégrade légèrement. Les résultats par P_{12} peuvent être très bons (annexe 21) et parfois beaucoup moins bons (annexe 22). C'est $P_{varégales}$ qui commence à donner les meilleurs résultats. Les tests pour la méthode par P_{12} sont moins bons que ceux de $P_{varégales}$. P_σ donne de meilleurs tests, que $P_{varégales}$, dans le second exemple.

Les choses se compliquent véritablement lorsque les variances théoriques sont égales à 8 (annexes 23 à 25). Les tests pour les paramètres issus de $P_{varégales}$ sont moins bons que ceux pour les deux autres méthodes, alors que les solutions sont très proches de celles espérées. La méthode par P_{12} n'est plus du tout fiable et trois cas de figures différents peuvent se produire dans l'utilisation de P_σ . En effet, alors que les tests sont toujours meilleurs que ceux de $P_{varégales}$, l'inversion peut-être bonne (annexe 23), mauvaise (annexe 24) ou même être faussée par un mauvais choix d'après le test T_6 (annexe 25).

Pour des variances théoriques supérieures, on constate encore ce type d'évolution. Dans les exemples où les variances théoriques sont égales à 16, (annexes 26 à 29), tous les cas sont possible. Les solutions issues de P_{12} ne sont jamais satisfaisantes, alors que celles issues de $P_{varégales}$ sont très fiables. Le système classique des moments donne des résultats parfois bons, parfois mauvais, qui ne sont, dans les quatre exemples, qu'une seule fois moins adaptés à notre échantillon, que ceux donnés par $P_{varégales}$, d'après T_6 .

En conclusion, à moins que l'on ne soit dans un cas simple, il faudrait trouver un moyen d'identifier le cas où les variances théoriques sont proches, pour pouvoir utiliser $P_{varégales}$.

CAS GENERAL (1)										
	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2		racines >0	acceptable	Test	Commentaires
Annexe 14	5.0	3.0	0.6	1.0	1.0		0.54601120	non		Les deux ensembles illustrent la différence entre les deux systèmes.
	4.98008269	3.01888086	0.60482854	0.98130809	0.96926662		3.82183292	oui	0.00173	
	4.92270225	3.10778865	0.61884743	0.92739623	0.89791942		3.76515225	oui	0.01335	
CAS GENERAL (2)										
	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2		racines >0	acceptable	Test	Commentaires
Annexe 15	1.0	3.0	0.7	3.0	1.0		0.79825763	oui	0.00140	
	1.04705111	3.03885134	0.71893196	3.03461871	0.96590948		1.10372983	oui	0.01191	
	1.07254045	3.06072655	0.73075294	3.05270097	0.91499783		0.77147585	oui	0.00183	
Annexe 16	1.0	3.0	0.7	3.0	1.0		0.72592143	oui	0.01524	La racine de R^- choisie par le test n'est apparemment pas la meilleure
	0.55836086	2.55239809	0.53078014	2.78588282	1.35792041		1.23574038	oui	0.00487	
	0.77195155	2.75498333	0.60739322	2.89014595	1.23370689		1.06939385	oui	0.00874	
Annexe 17	1.0	3.0	0.7	3.0	1.0					Pas de racine pour R^-
	1.07464487	3.09168564	0.72391464	3.01116070	0.91335100		0.74950208	oui	0.01721	






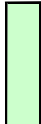
Légende :			
Paramètres théoriques :		Paramètres issus de R^- :	
Paramètres issus de $P12$:		Paramètres issus de $P12$:	
		Racine de R^- choisie :	
		Racine de $P12$ choisie :	

FIG. II.7. -

Le problème des moments dans le cas des moments empiriques.

VARIANCES EGALES - CAS NON SYMETRIQUE. (1)

	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2	racines > 0	acceptable	Test	Commentaires
Annexe 18	1.0	4.0	0.7	5.0	4.0	1.05058972	non		Les trois ensembles sont équivalents
	1.02820590	4.05828756	0.70865146	5.04579476	3.95605071	3.34675593	oui	0.00012	
	1.00517695	4.02746615	0.69849771	4.99209486	4.02746615	3.34757404	T6=0.00057 et T5=0.00264		
Annexe 19	1.02432768	4.05306643	0.70518348	5.03670172	3.96815123	3.34700817	oui	0.00045	Le test pour P varégales est meilleur que pour P.r.
	1.0	5.0	0.7	5.0	5.0	1.07793956	oui	0.05075	
	0.93823711	4.84155907	0.67951486	4.91973880	4.99224449	3.45223752	oui	0.00333	
Annexe 20	0.97321917	4.88801922	0.69161271	4.99747735	4.88801922	3.45406980	T6=0.00265 et T5=0.00332		La solution de S2 se dégrade.
	1.09609562	5.05903492	0.73295496	5.28323069	4.50280426	3.43159273	oui	0.01626	
	1.0	6.0	0.7	5.0	6.0	1.00058277	oui	0.02412	
Annexe 21	1.05018564	6.17043175	0.71368492	5.03736237	5.95786432	3.24849134	oui	0.00388	Le test pour P varégales est meilleur que pour P.r.
	1.00162274	6.10596915	0.69659531	4.92427708	6.10596915	3.25209268	T6=0.00329 et T5=0.00396		
	0.68342115	5.72178472	0.57628467	4.24324819	7.00078	3.09434719	oui	0.02578	
Annexe 22	1.0	7.0	0.7	5.0	7.0	1.15562866	oui	0.02206	Le test pour P varégales est moins bon que pour P.r.
	0.76793667	6.58736062	0.62650304	4.59319165	7.35733570	3.42397853	oui	0.00206	
	0.95230913	6.82366517	0.69177785	4.98949027	6.82366517	3.47525733	T6=0.00037 et T5=0.01206		
Annexe 22	0.98888887	6.87341949	0.70423430	5.07242153	6.71195707	3.47325805	oui	0.01474	Le test pour P varégales est moins bon que pour P.r.
	1.0	7.0	0.7	5.0	7.0	0.98441242	oui	0.02843	
	1.08923689	7.11376078	0.73336313	5.23650200	6.55838727	3.36327878	oui	0.00211	
Annexe 22	0.96523424	6.94282588	0.69124752	4.94841332	6.94282588	3.38613063	T6=0.00336 et T5=0.00985		Le test pour P varégales est moins bon que pour P.r.
	1.25151315	7.35874942	0.78573391	5.65509146	5.98442463	3.26467829	oui	0.01628	

Légende :

Paramètres théoriques : Paramètres issus de P.r. : Racine de P.r. choisie :

Paramètres issus de Parégales : Paramètres issus de P.12 : Racine de P.12 choisie :

FIG. II.8. -

Le problème des moments dans le cas des moments empiriques.

VARIANCES EGALES - CAS NON SYMETRIQUE. (2)

	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2	racines >0	acceptable	Test	Commentaires
Annexe 23	1.0	8.0	0.7	5.0	8.0	1.0699469	oui	0.01236	
	0.94217320	7.95842830	0.68263686	4.91308067	8.08486177	3.4160620	oui	0.00368	
	0.97149150	7.99725947	0.69282798	4.97869673	7.99725947	3.41735620	T6=0.00388 et T5=0.00181		
	3.62014200	9.85779265	0.57425068	0.29013280	7.14661509	2.71110500	oui	0.03298209	
Annexe 24	1.0	8.0	0.7	5.0	8.0	1.40176403	oui	0.00410	T6 privilegie P-r alors que P-variegales semble donner une meilleure solution.
	3.91029447	9.47813005	0.50107487	0.49796474	7.33651943	2.91098608	oui	0.00264	
	1.01816485	7.96808818	0.70316693	5.02591427	7.96808818	3.35252359	T6=0.00584 et T5=0.02534		
	3.17800505	10.5050670	0.68497068	0.09826500	6.59705860	2.04668202	oui	0.01194	
Annexe 25	1.0	8.0	0.7	5.0	8.0	1.11947921	oui	0.00461	Le test ne choisit pas la bonne racine de P-r.
	2.66871634	11.3150539	0.82725017	-0.13014102	5.08336347	3.07609454	oui	0.00486	
	1.10020339	8.27244657	0.72383286	5.02902276	8.27244657	3.08566302	T6=0.00537 et T5=0.00397		
	3.08359495	10.8453704	0.68494833	0.23206032	6.90304353	1.75467586	oui	0.02572	

Légende :	
Paramètres théoriques :	
Paramètres issus de Parégales :	
Paramètres issus de P-r :	
Racine de P-r choisie :	
Paramètres issus de P12 :	
Racine de P12 choisie :	

FIG. II.9. -

Le problème des moments dans le cas des moments empiriques.

VARIANCES EGALES - CAS NON SYMETRIQUE. (3)

	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2	racines >0	acceptable	Test	Commentaires
Annexe 26	1.0	16.0	0.7	5.0	16.0	1.17715817	oui	0.00072	Le test ne choisit pas la bonne racine de P _r .
	2.66716938	19.3210659	0.83000347	-0.22123104	12.8540652	3.20443221	oui	0.00272	
	1.05316579	16.1689338	0.71919590	5.06234368	16.1689338	3.229922404	T6=0.00289 et T5=0.00353		
	2.56049297	19.4206374	0.87030689	-0.40297715	11.6095229	0.99126648	oui	0.01221	
Annexe 27	1.0	16.0	0.7	5.0	16.0	0.82268609	oui	0.00541	T6 privilégie P _r alors que P _r égales semble donner une meilleure solution.
	1.60436444	16.89458054	0.88671179	6.88752314	13.3570709	2.80384831	oui	0.00083	
	1.07395556	16.01814534	0.72014096	5.10787360	16.01814534	3.27952536	T6=0.00094 et T5=0.03364		
	1.45017681	16.6007221	0.84296624	6.24345977	14.4074233	3.04137056	oui	0.01469	
Annexe 28	1.0	16.0	0.7	5.0	16.0				P _r n'a pas de racine positive.
	1.05428059	15.7939404	0.72336408	5.20853599	15.7939404	3.45343986	T6=0.00120 et T5=0.02328		
	3.42534264	18.2952048	0.61139389	0.28116721	14.7012961	2.34879036	oui	0.00564	
Annexe 29	1.0	16.0	0.7	5.0	16.0	0.84781430	oui	0.00389	T6 privilégie P _r égales qui semble donner une meilleure solution.
	1.54482790	16.9135776	0.88432916	6.89264712	13.247605	2.92544060	oui	0.00300	
	0.99999830	16.0002079	0.71639405	5.10222663	16.0002079	3.41906193	T6=0.00113 et T5=0.03503		
	3.30324882	18.503729	0.63463103	0.18356916	15.8330432	2.25669552	oui	0.05541677	

Légende :	
Paramètres théoriques :	
Paramètres issus de P _r :	
Paramètres issus de P _r égales :	
Paramètres issus de P _r 12 :	
Racine de P _r choisie :	
Racine de P _r 12 choisie :	

FIG. II.10. -

Le problème des moments dans le cas des moments empiriques.

10.4 Cas où x_1 et x_2 sont proches (voisinage de V_1).

Nous nous plaçons dans un cadre où les moyennes théoriques sont très proches. Nous commençons avec des variances théoriques très différentes, que nous rapprochons au fur et à mesure, jusqu'à tendre vers deux distributions identiques, (annexes 30 à 39).

Dans le premier exemple, où les moyennes théoriques sont égales à 2 et où les variances théoriques sont 1 et 10, les paramètres issus de l'inversion par P_σ ne sont pas très satisfaisants. Deux indices permettent de rejeter cette solution, puisque le moment d'ordre 3 est proche de 0 (avec $k_4 > 0$) et que le test T_6 est nettement moins bon que dans les exemples que nous avons étudiés jusqu'ici. La solution issue du second système des moments est nettement plus proche de celle espérée. Le fait que le test T_6 ne donne pas un bon résultat illustre l'utilité de remplacer la cinquième équation de moment par la sixième.

Lorsque l'on baisse à 5 la seconde variances théorique, annexes 31, 32 et 33, le troisième moment se rapproche davantage de 0 (toujours avec $k_4 > 0$). Les tests T_6 , lorsque la méthode classique donne une solution, ne sont pas très bons et les solutions proposées ne sont pas très satisfaisantes. La qualité de l'estimation par P_σ pour l'annexe 31, ne peut pas être mise en doute par la qualité du test, puisque $T_6 < 0.1$. L'utilisation de P_{12} , par contre, donne entière satisfaction. Les trois exemples illustrent les trois cas de figures rencontrés pour un tel mélange.

Une fois que les variances théoriques sont 1 et 1.5, le troisième moment centré est très proche de zéro, et le cinquième s'en rapproche, (annexes 34 à 36). La distribution de mélange tend vers une distribution symétrique. Même si, dans le premier exemple, les deux méthodes donnent des solutions équivalentes, il faut faire valoir que le troisième moment centré associé à $k_4 > 0$ incite à l'utilisation du second système des moments \mathfrak{S}_2 . Le troisième exemple, annexe 36, montre, cependant, que l'on s'approche aussi des limites de cette dernière méthode.

Le dernier cas abordé, annexes 37 à 39, est un cas extrême qui montre qu'il existe un seuil de proximité entre les deux distributions composantes au-delà duquel la distribution de mélange ne peut plus être décomposée fiablement. Dans le cas extrême de l'annexe 38 aucun des deux systèmes d'équations de moments ne donne de solution, et l'on doit recourir artificiellement à $P_{\text{varégales}}$.

En conclusion de ce paragraphe, si l'on est en présence d'un moment centré d'ordre trois proche de zéro (on peut, par exemple, se fixer comme seuil $|M_3^c| < 1$) avec $k_4 > 0$, il est nécessaire de privilégier l'utilisation du second système des moments.

10.5 Cas où les variables appartiennent à un voisinage de la variété V_2 .


Il s'agit ici de montrer que les craintes que nous manifestions plus haut dans l'utilisation du premier système des moments s'avèrent justifiées, et que \mathfrak{S}_2 pallie effectivement aux lacunes de \mathfrak{S}_1 .


Le premier exemple, annexe 40, illustre le cas où P_σ a deux racines proches qui engendrent deux ensembles de paramètres quasiment symétriques par rapport à l'ensemble des paramètres théoriques, et dont les tests, eux aussi, sont très proches. Le second exemple, annexe 41, illustre le cas où P_σ n'a pas de racine positive, et où P_{12} remplit son rôle de suppléant. Il est à noter que l'utilisation de P_{12} semble donner, dans les deux cas, une solution plus précise. Les exemples traités en annexe 42 et 43, reproduisent, pour un autre ensemble de paramètres, les mêmes problèmes.

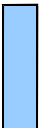
MOYENNES EGALES. (1)


	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2	racines >0	acceptable	Test	Commentaires
Annexe 30	2.0	1.0	0.7	2.0	10.0	0.00061207	oui	0.14433	Mc:3=0.30454
	1.99866243	2.03686989	0.86269476	2.07054562	13.9577671	0.00142259	non		
	1.99234068	0.98766572	0.70006352	2.04632440	9.94291514	0.00023072	non		
Annexe 31	2.0	1.0	0.7	2.0	5.0	0.00004656	oui	0.07170	Mc:3=-0.03661
	1.99778359	1.35166629	0.81875141	1.98006976	5.99496392	0.00010824	non		
	1.99929115	0.95644357	0.67654933	1.98470417	4.78025337	0.00001756	non		
Annexe 32	2.0	1.0	0.7	2.0	5.0	0.00016926	non		P r' n'offre pas de solution.
	2.01262252	1.03223945	0.71515389	1.98379385	5.17189508	0.00039347	non		
	2.01262252	1.03223945	0.71515389	1.98379385	5.17189508	0.00006384	non		
Annexe 33	2.0	1.0	0.7	2.0	5.0	8.579E-6	oui	1.69610	On ne retient pas la solution de P, le test n'étant pas concluant.
	2.00344428	0.90732569	0.66031863	1.99725986	4.65849666	0.00001994	non		
	2.00344428	0.90732569	0.66031863	1.99725986	4.65849666	3.235E-6	non		
						8.578E-6	oui	0.4528	
						0.00002336	non		


Légende :

Paramètres théoriques : 

Paramètres issus de Pr' : 

Racine de Pr choisie : 

Paramètres issus de Parégales : 

Paramètres issus de P12 : 


Racine de P12 choisie : 

FIG. II.11. -

Le problème des moments dans le cas des moments empiriques.

MOYENNES EGALES. (2)

	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2	racines >0	acceptable	Test	Commentaires
Annexe 34	2.0	1.0	0.7	2.0	1.5	0.00017709	oui	0.00044	Les deux ensembles de paramètres sont équivalents.
	2.00384168	1.06646236	0.88052823	1.96281222	1.78266622	0.00041134	non		
	2.00404351	1.06292532	0.87176105	1.96424523	1.75776568	0.00006660	non		
Annexe 35	2.0	1.0	0.7	2.0	1.5	0.00048132	non		Mc:3=-0.00928
	2.00100024	1.10105695	0.95154527	1.96288043	2.16626009	0.00006700	oui	0.01130	
	2.00429738	1.00869527	0.71659951	1.98624557	1.51672509	0.00015566	non		
Annexe 36	2.0	1.0	0.7	2.0	1.5	0.0002522	non		Les moyennes et variances sont très proches, l'inversion est difficile pour les deux systèmes
	1.99552815	1.23711980	0.92836941	2.03802810	0.23581681	0.00006691	oui	0.11358	
	1.98881301	1.39477828	0.55820917	2.01090363	0.87556735	0.00018211	non		
						0.00012011	oui	0.01460	
						0.00027940	non		
						0.00004539	non		
						0.00012035	oui	0.27044	
						0.00032765	non		

Légende :	
Paramètres théoriques :	
Paramètres issus de P12 :	
Paramètres issus de Pr :	
Paramètres issus de P12 :	
Racine de Pr choisie :	
Racine de P12 choisie :	

FIG. II.12. -

Le problème des moments dans le cas des moments empiriques.

MOYENNES EGALES. (3)

	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2	racines >0	acceptable	Test	Commentaires
Annexe 37	2.0	1.0	0.7	2.0	1.1	2.519E-6	oui	0.06385	Les moyennes et variances sont trop proches pour être différenciées
	2.00088082	1.03235467	0.99999982	-1.77179499	82.0626657	5.449E-6	non		
Annexe 38	2.00089840	1.03186889	0.99984037	1.88660797	4.16760552	7.813E-7	non		aucun des deux système ne donne de solution.
	2.0	1.0	0.7	2.0	1.1	2.085E-6	oui	0.47034	
Annexe 39	1.99183026	1.02730978	0.99860301	3.61424194	1.02730978	5.655E-6	non		de solution.
	2.0	1.0	0.7	2.0	1.1	0.00058529	non		
Annexe 39	1.98058864	1.01132847	0.99483680	3.91010559	0.2569210	0.00077491	non		de solution.
	2.0	1.0	0.7	2.0	1.1	0.00305778	non		
						0.00393450	T6=0.00285 et T5=0.03932		
						0.01912354	oui	0.22614	

Légende :	
Paramètres théoriques :	Paramètres issus de P_{σ} :
Paramètres issus de Paragales :	Paramètres issus de P12 :
	Racine de P_{σ} choisie :
	Racine de P12 choisie :

FIG. II.13. -

Le problème des moments dans le cas des moments empiriques.

VOISINAGE DE V_2

	Moyenne F1	Variance F1	P	Moyenne F2	Variance F2	racines >0	acceptable	Test	Commentaires
Annexe 40	2.09243444	1.0	0.5	-0.09243444	3.0	1.11536287	oui	0.00134	P σ a deux racines proches.
	2.10536756	0.97255443	0.48036213	-0.00847891	3.11530011	1.26575051	oui	0.00414	
	2.09133500	1.00641536	0.49508118	-0.05634097	3.06998975	1.15301642	oui	0.00056	
Annexe 41	2.09243444	1.0	0.5	-0.09243444	3.0				P σ n'a pas de racine positive.
	2.10050904	0.98264671	0.48999329	-0.05861019	3.01472671	1.16498215	oui	0.00067	
Annexe 42	0.77246780	1.0	0.5	-0.77246780	2.0	0.47636331	oui	0.00492	P σ a deux racines proches.
	0.68861584	1.09840019	0.60718480	-1.05777158	1.78554618	0.72576341	oui	0.00431	
	0.75338008	1.03561784	0.52207853	-0.82007115	1.95781281	0.61773036	oui	0.00553	
Annexe 43	0.77246780	1.0	0.5	-0.77246780	2.0				P σ n'a pas de racine positive.
	0.78827180	0.96021870	0.48314645	-0.73637990	2.03091248	0.58048035	oui	0.00490	






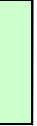
Légende :	
Paramètres théoriques :	
Paramètres issus de Paregales :	
Paramètres issus de P σ :	
Paramètres issus de P 12 :	
Racine de P σ choisie :	
Racine de P 12 choisie :	

FIG. II.14. -

Le problème des moments dans le cas des moments empiriques.

10.6 Cas où les variables appartiennent à un voisinage de la variété V_4 .

Nous nous cantonnerons, pour V_4 , au cas symétrique où les deux variances théoriques sont identiques et les proportions des deux distributions égales, (annexes 44 à 46). Cela viendra compléter l'étude faite au voisinage de V_1 . En effet, nous y avons envisagé le cas où $|M_3^c| < 1$ avec $k_4 > 0$ et l'exemple que nous traitons maintenant est celui où $k_4 < 0$. Cette remarque permet de mettre en concurrence, comme le montrent les trois exemples présentés, la méthode par P_σ et celle par $P_{varégales}$. La méthode par P_{12} n'aboutissant pas toujours, elle ne peut donner satisfaction.

10.7 Conclusion.

L'inversion par le polynôme P_{12} du second système semble très complémentaire de celle du premier système par le polynôme P_σ pour trouver la solution, lorsqu'elle existe, du problème des moments. Il apparaît, en effet que P_σ donne de très bonnes estimations lorsque les ensembles de paramètres sont éloignés des variétés V_1 et V_2 , alors que l'intérêt de l'estimation par la méthode P_{12} se précise lorsque l'on se rapproche de ces variétés. Le second système des moments doit cependant être utilisé avec précaution en complément du premier système, puisque nous savons que la précision des estimateurs des moments empiriques diminue avec l'augmentation de l'ordre de ces moments. Le chapitre suivant propose une synthèse de l'ensemble des contraintes théoriques et des problèmes particuliers soulevés par le mélange de deux distributions normales, débouchant sur une méthode fiable d'estimation des paramètres.

VARIANCES EGALES - CAS SYMETRIQUE (V4).

	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2	racines >0	acceptable	Test	Commentaires
Annexe 44	1.0	1.0	0.5	5.0	1.0	0.00011608	non		T6 préconise le choix de la solution de P σ .
	0.98848374	0.97962575	0.49755612	4.98878254	1.01665021	4.00050206	oui	0.00225	
	1.00235430	0.99809997	0.50102661	5.00267799	0.99809997	4.00063055	T6=0.00227 et T5=0.25035		
Annexe 45	1.0	1.0	0.5	5.0	1.0	0.00003336	non		T6 préconise le choix de la solution de P varégales.
	1.00578288	0.99692069	0.50020967	5.00584061	1.00055921	4.00011477	oui	0.001014	
	1.00714446	0.99873595	0.50055040	5.00720488	0.99873595	4.00011600	T6=0.001012 et T5=0.07754		
Annexe 46	1.0	1.0	0.5	5.0	1.0	0.00068181	non		Les tests sont mauvais, on ne retient pas ces solutions.
	1.00579630	1.03666526	0.50312787	5.00796237	0.97669277	4.00417662	oui	0.00022	
	0.98325364	1.00662488	0.49751277	4.9855918	1.00652488	4.00451331	T6=0.00027 et T5=0.20004		
	1.09106412	1.15170995	0.52455407	5.09424696	0.86276401	3.99670640	oui	0.76506	
						0.00349219	non		
						1.54463489	non		
						1.773844448	non		
						3.80174634	oui	5.00515	

Légende :

Paramètres théoriques : Paramètres issus de P σ : Racine de P σ choisie :

Paramètres issus de P varégales : Paramètres issus de P 12 : Racine de P 12 choisie :

FIG. II.15. -

Le problème des moments dans le cas des moments empiriques.

- CHAPITRE 11 -

Méthode d'estimation des paramètres d'une distribution de mélange de deux distributions normales inconnues.

11.1 Exposé de la méthode proposée.

Après avoir récapitulé les différentes méthodes d'estimation avec les différents systèmes, nous énoncerons la procédure générale d'estimation du mélange de deux distributions normales, par la méthode des moments.

11.1.1 Estimation des paramètres avec P_σ .

Par une méthode adéquate, on cherche les racines positives du polynôme :

$$P_\sigma(X) = 8X^9 + 28\kappa_4 X^7 - 12\mathfrak{M}_3^{c2} X^6 + (30\kappa_4^2 + 24\kappa_5 \mathfrak{M}_3^c) X^5 + (6\kappa_5^2 - 148\kappa_4 \mathfrak{M}_3^{c2}) X^4 + (-96\mathfrak{M}_3^{c4} + 36\kappa_4 \kappa_5 \mathfrak{M}_3^c - 9\kappa_4^3) X^3 + (21\kappa_4^2 + 24\mathfrak{M}_3^c \kappa_5) \mathfrak{M}_3^{c2} X^2 - 32\mathfrak{M}_3^{c4} \kappa_4 X + 8\mathfrak{M}_3^{c6},$$

où κ_4 , κ_5 et \mathfrak{M}_3 sont les estimateurs respectifs de k_4 , k_5 et M_3^c . Les moments et les cumulants sont estimés grâce aux estimateurs empiriques des moments.

Si X^* est la racine positive permettant d'obtenir les estimateurs des paramètres $(p^*, \mu_1^*, \mu_2^*, \sigma_1^*, \sigma_2^*)$, on calcule :

$$\begin{aligned} \sigma'^* &= \frac{2\mathfrak{M}_3^c X^{*3} + \kappa_5 X^{*2} - 3\kappa_4 X^* \mathfrak{M}_3^c + 2\mathfrak{M}_3^{c3}}{-2X^{*3} - 3\kappa_4 X^* + 4\mathfrak{M}_3^{c2}} \\ \text{et } \sigma^* &= \mathfrak{M}_2^c - X^*. \end{aligned}$$

Les paramètres sont alors estimés par :

$$\begin{aligned}
 p^* &= \frac{\sqrt{4X^{*3} + (\mathfrak{M}_3^c - 3\sigma'^*)^2} + |\mathfrak{M}_3^c - 3\sigma'^*|}{2\sqrt{4X^{*3} + (\mathfrak{M}_3^c - 3\sigma'^*)^2}}, \\
 \mu_1^* &= \mathfrak{M} - \operatorname{sgn}(\mathfrak{M}_3^c - 3\sigma'^*) \sqrt{\frac{\sqrt{4X^{*3} + (\mathfrak{M}_3^c - 3\sigma'^*)^2} - |\mathfrak{M}_3^c - 3\sigma'^*|}{\sqrt{4X^{*3} + (\mathfrak{M}_3^c - 3\sigma'^*)^2} + |\mathfrak{M}_3^c - 3\sigma'^*|}} X^*, \\
 \mu_2^* &= \mathfrak{M} + \operatorname{sgn}(\mathfrak{M}_3^c - 3\sigma'^*) \cdot \sqrt{\frac{\sqrt{4X^{*3} + (\mathfrak{M}_3^c - 3\sigma'^*)^2} + |\mathfrak{M}_3^c - 3\sigma'^*|}{\sqrt{4X^{*3} + (\mathfrak{M}_3^c - 3\sigma'^*)^2} - |\mathfrak{M}_3^c - 3\sigma'^*|}} X^*, \\
 \sigma_1^* &= \frac{\sigma'^* - \sigma^* (\mu_2^* - \mathfrak{M})}{p^* (\mu_1^* - \mu_2^*)} \\
 \text{et } \sigma_2^* &= \frac{\sigma'^* - \sigma^* (\mu_1^* - \mathfrak{M})}{(1 - p^*) (\mu_2^* - \mu_1^*)},
 \end{aligned}$$

où \mathfrak{M} et \mathfrak{M}_3^c , sont, respectivement, les estimateurs de M et M_3^c .

Lorsqu'il y a plusieurs solutions statistiquement raisonnables au problème des moments, nous choisirons celle qui minimisera le test :

$$T_6^* = \left| \frac{p^* (15 \sigma_1^{*3} + 45 \sigma_1^{*2} x_1^{*2} + 15 \sigma_1^* x_1^{*4} + x_1^{*6}) + (1 - p^*) (15 \sigma_2^{*3} x_2^{*4} + 45 \sigma_2^{*2} x_2^{*2} + 15 \sigma_2^{*3} + x_2^{*6}) - \mathfrak{M}_6^c}{2\mathfrak{M}_6^c} \right|,$$

où \mathfrak{M}_6^c est l'estimateur de M_6^c . Notons, enfin que nous rejèterons toute solution pour laquelle $T_6^* > 10$.

11.1.2 Cas où les variances sont égales.

On se ramène dans ce cas à la recherche des racines positives du polynôme de degré 3 :

$$P_{\text{varégales}}(X) = 2X^3 + \kappa_4 X - \mathfrak{M}_3^{c2}.$$

Si X^* est la racine, on calcule σ^* , p^* , μ_1^* et μ_2^* :

$$\begin{aligned}
 \sigma^* &= \mathfrak{M}_2^c - X^*, \\
 p^* &= \frac{\sqrt{4X^{*3} + \mathfrak{M}_3^{c2}} + |\mathfrak{M}_3^c|}{2\sqrt{4X^{*3} + \mathfrak{M}_3^{c2}}}, \\
 \mu_1^* &= \mathfrak{M} - \operatorname{sgn}(\mathfrak{M}_3^c) \sqrt{\frac{\sqrt{4X^{*3} + \mathfrak{M}_3^{c2}} - |\mathfrak{M}_3^c|}{\sqrt{4X^{*3} + \mathfrak{M}_3^{c2}} + |\mathfrak{M}_3^c|}} X^* \\
 \text{et } \mu_2^* &= \mathfrak{M} + \operatorname{sgn}(\mathfrak{M}_3^c) \cdot \sqrt{\frac{\sqrt{4X^{*3} + \mathfrak{M}_3^{c2}} + |\mathfrak{M}_3^c|}{\sqrt{4X^{*3} + \mathfrak{M}_3^{c2}} - |\mathfrak{M}_3^c|}} X^*.
 \end{aligned}$$

Dans le cas où plusieurs solutions seraient statistiquement acceptables, nous choisirions celle minimisant l'un des tests :

$$T_6^* = \left| \frac{p^* (15 \sigma_1^{*3} + 45 \sigma_1^{*2} x_1^{*2} + 15 \sigma_1^* x_1^{*4} + x_1^{*6}) + (1 - p^*) (15 \sigma_2^* x_2^{*4} + 45 \sigma_2^{*2} x_2^{*2} + 15 \sigma_2^{*3} + x_2^{*6}) - \mathfrak{M}_6^c}{2\mathfrak{M}_6^c} \right|,$$

où

$$T_5^* = \left| \frac{p^* (x_1^{*5} + 10 \sigma_1^* x_1^{*3} + 15 \sigma_1^{*2} x_1^*) + (1 - p^*) (x_2^{*5} + 10 \sigma_2^* x_2^{*3} + 15 \sigma_2^{*2} x_2^*) - \mathfrak{M}_5^c}{2\mathfrak{M}_5^c} \right|.$$

11.1.3 Cas où les moyennes sont les mêmes.

D'après Cohen, (voir [Cohen67]) les paramètres autres que les moyennes (égales à la moyenne empirique), seront estimés par :

$$\begin{aligned} \sigma_1^* &= \mathfrak{M}_2^c + \frac{1}{2} \left[\frac{\kappa_6}{5\kappa_4} - \sqrt{\left(\frac{\kappa_6}{5\kappa_4}\right)^2 + \frac{4\kappa_4}{3}} \right], \\ \sigma_2^* &= \mathfrak{M}_2^c + \frac{1}{2} \left[\frac{\kappa_6}{5\kappa_4} + \sqrt{\left(\frac{\kappa_6}{5\kappa_4}\right)^2 + \frac{4\kappa_4}{3}} \right] \\ \text{et } p^* &= \frac{\sigma_2^* - M_2}{\sigma_2^* + \sigma_1^*}. \end{aligned}$$

ou κ_6 est l'estimateur du sixième cumulant du mélange $k_6 = M_6^c - 15M_4^c M_2^c - 10M_3^{c2} + 30M_2^{c3}$.

11.1.4 Estimation des paramètres avec P_{12} .

Par une méthode adéquate, on cherche, comme pour P_σ , les racines positives du polynôme :

$$\begin{aligned} P_{12}(X) &= 96X^{12} + 384\kappa_4 X^{10} + 8(6\kappa_6 + 13 \mathfrak{M}_3^{c2})X^9 + 336\kappa_4^2 X^8 + 12\kappa_4(8\kappa_6 + 5\mathfrak{M}_3^{c2})X^7 \\ &\quad + 6(-16\kappa_4^3 + \kappa_6^2 + 4 \kappa_6 \mathfrak{M}_3^{c2} + 22 \mathfrak{M}_3^{c4})X^6 - 6\kappa_4^2(2\kappa_6 + 47\mathfrak{M}_3^{c2})X^5 \\ &\quad + 6\kappa_4(\kappa_4^3 - 2(4\kappa_6\mathfrak{M}_3^{c2} + 5\mathfrak{M}_3^{c4}))X^4 + (97\kappa_4^3\mathfrak{M}_3^{c2} + 48\mathfrak{M}_3^{c4}(\kappa_6 + 7\mathfrak{M}_3^{c2}))X^3 \\ &\quad - 141\kappa_4^2\mathfrak{M}_3^{c4}X^2 + 48 \kappa_4 \mathfrak{M}_3^{c6}X - 4 \mathfrak{M}_3^{c8}, \end{aligned}$$

où κ_4 , κ_5 , κ_6 et \mathfrak{M}_3^c sont les estimateurs respectifs de k_4 , k_5 , k_6 et M_3^c . Les moments et les cumulants sont estimés grâce aux estimateurs empiriques des moments.

Si X^* est la racine positive permettant d'obtenir les estimateurs des paramètres (p^* , μ_1^* , μ_2^* , σ_1^* , σ_2^*), on calcule maintenant:

$$\begin{aligned} \sigma^* &= \frac{4X^6 + 8X^4\kappa_4 + (\kappa_6 + 2\mathfrak{M}_3^{c2})X^3 - \kappa_4^2X^2 - 4\kappa_4\mathfrak{M}_3^{c2}X + 4\mathfrak{M}_3^{c4}}{-2X^3 - 7X\kappa_4 + 10\mathfrak{M}_3^{c2}} \\ \text{et } \sigma^* &= \mathfrak{M}_2^c - X^*. \end{aligned}$$

Les paramètres sont alors estimés, comme précédemment, par :

$$\begin{aligned}
 p^* &= \frac{\sqrt{4X^{*3} + (\mathfrak{M}_3^c - 3\sigma'^*)^2} + |\mathfrak{M}_3^c - 3\sigma'^*|}{2\sqrt{4X^{*3} + (\mathfrak{M}_3^c - 3\sigma'^*)^2}}, & [11.45] \\
 \mu_1^* &= \mathfrak{M} - \text{sgn}(\mathfrak{M}_3^c - 3\sigma'^*) \sqrt{\frac{\sqrt{4X^{*3} + (\mathfrak{M}_3^c - 3\sigma'^*)^2} - |\mathfrak{M}_3^c - 3\sigma'^*|}{\sqrt{4X^{*3} + (\mathfrak{M}_3^c - 3\sigma'^*)^2} + |\mathfrak{M}_3^c - 3\sigma'^*|}} X^*, \\
 \mu_2^* &= \mathfrak{M} + \text{sgn}(\mathfrak{M}_3^c - 3\sigma'^*) \cdot \sqrt{\frac{\sqrt{4X^{*3} + (\mathfrak{M}_3^c - 3\sigma'^*)^2} + |\mathfrak{M}_3^c - 3\sigma'^*|}{\sqrt{4X^{*3} + (\mathfrak{M}_3^c - 3\sigma'^*)^2} - |\mathfrak{M}_3^c - 3\sigma'^*|}} X^*, \\
 \sigma_1^* &= \frac{\sigma'^* - \sigma^* (\mu_2^* - \mathfrak{M})}{p^* (\mu_1^* - \mu_2^*)} \\
 \text{et } \sigma_2^* &= \frac{\sigma'^* - \sigma^* (\mu_1^* - \mathfrak{M})}{(1 - p^*) (\mu_2^* - \mu_1^*)},
 \end{aligned}$$

où \mathfrak{M} et \mathfrak{M}_5^c , sont, respectivement, les estimateurs de M et M_5^c .

S'il y a plusieurs solutions statistiquement acceptables au problème de moments, nous choisirons celle qui minimisera le test :

$$T_5^* = \left| \frac{p^* (x_1^{*5} + 10\sigma_1^* x_1^{*3} + 15\sigma_1^{*2} x_1^*) + (1 - p^*) (x_2^{*5} + 10\sigma_2^* x_2^{*3} + 15\sigma_2^{*2} x_2^*) - \mathfrak{M}_5^c}{2\mathfrak{M}_5^c} \right|.$$

Comme nous l'avons fait pour P_σ , nous fixons une limite supérieure à la validation d'une solution par ce test, et nous rejetons une solution pour laquelle $T_5^* > 10$.

11.1.5 Méthode générale d'estimation des paramètres d'un mélange de deux distributions normales.

Les problèmes d'unicité étudiés pour les systèmes de moments théoriques, \mathfrak{S}_1 et \mathfrak{S}_2 , se compliquent lors du passage au système de moments empiriques $\vec{M}_n = \mathfrak{F}(\hat{\theta})$. En effet, il n'y a pas continuité, en fonction des moments du mélange, des solutions des systèmes théoriques au voisinage de $V_1 \cup V_2$ ou V_3 selon les cas. La méthode d'estimation que nous proposons permet de choisir le système pour lequel l'estimation est la plus fiable, et permet de pallier aux cas d'instabilité présentés dans les articles de référence [Fryer70] et [Fryer72] dont nous rappelons le contenu. Dans leur article [Fryer70] consacré à l'étude de la stabilité de la méthode des moments pour le mélange de deux distributions normales inconnues Fryer et Robertson, envisagent neuf paramétrages différents pour le mélange de deux distributions normales. Ils mettent ainsi en évidence trois cas pour lesquels la méthode présente un biais très important, qu'ils ne parviennent pas à expliquer. Ces trois exemples sont situés dans un voisinage immédiat de V_2 , cas dont nous avons discuté la stabilisation plus haut. Nous sommes par conséquent en mesure de justifier le biais important constaté par Fryer et Robertson par le fait que \mathfrak{S}_1 n'est pas approprié à l'estimation des paramètres dans ces trois exemples. Il est à noter que la méthode de mesure du biais proposée par Fryer et Robertson ne fait pas apparaître de biais important dans le seul cas de mélange choisi au voisinage de V_1 . Les exemples que nous avons traités dans le chapitre précédent montrent pourtant une épaisseur non négligeable du voisinage de V_1 sur lequel l'utilisation de la méthode des moments basée sur \mathfrak{S}_1 n'est pas fiable. Dans l'article [Fryer72] de comparaison de différentes méthodes d'estimation des paramètres du mélange de deux distributions normales inconnues, Fryer et Robertson envisagent neuf nouveaux exemples de paramétrages. Ils constatent dans quatre cas, l'infériorité de

la méthode des moments par rapport à une méthode de maximisation de la vraisemblance et à une méthode de minimisation d'une distance du χ^2 , les cinq autres cas n'étant pas défavorables à la méthode des moments. Ces quatre cas, à l'origine de l'abandon progressif de la méthode des moments pour l'estimation des paramètres de distributions de mélanges, sont, d'après notre étude, à traiter avec le système \mathfrak{S}_2 . En effet, deux cas sont issus d'un proche voisinage de V_1 et deux autres d'un proche voisinage de V_2 .

La méthode d'estimation que nous proposons favorise le système des cinq premières équations de moments, et ne préconise le recours au second système que dans un voisinage de V_1 ou V_2 . Cette méthode tire les enseignements des quarante six exemples traités ici et des dix-huit cas étudiés dans les articles de Fryer et Robertson:

- ▶ 1. Si $|\mathfrak{M}_3^c| < 0.5$ et $\mathfrak{M}_3^c \neq 0$:
 - ▷ Si $\kappa_4 > 0$, l'estimation vient de \mathfrak{S}_2 .
 - ▷ Si $\kappa_4 \leq 0$, solution issue de P_σ ou $P_{varégales}$ minimisant le test \mathcal{T}_6 .
- ▶ 2. Si $\mathfrak{M}_3^c = 0$:
 - ▷ Si $\kappa_4 > 0$, cas particulier de \mathfrak{S}_2 où $\mu_1 = \mu_2$.
 - ▷ Si $\kappa_4 \leq 0$, choix de la solution issue de $P_{varégales}$.
- ▶ 3. On choisira la solution de \mathfrak{S}_2 si :
 - ▷ La solution issue de \mathfrak{S}_2 vérifie :

$$T_{V_2} = 10 \left(|x_1 - x_2| - (6\sqrt{6} - 9)^{1/4} \sqrt{|\sigma_1^2 - \sigma_2^2|} \right)^2 < 2,$$

- ▷ Si \mathfrak{S}_1 ne donne pas de solution acceptable.
- ▶ 4. Sinon la solution vient de \mathfrak{S}_1 .

- CHAPITRE 12 -

Estimateurs des moments d'un mélange exponentielle-exponentielle.

Le système des moments, non centrés, pour un mélange de deux distributions exponentielles, de paramètres respectifs m_1 et m_2 , est :

$$\mathfrak{S}_3 = \begin{cases} pm_1 + (1-p)m_2 = M_1 \\ 2pm_1^2 + 2(1-p)m_2^2 = M_2 \\ 6pm_1^3 + 6(1-p)m_2^3 = M_3 \end{cases} .$$

12.1 Calcul du jacobien de \mathfrak{S}_3 .

Le jacobien J_3 de \mathfrak{S}_3 est

$$J_3 = \begin{vmatrix} p & 1-p & m_1 - m_2 \\ 4pm_1 & 4(1-p)m_2 & 2m_1^2 - 2m_2^2 \\ 18pm_1^2 & 18(1-p)m_2^2 & 6m_1^3 - 6m_2^3 \end{vmatrix} ,$$

que l'on développe et factorise directement en :

$$J_3 = 12p(1-p)(m_1 - m_2)^4 .$$

Ce jacobien ne s'annule que si $m_1 = m_2$, c'est à dire si l'on n'est pas en présence d'un mélange. Il n'y a, donc, pas pour ce type de mélange, de problèmes équivalents à ceux signalés pour le mélange de deux distributions normales.

12.2 Résolution algébrique du système \mathfrak{S}_3 .

Cette étude a été effectuée par Rider (voir [Rider61] et [Rider62]). La résolution du système des moments (non centrés), donne deux triplets solutions calculables littéralement, lorsque

$$12 M_1^2 - 6 M_2 \neq 0 \text{ et } -9 (3 M_1^2 - 2 M_2) M_2^2 + 6 M_1 (4 M_1^2 - 3 M_2) M_3 + M_3^2 > 0.$$

Ces solutions sont symétriques (par la transformation $m_1 \mapsto m_2$, et $p \mapsto (1 - p)$). Ces solutions sont données ci-dessous (les M_i étant les moments d'ordre i) :

$$\begin{aligned} m_1 &= \frac{3 M_1 M_2 - M_3 + \sqrt{-9 (3 M_1^2 - 2 M_2) M_2^2 + 6 M_1 (4 M_1^2 - 3 M_2) M_3 + M_3^2}}{12 M_1^2 - 6 M_2}, \\ p &= \frac{1}{2} + \frac{12 M_1^3 - 9 M_1 M_2 + M_3}{2 \sqrt{-9 (3 M_1^2 - 2 M_2) M_2^2 + 6 M_1 (4 M_1^2 - 3 M_2) M_3 + M_3^2}}, \\ m_2 &= -\frac{-3 M_1 M_2 + M_3 + \sqrt{-9 (3 M_1^2 - 2 M_2) M_2^2 + 6 M_1 (4 M_1^2 - 3 M_2) M_3 + M_3^2}}{12 M_1^2 - 6 M_2}, \end{aligned}$$

ainsi que

$$\begin{aligned} m_1 &= -\frac{-3 M_1 M_2 + M_3 + \sqrt{-9 (3 M_1^2 - 2 M_2) M_2^2 + 6 M_1 (4 M_1^2 - 3 M_2) M_3 + M_3^2}}{12 M_1^2 - 6 M_2}, \\ p &= \frac{1}{2} + \frac{12 M_1^3 - 9 M_1 M_2 + M_3}{2 \sqrt{-9 (3 M_1^2 - 2 M_2) M_2^2 + 6 M_1 (4 M_1^2 - 3 M_2) M_3 + M_3^2}}, \\ m_2 &= \frac{3 M_1 M_2 - M_3 + \sqrt{-9 (3 M_1^2 - 2 M_2) M_2^2 + 6 M_1 (4 M_1^2 - 3 M_2) M_3 + M_3^2}}{12 M_1^2 - 6 M_2}. \end{aligned}$$

L'inversion de ce système donne une solution unique, il n'est donc pas utile de définir un critère de choix pour distinguer des solutions différentes. La forme des solutions n'assure pas, pour autant, que les solutions obtenues par inversion du système soient statistiquement acceptables.

12.3 Exemples d'inversion du système \mathfrak{S}_3 .

L'étude de l'inversion doit se concentrer autour du cas annulant le jacobien, c'est à dire lorsque les deux distributions composant le mélange sont proches l'une de l'autre. Les résultats des exemples traités dans les annexes 47 à 55, sont récapitulés dans le tableau FIG. II.16.

Les résultats des exemples traités pour les moments théoriques, cités en annexe 47 et 48, montrent que d'éventuelles erreurs dans la résolution par le système empirique, ne viendraient que d'un manque de précision des moments empiriques.

Le cas général, pour les moments empiriques, cité en annexe 49, est bien estimé par la méthode. Les problèmes des annexes 50 à 52, sont essentiellement dus au fait que la distribution de mélange se rapproche trop d'une distribution exponentielle simple. Les annexes 53 à 54, montrent, dans un cas extrême, les dégénérescences liées à la méthode.

MELANGES EXPONENTIELLE-EXPONENTIELLE.								
	Paramètre F1	p	Paramètre F2	Commentaires				
Annexe 47	3.0	0.7	1.0	Système théorique				
	3.0	0.7	1.0					
Annexe 48	1.0001	0.7	1.0	Système théorique				
	1.0001	0.7	1.0					
Annexe 49	2.0	0.3	1.0	cas général				
	2.046887687	0.281250057	1.010775899					
Annexe 50	1.2	0.3	1	Tous les paramètres sont bien estimés.				
	1.230788933	0.2578908341	0.9972737237					
Annexe 51	1.2	0.3	1	La distribution de mélange est trop proche d'une distribution exponentielle.				
	2.175493119	0.00951502503	1.043919526					
Annexe 52	1.2	0.5	1	p est mal estimé.				
	1.2720162	0.2248471473	1.045829274					
Annexe 53	1.1	0.3	1	La distribution de mélange est trop proche d'une distribution exponentielle.				
	10.69721687	2.72E-6	1.028152786					
Annexe 54	1.1	0.3	1	p est supérieur à 1				
	1.033271822	1.004360216	1.959452341					
Annexe 55	1.1	0.3	1	certains paramètres sont négatifs				
	-0.5128172088	-0.001361234	1.030250451					
<table border="1" style="width: 100%;"> <tr> <td style="width: 80%;">Paramètres théoriques :</td> <td style="width: 20%;"></td> </tr> <tr> <td>Paramètres estimés :</td> <td></td> </tr> </table>					Paramètres théoriques :		Paramètres estimés :	
Paramètres théoriques :								
Paramètres estimés :								

FIG. II.16. -

Estimateurs des moments d'un mélange exponentielle-exponentielle.

- CHAPITRE 13 -

Estimateurs des moments d'un mélange normale-exponentielle.

Le système des moments, non centrés, pour le mélange d'une distributions normale $N(\mu_1, \sigma_1)$ et d'une distributions exponentielles, de paramètre m_2 , s'écrit :

$$\mathfrak{S}_4 = \begin{cases} M_1 = p\mu_1 + (1-p)m_2 \\ M_2 = p(\mu_1^2 + \sigma_1) + 2(1-p)m_2^2 \\ M_3 = p(\mu_1^3 + 3\mu_1\sigma_1) + 6(1-p)m_2^3 \\ M_4 = p(\mu_1^4 + 6\mu_1^2\sigma_1 + 3\sigma_1^2) + 24(1-p)m_2^4 \end{cases} .$$

13.1 Calcul du jacobien.

Le jacobien J_4 de \mathfrak{S}_4 est, si l'on factorise p dans les premières et dernières colonnes, et $(1-p)$ dans la deuxième,

$$J_4 = p^2(1-p) \begin{vmatrix} 1 & 1 & \mu_1 - m_2 & 0 \\ 2\mu_1 & 4m_2 & \mu_1^2 + \sigma_1 - 2m_2^2 & 1 \\ 3(\mu_1^2 + \sigma_1) & 18m_2^2 & \mu_1^3 + 3\sigma_1\mu_1 - 6m_2^3 & 3\mu_1 \\ 4(\mu_1^3 + 3\sigma_1\mu_1) & 96m_2^3 & \mu_1^4 + 6\sigma_1\mu_1^2 + 3\sigma_1^2 - 24m_2^4 & 6(\mu_1^2 + \sigma_1) \end{vmatrix},$$

en ajoutant à la troisième colonne m_2 fois la deuxième, on a

$$J_4 = p^2(1-p) \begin{vmatrix} 1 & 1 & \mu_1 & 0 \\ 2\mu_1 & 4m_2 & \mu_1^2 + \sigma_1 + 2m_2^2 & 1 \\ 3(\mu_1^2 + \sigma_1) & 18m_2^2 & \mu_1^3 + 3\sigma_1\mu_1 + 12m_2^3 & 3\mu_1 \\ 4(\mu_1^3 + 3\sigma_1\mu_1) & 96m_2^3 & \mu_1^4 + 6\sigma_1\mu_1^2 + 3\sigma_1^2 + 72m_2^4 & 6(\mu_1^2 + \sigma_1) \end{vmatrix}.$$

Estimateurs des moments d'un mélange normale-exponentielle.

Si l'on retranche, alors, à la troisième colonne μ_1 fois la première, on a

$$J_4 = p^2(1-p) \begin{vmatrix} 1 & 1 & 0 & 0 \\ 2\mu_1 & 4m_2 & -\mu_1^2 + \sigma_1 + 2m_2^2 & 1 \\ 3(\mu_1^2 + \sigma_1) & 18m_2^2 & -2\mu_1^3 + 12m_2^3 & 3\mu_1 \\ 4(\mu_1^3 + 3\sigma_1\mu_1) & 96m_2^3 & -3\mu_1^4 - 6\sigma_1\mu_1^2 + 3\sigma_1^2 + 72m_2^4 & 6(\mu_1^2 + \sigma_1) \end{vmatrix}.$$

Afin de pouvoir développer le déterminant suivant la première ligne, dans de bonnes conditions, on retire la deuxième colonne à la première pour obtenir :

$$J_4 = p^2(1-p) \begin{vmatrix} 0 & 1 & 0 & 0 \\ 2\mu_1 - 4m_2 & 4m_2 & -\mu_1^2 + \sigma_1 + 2m_2^2 & 1 \\ 3(\mu_1^2 + \sigma_1) - 18m_2^2 & 18m_2^2 & -2\mu_1^3 + 12m_2^3 & 3\mu_1 \\ 4(\mu_1^3 + 3\sigma_1\mu_1) - 96m_2^3 & 96m_2^3 & -3\mu_1^4 - 6\sigma_1\mu_1^2 + 3\sigma_1^2 + 72m_2^4 & 6(\mu_1^2 + \sigma_1) \end{vmatrix}$$

et après développement :

$$J_4 = -p^2(1-p) \begin{vmatrix} 2\mu_1 - 4m_2 & -\mu_1^2 + \sigma_1 + 2m_2^2 & 1 \\ 3(\mu_1^2 + \sigma_1) - 18m_2^2 & -2\mu_1^3 + 12m_2^3 & 3\mu_1 \\ 4(\mu_1^3 + 3\sigma_1\mu_1) - 96m_2^3 & -3\mu_1^4 - 6\sigma_1\mu_1^2 + 3\sigma_1^2 + 72m_2^4 & 6(\mu_1^2 + \sigma_1) \end{vmatrix},$$

ce qui permet d'obtenir :

$$J_4 = -p^2(1-p) \left(+3 \left(\begin{array}{l} -\mu_1^6 + 12 \mu_1^5 m_2 - 66 \mu_1^4 m_2^2 + 192 \mu_1^3 m_2^3 - 288 \mu_1^2 m_2^4 + 288 \mu_1 m_2^5 - 144 m_2^6 \\ (\mu_1^4 - 8 \mu_1^3 m_2 + 36 \mu_1^2 m_2^2 - 96 \mu_1 m_2^3 + 48 m_2^4) \sigma_1 - 9 (\mu_1^2 - 4 \mu_1 m_2 - 2 m_2^2) \sigma_1^2 - 9 \sigma_1^3 \end{array} \right) \right). \quad [13.46]$$

Afin de factoriser ce déterminant, il nous faut mettre le polynôme de degré trois en σ_1 , sous la forme de Cardan. Posons pour cela :

$$X = \sigma_1 - \frac{(\mu_1^2 - 4\mu_1 m_2 - 2m_2^2)}{3}$$

et [13.46] devient :

$$J_4 = 9p^2(1-p) \left(\begin{array}{l} \overbrace{X^3 + \frac{-2}{3} (\mu_1^4 - 8 \mu_1^3 m_2 + 24 \mu_1^2 m_2^2 - 40 \mu_1 m_2^3 + 26 m_2^4) X}^A \\ + \frac{8}{27} \underbrace{(\mu_1^6 - 12 \mu_1^5 m_2 + 60 \mu_1^4 m_2^2 - 160 \mu_1^3 m_2^3 + 222 \mu_1^2 m_2^4 - 120 \mu_1 m_2^5 + 16 m_2^6)}_B \end{array} \right). \quad [13.47]$$

Le déterminant de Cardan est alors :

$$\begin{aligned} \Delta &= 4A^3 + 27B^2, \\ &= \frac{32}{27} \left(\begin{array}{l} \mu_1^{12} - 24 \mu_1^{11} m_2 + 264 \mu_1^{10} m_2^2 - 1736 \mu_1^9 m_2^3 + 7434 \mu_1^8 m_2^4 - 21024 \mu_1^7 m_2^5 + 36864 \mu_1^6 m_2^6 - 27936 \mu_1^5 m_2^7 \\ -32868 \mu_1^4 m_2^8 + 113184 \mu_1^3 m_2^9 - 130464 \mu_1^2 m_2^{10} + 73440 \mu_1 m_2^{11} - 17064 m_2^{12} \end{array} \right), \end{aligned}$$

dont le signe est très difficile à établir. On peut vérifier, sur des exemples que ce déterminant n'est pas de signe constant. Le nombre de racines réelles du polynôme en X de [13.47] étant fonction du signe de ce déterminant, nous ne pouvons donner de factorisation plus précise du jacobien. La figure FIG. II.17 représente une partie de la surface Δ .

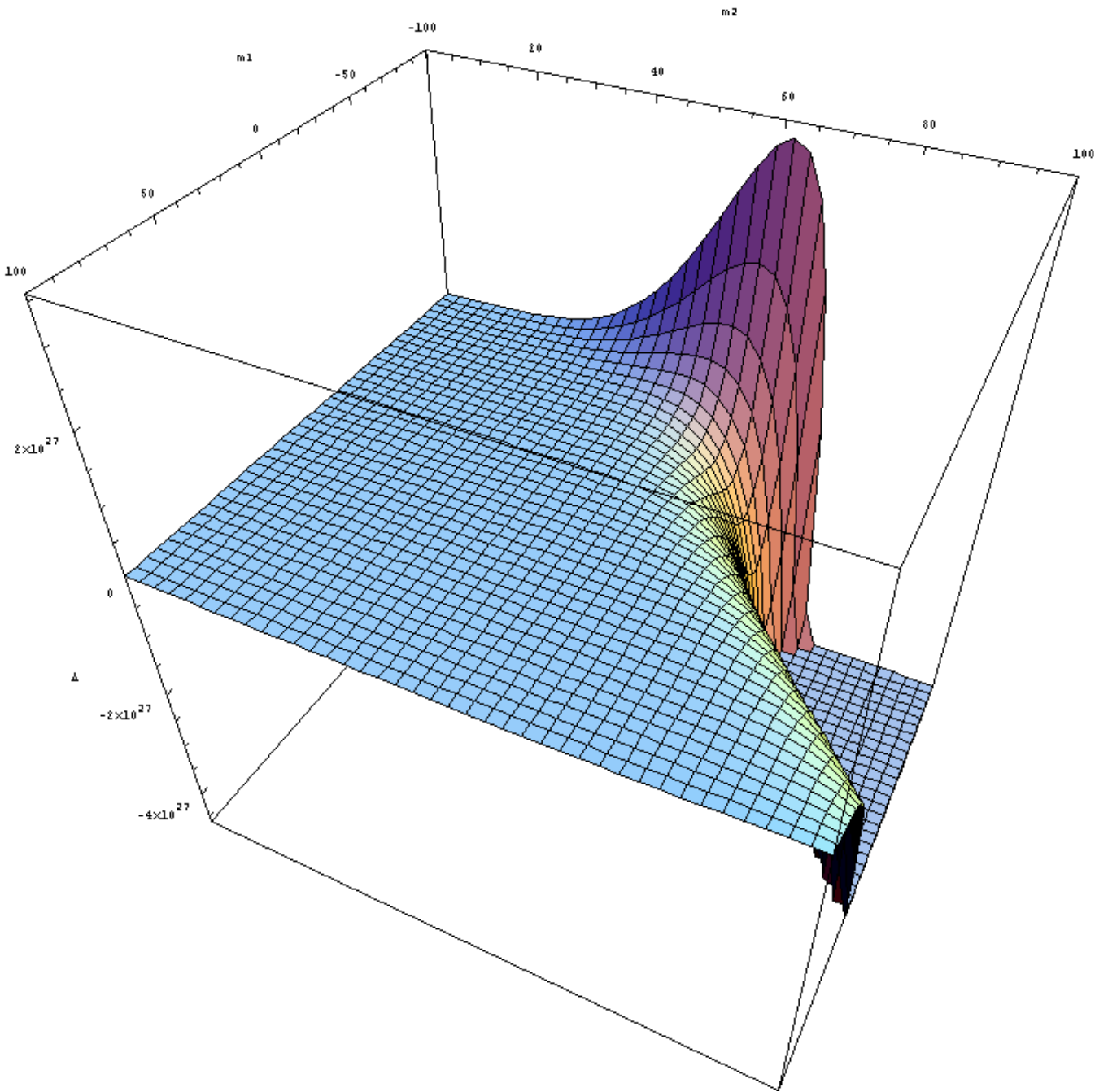


FIG.II.17. - PARTIE DE LA SURFACE Δ .

Estimateurs des moments d'un mélange normale-exponentielle.

Pour trouver un exemple d'ensemble de paramètres annulant le jacobien, il est nécessaire d'en fixer deux, ici $m_1 = 0$ et $m_2 = 1$, et trouver les racines convenables du polynôme en le dernier paramètre, ici

$$\begin{aligned}\sigma_{11} &= \frac{2}{3} - \frac{2}{3}\sqrt{13} \cos \left[\frac{1}{3} \left(\pi - \arctan \left(\frac{3\sqrt{237}}{8} \right) \right) \right] + 2\sqrt{\frac{13}{3}} \sin \left[\frac{1}{3} \left(\pi - \arctan \left(\frac{3\sqrt{237}}{8} \right) \right) \right] \approx 0.941367 \\ \text{et } \sigma_{12} &= \frac{2}{3} + \frac{4}{3}\sqrt{13} \cos \left[\frac{1}{3} \left(\pi - \arctan \left(\frac{3\sqrt{237}}{8} \right) \right) \right] \approx 4.68585.\end{aligned}$$

13.2 Inversion algébrique du système.

Des deux premières équations de \mathfrak{S}_4 , on tire :

$$\mu_1 = -\frac{m_2 - pm_2 - M_1}{p} \text{ et } \sigma_1 = -\frac{(1-p^2)m_2^2 + 2(-1+p)m_2M_1 + M_1^2 - pM_2}{p^2}, \quad [13.48]$$

par substitution les deux dernières équations deviennent :

$$0 = -2(-1+p^3)m_2^3 + 6(-1+p)m_2^2M_1 - 2M_1^3 + 3pM_1M_2 + 3(-1+p)m_2(-2M_1^2 + pM_2) - p^2M_3 \quad [13.49]$$

et

$$\begin{aligned}0 &= -2(1-4p-4p^3+7p^4)m_2^4 - 8(-1+p)^3m_2^3M_1 - 8(-1+p)m_2M_1^3 - 2M_1^4 + 3p^2M_2^2 \\ &\quad + 12(-1+p)m_2^2(-(-1+p)M_1^2 + p^2M_2) - p^3M_4.\end{aligned} \quad [13.50]$$

Grâce à [13.49] et [13.50], on peut faire baisser le degré des équations en m_2 . En effet, on tire de [13.49]

$$m_2^3 = \frac{6(-1+p)m_2^2M_1 + 3(-1+p)m_2(-2M_1^2 + pM_2) - p^2M_3 - 2M_1^3 + 3pM_1M_2}{2(-1+p^3)},$$

ce qui permet d'exprimer les monômes en m_2 de degré $n \geq 3$ en fonction de monômes de degré n , $n-1$ et $n-2$. Par exemple :

$$m_2^4 = \frac{6(-1+p)m_2^3M_1 + 3(-1+p)m_2^2(-2M_1^2 + pM_2) + m_2(-p^2M_3 - 2M_1^3 + 3pM_1M_2)}{2(-1+p^3)},$$

que l'on peut substituer dans [13.50] qui devient :

$$\begin{aligned}&m_2^3 (-2M_1 - 8pM_1 - 8p^2M_1 - 16p^3M_1 + 26p^4M_1 + 8p^5M_1) \\ &m_2^2 (6M_1^2 + 12pM_1^2 + 12p^3M_1^2 - 30p^4M_1^2 + 3pM_2 - 12p^4M_2 + 9p^5M_2) \\ &+ m_2 \left(\begin{array}{l} -6M_1^3 - 6pM_1^3 - 6p^2M_1^3 - 6p^3M_1^3 - 3pM_1M_2 + 9p^2M_1M_2 + 9p^3M_1M_2 \\ + 21p^4M_1M_2 + p^2M_3 - 3p^3M_3 - 3p^4M_3 - 7p^5M_3 \end{array} \right) \\ &+ 2(1+p+p^2)(M_1^4 - 3p^2M_2^2 + p^3M_4) = 0,\end{aligned} \quad [13.51]$$

et en renouvelant l'opération, [13.51] devient :

$$\begin{aligned}
& m_2^2 \left(-3 (-1+p)^2 p (-2 (1+p (3+p (9+5 p)))) M_1^2 + (1+p+p^2) (1+p (2+3 p)) M_2 \right) \\
& + m_2 p \left(\begin{array}{l} 6 (-2+p (-1+5 p (-1+p (3+p)))) M_1^3 - 3 (-2+p (-2+p (1+p (5+p (23+11 p)))) M_1 M_2 \\ +p (1+p+p^2) (-1+p (3+p (3+7 p))) M_3 \end{array} \right) \\
& + p \left(\begin{array}{l} 6 (1+p (2+p (5+p))) M_1^4 - 3 (1+p+4 p^2) (1+p (4+p)) M_1^2 M_2 \\ +p (1+p+4 p^2) (1+p (4+p)) M_1 M_3 + p (1+p+p^2)^2 (3 M_2^2 - p M_4) \end{array} \right) = 0,
\end{aligned} \tag{13.52}$$

d'où l'on peut tirer m_2^2 en fonction de m_2 , si

$$-3 (-1+p)^2 p (-2 (1+p (3+p (9+5 p)))) M_1^2 + (1+p+p^2) (1+p (2+3 p)) M_2 \neq 0,$$

c'est à dire

$$T_{un}(p) = +3 p^4 M_2 + p^3 (-10 M_1^2 + 5 M_2) + p^2 (-18 M_1^2 + 6 M_2) + p (-6 M_1^2 + 3 M_2) - 2 M_1^2 + M_2 \neq 0. \tag{13.53}$$

Dans le cas où $T_{un}(p) = 0$, [13.52] devient

$$\begin{aligned}
& m_2 p \left(\begin{array}{l} 6 (-2+p (-1+5 p (-1+p (3+p)))) M_1^3 - 3 (-2+p (-2+p (1+p (5+p (23+11 p)))) M_1 M_2 \\ +p (1+p+p^2) (-1+p (3+p (3+7 p))) M_3 \end{array} \right) \\
& + p \left(\begin{array}{l} 6 (1+p (2+p (5+p))) M_1^4 - 3 (1+p+4 p^2) (1+p (4+p)) M_1^2 M_2 \\ +p (1+p+4 p^2) (1+p (4+p)) M_1 M_3 + p (1+p+p^2)^2 (3 M_2^2 - p M_4) \end{array} \right) = 0,
\end{aligned} \tag{13.54}$$

que l'on ramène, en exploitant à trois reprises le fait que, d'après [13.53]

$$p^4 = \frac{2 M_1^2 - M_2 - p (-6 M_1^2 + 3 M_2) - p^3 (-10 M_1^2 + 5 M_2) - p^2 (-18 M_1^2 + 6 M_2)}{3 M_2},$$

à

$$\begin{aligned}
& \underbrace{\left(\begin{array}{l} -1440 M_1^5 M_2^2 + 144 M_1^3 M_2^3 + 288 M_1 M_2^4 \\ +1400 M_1^6 M_3 - 744 M_1^4 M_2 M_3 + 54 M_1^2 M_2^2 M_3 - 16 M_2^3 M_3 \\ +p \left(\begin{array}{l} -4320 M_1^5 M_2^2 + 648 M_1^3 M_2^3 + 837 M_1 M_2^4 + 4200 M_1^6 M_3 \\ -1812 M_1^4 M_2 M_3 - 78 M_1^2 M_2^2 M_3 - 60 M_2^3 M_3 \end{array} \right) \\ +p^2 \left(\begin{array}{l} -12960 M_1^5 M_2^2 - 540 M_1^3 M_2^3 + 1566 M_1 M_2^4 \\ +12600 M_1^6 M_3 - 3336 M_1^4 M_2 M_3 - 174 M_1^2 M_2^2 M_3 - 60 M_2^3 M_3 \end{array} \right) \\ +p^3 \left(\begin{array}{l} -7200 M_1^5 M_2^2 - 576 M_1^3 M_2^3 + 2007 M_1 M_2^4 + 7000 M_1^6 M_3 \\ +60 M_1^4 M_2 M_3 - 882 M_1^2 M_2^2 M_3 - 44 M_2^3 M_3 \end{array} \right) \end{array} \right)}_{T'_{un}(p)} \\
& + p \left(\begin{array}{l} -126 M_1^4 M_2^3 + 45 M_1^2 M_2^4 + 9 M_2^5 - 240 M_1^5 M_2 M_3 - 66 M_1^3 M_2^2 M_3 \\ +93 M_1 M_2^3 M_3 + 200 M_1^6 M_4 - 72 M_1^4 M_2 M_4 - 6 M_1^2 M_2^2 M_4 - 4 M_2^3 M_4 \\ -216 M_1^4 M_2^3 + 243 M_1^2 M_2^4 - 27 M_2^5 - 720 M_1^5 M_2 M_3 - 270 M_1^3 M_2^2 M_3 \end{array} \right) \\
& + p^2 \left(\begin{array}{l} +288 M_1 M_2^3 M_3 + 600 M_1^6 M_4 - 156 M_1^4 M_2 M_4 - 42 M_1^2 M_2^2 M_4 - 15 M_2^3 M_4 \\ -486 M_1^4 M_2^3 + 297 M_1^2 M_2^4 - 27 M_2^5 - 2160 M_1^5 M_2 M_3 - 1170 M_1^3 M_2^2 M_3 \end{array} \right) \\
& + p^3 \left(\begin{array}{l} +531 M_1 M_2^3 M_3 + 1800 M_1^6 M_4 - 168 M_1^4 M_2 M_4 - 66 M_1^2 M_2^2 M_4 - 15 M_2^3 M_4 \\ 18 M_1^4 M_2^3 + 711 M_1^2 M_2^4 - 36 M_2^5 - 1200 M_1^5 M_2 M_3 - 978 M_1^3 M_2^2 M_3 \\ +438 M_1 M_2^3 M_3 + 1000 M_1^6 M_4 + 180 M_1^4 M_2 M_4 - 102 M_1^2 M_2^2 M_4 - 11 M_2^3 M_4 \end{array} \right) \Bigg) = 0.
\end{aligned} \tag{13.55}$$

Estimateurs des moments d'un mélange normale-exponentielle.

Si le coefficient $T'_{un}(p)$ de m_2 n'est pas nul, on tire de [13.55] l'expression de m_2 en fonction de p , paramètre appartenant à $]0; 1[$ et racine de [13.53]. Tous les paramètres sont ainsi déterminés, et doivent être comparés à ceux issus de P_{Normex} , grâce au test T_5^{Normex} .

Si le coefficient $T'_{un}(p)$ de m_2 est nul, le second terme, $T''_{un}(p)$ de [13.55] est nul aussi (puisque $m_2 > 0$). Nous sommes dans le cas particulier où $T_{un}(p)$, $T'_{un}(p)$ et $T''_{un}(p)$ ont une racine, $p \in]0; 1[$, commune. Ceci est équivalent à l'existence d'une telle racine pour le polynôme :

$$P_{un}(p) = T_{un}^2(p) + T_{un}'^2(p) + T_{un}''^2(p),$$

qui s'écrit encore :

$$\begin{aligned}
P_{un}(p) = & +6 p^7 M_2 (-10 M_1^2 + 5 M_2) + 9 p^8 M_2^2 \\
& (-2 M_1^2 + M_2)^2 \left(\begin{aligned} & 1 + M_2^4 (256 M_3^2 + (9 M_2^2 - 4 M_4)^2) + 6 M_1 M_2^4 M_3 (-1536 M_2 + 279 M_2^2 - 124 M_4) \\ & + 54 M_1^3 M_2^3 M_3 (M_2 (-192 + 257 M_2) - 66 M_4) \\ & + 360 M_1^5 M_2^2 M_3 (6 M_2 (-172 + 7 M_2) - 61 M_4) \\ & - 24000 M_1^7 M_2 M_3 (42 M_2 + M_4) + 10000 M_1^8 (49 M_2^3 + M_4^2) \\ & + 200 M_1^6 M_2 (-154 M_2^3 + 72 M_2 (36 M_2^2 + M_3^2) - 63 M_2^2 M_4 + 14 M_4^2) \\ & + M_1^2 M_2^3 (82944 M_2^3 + 1134 M_2^4 - 704 M_2^5 + 8649 M_2 M_3^2 - 756 M_2^2 M_4 + 112 M_4^2) \\ & + 3 M_1^4 M_2^2 (138240 M_2^3 + 1323 M_2^4 + 7628 M_2^5 + 7440 M_2 M_3^2 - 1188 M_2^2 M_4 + 332 M_4^2) \\ & + 240 M_1^7 M_2^2 M_3 (3 M_2 (298 + 33 M_2) - 103 M_4) \end{aligned} \right) \\
& +6 (2 M_1^2 - M_2)p \left(\begin{aligned} & -48000 M_1^9 M_2 M_3 (42 M_2 + M_4) + 3 M_1 M_2^5 M_3 (3 (1136 - 3 M_2) M_2 + 283 M_4) \\ & + 18 M_1^5 M_2^3 M_3 (20010 M_2 + 267 M_2^2 + 772 M_4) + 20000 M_1^{10} (49 M_2^3 + M_4^2) \\ & + 3 M_1^3 M_2^4 M_3 (3 (2134 - 1299 M_2) M_2 + 962 M_4) \\ & + 8 M_1^6 M_2^2 (81 M_2^3 (400 + 7 M_2) + 2186 M_2^4 + 4140 M_2 M_3^2 + 1071 M_2^2 M_4 - 166 M_4^2) \\ & - M_1^4 M_2^3 (81 M_2^3 (3248 + 55 M_2) + 23092 M_2^4 + 3150 M_2 M_3^2 - 1296 M_2^2 M_4 + 904 M_4^2) \\ & + M_2 (-1 + 81 M_2^8 + 9 M_2^9 M_4 - 20 M_2^4 (16 M_2^3 + M_4^2)) \\ & + 600 M_1^8 M_2 (1728 M_2^3 + 48 M_2 M_3^2 - 33 M_2^2 M_4 - 4 (189 M_2^3 + M_4^2)) \\ & + M_1^2 (2 - 3 M_2^4 (26784 M_2^3 + 54 M_2^4 - 8 M_2^5 + 2976 M_2 M_3^2 - 213 M_2^2 M_4 + 42 M_4^2)) \\ & - 48 M_1^7 M_2^2 M_3 (3 M_2 (-36905 + 489 M_2) - 8929 M_4) \end{aligned} \right) \\
& +3p^2 \left(\begin{aligned} & + 12 M_1^3 M_2^5 M_3 (3 M_2 (-1291 + 2223 M_2) - 650 M_4) + 360000 M_1^{12} (49 M_2^3 + M_4^2) \\ & + 480 M_1^9 M_2^2 M_3 (6 M_2 (6031 + 126 M_2) - 287 M_4) - 864000 M_1^{11} M_2 M_3 (42 M_2 + M_4) \\ & - 6 M_1 M_2^6 M_3 (12 M_2 (857 + 51 M_2) + 871 M_4) - 36 M_1^5 M_2^4 M_3 (3 M_2 (19718 + 1887 M_2) + 982 M_4) \\ & + 8 M_1^8 M_2^2 (-324000 M_2^3 + 7047 M_2^4 + 352738 M_2^5 + 51480 M_2 M_3^2 + 31338 M_2^2 M_4 - 2078 M_4^2) \\ & + 1600 M_1^{10} M_2 (-9023 M_2^3 + 324 M_2 (36 M_2^2 + M_3^2) - 189 M_2^2 M_4 - 107 M_4^2) \\ & - 4 M_1^6 M_2^3 (1578528 M_2^3 + 18630 M_2^4 + 74492 M_2^5 + 70335 M_2 M_3^2 + 8658 M_2^2 M_4 + 1148 M_4^2) \\ & + M_2^2 (7 + 81 M_2^8 + 252 M_2^9 M_4 + 115 M_2^4 (16 M_2^3 + M_4^2)) \\ & + M_1^2 M_2 (-32 + M_2^4 (534195 M_2^3 - 3402 M_2^4 + 2816 M_2^5 + 60570 M_2 M_3^2 - 3204 M_2^2 M_4 + 656 M_4^2)) \\ & + M_1^4 (36 + M_2^4 (133588 M_2^3 + 243 M_2 (M_2^2 (1680 + 131 M_2) - 608 M_2^3) - 2160 M_2^2 M_4 + 3580 M_4^2)) \\ & + 3 M_1 M_2^6 M_3 (-3 M_2 (20996 + 2391 M_2) - 5020 M_4) \\ & + 6 M_1^3 M_2^5 M_3 (12 M_2 (-5792 + 4623 M_2) - 4531 M_4) \\ & + 480 M_1^9 M_2^2 M_3 (102633 M_2 + 2007 M_2^2 - 1976 M_4) - 3072000 M_1^{11} M_2 M_3 (42 M_2 + M_4) \\ & - 6 M_1^5 M_2^4 M_3 (1019856 M_2 + 114669 M_2^2 + 2636 M_4) \\ & + 12 M_1^7 M_2^3 M_3 (3 (662416 - 12705 M_2) M_2 + 122828 M_4) + 1280000 M_1^{12} (49 M_2^3 + M_4^2) \\ & + 12 M_1^8 M_2^2 (-522720 M_2^3 + 8559 M_2^4 + 285816 M_2^5 + 144960 M_2 M_3^2 + 74244 M_2^2 M_4 - 10696 M_4^2) \\ & + 1200 M_1^{10} M_2 (-34972 M_2^3 + 1536 M_2 (36 M_2^2 + M_3^2) - 669 M_2^2 M_4 - 348 M_4^2) \\ & - 2 M_1^6 M_2^3 (11504592 M_2^3 + 135513 M_2^4 - 26672 M_2^5 + 420336 M_2 M_3^2 + 72918 M_2^2 M_4 + 9292 M_4^2) \\ & + M_2^2 (23 + 405 M_2^8 + 855 M_2^9 M_4 + 269 M_2^4 (16 M_2^3 + M_4^2)) \\ & + M_1^2 M_2 (-110 + 3 M_2^4 (629586 M_2^3 - 3267 M_2^4 + 8952 M_2^5 + 64554 M_2 M_3^2 - 3075 M_2^2 M_4 + 698 M_4^2)) \\ & + 2 M_1^4 (64 + 3 M_2^4 (114318 M_2^3 + 21303 M_2^4 + 51100 M_2^5 - 100032 M_2 M_3^2 - 2094 M_2^2 M_4 + 1386 M_4^2)) \end{aligned} \right) \\
& +2p^3 \left(\begin{aligned} & + 12 M_1^3 M_2^5 M_3 (3 M_2 (-1291 + 2223 M_2) - 650 M_4) + 360000 M_1^{12} (49 M_2^3 + M_4^2) \\ & + 480 M_1^9 M_2^2 M_3 (6 M_2 (6031 + 126 M_2) - 287 M_4) - 864000 M_1^{11} M_2 M_3 (42 M_2 + M_4) \\ & - 6 M_1 M_2^6 M_3 (12 M_2 (857 + 51 M_2) + 871 M_4) - 36 M_1^5 M_2^4 M_3 (3 M_2 (19718 + 1887 M_2) + 982 M_4) \\ & + 8 M_1^8 M_2^2 (-324000 M_2^3 + 7047 M_2^4 + 352738 M_2^5 + 51480 M_2 M_3^2 + 31338 M_2^2 M_4 - 2078 M_4^2) \\ & + 1600 M_1^{10} M_2 (-9023 M_2^3 + 324 M_2 (36 M_2^2 + M_3^2) - 189 M_2^2 M_4 - 107 M_4^2) \\ & - 4 M_1^6 M_2^3 (1578528 M_2^3 + 18630 M_2^4 + 74492 M_2^5 + 70335 M_2 M_3^2 + 8658 M_2^2 M_4 + 1148 M_4^2) \\ & + M_2^2 (7 + 81 M_2^8 + 252 M_2^9 M_4 + 115 M_2^4 (16 M_2^3 + M_4^2)) \\ & + M_1^2 M_2 (-32 + M_2^4 (534195 M_2^3 - 3402 M_2^4 + 2816 M_2^5 + 60570 M_2 M_3^2 - 3204 M_2^2 M_4 + 656 M_4^2)) \\ & + M_1^4 (36 + M_2^4 (133588 M_2^3 + 243 M_2 (M_2^2 (1680 + 131 M_2) - 608 M_2^3) - 2160 M_2^2 M_4 + 3580 M_4^2)) \\ & + 3 M_1 M_2^6 M_3 (-3 M_2 (20996 + 2391 M_2) - 5020 M_4) \\ & + 6 M_1^3 M_2^5 M_3 (12 M_2 (-5792 + 4623 M_2) - 4531 M_4) \\ & + 480 M_1^9 M_2^2 M_3 (102633 M_2 + 2007 M_2^2 - 1976 M_4) - 3072000 M_1^{11} M_2 M_3 (42 M_2 + M_4) \\ & - 6 M_1^5 M_2^4 M_3 (1019856 M_2 + 114669 M_2^2 + 2636 M_4) \\ & + 12 M_1^7 M_2^3 M_3 (3 (662416 - 12705 M_2) M_2 + 122828 M_4) + 1280000 M_1^{12} (49 M_2^3 + M_4^2) \\ & + 12 M_1^8 M_2^2 (-522720 M_2^3 + 8559 M_2^4 + 285816 M_2^5 + 144960 M_2 M_3^2 + 74244 M_2^2 M_4 - 10696 M_4^2) \\ & + 1200 M_1^{10} M_2 (-34972 M_2^3 + 1536 M_2 (36 M_2^2 + M_3^2) - 669 M_2^2 M_4 - 348 M_4^2) \\ & - 2 M_1^6 M_2^3 (11504592 M_2^3 + 135513 M_2^4 - 26672 M_2^5 + 420336 M_2 M_3^2 + 72918 M_2^2 M_4 + 9292 M_4^2) \\ & + M_2^2 (23 + 405 M_2^8 + 855 M_2^9 M_4 + 269 M_2^4 (16 M_2^3 + M_4^2)) \\ & + M_1^2 M_2 (-110 + 3 M_2^4 (629586 M_2^3 - 3267 M_2^4 + 8952 M_2^5 + 64554 M_2 M_3^2 - 3075 M_2^2 M_4 + 698 M_4^2)) \\ & + 2 M_1^4 (64 + 3 M_2^4 (114318 M_2^3 + 21303 M_2^4 + 51100 M_2^5 - 100032 M_2 M_3^2 - 2094 M_2^2 M_4 + 1386 M_4^2)) \end{aligned} \right)
\end{aligned}$$

$$\begin{aligned}
& +p^4 \left(\begin{aligned} & -12 M_1^2 M_2 + 6 M_2^2 + 36 (-3 M_1^2 + M_2)^2 + 30 (-2 M_1^2 + M_2)^2 \\ & +36 (-9 M_1 M_2^2 (240M_1^4 + 10 M_1^2 M_2 - 29M_2^2) + (2100M_1^6 - 556 M_1^4 M_2 - 29 M_1^2 M_2^2 - 10 M_2^3) M_3)^2 \\ & + 6(-9 M_1 M_2^2 (800M_1^4 + 64 M_1^2 M_2 - 223M_2^2) + 2 (3500M_1^6 + 30 M_1^4 M_2 - 441 M_1^2 M_2^2 - 22 M_2^3) M_3) \\ & (-9 M_1 M_2^2 (160 M_1^4 - 24 M_1^2 M_2 - 31 M_2^2) + 2 (700 M_1^6 - 302 M_1^4 M_2 - 13 M_1^2 M_2^2 - 10 M_2^3) M_3) \\ & + 6 (9 M_2^3 (2 M_1^4 + 79 M_1^2 M_2 - 4 M_2^2) - 6 M_1 M_2 (200 M_1^4 + 163 M_1^2 M_2 - 73 M_2^2) M_3 \\ & + (10 M_1^2 + M_2) (100 M_1^4 + 8 M_1^2 M_2 - 11 M_2^2) M_4) (-9 M_2^3 (8 M_1^4 - 9 M_1^2 M_2 + M_2^2) \\ & -6 M_1 M_2 (40 M_1^4 + 15 M_1^2 M_2 - 16 M_2^2) M_3 + (2 M_1^2 - M_2) (100 M_1^4 + 24 M_1^2 M_2 + 5 M_2^2) M_4) \\ & + 9 \left(\begin{aligned} & 9 M_2^3 (18 M_1^4 - 11 M_1^2 M_2 + M_2^2) + 3 M_1 M_2 (240 M_1^4 + 130 M_1^2 M_2 - 59 M_2^2) M_3 \\ & + (-600 M_1^6 + 56 M_1^4 M_2 + 22 M_1^2 M_2^2 + 5 M_2^3) M_4 \end{aligned} \right)^2 \end{aligned} \right) \\
& +2p^5 \left(\begin{aligned} & 9 M_2 (-2 M_1^2 + M_2) + 30(6 M_1^4 - 5 M_1^2 M_2 + M_2^2) \\ & +6 (-9 M_1 M_2^2 (800 M_1^4 + 64 M_1^2 M_2 - 223M_2^2) + 2(3500M_1^6 + 30M_1^4 M_2 - 441M_1^2 M_2^2 - 22M_2^3)M_3) \\ & (-9 M_1 M_2^2 (240M_1^4 + 10 M_1^2 M_2 - 29 M_2^2) + (2100M_1^6 - 556 M_1^4 M_2 - 29 M_1^2 M_2^2 - 10 M_2^3) M_3) \\ & + 3 (9 M_2^3 (2 M_1^4 + 79 M_1^2 M_2 - 4 M_2^2) - 6 M_1 M_2 (200 M_1^4 + 163 M_1^2 M_2 - 73 M_2^2) M_3 \\ & + (10 M_1^2 + M_2) (100 M_1^4 + 8 M_1^2 M_2 - 11 M_2^2) M_4) \\ & (-9 M_2^3 (18 M_1^4 - 11 M_1^2 M_2 + M_2^2) - 3 M_1 M_2 (240 M_1^4 + 130 M_1^2 M_2 - 59 M_2^2) M_3 \\ & + (600 M_1^6 - 56 M_1^4 M_2 - 22 M_1^2 M_2^2 - 5 M_2^3) M_4) \\ & 36 M_2 (-3 M_1^2 + M_2) + 25 (-2 M_1^2 + M_2)^2 \end{aligned} \right) \\
& +p^6 \left(\begin{aligned} & + (9 M_1 M_2^2 (800M_1^4 + 64M_1^2 M_2 - 223M_2^2) - 2 (3500M_1^6 + 30M_1^4 M_2 - 441 M_1^2 M_2^2 - 22 M_2^3) M_3)^2 \\ & + \left(\begin{aligned} & 9 M_2^3 (-2 M_1^4 - 79 M_1^2 M_2 + 4 M_2^2) + 6 M_1 M_2 (200 M_1^4 + 163 M_1^2 M_2 - 73 M_2^2) M_3 \\ & - (10 M_1^2 + M_2) (100 M_1^4 + 8 M_1^2 M_2 - 11 M_2^2) M_4 \end{aligned} \right)^2 \end{aligned} \right).
\end{aligned}$$

Ce cas correspond à une redondance dans les équations du système. En effet, les quatre équations ne permettent d'exprimer que trois paramètres, puisque, si l'on tire μ_1 , σ_1 et p , des trois premières équations, la quatrième sera implicitement vérifiée, pour toutes les valeurs de m_2 . Si ce cas se produit, il est donc nécessaire de recourir à la cinquième équation de moments. Cette équation s'écrit :

$$-120 (-1 + p) m_2^5 + p (\mu_1^5 + 10 \mu_1^3 \sigma_1 + 15 \mu_1 \sigma_1^2) = M_5,$$

ou encore, compte tenu de [13.48],

$$\begin{aligned}
P_{trois}(p) = & m_2^5(-2(3 + p(5 - 20p - 35p^3 + 47p^4))) + m_2^4(-30(-1 + p)^2(-1 + (-2 + p)p)M_1) \\
& + m_2^3(20(-1 + p)^2(-3(1 + p)M_1^2 + p(1 + 2p)M_2)) + m_2^2(20(-1 + p)M_1((-3 + p)M_1^2 + 3pM_2)) = 0, \quad [13.56] \\
& + m_2(15(-1 + p)(2M_1^4 - 4pM_1^2 M_2 + p^2 M_2^2)) + (6M_1^5 - 20pM_1^3 M_2 + 15p^2 M_1 M_2^2 - p^4 M_5)
\end{aligned}$$

qui est un polynôme en m_2 puisque p est connu en tant que racine convenable de P_{un} . Une solution éventuellement issue de cette méthode sera comparée avec les précédentes grâce au test T_6^{Normex} , construit sur la sixième équation des moments centrés :

$$T_6^{Normex} = \left| \frac{\left(\begin{aligned} & (1 - p)(720m_2^6 - 720m_2^5 M_1 + 360m_2^4 M_1^2 - 120m_2^3 M_1^3 + 30m_2^2 M_1^4 - 6m_2 M_1^5 + M_1^6) \\ & + p(15\sigma_1^3 + 45\sigma_1^2 x_1^2 + 15\sigma_1 x_1^4 + x_1^6) \end{aligned} \right) - M_6^c}{2M_6^c} \right|.$$

Estimateurs des moments d'un mélange normale-exponentielle.

Nous nous plaçons, désormais dans le cas où $T_{un}(p) \neq 0$. On peut, en substituant comme précédemment, m_2^2 dans [13.49] abaisser le degré de cette équation, pour obtenir dans un premier temps :

$$\begin{aligned}
& 3(1-p)(-2(1+p(3+p(9+5p)))M_1^2 + (1+p+p^2)(1+p(2+3p))M_2)(2M_1^3 - 3pM_1M_2 + p^2M_3) \\
& + m_2^2 \left(\begin{array}{l} -2(6(1+p^2(4+p(-21+p(12+5p(7+p))))))M_1^3 - 3(1+p+p^2) \\ (1+p(-2+p(1+p(-7+p(32+11p))))M_1M_2 + p(1+p+p^2)^2(-1+p(3+p(3+7p)))M_3 \end{array} \right) \\
& + m_2 \left(\begin{array}{l} 12(2+p^2(4+(-2+p)p(19+14p)))M_1^4 \\ -6(-2+(-2+p)p)(1+p+4p^2)(-1+p(2+5p))M_1^2M_2 \\ -6(-2+(-2+p)p)(1+p+4p^2)(-1+p(2+5p))M_1^2M_2 \\ -2p(1+p+p^2)(1+p+4p^2)(1+p(4+p))M_1M_3 \\ +p(3(1+p+p^2)(1+p(-4+p(-6+p(-16+7p))))M_2^2 + 2p(1+p+p^2)^3M_4) \end{array} \right) = 0,
\end{aligned}$$

et dans un second temps :

$$\begin{aligned}
& 72(-1+p(-1+p(-7+15p(1+2p))))M_1^7 \\
& -36(-3+p(-5+p(-13+p(27+2p(55+68p))))M_1^5M_2 \\
& +12p(1+p(5+p(3+p(23+2p(59+51p))))M_1^4M_3 \\
& -36p(1+p)(-1+p+28p^3+17p^4+27p^5)M_1^2M_2M_3 \\
& +pM_3 \left(\begin{array}{l} 3(-5+p(-1+p(9+p(61+p(59+p(63+p(-11+41p))))))M_2^2 \\ -2p(1+p+p^2)^2(-1+p(3+p(3+7p)))M_4 \end{array} \right) \\
& +6M_1^3 \left(\begin{array}{l} 3(-3+p(-10+p(-11+p(18+p(169+p(188+153p))))M_2^2 \\ -2p(1+p^2(4+p(-21+p(12+5p(7+p))))M_4 \end{array} \right) \\
& +M_1 \left(\begin{array}{l} -9(-1+p(-5+p(-3+p(5+p(55+p(75+p(41+49p))))))M_2^3 \\ +2p(1+p+4p^2)(1+p(4+p))(-1+p(3+p(3+7p)))M_3^2 \\ +6p(1+p+p^2)(1+p(-2+p(1+p(-7+p(32+11p))))M_2M_4 \end{array} \right) \\
& +m_2 \left(\begin{array}{l} 72(1+p(-3+p(2+p(-48+p+95p^2))))M_1^6 \\ -108(1+p(-4+p(-4+p(-44+3p(1+4p(5+7p))))))M_1^4M_2 \\ -9(-1+p)^2(1+p(2+3p))(1+p(-4+p(-6+p(-16+7p))))M_2^3 \\ +24p(-3+p(-10+p(-13+p(-28+p(31+p(142+25p))))))M_1^3M_3 \\ -12p(-2+p(-13+p(-12+p(-35+p(86+p(129+p(208+71p))))))M_1M_2M_3 \\ +2p(1+p+p^2)(-1+p(3+p(3+7p)))^2M_3^2 - 6p(-1+p^3)^2(1+p(2+3p))M_2M_4 \\ +6M_1^2 \left(\begin{array}{l} 9(1+p(-4+p(-7+p(-36+p(7+p(64+p(47+72p))))))M_2^2 \\ +2(-1+p)^2p(1+p+p^2)(1+p(3+p(9+5p)))M_4 \end{array} \right) \end{array} \right) = 0,
\end{aligned}$$

[13.57]

Ce qui permet d'établir l'expression littérale de m_2 , lorsque

$$\left(\begin{array}{l} 72(1+p(-3+p(2+p(-48+p+95p^2))))M_1^6 \\ -108(1+p(-4+p(-4+p(-44+3p(1+4p(5+7p))))))M_1^4M_2 \\ -9(-1+p)^2(1+p(2+3p))(1+p(-4+p(-6+p(-16+7p))))M_2^3 \\ +24p(-3+p(-10+p(-13+p(-28+p(31+p(142+25p))))))M_1^3M_3 \\ -12p(-2+p(-13+p(-12+p(-35+p(86+p(129+p(208+71p))))))M_1M_2M_3 \\ +2p(1+p+p^2)(-1+p(3+p(3+7p)))^2M_3^2 - 6p(-1+p^3)^2(1+p(2+3p))M_2M_4 \\ +6M_1^2 \left(\begin{array}{l} 9(1+p(-4+p(-7+p(-36+p(7+p(64+p(47+72p))))))M_2^2 \\ +2(-1+p)^2p(1+p+p^2)(1+p(3+p(9+5p)))M_4 \end{array} \right) \end{array} \right) \neq 0,$$

c'est à dire lorsque

$$T_{deux}(p) = \left(\begin{array}{c} 72 M_1^6 - 108 M_1^4 M_2 + 54 M_1^2 M_2^2 - 9 M_2^3 \\ +p \left(\begin{array}{c} -216 M_1^6 + 432 M_1^4 M_2 - 216 M_1^2 M_2^2 + 36 M_2^3 - 72 M_1^3 M_3 \\ +24 M_1 M_2 M_3 + 2 M_3^2 + 12 M_1^2 M_4 - 6 M_2 M_4 \end{array} \right) \\ +p^2 \left(\begin{array}{c} 144 M_1^6 + 432 M_1^4 M_2 - 378 M_1^2 M_2^2 + 54 M_2^3 - 240 M_1^3 M_3 \\ +156 M_1 M_2 M_3 - 10 M_3^2 + 24 M_1^2 M_4 - 12 M_2 M_4 \end{array} \right) \\ +p^3 \left(\begin{array}{c} -3456 M_1^6 + 4752 M_1^4 M_2 - 1944 M_1^2 M_2^2 + 180 M_2^3 - 312 M_1^3 M_3 \\ +144 M_1 M_2 M_3 - 4 M_3^2 + 72 M_1^2 M_4 - 18 M_2 M_4 \end{array} \right) \\ +p^4 \left(\begin{array}{c} 72 M_1^6 - 324 M_1^4 M_2 + 378 M_1^2 M_2^2 - 234 M_2^3 - 672 M_1^3 M_3 \\ +420 M_1 M_2 M_3 + 2 M_3^2 - 60 M_1^2 M_4 + 12 M_2 M_4 \end{array} \right) \\ +p^5 \left(\begin{array}{c} 6840 M_1^6 - 6480 M_1^4 M_2 + 3456 M_1^2 M_2^2 - 108 M_2^3 + 744 M_1^3 M_3 \\ -1032 M_1 M_2 M_3 + 116 M_3^2 - 84 M_1^2 M_4 + 24 M_2 M_4 \end{array} \right) \\ +p^6 \left(\begin{array}{c} -9072 M_1^4 M_2 + 2538 M_1^2 M_2^2 - 414 M_2^3 + 3408 M_1^3 M_3 \\ -1548 M_1 M_2 M_3 + 194 M_3^2 - 72 M_1^2 M_4 + 36 M_2 M_4 \end{array} \right) \\ +p^7 \left(\begin{array}{c} 3888 M_1^2 M_2^2 + 684 M_2^3 + 600 M_1^3 M_3 - 2496 M_1 M_2 M_3 \\ +284 M_3^2 + 48 M_1^2 M_4 - 6 M_2 M_4 \end{array} \right) \\ +p^8 (-189 M_2^3 - 852 M_1 M_2 M_3 + 182 M_3^2 + 60 M_1^2 M_4 - 12 M_2 M_4) \\ +p^9 (98 M_3^2 - 18 M_2 M_4) \end{array} \right) \neq 0,$$

Comme nous disposons désormais des expressions littérales de μ_1 , m_2 et σ_1 en fonction de p et des 4 premiers moments, nous pouvons substituer ces trois paramètres dans les troisième et quatrième équations de \mathfrak{S}_4 . Ces substitutions nous donnent deux polynômes, issus respectivement de la troisième et de la quatrième équation, de degrés respectifs 30 et 39. Ces polynômes, dont nous tenons l'expression à la disponibilité du lecteur, nous permettent, en calculant leur *PGCD*, d'obtenir un troisième polynôme, de degré 12, cette fois, dont l'une des racines réelle, comprise entre 0 et 1, permettra l'inversion du système \mathfrak{S}_4 . Ce polynôme, P_{Normex} , est :

Estimateurs des moments d'un mélange normale-exponentielle.

$$\begin{aligned}
P_{Normex}(p) = & 27 (-2 M_1^2 + M_2)^6 \\
+ p & \left(9 (2 M_1^2 - M_2)^3 (72 M_1^6 - 180 M_1^4 M_2 + 27 M_2^3 + 48 M_1^3 M_3 - 72 M_1 M_2 M_3) \right. \\
& \quad \left. + 8 M_3^3 - 3 M_2 M_4 + 6 M_1^2 (15 M_2^2 + M_4) \right) \\
+ p^2 & \left(324 (72 M_1^{12} - 240 M_1^{10} M_2 + 336 M_1^8 M_2^2 - 228 M_1^6 M_2^3 + 71 M_1^4 M_2^4 - 10 M_1^2 M_2^5 + M_2^6) \right. \\
& \quad + 216 M_1 (40 M_1^8 - 120 M_1^6 M_2 + 104 M_1^4 M_2^2 - 20 M_1^2 M_2^3 - 3 M_2^4) M_3 \\
& \quad - 2 M_3^4 + 36 (92 M_1^6 - 148 M_1^4 M_2 + 42 M_1^2 M_2^2 + 3 M_2^3) M_3^2 + 48 M_1 (7 M_1^2 - 3 M_2) M_3^3 \\
& \quad \left. - 3 (2 M_1^2 - M_2) (9 (40 M_1^6 - 76 M_1^4 M_2 + 38 M_1^2 M_2^2 - M_2^3) + 96 M_1 (M_1^2 - M_2) M_3 + 8 M_3^2) M_4 \right) \\
+ 2p^3 & \left(54 (232 M_1^{12} - 1248 M_1^{10} M_2 + 1920 M_1^8 M_2^2 - 1216 M_1^6 M_2^3 + 291 M_1^4 M_2^4 - 30 M_1^2 M_2^5 + 3 M_2^6) \right. \\
& \quad + 108 M_1 (216 M_1^8 - 544 M_1^6 M_2 + 452 M_1^4 M_2^2 - 108 M_1^2 M_2^3 + 9 M_2^4) M_3 \\
& \quad + 36 (106 M_1^6 - 156 M_1^4 M_2 - 9 M_1^2 M_2^2 + 3 M_2^3) M_3^2 + 24 M_1 (23 M_1^2 + 6 M_2) M_3^3 + 9 M_3^4 \\
& \quad - 12 M_4 \left(9 (8 M_1^8 - 48 M_1^6 M_2 + 56 M_1^4 M_2^2 - 17 M_1^2 M_2^3 + 2 M_2^4) \right. \\
& \quad \quad \left. + 3 M_1 (44 M_1^4 - 68 M_1^2 M_2 + 7 M_2^2) M_3 + (26 M_1^2 + 3 M_2) M_3^2 \right) \\
& \quad \left. + 12 (3 (4 M_1^4 - 6 M_1^2 M_2 + M_2^2) + 4 M_1 M_3) M_4^2 - 2 M_4^3 \right) \\
- 6p^4 & \left(27 M_2 (576 M_1^{10} - 1472 M_1^8 M_2 + 1424 M_1^6 M_2^2 - 686 M_1^4 M_2^3 + 212 M_1^2 M_2^4 - 27 M_2^5) \right. \\
& \quad - 36 M_1 (248 M_1^8 - 528 M_1^6 M_2 + 272 M_1^4 M_2^2 + 88 M_1^2 M_2^3 - 27 M_2^4) M_3 \\
& \quad - 144 (30 M_1^6 - 41 M_1^4 M_2 + M_2^2) M_3^2 - 24 M_1 (17 M_1^2 + 6 M_2) M_3^3 + 6 M_3^4 \\
& \quad + 4 M_4 \left(9 (152 M_1^8 - 312 M_1^6 M_2 + 236 M_1^4 M_2^2 - 59 M_1^2 M_2^3 + 5 M_2^4) \right. \\
& \quad \quad \left. + 3 M_1 (92 M_1^4 - 180 M_1^2 M_2 + 31 M_2^2) M_3 + (62 M_1^2 + 3 M_2) M_3^2 \right) \\
& \quad \left. - 2 (24 M_1^4 - 72 M_1^2 M_2 + 15 M_2^2 + 20 M_1 M_3) M_4^2 + 2 M_4^3 \right) \\
- 6p^5 & \left(189 M_2^6 + 288 M_1^9 M_3 + 19008 M_1^7 M_2 M_3 + 36 M_3^2 M_3^2 + 2 M_3^4 - 384 M_1^5 M_3 (51 M_2^2 - 11 M_4) \right. \\
& \quad - 144 M_1^8 (180 M_2^2 - 11 M_4) + 315 M_2^4 M_4 + 48 M_2 M_3^2 M_4 - 66 M_2^2 M_4^2 + 4 M_4^3 \\
& \quad + 48 M_1^6 (729 M_2^3 - 92 M_3^2 - 171 M_2 M_4) - 8 M_1^3 M_3 (-1062 M_2^3 + 91 M_3^2 + 504 M_2 M_4) \\
& \quad - 4 M_1 M_3 (837 M_2^4 + 84 M_2 M_3^2 - 39 M_2^2 M_4 + 22 M_4^2) \\
& \quad - 6 M_1^4 (3249 M_2^4 - 184 M_2 M_3^2 - 1242 M_2^2 M_4 + 80 M_4^2) \\
& \quad \left. + 4 M_1^2 (1107 M_2^5 + 792 M_2^2 M_3^2 - 504 M_2^3 M_4 + 125 M_3^2 M_4 + 126 M_2 M_4^2) \right) \\
+ 6p^6 & \left(162 M_2^6 - 19296 M_1^7 M_2 M_3 - 288 M_3^2 M_3^2 + 96 M_1^5 M_3 (411 M_2^2 - 17 M_4) + 7632 M_1^8 M_4 \right. \\
& \quad + 63 M_2^4 M_4 + 60 M_2 M_3^2 M_4 + 30 M_2^2 M_4^2 + 8 M_1^3 M_3 (-2754 M_2^3 + 41 M_3^2 + 246 M_2 M_4) \\
& \quad - 24 M_1^6 (543 M_2^3 - 37 M_3^2 + 288 M_2 M_4) + 4 M_1 M_3 (297 M_2^4 + 48 M_2 M_3^2 - 237 M_2^2 M_4 + 16 M_4^2) \\
& \quad + 6 M_1^4 (783 M_2^4 - 876 M_2 M_3^2 + 282 M_2^2 M_4 + 28 M_4^2) \\
& \quad + 4 M_1^2 (513 M_2^5 + 981 M_2^2 M_3^2 + 207 M_2^3 M_4 - 95 M_3^2 M_4 - 36 M_2 M_4^2) - 4 (6 M_3^4 + M_4^3) \\
& \quad \left. 27 M_2^4 (111 M_1^4 - 50 M_1^2 M_2 + 33 M_2^2) - 18 M_1 M_2^2 (940 M_1^4 - 356 M_1^2 M_2 + 105 M_2^2) M_3 \right) \\
- 12p^7 & \left(-12 M_1 (19 M_1^2 - 42 M_2) M_3^3 - 24 M_3^4 + 2 (9 M_2 (416 M_1^6 - 170 M_1^4 M_2 + 85 M_1^2 M_2^2 - 4 M_2^3) \right. \\
& \quad - 3 M_1 (524 M_1^4 + 36 M_1^2 M_2 - 13 M_2^2) M_3 + 5 (10 M_1^2 - 3 M_2) M_3^2) M_4 \\
& \quad \left. + (348 M_1^4 - 108 M_1^2 M_2 + 3 M_2^2 - 8 M_1 M_3) M_4^2 + M_4^3 \right) \\
+ 3p^8 & \left(2457 M_2^6 - 88 M_3^2 M_3^2 + 172 M_3^4 + 2592 M_1^5 M_3 M_4 + 24 M_2 (57 M_2^3 + 13 M_3^2) M_4 \right. \\
& \quad - 156 M_2^2 M_4^2 + 4 M_4^3 + 16 M_1^3 M_3 (-2088 M_2^3 + 253 M_3^2 - 846 M_2 M_4) \\
& \quad + 24 M_1^4 (-644 M_2 M_3^2 + 894 M_2^2 M_4 - 13 M_4^2) \\
& \quad - 8 M_1 M_3 (1539 M_2^4 + 564 M_2 M_3^2 + 111 M_2^2 M_4 + 20 M_4^2) \\
& \quad \left. + 4 M_1^2 (3213 M_2^5 + 7128 M_2^2 M_3^2 - 126 M_2^3 M_4 + 308 M_3^2 M_4 + 252 M_2 M_4^2) \right) \\
+ 3p^9 & \left(-441 M_2^6 - 9 M_2^4 (336 M_1 M_3 + 79 M_4) + 12 M_2^2 (1308 M_1^2 M_3^2 + 226 M_1 M_3 M_4 - 7 M_4^2) \right. \\
& \quad + 72 M_2^3 (16 M_3^2 - 67 M_1^2 M_4) - 48 M_2 (118 M_1 M_3^2 + 2 M_3 (58 M_1^3 + M_3) M_4 - 9 M_1^2 M_4^2) \\
& \quad \left. + 4 (M_3^3 (212 M_1^3 + 105 M_3) + 242 M_1^2 M_3^2 M_4 + 2 M_1 (27 M_1^3 - 20 M_3) M_4^2 + 2 M_4^3) \right) \\
+ p^{10} & \left(18 M_3^2 (58 M_2^3 - 504 M_1 M_2 M_3 + 49 M_3^2) - 48 (3 M_2 (3 M_1^2 + M_2) + 7 M_1 M_3) M_4^2 + 24 M_4^3 \right. \\
& \quad \left. + 3 (63 M_2^4 + 1704 M_1 M_2^2 M_3 + 8 (85 M_1^2 - 12 M_2) M_3^2) M_4 \right) \\
+ 2p^{11} & (343 M_3^4 - 336 M_2 M_3^2 M_4 + 12 (3 M_2^2 - 2 M_1 M_3) M_4^2 + 6 M_4^3) + 4 p^{12} M_4^3.
\end{aligned}$$

Il est nécessaire, avant de rechercher les racines de P_{Normex} de connaître les racines de T_{un} et T_{deux} . En effet, nous ne pouvons pas exclure, d'après ce qui précède, que les racines réelles, appartenant à $]0; 1[$, de ces polynômes,

permettent l'inversion du système \mathfrak{S}_4 . Les éventuelles solutions issues de ces deux polynômes seront comparées avec celles issues des racines statistiquement convenables de P_{Normex} au moyen du test, basé sur la cinquième équation de moment centrés :

$$T_5^{Normex} = \left| \frac{((1-p)(120m_2^5 - 120m_2^4M_1 + 60m_2^3M_1^2 - 20m_2^2M_1^3 + 5m_2M_1^4 - M_1^5) + 15p\sigma_1^2x_1 + 10p\sigma_1x_1^3 + px_1^5) - M_5^c}{2M_5^c} \right|.$$

Le fait d'utiliser les moments centrés dans ce test nous permettra de développer naturellement une méthode pour les mélanges généralisés de la famille réunissant les distributions normales et exponentielles.

Dans le cas où $T_{deux}(p) = 0$, avec $T_{un}(p) \neq 0$, on aura, d'après [13.57], comme $m_2 > 0$, pour la même valeur de p :

$$T_{trois}(p) = \left(\begin{array}{l} -72 M_1^7 + 108 M_1^5 M_2 - 54 M_1^3 M_2^2 + 9 M_1 M_2^3 \\ +p \left(\begin{array}{l} -72 M_1^7 + 180 M_1^5 M_2 - 180 M_1^3 M_2^2 + 45 M_1 M_2^3 + 12 M_1^4 M_3 \\ +36 M_1^2 M_2 M_3 - 15 M_2^2 M_3 - 2 M_1 M_2^3 - 12 M_1^3 M_4 + 6 M_1 M_2 M_4 \end{array} \right) \\ +p^2 \left(\begin{array}{l} -504 M_1^7 + 468 M_1^5 M_2 - 198 M_1^3 M_2^2 + 27 M_1 M_2^3 + 60 M_1^4 M_3 \\ -3 M_2^2 M_3 - 4 M_1 M_2^3 - 6 M_1 M_2 M_4 + 2 M_3 M_4 \end{array} \right) \\ +p^3 \left(\begin{array}{l} 1080 M_1^7 - 972 M_1^5 M_2 + 324 M_1^3 M_2^2 - 45 M_1 M_2^3 \\ +36 M_1^4 M_3 - 36 M_1^2 M_2 M_3 + 27 M_2^2 M_3 + 18 M_1 M_2^3 - 48 M_1^3 M_4 - 2 M_3 M_4 \end{array} \right) \\ +p^4 \left(\begin{array}{l} 2160 M_1^7 - 3960 M_1^5 M_2 + 3042 M_1^3 M_2^2 - 495 M_1 M_2^3 + 276 M_1^4 M_3 - 1008 M_1^2 M_2 M_3 \\ +183 M_2^2 M_3 + 64 M_1 M_2^3 + 252 M_1^3 M_4 - 48 M_1 M_2 M_4 - 12 M_3 M_4 \end{array} \right) \\ +p^5 \left(\begin{array}{l} -4896 M_1^5 M_2 + 3384 M_1^3 M_2^2 - 675 M_1 M_2^3 + 1416 M_1^4 M_3 - 1620 M_1^2 M_2 M_3 \\ +177 M_2^2 M_3 + 218 M_1 M_2^3 - 144 M_1^3 M_4 + 156 M_1 M_2 M_4 - 40 M_3 M_4 \end{array} \right) \\ +p^6 \left(\begin{array}{l} 2754 M_1^3 M_2^2 - 369 M_1 M_2^3 + 1224 M_1^4 M_3 - 1584 M_1^2 M_2 M_3 \\ +189 M_2^2 M_3 + 252 M_1 M_2^3 - 420 M_1^3 M_4 + 216 M_1 M_2 M_4 - 56 M_3 M_4 \end{array} \right) \\ +p^7 \left(\begin{array}{l} -441 M_1 M_2^3 - 972 M_1^2 M_2 M_3 - 33 M_2^2 M_3 + 262 M_1 M_2^3 \\ -60 M_1^3 M_4 + 258 M_1 M_2 M_4 - 60 M_3 M_4 \end{array} \right) \\ +p^8 (123 M_2^2 M_3 + 56 M_1 M_2^3 + 66 M_1 M_2 M_4 - 34 M_3 M_4) \\ -14 p^9 M_3 M_4 \end{array} \right) = 0.$$

Comme précédemment, la nullité simultanée de ces deux polynômes est équivalente à la nullité de la somme de leurs carrés. Si l'on note

$$P_{deux}(p) = T_{deux}(p)^2 + T_{trois}(p)^2.$$

T_{deux} et T_{trois} seront simultanément nuls s'il existe une racine, p , appartenant à $]0; 1[$, du polynôme :

$$\begin{aligned}
 P_{deux}(p) = & 81(1 + M_1^2)(-2M_1^2 + M_2)^6 \\
 & + p18(2M_1^2 - M_2)^3 \left(\begin{aligned} & 72M_1^8 - 36M_1^6(6 + 5M_2) - 12M_1^5M_3 - 36M_1^3(2 + M_2)M_3 \\ & + 3M_1M_2(8 + 5M_2)M_3 + 12M_1^4(3M_2(12 + 5M_2) + M_4) \\ & + M_1^2(-216M_2^2 - 45M_2^3 + 2M_3^3 - 6(-2 + M_2)M_4) + 2(18M_2^3 + M_3^3 - 3M_2M_4) \end{aligned} \right) \\
 & + p^2 \left(\begin{aligned} & 77760M_1^{14} + 324M_2^6 - 67392M_1^{12}(-1 + 3M_2) - 10368M_1^{11}M_3 + 1728M_1^9(-2 + 7M_2)M_3 + 324M_2^3M_3^2 \\ & + 225M_2^4M_3^2 + 4M_3^4 + 144M_1^7M_3(3M_2(4 + 11M_2) - 4M_4) - 24M_2(9M_2^3 + M_3^2)M_4 + 36M_2^2M_4^2 \\ & + 432M_1^{10}(3M_2(-120 + 187M_2) + 4M_4) - 48M_1^5M_3(162M_2^2 + 351M_2^3 + M_3^2 + 6(6 + M_2)M_4) - \\ & 144M_1^8(-1026M_2^2 + 1233M_2^3 - 7M_3^2 + 6(2 + 5M_2)M_4) \\ & + 144M_1^6(9M_2^3(-53 + 65M_2) + 20M_3^2 - 5M_2M_3^2 + 6M_2(7 + 6M_2)M_4 + M_4^2) - \\ & 24M_1^4(27M_2^4(-27 + 35M_2) - 3M_2(6 + 29M_2)M_3^2 + 2(54M_2^2(2 + M_2) - M_3^2)M_4 + 6(-1 + M_2)M_4^2) + \\ & M_1^2(-2916M_2^5 + 2511M_2^6 - 1368M_2^2M_3^2 - 1332M_2^3M_3^2 + 4M_3^4 + \\ & 24(18M_2^3(4 + M_2) - (-2 + M_2)M_3^2)M_4 + 36(-4 + M_2)M_2M_4^2) + \\ & 12M_1M_2M_3(8M_3^2 - M_2(90M_2^2 + 117M_2^3 - 5M_3^2 + 12(2 + M_2)M_4)) + \\ & 36M_1^3M_3(-8M_3^2 + M_2(156M_2^2 + 249M_2^3 - 4M_3^2 + 8(5 + 2M_2)M_4)) \end{aligned} \right) \\
 & - 2p^3 \left(\begin{aligned} & 41472M_1^{14} - 324M_2^6 - 31104M_1^{12}(-9 + 2M_2) + 12960M_1^{11}M_3 - 4752M_1^9(4 + M_2)M_3 \\ & + 216M_2^3M_3^2 - 45M_2^4M_3^2 + 20M_3^4 - 72M_1^7M_3(-1680M_2 + 51M_2^2 - 10M_4) + \\ & 6M_2(99M_2^3 + (-6 + 5M_2)M_3^2)M_4 - 72M_2^2M_4^2 - 864M_1^{10}(3M_2(264 + M_2) + 11M_4) + \\ & 24M_1^5M_3(7M_3^2 + 18M_2(3M_2(-96 + 5M_2) - M_4) + 192M_4) + \\ & 144M_1^8(4185M_2^2 + 306M_2^3 - 5M_3^2 + 3(-4 + 31M_2)M_4) - \\ & 12M_1^3M_3(9M_2^3(-414 + 25M_2) + 20M_3^2 - 12M_2M_3^2 - 36(-11 + M_2)M_2M_4 - 2M_4^2) + \\ & M_1M_3(27M_2^4(-208 + 11M_2) - 6M_2(12 + 11M_2)M_3^2 - 4(18M_2^2(-17 + 2M_2) - M_3^2)M_4 - 12M_2M_4^2) + \\ & 72M_1^4(9M_2^4(72 + 13M_2) + M_2(278 + 11M_2)M_3^2 + (6M_2^2(27 + 2M_2) - M_3^2)M_4 - (4 + M_2)M_4^2) + \\ & 2M_1^2(-972M_2^5 - 405M_2^6 - 2466M_2^2M_3^2 + 90M_2^3M_3^2 - 4M_3^4 \\ & + 6(9(-44 + M_2)M_2^3 - (-6 + 5M_2)M_3^2)M_4 + 18M_2(8 + M_2)M_4^2) \\ & - 72M_1^6(270M_3^2 + M_2(3438M_2^2 + 423M_2^3 + 34M_3^2 + 3(38 + 31M_2)M_4)) \end{aligned} \right) \\
 & + p^4 \left(\begin{aligned} & -212544M_1^{14} + 20088M_2^6 + 5184M_1^{12}(294 + 211M_2) - 79488M_1^{11}M_3 + 1728M_1^9(270 + 193M_2)M_3 - \\ & 9M_2^3(76 + 89M_2)M_3^2 + 84M_3^4 - 24M_2(207M_2^3 - (9 + 2M_2)M_3^2)M_4 + 4(90M_2^2 + M_3^2)M_4^2 \\ & - 288M_1^7M_3(3M_2(1318 + 485M_2) + 7M_4) - 432M_1^{10}(9M_2(1280 + 487M_2) + 128M_4) + \\ & 144M_1^8(46170M_2^2 + 10782M_2^3 - 53M_3^2 + 12(-67 + 50M_2)M_4) - \\ & 36M_1^6(9M_2^3(13956 + 2045M_2) - 2440M_3^2 - 616M_2M_3^2 + 12M_2(-596 + 105M_2)M_4 - 32M_4^2) + \\ & 48M_1^3M_3(-9M_2^3(672 + 145M_2) + 86M_3^2 + 30M_2M_3^2 + 3M_2(184 + 11M_2)M_4 + M_4^2) + \\ & 12M_1^4(243M_2^4(557 + 49M_2) - 6M_2(1446 + 229M_2)M_3^2 \\ & + 2(9M_2^2(-870 + 59M_2) - 2M_3^2)M_4 - 48(-4 + M_2)M_4^2) + \\ & 8M_1M_3(27M_2^4(131 + 32M_2) - 6M_2(57 + 13M_2)M_3^2 + (9(-88 + M_2)M_2^2 - M_3^2)M_4 - 6M_2M_4^2) + \\ & M_1^2(-290628M_2^5 - 12231M_2^6 + 32976M_2^2M_3^2 + 5760M_2^3M_3^2 - 56M_3^4 \\ & - 24(9M_2^3(-253 + 8M_2) - (-12 + 5M_2)M_3^2)M_4 + 36(-52 + M_2)M_2M_4^2) + \\ & 48M_1^5M_3(45M_2(M_2(426 + 107M_2) - 2M_4) - 4(M_3^2 + 153M_4)) \end{aligned} \right)
 \end{aligned}$$

TSVP

→

Estimateurs des moments d'un mélange normale-exponentielle.

$$\begin{aligned}
 & \left(\begin{aligned}
 & 699840M_1^{14} - 10368M_1^{12}(-2 + 195M_2) + 49248M_1^{11}M_3 - 432M_1^9(2264 + 923M_2)M_3 + \\
 & 144M_1^7M_3(21M_2(762 + 221M_2) - 26M_4) + 432M_1^{10}(3M_2(1476 + 2257M_2) + 22M_4) - \\
 & 144M_1^5M_3(12717M_2^2 + 3681M_2^3 + 8M_3^2 + 22(-9 + 4M_2)M_4) - \\
 & 288M_1^8(14445M_2^2 + 8568M_2^3 - 113M_3^2 + 3(-75 + 62M_2)M_4) + \\
 & 6M_1^3M_3(9M_2^3(10540 + 3207M_2) - 24(18 + 31M_2)M_3^2 + 6M_2(-820 + 223M_2)M_4 - 8M_4^2) + \\
 & 72M_1^6(9M_2^3(5097 + 1892M_2) - 2296M_3^2 - 708M_2M_3^2 + 6M_2(-375 + 172M_2)M_4 + 42M_4^2) + \\
 & M_1M_3(-297M_2^4(184 + 61M_2) + 24M_2(49 + 62M_2)M_3^2 \\
 & - 4(9M_2^2(-188 + 17M_2) + 17M_3^2)M_4 + 60M_2M_4^2) \\
 & + M_1^2(145800M_2^5 + 29565M_2^6 - 60084M_2^2M_3^2 - 28314M_2^3M_3^2 + 200M_3^4 \\
 & + 12(9M_2^3(-283 + 32M_2) + (76 + 11M_2)M_3^2)M_4 + 72M_2(11 + 4M_2)M_4^2) - \\
 & 6M_1^4(243M_2^4(776 + 213M_2) - 54M_2(590 + 219M_2)M_3^2 \\
 & + 2(153M_2^2(-66 + 17M_2) - 106M_3^2)M_4 + 12(14 + 33M_2)M_4^2) + \\
 & 2(-1134M_2^6 + 1728M_2^3M_3^2 + 1413M_2^4M_3^2 - 22M_3^4 \\
 & + 12M_2(63M_3^2 - 10(1 + M_2)M_3^2)M_4 - 2(36M_2^2 - M_3^2)M_4^2)
 \end{aligned} \right) \\
 -2p^5 & \\
 & \left(\begin{aligned}
 & -1010880M_1^{14} + 6804M_2^6 + 57024M_1^{12}(158 + 81M_2) - 321408M_1^{11}M_3 + 6912M_1^9(161 + 256M_2)M_3 \\
 & + 7884M_2^3M_3^2 - 5679M_2^4M_3^2 + 440M_3^4 - 576M_1^7M_3(3M_2(1650 + 1519M_2) - 34M_4) + \\
 & 24M_2(117M_3^2 + (-60 + 79M_2)M_3^2)M_4 - \\
 & 4(63M_2^2 + 11M_3^2)M_4^2 - 432M_1^{10}(3M_2(19656 + 6659M_2) + 640M_4) + \\
 & 288M_1^5M_3(6426M_2^2 + 6435M_2^3 + 22M_3^2 - (64 + 21M_2)M_4) + \\
 & 144M_1^8(194940M_2^2 + 55764M_2^3 - 293M_3^2 + 24(-93 + 79M_2)M_4) - \\
 & 288M_1^6(9M_2^3(5583 + 1531M_2) - 3(399 + 172M_2)M_3^2 + 3M_2(-513 + 79M_2)M_4 - 20M_4^2) + \\
 & 12M_1^3M_3(-27M_2^3(1172 + 1645M_2) + 104M_3^2 + 2412M_2M_3^2 + 6M_2(36 + 499M_2)M_4 + 180M_4^2) \\
 & + 4M_1M_3(27M_2^4(-254 + 463M_2) - 3M_2(624 + 677M_2)M_3^2 \\
 & - 2(36M_2^2(-7 + 40M_2) - 55M_3^2)M_4 - 132M_2M_4^2) + \\
 & M_1^2(-277020M_2^5 - 92097M_2^6 + 93384M_2^2M_3^2 + 98100M_2^3M_3^2 - 1060M_3^4 \\
 & + 24(9M_2^3(22 + 61M_2) - (-132 + 215M_2)M_3^2)M_4 + 144M_2(7 + 17M_2)M_4^2) + \\
 & 48M_1^4(243M_2^4(295 + 83M_2) - 6M_2(1091 + 1002M_2)M_3^2 \\
 & - (81M_2^2(43 + 11M_2) + 205M_3^2)M_4 - 3(-2 + 59M_2)M_4^2) \\
 & 2332800M_1^{14} - 69012M_2^6 - 10368M_1^{12}(-71 + 377M_2) - \\
 & 425952M_1^{11}M_3 + 432M_1^9(168 + 29M_2)M_3 + 45M_2^3(272 + 35M_2)M_3^2 - \\
 & 780M_3^4 + 576M_1^7M_3(3M_2(276 + 37M_2) - 37M_4) + 6M_2(2133M_3^2 + 52(-9 + 2M_2)M_3^2)M_4 - \\
 & 8(90M_2^2 + 7M_3^2)M_4^2 + 864M_1^{10}(36M_2(175 + 117M_2) + 319M_4) + \\
 & 144M_1^5M_3(-6480M_2^2 + 1269M_2^3 + 106M_3^2 + 6(63 + 31M_2)M_4) - \\
 & 144M_1^8(83754M_2^2 + 9891M_2^3 - 491M_3^2 + 6(-444 + 427M_2)M_4) - \\
 & 24M_1^3M_3(27M_2^3(-915 + 79M_2) + 1682M_3^2 + 72M_2M_3^2 - 12M_2(-229 + 105M_2)M_4 - 19M_4^2) - \\
 & 72M_1^6(36M_2^3(-3981 + 32M_2) + 6(682 + 131M_2)M_3^2 - 3M_2(-3588 + 1045M_2)M_4 + 98M_4^2) + \\
 & M_1M_3(-81M_2^4(912 + 13M_2) - 6M_2(-4764 + 83M_2)M_3^2 \\
 & - 4(9M_2^2(-568 + 247M_2) - 91M_3^2)M_4 + 312M_2M_4^2) + \\
 & 2M_1^2(-486M_2^5(-960 + 17M_2) + 306M_2^2(-397 + 21M_2)M_3^2 - 112M_3^4 + \\
 & 3(27M_2^3(-906 + 125M_2) - 2(-516 + 377M_2)M_3^2)M_4 + 36M_2(61 + 5M_2)M_4^2) + \\
 & 72M_1^4(27M_2^4(-2328 + 73M_2) - 6M_2(-1134 + 145M_2)M_3^2 \\
 & - (9M_2^2(-816 + 169M_2) + 19M_3^2)M_4 - (100 + 27M_2)M_4^2)
 \end{aligned} \right) \\
 +p^6 & \\
 & \left(\begin{aligned}
 & 425952M_1^{11}M_3 + 432M_1^9(168 + 29M_2)M_3 + 45M_2^3(272 + 35M_2)M_3^2 - \\
 & 780M_3^4 + 576M_1^7M_3(3M_2(276 + 37M_2) - 37M_4) + 6M_2(2133M_3^2 + 52(-9 + 2M_2)M_3^2)M_4 - \\
 & 8(90M_2^2 + 7M_3^2)M_4^2 + 864M_1^{10}(36M_2(175 + 117M_2) + 319M_4) + \\
 & 144M_1^5M_3(-6480M_2^2 + 1269M_2^3 + 106M_3^2 + 6(63 + 31M_2)M_4) - \\
 & 144M_1^8(83754M_2^2 + 9891M_2^3 - 491M_3^2 + 6(-444 + 427M_2)M_4) - \\
 & 24M_1^3M_3(27M_2^3(-915 + 79M_2) + 1682M_3^2 + 72M_2M_3^2 - 12M_2(-229 + 105M_2)M_4 - 19M_4^2) - \\
 & 72M_1^6(36M_2^3(-3981 + 32M_2) + 6(682 + 131M_2)M_3^2 - 3M_2(-3588 + 1045M_2)M_4 + 98M_4^2) + \\
 & M_1M_3(-81M_2^4(912 + 13M_2) - 6M_2(-4764 + 83M_2)M_3^2 \\
 & - 4(9M_2^2(-568 + 247M_2) - 91M_3^2)M_4 + 312M_2M_4^2) + \\
 & 2M_1^2(-486M_2^5(-960 + 17M_2) + 306M_2^2(-397 + 21M_2)M_3^2 - 112M_3^4 + \\
 & 3(27M_2^3(-906 + 125M_2) - 2(-516 + 377M_2)M_3^2)M_4 + 36M_2(61 + 5M_2)M_4^2) + \\
 & 72M_1^4(27M_2^4(-2328 + 73M_2) - 6M_2(-1134 + 145M_2)M_3^2 \\
 & - (9M_2^2(-816 + 169M_2) + 19M_3^2)M_4 - (100 + 27M_2)M_4^2)
 \end{aligned} \right) \\
 +2p^7 &
 \end{aligned}$$

TSVP →

Estimateurs des moments d'un mélange normale-exponentielle.

$$\begin{aligned}
& \left(\begin{aligned}
& 4665600M_1^{14} + 23814M_2^6 - 5184M_1^{12}(9119 + 5340M_2) + 3017088M_1^{11}M_3 - 1728M_1^9(5084 + 5939M_2)M_3 \\
& + 90828M_2^3M_3^2 + 42903M_2^4M_3^2 - 3668M_3^4 + 288M_1^7M_3(3M_2(33612 + 15403M_2) - 442M_4) + \\
& \quad 24M_2(513M_2^3 - (547 + 182M_2)M_3^2)M_4 - 8(189M_2^2 - 10M_3^2)M_4^2 \\
& \quad + 864M_1^{10}(3M_2(41334 + 16541M_2) + 1400M_4) - \\
& \quad 864M_1^5M_3(32706M_2^2 + 10131M_2^3 - 149M_3^2 + 3(-172 + 81M_2)M_4) - \\
& \quad 144M_1^8(838161M_2^2 + 239076M_2^3 - 5365M_3^2 + 12(-882 + 1039M_2)M_4) + \\
& \quad 144M_1^3M_3(9M_2^3(7424 + 1779M_2) - 6(325 + 194M_2)M_3^2 + 3M_2(-1252 + 301M_2)M_4 - 13M_4^2) + \\
& \quad 36M_1^6(9M_2^3(225120 + 47917M_2) - 24(3046 + 2029M_2)M_3^2 + 24M_2(-2964 + 1955M_2)M_4 + 2188M_4^2) + \\
& \quad 8M_1M_3(-81M_2^4(1585 + 322M_2) + 6M_2(2885 + 642M_2)M_3^2 - \\
& \quad (27M_2^2(-612 + 61M_2) + 269M_3^2)M_4 + 168M_2M_4^2) + \\
& \quad 3M_1^2(679428M_2^5 + 82053M_2^6 - 495216M_2^2M_3^2 - 165744M_2^3M_3^2 + 2960M_3^4 + \\
& \quad 8(27M_2^3(-569 + 106M_2) + 2(727 + 411M_2)M_3^2)M_4 + 72M_2(36 + 13M_2)M_4^2) - \\
& \quad 36M_1^4(135M_2^4(4281 + 679M_2) - 72M_2(1619 + 655M_2)M_3^2 + \\
& \quad 8(27M_2^2(-223 + 77M_2) + 52M_3^2)M_4 + 60(5 + 19M_2)M_4^2) \\
& \quad 19116 M_2^6 + 25920 M_1^{12} (-19 + 408 M_2) - 4380480 M_1^{11} M_3 + \\
& \quad 1728 M_1^9 (9395 + 7447 M_2) M_3 - 108 M_2^3 (213 + 331 M_2) M_3^2 + 3020 M_3^4 \\
& -432 M_1^{10} (3 M_2 (22122 + 22895 M_2) - 1700 M_4) + 18 M_2 (36 M_2^3 + 5 (74 + 119 M_2) M_3^2) M_4 \\
& + 8 (27 M_2^2 - 59 M_3^2) M_4^2 - 144 M_1^7 M_3 (217854 M_2 + 130323 M_2^2 + 878 M_4) \\
& + 432 M_1^5 M_3 (15 M_2^2 (3567 + 1777 M_2) - 459 M_3^2 - 2 (322 + 73 M_2) M_4) \\
& + 72 M_1^8 (644274 M_2^2 + 452853 M_2^3 - 14470 M_3^2 - 12 (-201 + 176 M_2) M_4) - \\
& \quad 72 M_1^6 (9 M_2^3 (42853 + 25363 M_2) - 3 (10866 + 15383 M_2) M_3^2 \\
& \quad - 3 M_2 (2652 + 1397 M_2) M_4 - 224 M_4^2) + \\
& \quad 18 M_1^3 M_3 (-27 M_2^3 (13772 + 5799 M_2) + 13216 M_3^2 + \\
& \quad 19728 M_2 M_3^2 + 48 M_2 (279 + 289 M_2) M_4 + 252 M_4^2) + \\
& \quad 3 M_1 M_3 (27 M_2^4 (2536 + 2761 M_2) - 24 M_2 (2121 + 829 M_2) M_3^2 - \\
& \quad 8 (9 M_2^2 (299 + 241 M_2) - 244 M_3^2) M_4 - 96 M_2 M_4^2) + \\
& \quad 3 M_1^2 (-134784 M_2^5 - 112941 M_2^6 + 527652 M_2^2 M_3^2 + 224946 M_2^3 M_3^2 - 5776 M_3^4 \\
& + 6 (9 M_2^3 (652 + 259 M_2) - 4 (219 + 872 M_2) M_3^2) M_4 + 72 M_2 (-1 + 40 M_2) M_4^2) + \\
& \quad 18 M_1^4 (27 M_2^4 (14264 + 7993 M_2) - 6 M_2 (35482 + 27651 M_2) M_3^2 - \\
& \quad 4 (18 M_2^2 (300 + 197 M_2) - 151 M_3^2) M_4 - 48 (2 + 41 M_2) M_4^2) \\
& \quad 15595200M_1^{12} + 143748M_2^6 + 1762560M_1^{11}M_3 - 576M_1^9(-3774 + 18809M_2)M_3 \\
& + 3M_2^3(1084 + 10887M_2)M_3^2 + 2904M_3^4 + 48M_1^7M_3(3M_2(20164 + 114513M_2) - 2516M_4) + \\
& 1728M_1^{10}(M_2(-17352 + 6919M_2) - 375M_4) - 24M_2(789M_2^3 + 4(18 + 143M_2)M_3^2)M_4 + \\
& 8(90M_2^2 + 127M_3^2)M_4^2 - 144M_1^5M_3(42018M_2^2 + 76835M_2^3 - 2173M_3^2 - 4(-237 + 1156M_2)M_4) + \\
& -48M_1^8(-476658M_2^2 + 399267M_2^3 - 29714M_3^2 - 6(-1706 + 7265M_2)M_4) + \\
& 48M_1^6(9M_2^3(-17485 + 29781M_2) - 32912M_3^2 - 87411M_2M_3^2 - 6M_2(-3601 + 6781M_2)M_4 - 1286M_4^2) + \\
& -12M_1^3M_3(-9M_2^3(12276 + 24395M_2) + 9944M_3^2 + 40860M_2M_3^2 + 24M_2(11 + 1869M_2)M_4 + 8M_4^2) + \\
& \quad 4M_1M_3(-9M_2^4(2342 + 5327M_2) + 3M_2(2712 + 5701M_2)M_3^2 + \\
& \quad 8(45M_2^2(11 + 58M_2) - 347M_3^2)M_4 - 1068M_2M_4^2) + \\
& \quad M_1^2(-1121364M_2^5 + 286875M_2^6 - 525528M_2^2M_3^2 - 737268M_2^3M_3^2 + 29588M_3^4 \\
& - 24(12M_2^3(-427 + 473M_2) - (102 + 5101M_2)M_3^2)M_4 + 72M_2(-70 + 13M_2)M_4^2) - \\
& \quad 72M_1^4(3M_2^4(-16971 + 15163M_2) - M_2(31586 + 48437M_2)M_3^2 - \\
& \quad (M_2^2(-2886 + 6293M_2) - 461M_3^2)M_4 - 2(59 + 184M_2)M_4^2)
\end{aligned} \right)
\end{aligned}$$

TSVP →

Estimateurs des moments d'un mélange normale-exponentielle.

$$\begin{aligned}
 & \left(\begin{aligned}
 & 149364 M_2^6 + 5184 M_1^9 (-4505 + 1561 M_2) M_3 + 95256 M_2^3 M_3^2 - 30735 M_2^4 M_3^2 - 21364 M_3^4 + \\
 & \quad 6 M_2 (657 M_2^3 + (-1334 + 4863 M_2) M_3^2) M_4 \\
 & \quad - 24 (51 M_2^2 + 125 M_3^2) M_4^2 + 25920 M_1^{10} (2394 M_2 + 5 M_4) \\
 & \quad + 432 M_1^8 (-5462 M_3^2 + 3 M_2 (3 M_2 (-19657 + 3713 M_2) - 2255 M_4) + 1612 M_4) \\
 & \quad + 72 M_1^7 M_3 (3 (170608 - 91239 M_2) M_2 + 13150 M_4) \\
 & \quad + 24 M_1^5 M_3 (-29083 M_3^2 + 6 M_2 (324 M_2 (-588 + 325 M_2) - 12727 M_4) + 17784 M_4) - \\
 & \quad \quad 36 M_1^3 M_3 (-190014 M_2^3 + 101643 M_2^4 + 7492 M_3^2 - \\
 & \quad \quad 30052 M_2 M_3^2 - 2 M_2 (-6242 + 19861 M_2) M_4 + 336 M_4^2) - \\
 & \quad \quad 72 M_1^6 (9 M_2^3 (-78994 + 19865 M_2) + 39790 M_3^2 - \\
 & \quad \quad 94720 M_2 M_3^2 - 6 M_2 (-4127 + 9592 M_2) M_4 + 630 M_4^2) + \\
 & \quad \quad 3 M_1 M_3 (9 M_2^4 (-47600 + 9733 M_2) - 14 M_2 (-5252 + 3223 M_2) M_3^2 - \\
 & \quad \quad 4 (6 M_2^2 (-1333 + 2875 M_2) - 2495 M_3^2) M_4 + 5880 M_2 M_4^2) \\
 & \quad + 2 M_1^2 (994356 M_2^5 - 233685 M_2^6 - 557838 M_2^2 M_3^2 + 573282 M_2^3 M_3^2 - 36356 M_3^4 + \\
 & \quad 6 (9 M_2^3 (-2290 + 2897 M_2) - (-2198 + 26581 M_2) M_3^2) M_4 - 72 M_2 (-43 + 148 M_2) M_4^2) + \\
 & \quad \quad 24 M_1^4 (54 M_2^4 (-9864 + 3433 M_2) - 3 M_2 (-55110 + 87413 M_2) M_3^2 - \\
 & \quad \quad (27 M_2^2 (-2018 + 3429 M_2) - 8968 M_3^2) M_4 + 3 (-104 + 443 M_2) M_4^2) \\
 & \quad 112104 M_2^6 + 8208000 M_1^9 M_3 - 113940 M_2^2 M_3^2 + 69057 M_2^4 M_3^2 + \\
 & \quad 103468 M_3^4 + 288 M_1^7 M_3 (6 M_2 (-60115 + 9717 M_2) - 4775 M_4) - \\
 & \quad 24 M_2 (45 M_2^3 + (-1131 + 2330 M_2) M_3^2) M_4 + 24 (57 M_2^2 + 367 M_3^2) M_4^2 + \\
 & \quad 1728 M_1^8 (78408 M_2^2 + 1007 M_3^2 + 5(77 + 101 M_2) M_4) \\
 & \quad + 48 M_1^5 M_3 (1837890 M_2^2 - 377217 M_3^2 + 28954 M_3^2 + 12(-1043 + 6918 M_2) M_4) + \\
 & \quad 48 M_1^3 M_3 (36 M_2^3 (-20520 + 3169 M_2) - 2(-16441 + 22749 M_2) M_3^2 - \\
 & \quad \quad 18 M_2 (-1144 + 3467 M_2) M_4 + 1111 M_4^2) + \\
 & \quad 36 M_1^6 (9 M_2^3 (-268852 + 36737 M_2) + 437256 M_3^2 - \\
 & \quad 267760 M_2 M_3^2 - 12 M_2 (-1588 + 13353 M_2) M_4 + 5380 M_4^2) + \\
 & \quad 8 M_1 M_3 (-27 M_2^4 (-3005 + 1726 M_2) + 6 M_2 (-33087 + 4372 M_2) M_3^2 + \\
 & \quad (9 M_2^2 (-3536 + 5105 M_2) - 10193 M_3^2) M_4 - 7734 M_2 M_4^2) - \\
 & \quad 36 M_1^4 (9 M_2^4 (-99027 + 15485 M_2) - 2 M_2 (-288134 + 133297 M_2) M_3^2 - \\
 & \quad 2(3 M_2^2 (-6390 + 17849 M_2) - 8498 M_3^2) M_4 + 20(14 + 335 M_2) M_4^2) + \\
 & \quad M_1^2 (1643652 M_2^5 + 731511 M_2^6 + 10438704 M_2^2 M_3^2 - 1871280 M_2^3 M_3^2 + 184904 M_3^4 - \\
 & \quad 24(9 M_2^3 (-1471 + 3290 M_2) - (-3000 + 34387 M_2) M_3^2) M_4 + 144 M_2 (-11 + 839 M_2) M_4^2) \\
 & \quad 6 (-43794 M_2^6 + 3 M_2^3 (-2744 + 863 M_2) M_3^2 + 12734 M_3^4) + \\
 & \quad 410400 M_1^8 M_4 + 24 M_2 (1170 M_2^3 - (-549 + 958 M_2) M_3^2) M_4 - \\
 & \quad 8 (90 M_2^2 - 611 M_3^2) M_4^2 - 4320 M_1^7 M_3 (2609 M_2 + 24 M_4) \\
 & \quad + 48 M_1^5 M_3 (894537 M_2^2 - 79560 M_2^3 + 8333 M_3^2 + 18 (191 + 1313 M_2) M_4) - \\
 & \quad 72 M_1^6 (507843 M_2^3 - 45788 M_2^3 + 20332 M_2 M_3^2 + 3 M_2 (4202 + 2261 M_2) M_4 - 350 M_4^2) + \\
 & \quad 6 M_1^3 M_3 (9 M_2^3 (-247332 + 25603 M_2) + 192304 M_3^2 - \\
 & \quad 125112 M_2 M_3^2 - 6 M_2 (-3316 + 29545 M_2) M_4 + 4988 M_4^2) + \\
 & \quad M_1 M_3 (-27 M_2^4 (-9688 + 5711 M_2) + 24 M_2 (-48723 + 3247 M_2) M_3^2 + \\
 & \quad 4 (9 M_2^2 (-2672 + 4145 M_2) - 10085 M_3^2) M_4 - 34680 M_2 M_4^2) + \\
 & \quad M_1^2 (-526824 M_2^5 + 162729 M_2^6 + 6884172 M_2^2 M_3^2 - 576306 M_2^3 M_3^2 + 78232 M_3^4 - \\
 & \quad 12 (9 M_2^3 (-155 + 2176 M_2) - 11 (-192 + 2753 M_2) M_3^2) M_4 + 72 M_2 (79 + 917 M_2) M_4^2) - \\
 & \quad 6 M_1^4 (243 M_2^4 (-3352 + 833 M_2) - 6 M_2 (-384906 + 71791 M_2) M_3^2 - \\
 & \quad 2 (9 M_2^2 (1734 + 10567 M_2) - 21556 M_3^2) M_4 + 12 (118 + 1817 M_2) M_4^2) \\
 \end{aligned} \right)
 \end{aligned}$$

TSVP →

Estimateurs des moments d'un mélange normale-exponentielle.

$$\begin{aligned}
& +p^{14} \left(\begin{aligned} & 624348M_2^6 + 143316M_2^3M_3^2 + 47583M_2^4M_3^2 + 174008M_3^4 - \\ & 24M_2(333M_2^3 + (-232 + 1151M_2)M_3^2)M_4 - 4(423M_2^2 - 2132M_3^2)M_4^2 + \\ & 576M_1^5M_3(34938M_2^2 + 238M_3^2 + 810M_4 + 721M_2M_4) + \\ & 12M_1^3M_3(9M_2^3(-224084 + 14211M_2) - \\ & 12(-11994 + 4769M_2)M_3^2 - 6M_2(5812 + 15167M_2)M_4 + 3316M_4^2) + \\ & 4M_1M_3(-27M_2^4(19666 + 571M_2) + 3M_2(-209504 + 5489M_2)M_3^2 + \\ & 2(18M_2^2(298 + 1097M_2) - 7619M_3^2)M_4 - 14352M_2M_4^2) + \\ & 48M_1^4(18M_2(21465M_3^2 + (-15480 + 1799M_2)M_3^2) + \\ & (9M_2^2(2613 + 964M_2) - 4195M_3^2)M_4 - 12(11 + 150M_2)M_4^2) + \\ & M_1^2(4359420M_2^5 + 194481M_2^6 + 12677400M_2^2M_3^2 - 597924M_2^3M_3^2 + \\ & 96868M_3^4 - 24(9M_2^3(874 + 1279M_2) - (328 + 19265M_2)M_3^2)M_4 + \\ & 36M_2(236 + 2641M_2)M_4^2) + 720M_1^6(2362M_3^2 + M_4(-1854M_2 + 5M_4)) \end{aligned} \right) \\
& +2p^{15} \left(\begin{aligned} & -14M_3M_4(9(M_1(306M_1^2 - 41M_2)M_2^2 + (136M_1^4 - 176M_1^2M_2 + 21M_2^2)M_3 + 28M_1M_3^2) - \\ & 4(105M_1^3 - 54M_1M_2 + 14M_3)M_4) - \\ & 2(342M_2^3 + 300M_1^3M_3 - 1248M_1M_2M_3 + 142M_3^2 - 3M_2M_4 + 24M_1^2(81M_2^2 + M_4)) \\ & (189M_2^3 - 182M_3^2 - 60M_1^2M_4 + 12M_2(71M_1M_3 + M_4)) + \\ & 4(-49M_3^2 + 9M_2M_4)(4536M_1^4M_2 + 207M_2^3 - 1704M_1^3M_3 + \\ & 774M_1M_2M_3 - 97M_3^2 - 18M_2M_4 + M_1^2(-1269M_2^2 + 36M_4)) - \\ & (123M_2^2M_3 + 56M_1M_3^2 + 66M_1M_2M_4 - 34M_3M_4) \\ & (972M_1^2M_2M_3 + 60M_1^3M_4 + 3M_3(11M_2^2 + 20M_4) + M_1(441M_3^2 - 262M_3^2 - 258M_2M_4)) \end{aligned} \right) \\
& +p^{16} \left(\begin{aligned} & (123M_2^2M_3 + 56M_1M_3^2 + 66M_1M_2M_4 - 34M_3M_4)^2 - \\ & 8(-49M_3^2 + 9M_2M_4)(342M_2^3 + 300M_1^3M_3 - 1248M_1M_2M_3 + 142M_3^2 - 3M_2M_4 + 24M_1^2(81M_2^2 + M_4)) \\ & + (189M_2^3 - 182M_3^2 - 60M_1^2M_4 + 12M_2(71M_1M_3 + M_4))^2 + \\ & 28M_3M_4(972M_1^2M_2M_3 + 60M_1^3M_4 + 3M_3(11M_2^2 + 20M_4) + M_1(441M_3^2 - 262M_3^2 - 258M_2M_4)) \end{aligned} \right) \\
& +p^{17} \left(\begin{aligned} & -28M_3M_4(123M_2^2M_3 + 56M_1M_3^2 + 66M_1M_2M_4 - 34M_3M_4) \\ & +4(-49M_3^2 + 9M_2M_4)(189M_2^3 - 182M_3^2 - 60M_1^2M_4 + 12M_2(71M_1M_3 + M_4)) \end{aligned} \right) \\
& +p^{18} (196M_3^2M_4^2 + (98M_3^2 - 18M_2M_4)^2)
\end{aligned}$$

Ce cas correspond à nouveau à une redondance dans les équations du système, puisque les quatre équations ne permettent d'exprimer que trois paramètres. Il faut donc recourir à la cinquième équation de moments et comparer une solution éventuellement issue de cette méthode avec les précédentes grâce au test T_6^{Normex} .

13.3 Méthode d'inversion du système \mathfrak{S}_4 .

La méthode d'inversion du système des moments, du mélange d'une distribution normale et d'une distribution exponentielle, se déroule en trois étapes.

►1. Même si nous ne sommes pas sûr de la possibilité théorique qu'elles existent (nous n'avons trouvé aucun exemple), nous cherchons les racines réelles appartenant à $]0; 1[$ des polynômes P_{un} et P_{deux} . Si p_0 est une telle racine, nous tirons, dans un premier temps, m_2 comme racine réelle strictement positive du polynôme [13.56] où p est remplacé par p_0 . Si un tel $m_2 > 0$ existe, il ne nous reste qu'à reporter les valeurs de p_0 et m_2 ainsi obtenues, dans les expressions littérales de μ_1 et σ_1 . Nous ne conservons, bien sûr, que les ensembles de paramètres solutions statistiquement acceptables.

►2. Nous cherchons les racines réelles appartenant à $]0; 1[$ du polynôme P_{Normex} , ce qui nous donnera les différents p candidats solutions. Nous portons ces résultats dans les expressions littérales de m_2 , μ_1 et σ_1 [13.48] et ne retenons que les ensembles solutions vérifiant $m_2 > 0$ et $\sigma_1 > 0$. Cette phase de l'inversion s'achève, dans le cas de non unicité des solutions statistiquement acceptables, par le choix de la solution minimisant le test T_{Normex}^5 .

►3. Si 1. et 2. fournissent des solutions, nous choisissons celle de toutes les solutions qui minimisera le test T_{Normex}^6 .

13.4 Exemples d'inversions du système \mathfrak{S}_4 .

Nous n'avons pas une connaissance aussi précise du jacobien de ce système, que celles que nous avons pour les systèmes relatifs au mélange de deux distributions normales. Le choix des exemples en devient peut-être moins pertinents. Nous ne donnerons, donc, d'exemple, que dans le cas général, et dans le cas d'annulation du jacobien que nous avons cité en exemple plus haut.

13.4.1 Exemples d'inversion théorique.

Les résultats que nous commenterons ici, sont récapitulés dans le tableau FIG. II.18.

Les deux premiers exemples, (annexe 56 et 57), illustrent les deux formes unimodales que peuvent prendre un mélange normale-exponentielle. L'exemple de l'annexe 58, est celui d'un mélange de deux distributions de même moyennes et de même variances théoriques. Il n'y a pas de difficulté dans ces trois exemples.

Les deux cas suivants, annexe 59 et annexe 60, sont ceux correspondant aux cas d'annulation du jacobien. Ils reproduisent deux problèmes que nous avons déjà rencontrés pour le mélange de deux distributions normales, puisque le premier est un cas de racine double du polynôme P_{Normex} , et que le second est celui d'une solution qui n'est pas atteignable par ce système, et nécessite, a priori, le recours à une équation supplémentaire.

13.4.2 Exemples d'estimation des paramètres d'un échantillon.

Les résultats obtenus pour les moments empiriques et présentés en annexe 61 à 70, sont résumés dans les tableaux FIG. II.19 et FIG. II.20. Nous nous concentrons, après avoir détaillé dans les annexes 61 à 63, le pendant des exemples généraux traités dans le cadre du système théorique, sur l'étude de l'inversion du système empirique, au voisinage de deux cas d'annulation du jacobien. Les exemples des annexes 64, 65 et 66 montrent que dans le cas induisant théoriquement une racine double, le "décalage" provoqué par les moments empiriques donne deux solutions proches ou pas de solution. De même, comme le montrent les résultats cités en annexes 67 à 70, le cas "d'insuffisance" du système, engendre, lors du passage aux moment empiriques, des situations critiques du type de celles que nous avons déjà rencontrées.

13.4.3 Conclusion.

Compte tenu de la complexité d'écriture des polynômes que nous considérons, nous ne développerons pas de paragraphe équivalent, pour le mélange normale-exponentielle, à ceux présentant, dans les autres cas, la méthode d'estimation avec les moments empiriques. Nous en sommes à un point où nous connaissons une méthode d'estimation des paramètres d'un mélange d'une distribution normale et d'une distribution exponentielle, dont nous avons identifié, sous une forme certe complexe, les domaines de non application. Cette méthode pourra donc être appliquée, comme la méthode initiale de Pearson, aux problèmes de continuité des solutions dans le voisinage de la variété annulant le jacobien.

INVERSION DE S4 POUR DES MOMENTS THEORIQUES

	Moyenne F1	Variance F1	p	Paramètre F2	1>racines=0	acceptable	Test Tsnormet	Commentaires
	Annexe 56	0.0	2.0	0.3	1.0	0.10290133	oui	
	0.0	2.0	0.3	1.0	0.36081114	oui	0.322049	
					0.98967385	oui	1.317914	
Annexe 57	1.0	2.0	0.5	4.0	0.10165285	oui	0.099264	
	1.0	2.0	0.5	4.0	0.5	oui	0.0	
Annexe 58	1.0	1.0	0.5	1.0				
	1.0	1.0	0.5	1.0	0.5	oui	0.0	
Annexe 59	0.0	0.94136684	0.7	1.0				Il y a une racine double
					0.51675553	oui	0.332803	
					0.52344703	oui	0.044653	
					0.7	oui	0.0	
					0.7	oui	0.0	
Annexe 60	0.0	4.68584617	0.5	1.0	0.16757485	oui	0.323474	La solution ne peut être atteinte.
	-2.57486722	0.87873829	0.16757485	1.11899910	0.96222234	non		

Légende :

Paramètres théoriques : Racine retenue pour l'inversion :

Paramètres calculés :

FIG. II.18. -

Estimateurs des moments d'un mélange normale-exponentielle.

INVERSION DE S₄ POUR DES MOMENTS EMPIRIQUES 1

	p			Paramètre F ₂	1>racines>0			Commentaires
	Moyenne F ₁	Variance F ₁	p		acceptable	Test T _{5normex}		
Annexe 61	0.0	2.0	0.3	1.0				
	0.05477885	2.06568014	0.31039688	0.98808554	0.09447871	oui	0.118002	
					0.31039688	oui	0.007989	
					0.36396935	oui	0.328589	
					0.99281194	oui	1.616472	
	1.0	2.0	0.5	4.0				
Annexe 62	0.98817036	1.99177643	0.50000254	4.02668958				
	1.0	1.0	0.5	1.0	0.10290405	oui	0.097352	
					0.50000254	oui	0.003090	
Annexe 63	1.01732736	1.02979712	0.48367114	0.98803143				
	0.0	0.94136684	0.5	1.0	3.0795E-7	non		
					0.48367114	oui	0.007055	
Annexe 64	-0.51230776	0.66229771	0.29223659	0.91541446				
	0.0	0.94136684	0.5	1.0	0.29223659	oui	0.013325	l'estimation n'est pas bonne.
					0.44448145	oui	0.277748	
Annexe 65	0.0	0.94136684	0.3	1.0				
	0.37595574	0.93336298	0.55713404	1.10887081	0.29023618	oui	0.224985	l'estimation n'est pas bonne.
					0.55713404	oui	0.063909	
Annexe 66	0.0	0.94136684	0.3	1.0				
	0.02546270	0.94798025	0.30571448	1.000311	0.27891681	oui	0.005002	
					0.29549171	oui	0.223447	
					0.30571448	oui	0.000616	
					0.45347870	oui	0.036774	

FIG. II.19. -

Estimateurs des moments d'un mélange normale-exponentielle.

Légende :

Paramètres théoriques :

Racine retenue pour l'inversion

Paramètres calculés :



Il ya deux racines proches

INVERSION DE S_4 POUR DES MOMENTS EMPIRIQUES 2

	Moyenne F1	Variance F1	p	Paramètre F2	1>racines>0	acceptable	Test Tsnormex	Commentaires
	Annexe 67	0.0	4.68584617	0.3	1.0	0.10636466 0.28572474 0.31665744 0.99841621	oui oui oui non	0.352289 0.030084 0.184602
Annexe 68	0.0	4.68584617	0.3	1.0	0.09941814 0.99711692	oui non	0.423201	l'estimation est mauvaise.
Annexe 69	0.0	4.68584617	0.5	1.0	0.17005080 0.47942449 0.52499534 0.97928108	oui oui oui non	0.170210 0.048000 0.180466	Il y a deux ensembles de paramètres proches
Annexe 70	0.0	4.68584617	0.5	1.0	0.16587310 0.95811690	oui non	0.336006	l'estimation est mauvaise.
	-2.59764579	0.84669134	0.16587310	1.11710361				

Légende :	
Paramètres théoriques :	[lightgray box]
Paramètres calculés :	[darkgray box]
Racine retenue pour l'inversion :	[yellow box]

FIG. II.20. -

Estimateurs des moments d'un mélange normale-exponentielle.

- CHAPITRE 14 -

Estimation des paramètres d'un mélange généralisé simple.

14.1 Enoncé de la méthode.

Nous avons présenté, jusqu'ici, les méthodes d'estimation des paramètres dans les trois types de mélanges de deux distributions rencontrés pour la famille réunissant les distributions normales et exponentielles. Nous savons traiter indépendamment chacun de ces cas, si nous savons l'identifier, ce qui n'est, par définition, pas le cas dans le cadre des mélanges généralisés.

La méthode d'estimation des paramètres d'un mélange généralisé, que nous proposons, consiste à traiter, dans chacun des trois cas, normale-normale, normale-exponentielle et exponentielle-exponentielle, avant de choisir, parmi les trois couples (type de mélange-paramètres estimés), le mieux adapté au mélange étudié. Il semble naturel qu'une fois de plus, le critère de choix se base sur les équations de moments inutilisées lors de l'estimation par la méthode des moments. Nous proposons, pour cela d'utiliser la sixième équation de moments centrés, si l'estimation du mélange, supposé être un mélange normale-normale, est faite par \mathfrak{S}_1 , et la cinquième équation de moments centrés, lorsque c'est \mathfrak{S}_2 qui est utilisé.

Nous disposons des test T_5 (resp. T_6) et T_5^{Normex} (resp. T_6^{Normex}) pour les mélanges normale-normale et normale-exponentielle, et il suffit de définir leur équivalent pour les mélanges exponentielle-exponentielle. Nous définissons ainsi :

$$T_5^{Exp} = \left| \frac{\left(\begin{array}{l} p(120m_1^5 - 120m_1^4M_1 + 60m_1^3M_1^2 - 20m_1^2M_1^3 + 5m_1M_1^4 - M_1^5) \\ + (1-p)(120m_2^5 - 120m_2^4M_1 + 60m_2^3M_1^2 - 20m_2^2M_1^3 + 5m_2M_1^4 - M_1^5) \end{array} \right) - M_5^c}{2M_5^c} \right|$$

$$\text{et } T_6^{Exp0} = \left| \frac{\left(\begin{array}{l} +p(720m_1^6 - 720m_1^5M_1 + 360m_1^4M_1^2 - 120m_1^3M_1^3 + 30m_1^2M_1^4 - 6m_1M_1^5 + M_1^6) \\ +(1-p)(720m_2^6 - 720m_2^5M_1 + 360m_2^4M_1^2 - 120m_2^3M_1^3 + 30m_2^2M_1^4 - 6m_2M_1^5 + M_1^6) \end{array} \right) - M_6^c}{2M_6^c} \right|,$$

ce qui permet d'établir la méthode. Les résultats sur les mélanges théoriques, donnent, comme nous l'avons vu précédemment, des test nuls pour les bons ensembles de paramètres associés au bon type de mélange, ce qui justifie cette méthode de choix.

14.2 Exemples d'applications.

Nous ne disposons pas de règle, a priori, définissant des mélanges généralisés pour lesquels les tests de choix du type de mélange peuvent être mis en défaut. Les exemples qui suivent ont été choisis, après l'étude de nombreux mélanges de types et de paramètres différents, pour l'illustration qu'il donnent des réussites et des échecs de la méthode.

14.2.1 Exemples dans le cas théorique.

Il n'y a pas de surprise dans les exemples sur les moments théoriques, puisque les bons paramètres et les bons types de mélanges sont toujours choisis. Nous constatons, aussi, que les mauvais types de mélanges ne donnent, souvent, pas d'estimation des paramètres, puisque les cas où seul le polynôme $P_{varégales}$ donne une solution pour le mélange normale-normale, peuvent être considérés comme inappropriés, ce polynôme donnant, quels que soient les moments du mélange, une solution acceptable. Les premiers moments suffisent, donc dans ces cas là, à exclure certain types de mélanges, sans avoir à recourir à des critères de choix.

Les tableaux récapitulatifs, FIG. II.21, FIG. II.22 et FIG. II.23, résument les 10 études de mélanges que nous avons retenues. Les annexes 71, 72 et 73, sont consacrées à des mélanges exponentielle-exponentielle, et soulignent que ce type de mélange semble ne pas pouvoir être confondu avec un mélange normale-normale. Ce ne semble pas être aussi évident pour les mélanges normale-exponentielle, dont les tests sur les moments centrés d'ordre 5 et 6 sont très proches de zéro. Les trois annexes suivantes, 74, 75 et 76, montrent que, réciproquement, le mélange normale-normale ne peut pas être approché par un mélange exponentielle-exponentielle. Il peut, cependant, y avoir une confusion avec les mélanges normale-exponentielle, dont les tests sont, dans certains cas, proches de zéro. Nous présentons, enfin, dans les annexes de 77 à 80, le cas des mélanges normale-exponentielle, pour lesquels, le type exponentielle-exponentielle ne donne pas de solution. Dans les deux cas des annexes 78 et 80, où P_{12} donne une solution, le type normale-normale donne des tests proches de zéro, qui peuvent entraîner une confusion dans le cas empirique auquel nous nous intéressons dans le paragraphe suivant.

14.2.2 Exemples dans le cas empirique.

Le tableau FIG.II.24, regroupant les annexes de 81 à 85, met en évidence, que dans le cas des mélanges exponentielle-exponentielle, la taille de l'échantillon est primordiale pour une bonne identification du type du mélange. Il semble, au vu des expériences, que la taille typique de l'échantillon permettant d'éviter les confusions, avec le mélange de type normale-exponentielle, soit de 1000000 d'individus. D'après les résultats présentés dans le tableau FIG.II.25, même si les tests des types normale-normale et normale-exponentielle sont proches, la taille de l'échantillon, de 100000 individus, semble suffisante pour identifier le bon type de mélange. Le cas des mélanges normale-exponentielle, présentés dans les annexes de 89 à 95, et dont les résultats sont présentés dans les tableaux FIG.II.26 et FIG.II.27, est un peu plus délicat. En effet, la taille de l'échantillon est très importante, pour les choix du bon type de mélange, et le dernier cas montre que même pour un échantillon de 1000000 d'individus, le choix du type n'est pas assuré.

Le problème des mélanges généralisés, ajoute une dimension supplémentaire à l'ensemble des difficultés que nous avons pu rencontrer précédemment, dans le cadre des mélanges de type identifié. Il semble aussi, et c'est ce que laisse entendre le cas du type normale-normale, qu'une parfaite connaissance de chacun des types particuliers, permette de contourner la majorité des difficultés liées aux mélanges généralisés. C'est, en tous cas sur ce point, qu'il paraît nécessaire de progresser, dans un premier temps.

Estimation des paramètres d'un mélange généralisé Cas théorique du mélange Exponentielle-Exponentielle

	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2	Test T5	Test T6	Commentaires
Annexe 71	1.0	/	0.6	0.3	/	0.0	0.0	
	0.58901744	/	0.6	0.3	/	0.120507	0.245858	seul Pvarégales donne une solution
	0.27734034	0.05822896	0.38209893	0.99373279	/	0.003602	0.006402	les tests sont proches de zéro
Annexe 72	1.0	/	0.6	3.0	/	0.0	0.0	
	1.0	/	0.6	3.0	/	0.118648	0.247520	seul Pvarégales donne une solution
	1.50668887	1.92565543	0.97407264	12.82325119	1.92565543	0.008969	0.017393	les tests sont proches de zéro
Annexe 73	2.0	/	0.6	3.0	/	0.0	0.0	
	2.0	/	0.6	3.0	/	0.259503	0.131354	seul Pvarégales donne une solution
	2.03503477	2.60650464	0.96463754	12.25573018	2.60650464	0.018751	0.041553	
	1.43898267	1.39337701	0.26262660	2.74228075	/			

Légende :			
Paramètres théoriques :		Paramètres Normale-Normale :	
Paramètres Expo-Expo :		Paramètres Normale-Expo :	
		Mélange choisi :	

Fig. II.21. -

Estimation des paramètres d'un mélange généralisé simple.

Estimation des paramètres d'un mélange généralisé Cas théorique du mélange Normale-Normale

	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2	Test T5	Test T6	Commentaires
Annexe 74	1.0	2.0	0.6	3.0	1.0			
	1.0	/				0.0	0.0	pas de solution
Annexe 75	1.0	1.0	0.6	2.0	4.0			
	1.0	/				0.0	0.0	pas de solution
Annexe 76	1.58803706	3.30391324	0.51698669	1.19873710	/	0.027357	0.101880	les tests sont proches de zéro
	3.0	20.0	0.6	1.0	1.0			
	1.0	/				0.0	0.0	pas de solution
	2.97382815	20.06552691	0.59948897	1.04172467	/	0.011506	0.001524	les tests sont proches de zéro

Légende :	
Paramètres théoriques :	Paramètres Normale-Normale
Paramètres Expo-Expo :	Paramètres Normale-Expo :
	Mélange choisi :

FIG. II.22. -

Estimation des paramètres d'un mélange généralisé simple.

Estimation des paramètres d'un mélange généralisé Cas théorique du mélange Normale-Exponentielle

	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2	Test T5	Test T6	Commentaires
Annexe 77	1.0	4.0	0.6	3.0	/			
	1.58036286	4.40398875	0.98147626	13.43742727	4.40398875	0.190041	0.082410	seul Pvarégales donne une solution
	1.0	4.0	0.6	3.0	/	0.0	0.0	
Annexe 78	1.0	4.0	0.6	1.0	/			
	1.135446428	4.28920267	0.56570558	0.82354621	0.80514722	0.080779		il n'y a pas de solution on retient la solution de P12
	1.0	4.0	0.6	1.0	/	0.0	0.0	
Annexe 79	1.0	1.0	0.6	1.0	/			
	0.92182890	0.74685223	0.97643001	4.23838059	0.74685223	0.200747	0.091779	seul Pvarégales donne une solution
	1.0	1.0	0.6	1.0	/	0.0	0.0	
Annexe 80	0.0	1.0	0.6	1.0	/			
	0.33674095	0.96653071	0.95363122	1.70100051	5.08933166	0.004000		il n'y a pas de solution c'est la solution de P12- T5 proche de 0
	0.0	1.0	0.6	1.0	/	0.0	0.0	

Légende :	
Paramètres théoriques :	Paramètres Normale-Normale
Paramètres Expo-Expo :	Mélange choisi
	Paramètres Normale-Expo :

FIG. II.23. -

Estimation des paramètres d'un mélange généralisé simple.

Estimation des paramètres d'un mélange généralisé Cas empirique du mélange Exponentielle-Exponentielle

	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2	Test T5	Test T6	Commentaires
Annexe 81	1.0	/	0.4	0.3	/			
	1.00616939	/	0.38910866	0.30509304	/	0.015402	0.026842	l'échantillon est trop petit pour
	0.48286240	0.21090567	0.97531105	4.33177176	0.21090567	0.122101	0.253074	que l'on reconnaisse le type du mélange.
	0.28897659	0.06742860	0.58296923	0.99877876	/	0.014252	0.028691	
Annexe 82	1.0	/	0.4	0.3	/			
	1.0084773	/	0.39018254	0.30401388	/	0.001544	0.003107	échantillon de 1000000 d'individus
	0.48169344	0.20862924	0.97461386	4.31015063	0.20862924	0.117735	0.247093	
	0.28160786	0.06048700	0.58302999	0.99454964	/	0.007782	0.015534	
Annexe 83	1.0	/	0.4	3.0	/			
	0.99801788	/	0.40293023	2.99550292	/	0.018905	0.039580	l'échantillon est trop petit pour
	1.81157453	2.56100805	0.96659205	13.15861277	2.56100805	0.127883	0.259177	que l'on reconnaisse le type du mélange.
	0.95285147	0.74896525	0.39274032	2.99119684	/	0.015325	0.036278	
Annexe 84	1.0	/	0.4	3.0	/			
	0.98722375	/	0.39334927	2.98732371	/	0.000707	0.002007	échantillon de 1000000 d'individus
	1.80979153	2.52437762	0.96513330	13.01803117	2.52437762	0.120711	0.245375	
	0.89643206	0.59423949	0.38866601	2.96214342	/	0.004121	0.006592	
Annexe 85	3.0	/	0.4	2.0	/			
	2.98032949	/	0.38213172	2.007375	/	0.009479	0.022308	l'échantillon est assez grand pour
	2.01306670	2.53131300	0.96387443	12.1473029	2.53131300	0.128986	0.253619	que l'on reconnaisse le type du mélange.
	1.40784371	1.29864702	0.24580622	2.69574622	/	0.014866	0.029218	

Légende :

Paramètres théoriques : Paramètres Normale-Normale Mélange choisi :

Paramètres Expo-Expo : Paramètres Normale-Expo :

FIG. II.24. -

Estimation des paramètres d'un mélange généralisé simple.

Estimation des paramètres d'un mélange généralisé Cas empirique du mélange Normale-Normale

	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2	Test T5	Test T6	Commentaires
Annexe 86	1.0	2.0	0.6	3.0	1.0			
	0.99203358	1.96799683	0.59534566	3.0025707	0.99724281	/	0.001345	pas de solution
								pas de solution
Annexe 87	1.0	1.0	0.6	2.0	4.0			
	1.00030375	1.02069346	0.60996276	2.00251641	4.08666255	/	0.004711	pas de solution
	1.57950366	3.37521160	0.50766700	1.19706000	/		0.091513	
Annexe 88	3.0	20.0	0.6	1.0	1.0			
	2.95261012	19.83734829	0.61168198	1.01816235	0.69145387	/	0.001311	pas de solution
	2.94830504	20.08697924	0.59645213	1.09753116	/		0.004263	les tests sont proches.






Légende :			
Paramètres théoriques :		Paramètres Normale-Normale	
Paramètres Expo-Expo :		Paramètres Normale-Expo :	
		Mélange choisi :	

FIG. II.25. -

Estimation des paramètres d'un mélange généralisé simple.

Estimation des paramètres d'un mélange généralisé Cas empirique du mélange Normale-Exponentielle (1).

	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2	Test T5	Test T6	Commentaires
Annexe 89	1.0	/	0.6	3.0	/			
	1.57481229	4.40747573	0.98173104	13.43384465	4.40747573	0.074586	0.172749	il n'y a pas de solution
	1.00362061	4.0118766	0.60511885	2.99876241	/	0.008974	0.027084	seul Pvarégales donne une solution
Annexe 90	1.0	4.0	0.6	1.0	/			
	1.11650971	4.08180042	0.63488381	0.80856695	0.54167623	0.190513		les tests sont proches,
	0.97934946	3.91105593	0.61015857	1.04277362	/	0.228123		on choisit le mauvais type de mélange
Annexe 91	1.0	4.0	0.6	1.0	/			
	1.10994187	4.32503767	0.55040746	0.83042874	0.86590947	0.166824		les tests sont proches,
	0.95365402	3.94210364	0.59933462	1.03007896	/	0.111515		on choisit le bon type de mélange
Annexe 92	1.0	4.0	0.6	1.0	/			
	1.14451460	4.33937999	0.54985322	0.82708744	0.86343358	0.062990		échantillon de 1000000 d'individus
	1.00837351	4.01531550	0.59904759	0.99154427	/	0.033230		

Légende :

Paramètres théoriques : Paramètres Normale-Normale Mélange choisi :

Paramètres Expo-Expo : Paramètres Normale-Expo :

FIG. II.26. -

Estimation des paramètres d'un mélange généralisé simple.

Estimation des paramètres d'un mélange généralisé Cas empirique du mélange Normale-Exponentielle (2).

	Moyenne F1	Variance F1	p	Moyenne F2	Variance F2	Test T5	Test T6	Commentaires
	1.0	1.0	0.6	1.0	/			
Annexe 93	0.91290762	0.73360771	0.97416129	4.086380082	0.73360771	0.084801	0.183520	il n'y a pas de solution seul Pvarégales donne une solution
	1.00900914	1.01247108	0.58893008	0.97467774	/	0.005764	0.025836	
	0.0	1.0	0.6	1.0	/			
Annexe 94	0.32759780	0.92605890	0.93941170	1.48530856	4.71884674	0.024195		les tests sont proches, on choisit le mauvais type de mélange
	-0.45710937	0.76668786	0.37216725	0.90448078	/	0.030061		
	0.0	1.0	0.6	1.0	/			
Annexe 95	0.33356301	0.94737999	0.94709383	1.60406416	4.92810563	0.014028		échantillon de 1000000 d'individus les tests sont proches,
	0.06336611	1.01099088	0.65029367	1.02821637	/	0.022567		on choisit le mauvais type de mélange

Légende :	
Paramètres théoriques :	Paramètres Normale-Normale
Paramètres Expo-Expo :	Paramètres Normale-Expo :
	Mélange choisi :

FIG. II.27. -

Estimation des paramètres d'un mélange généralisé simple.

TROISIÈME PARTIE.

**SEGMENTATION STATISTIQUE
D'IMAGES.**

- CHAPITRE 15 -

Classification bayésienne d'images - Modèles de Markov cachés.

Nous nous plaçons dans le cadre général de la segmentation statistique de l'image. L'image que nous souhaitons traiter est la discrétisation, sur le plan de la couleur, d'une vérité naturelle continue. Deux pixels appartenant à un même élément naturel peuvent avoir une couleur différente, et notre travail consiste à associer à chaque pixel la couleur correspondant à l'élément naturel auquel il appartient et que nous appellerons sa classe. Une difficulté supplémentaire vient s'ajouter à celles liées à la variabilité naturelle du terrain. En effet, un ensemble de perturbations, dues, par exemple, en imagerie satellitaire, à la météorologie ou au protocole d'acquisition de l'image contribuent à altérer la qualité de l'observation.

D'un point de vue statistique, nous séparerons l'observation, réalisation d'un vecteur aléatoire Y , continue et dont les composantes sont à valeurs dans \mathbb{R} , de l'image de classes, réalisation d'un vecteur aléatoire discret, X , dont les composantes sont à valeurs dans l'ensemble des classes. En fait, les réalisations des composantes de Y sont les observations des réalisations bruitées des composantes correspondantes, dites cachées, de X .

La modélisation statistique de l'ensemble des perturbations, que l'on appellera le bruit, s'effectue de la manière suivante : les composantes de Y , correspondant à des composantes de X appartenant à une même classe, suivent toutes une même loi de probabilité continue. L'échantillon statistique constitué des réalisations de toutes les composantes de Y suit, par conséquent, une distribution de mélange.

La segmentation statistique classique de l'image se fait en deux étapes. La première consiste à établir quelles sont les distributions présentes dans le mélange et en quelles proportions. La seconde permet d'estimer les réalisations des composantes de X en fonction de celles de Y . La première phase sera appelée estimation des paramètres et la seconde classification.

15.1 Démarche bayésienne.

Soit le modèle statistique $(\mathbb{R}, \mathcal{Y}, P_{\omega \in \Omega})$, où $\Omega = \{\omega_1, \omega_2, \dots, \omega_k\}$ est l'ensemble des classes (supposé fini).

Nous considérerons que les ensembles de composantes des deux variables aléatoires, Y et X , associées à une image sont indexés par l'ensemble, noté S , des pixels de cette image. Nous observons, alors, la réalisation y de la

variable $Y = (Y_s)_{s \in S}$, continue et à valeurs dans $\mathbb{R}^{|S|}$, et souhaitons estimer la valeur ignorée de la variable cachée $X = (X_s)_{s \in S}$, discrète et à valeurs dans $\Omega^{|S|}$ fini.

15.1.1 Modélisation du bruit - lois de Y conditionnelles à X .

Comme nous l'avons signalé plus haut, le bruit peut être modélisé par la somme de deux bruits, le premier lié aux textures des classes présentes dans l'image et le second, résultant des difficultés techniques d'acquisition de l'image. Ce bruit, qui sera supposé stationnaire dans la suite, est modélisé habituellement par des lois gaussiennes.

Le type de modélisation choisie pour le bruit incite à distinguer deux types de corrélations entre les variables de Y conditionnellement à X . En effet, pour deux pixels distincts, s et t , deux cas peuvent se présenter pour les réalisations respectives, x_s et x_t , de X_s et X_t . Si $x_s \neq x_t$, la corrélation entre Y_s et Y_t peut être considérée comme principalement liée à l'acquisition de l'information. Si $x_s = x_t$, il faut tenir compte, en plus, d'une corrélation issue de la variabilité naturelle des classes. Il est raisonnable de choisir, pour deux pixels distincts, s et t , suffisamment éloignés, une corrélation naturelle nulle, c'est le principe des corrélations à portée finie.

Dans ce qui va suivre, la loi conditionnelle, pour un pixel s , donné, de Y_s par rapport à X_s sera une distribution de probabilité, admettant une densité par rapport à la mesure de Lebesgue, dont la forme ainsi que l'ensemble des paramètres sont uniquement déterminés par la classe de la réalisation de X_s .

15.1.2 Aspect bayésien.

Cette démarche d'estimation nécessite le calcul de la loi de X conditionnelle à Y , dite loi a posteriori. Considérons, comme exemple, le cas simple d'un pixel. Si la fréquence d'apparition de ω_i dans la nature est connue, on peut introduire une probabilité a priori Π sur l'ensemble, Ω , des classes :

$$\Pi(\{\omega_i\}) = \Pi_i. \quad [15.58]$$

On définit, alors, une probabilité sur $\Omega \times \mathbb{R}$ en considérant les P_{ω_i} , qui définissent les distributions du bruit associées respectivement aux ω_i . Cette dernière donne alors :

$$P_{\omega_i}(y_s) = P(X_s = \omega_i / Y_s = y_s) = \Pi^{y_s}(\{\omega_i\}) = \Pi_i^{y_s}. \quad [15.59]$$

Nous pouvons ainsi développer l'aspect de la problématique lié à la classification.

15.1.3 La classification bayésienne.

Pour cela, on introduit une stratégie de classification $\widehat{S} : \mathbb{R}^{|S|} \longrightarrow \Omega^{|S|}$, application déterministe et indépendante des réalisations de (X, Y) , permettant d'attribuer un vecteur de classes dans $\Omega^{|S|}$, à toutes les réalisations possibles de Y .

Afin d'établir la fiabilité de la stratégie de classification, et d'optimiser celle-ci, on se donne une fonction de perte $L : \Omega^{|S|} \times \Omega^{|S|} \longrightarrow \mathbb{R}^+$, qui permet de quantifier la gravité de l'erreur commise lors de la classification \widehat{S} . L'espérance mathématique de $L(\widehat{S}(Y), X)$ donne alors une mesure de la fiabilité de la méthode de classification.

La stratégie bayésienne, \widehat{S}_B , est celle qui minimise, pour une fonction de perte donnée, L , cette espérance.

\widehat{S}_B se définit donc par:

$$E \left[L \left(\widehat{S}_B(Y), X \right) \right] = \min_{\widehat{S}} E \left[L \left(\widehat{S}(Y), X \right) \right]. \quad [15.60]$$

Remarquons que pour une image classique de 256×256 pixels que l'on souhaite segmenter suivant k classes, l'ensemble des réalisations possibles de X est $k^{256 \times 256}$, ce qui, pour une segmentation suivant deux classes, donne un ordre de grandeur de 10^{19661} possibilités de configurations différentes du vecteur. Ceci interdit le calcul analytique direct de la segmentation dans le cadre d'un tel modèle.

Afin de réduire les calculs, on pourrait supposer que les variables aléatoires indexées par les pixels sont indépendantes. Cette hypothèse, très forte, semble ne pas être très raisonnable et serait, en tout état de cause, difficilement justifiable. Au contraire, bien modélisée, la relation entre les pixels de l'image permet d'exploiter au mieux l'ensemble des informations contenues dans l'observation. Les modélisations markoviennes permettront de donner une structure pratique à cette dépendance, induisant une simplification des calculs bruts.

15.2 Méthodes de classification bayésienne.

15.2.1 Position du problème

Soit $A \subseteq B \subseteq S$. On désigne par X_A (resp. Y_B) la restriction de X (resp. Y) à A (resp. B). D'une manière générale, la segmentation bayésienne, pour une fonction de perte 0-1, la plus couramment utilisée, consiste à estimer X_A par la configuration \widehat{X}_A pour laquelle la probabilité conditionnelle à $Y_B = y_B$ (probabilité a posteriori) est maximale.

Les différences entre les méthodes de segmentation bayésiennes viennent, donc, du choix que l'on fait pour A et B .

15.2.2 Méthodes locales

C'est le cas où $A = \{s\}$ et $B = V$, (V contenant s et un petit nombre de pixels voisins). Signalons que l'on peut distinguer les méthodes aveugles, où B est un singleton, des méthodes contextuelles, qui prennent en compte une corrélation entre les composantes de Y_B . Il n'est pas nécessaire de supposer X markovien. En effet, si on note $|V|$ le cardinal de V , pour tout pixel $s \in S$ et tout $\omega_i \in \Omega$ on définit les fonction discriminantes :

$$FD(\omega_i, y_V) = \sum_{\omega \in \Omega^{|V|-1}} P[X_s = \omega_i, X_{V \setminus \{s\}} = \omega, Y_V = y_V]. \quad [15.61]$$

La segmentation est alors obtenue en appliquant à chaque pixel

$$\left[\widehat{X}_s = \omega_i \right] \iff \left[FD(\omega_i, y_V) = \max_{\omega_j \in \Omega} FD(\omega_j, y_V) \right]. \quad [15.62]$$

En situations réelles, la distribution de (X_V, Y_V) peut être inconnue; elle devra, alors, être estimée à partir de l'observation $Y = y$. Les méthodes locales ont été très étudiées, en imagerie, avant l'introduction des modélisations markoviennes, (voir [Haralick86], [Swain81]). Elles continuent, d'ailleurs, à susciter l'intérêt, (voir [Marhic91], [Liang92], [Braathen93], [Masson93], [Peng95] ou [Salzenstein96]).

15.2.3 Méthodes globales

C'est le cas où $B = S$. Dans le cas où $A = \{s\}$, la méthode bayésienne, associée à la fonction de perte 0-1, est le *MPM*. Dans le cas où $A = S$, la méthode bayésienne, toujours associée à la même fonction de perte, est le *MAP*. Les solutions de ces méthodes seront approchées dans la suite, en fonction du modèle utilisé, par des algorithmes appropriés que nous détaillerons.

15.3 Les champs de Markov cachés.

Nous considérons deux champs aléatoires $X = (X_s)_{s \in S}$ et $Y = (Y_s)_{s \in S}$ où S est l'ensemble des pixels. Chaque X_s est à valeurs dans l'ensemble des classes $\Omega = \{\omega_1, \dots, \omega_k\}$ et Y_s est à valeurs dans \mathbb{R} .

15.3.1 Définition.

DÉFINITION 15.1: *Le champ X est markovien pour le voisinage V si*

$$\text{pour tout } s \in S, \quad \mathcal{L}(X_s/X_{t, t \neq s}) = \mathcal{L}(X_s/X_{t, t \in V_s}), \quad [15.63]$$

où V_s est un voisinage de s de forme géométrique V indépendante de s .

Par cette définition, nous acceptons de restreindre la dépendance conditionnelle de l'une des variables, par rapport à l'ensemble des variables aléatoire du champs de Markov, à une dépendance conditionnellement à un voisinage de la variable concernée. Les formes de champs de Markov les plus utilisées pour une structure bidimensionnelle comme celle des images, sont, conformément à l'intuition que l'on peut en avoir et en tenant compte de la durée du temps de calcul, celles relatives aux quatre, voir, aux huit plus proches voisins.

15.3.2 Théorème de Hammersley-Clifford.

On peut donner, (voir [Besag74]), une définition analytique des champs de Markov, pour cela, introduisons la notion de clique.

DÉFINITION 15.2: *Soit V un voisinage de forme géométrique fixée. Un sous-ensemble C de S est une clique, pour le voisinage V , si C est un singleton ou si les éléments de C sont mutuellement voisins.*

EXEMPLE 15.1: *Pour un champs de Markov par rapport aux quatre plus proches voisins, il y a trois types de cliques différents. Le singleton est la seule clique d'ordre 1 et les cliques d'ordre 2 sont horizontales ou verticales.*

Si l'on suppose que toutes les configurations du champs X sont susceptibles de se produire avec une probabilité non nulle, on peut donner une expression analytique du champs en fonction de la forme du voisinage:

THÉORÈME 15.1 (DE HAMMERSLEY-CLIFFORD.): *Le champ X est markovien, pour le voisinage V , avec une probabilité non nulle d'existence de toutes ses réalisations si et seulement si :*

$$P_X(x) = \gamma e^{-U(x)}, \quad (\gamma \text{ est la constante de normalisation}), \quad [15.64]$$

avec une fonction énergie U donnée par

$$U(x) = \sum_{t \in T} \sum_{C \in D_t} \varphi_t(x_C), \quad [15.65]$$

où T est l'ensemble de tous les types de cliques associés au voisinage V , D_t est l'ensemble des cliques du champs de la sorte $t \in T$ et φ_t est une application qui associe à chaque $x_C = (x_s)_{s \in C}$, avec $C \in D_t$, un nombre réel.

Grâce au théorème de Hammersley-Clifford, on peut, pour un type de voisinage fixé et si l'on se donne des fonctions φ_t , établir une expression de la loi conditionnelle de X a priori. En fonction des φ_t , indépendantes de la position de la clique C dans S :

$$P[X_s = \omega_i / X_V = x_V] = \frac{e^{-\sum_{t \in T} \sum_{C \in D_t} \varphi_t(x_C^i)}}{\sum_{j=1}^k e^{-\sum_{t \in T} \sum_{C \in D_t} \varphi_t(x_C^j)}}, \quad [15.66]$$

où x_C^i est l'ensemble des réalisations x_C pour lesquelles $x_s = \omega_i$.

La difficulté, pour le choix de la fonction d'énergie, consiste à trouver un bon compromis entre la prise en compte d'un maximum de types de cliques et de la possibilité d'effectuer les calculs sur le modèle. Si la forme des voisinages, et donc celle des cliques, est simple, on connaît la loi de X à une constante γ (que l'on ne peut pas calculer) près. Par contre, il est possible de calculer explicitement les lois conditionnelles de X_s sachant X_{V_s} , et simuler, grâce à l'échantillonneur de Gibbs, (voir [Geman84]), des réalisations du champs X . D'après [Salzenstein96] il est raisonnable de se restreindre aux voisinages d'ordre quatre, et de ne pas considérer, par conséquent, les cliques d'ordre supérieur à deux.

15.4 Utilisation des champs de Markov cachés en imagerie.

15.4.1 Loi de (X, Y) .

Le calcul de la loi du couple (X, Y) est permis par deux concessions concernant la modélisation du bruit. Nous supposons, ainsi, que :

- $(Y_s)_{s \in S}$ est une famille de variables indépendantes conditionnellement à toute réalisation de X .
- $\mathcal{L}(Y_s/X)$ est assimilable à $\mathcal{L}(Y_s/X_s)$, pour tout s appartenant à S .

Sous ces hypothèses, la distribution modélisant le bruit :

$$P_X(Y) = P(Y | X),$$

s'écrit :

$$P_x(y) = \prod_{s \in S} P(Y_s = y_s | X_s = x_s) = \prod_{s \in S} f_{x_s}(y_s),$$

où $f_{x_s}(y_s)$ est la densité modélisant le bruit pour le pixel s .

La loi du couple (X, Y) est alors :

$$P[(X, Y) = (x, y)] = ce^{-U(x) + \sum_{s \in S} \ln(f_{x_s}(y_s))}, \quad [15.67]$$

où U , la fonction d'énergie, a été décrite plus haut et c est une constante de normalisation.

L'expression [15.67] permet de déduire que $P[X = x | Y = y]$ est une distribution de Markov, ce qui autorise la simulation des réalisations de X selon la loi a posteriori, par des algorithmes appropriés.

15.4.2 Mode des marginales a posteriori (MPM)

Le MPM, que nous avons défini plus haut, s'écrit :

$$\forall s, \hat{x}_s^{MPM} = \arg \max_{\hat{x}_s} P(X_s = x_s / Y = y). \quad [15.68]$$

L'avantage de cette méthode est que la fonction de perte pénalise l'erreur commise proportionnellement au nombre de sites erronés. Comme, le calcul direct des $P(X_s = x_s / Y = y)$ est techniquement impossible, l'algorithme de Marroquin et al, (voir [Marroquin87]), permet d'approcher le MPM, par une application directe de l'échantillonneur de Gibbs.

L'échantillonneur de Gibbs.

C'est une procédure itérative et stochastique, développée par Geman et Geman dans [Geman84], permettant de simuler une suite de réalisations de X convergent en loi vers une distribution de Gibbs.

Rappelons que l'on considère deux champs aléatoires $X = (X_s)_{s \in S}$ et $Y = (Y_s)_{s \in S}$ où S est l'ensemble des pixels. Chaque X_s est à valeurs dans l'ensemble des classes $\Omega = \{\omega_1, \dots, \omega_k\}$ et Y_s est à valeurs dans \mathbb{R} .

Après la $n^{ième}$ itération de l'échantillonneur de Gibbs, la configuration du champs simulé sera notée :

$$x(n) = \{x_s(n), s \in S\},$$

et l'on cherche à passer à la configuration $x(n+1)$. La configuration $x(n+1)$ sera initialisée à $x(n)$.

On notera, de plus, $U_t(x)$, la restriction de la fonction énergie $U(x)$ aux termes de la somme contenant le pixel t . L'algorithme, présenté maintenant, permet de simuler un champ X suivant la loi conditionnelle à $Y = y$:

ECHANTILLONNEUR DE GIBBS :

Les étapes de [G₂] à [G₅] sont présentées pour le pixel s_0 à l'itération $n+1$ et sont à répéter, à chaque itération pour tous les pixels de l'image et jusqu'à ce qu'un critère d'arrêt prédéfini soit vérifié.

[G₁] Création d'une configuration initiale $x(0)$.

[G₂] Pour tout $\omega_i \in \Omega$, on calcule, conditionnellement à $Y = y$:

$$p_i = e^{-U_{s_0}(x(n+1)) + \ln(f_{\omega_i}(y_{s_0}))}.$$

[G₃] Tirage de $q \in [0; 1]$ suivant la loi uniforme sur cet ensemble.

[G₄] Détermination de l'unique j tel que $\sum_{i=1}^{j-1} p_i \leq q < \sum_{i=1}^j p_i$.

[G₅] Alors $x_{s_0}(n+1) = \omega_j$ et on passe au pixel suivant.

L'échantillonneur de Gibbs est, avec l'algorithme de Metropolis, (voir [Metropolis53]), que nous n'utiliserons pas ici, le point central de nombreuses procédures d'estimation dans le cadre de problèmes faisant intervenir une modélisation par les champs de Markov

L'algorithme de Maroquin et al.

Grâce aux hypothèses consenties lors de la modélisation, qui permettent l'utilisation de l'échantillonneur de Gibbs, il devient possible de mettre en place un algorithme stochastique de classification. Cet algorithme, développé dans le cadre de la reconstruction d'images, se déroule en trois étapes, la première étant la plus longue.

ALGORITHME DE MAROQUIN ET AL.

Le nombre N de réalisations de X , simulées selon la loi conditionnelle à $Y = y$, lors de l'étape [M₁] est déterminé par compromis entre la nécessité de convergence de l'algorithme de Maroquin et al. et le coût en temps calcul de chaque utilisation de l'échantillonneur de Gibbs.

[M₁] Simulation, par l'échantillonneur de Gibbs, de N réalisations, x^1, \dots, x^N du champ X suivant la loi conditionnelle à $Y = y$.

[M₂] Estimation, pour tout $s \in S$, de la loi de X_s , par les fréquences d'apparition des classes dans l'échantillon : x_s^1, \dots, x_s^N .

[M₃] Estimation, pour tout $s \in S$, de X_s par la classe maximisant la loi obtenue en [M₂].

15.4.3 Maximum a posteriori (*MAP*)

Il s'écrit :

$$\hat{x}^{MAP} = \arg \max_x P(X = x / Y = y). \quad [15.69]$$

Le *MAP* a pour inconvénient majeur de pénaliser aveuglément les erreurs, indépendamment du nombre de sites erronés.

Plusieurs méthodes ont été proposées pour approcher le *MAP*, parmi lesquelles *ICM* (Iterative Conditional Mode) de Besag (voir [Besag86]) et le recuit simulé (voir [Geman84]).

ICM.

C'est une méthode déterministe et itérative de calcul du *MAP*, introduite par Besag dans [Besag86], pour pallier aux problèmes liés aux besoins calculatoires des algorithmes stochastiques de calcul du *MAP*. Chaque itération inclut au moins un balayage complet de l'image, la stratégie de balayage n'étant pas unique. Un balayage, ligne par ligne, en aller-retour, peut constituer, en la matière, un choix raisonnablement efficace. A l'itération $n + 1$, il s'agit de minimiser, en chaque site successivement, la fonction énergie calculée avec la classification obtenue à l'itération n , et en tenant compte de la réactualisation progressive des sites déjà visités.

Le choix de la configuration initiale influence fortement l'estimateur obtenu par convergence de cet algorithme. Il semble nécessaire d'avoir recours à une méthode de "pré-estimation" permettant une initialisation raisonnable de l'*ICM*. Le choix d'un tel algorithme est favorisé par les classifications multi-résolution, dans lesquelles l'information initiale est plus importante.

Algorithme du recuit simulé.

Si l'on introduit dans l'énergie du couple (X, Y) , utilisée en [15.67], le paramètre T appelé température, la nouvelle énergie sera :

$$U_T = -\frac{1}{T} \left[U(x) - \sum_{s \in S} \ln(f_{x_s}(y_s)) \right],$$

et la loi de la distribution markovienne d'énergie U_T ,

$$P_T [(X, Y) = (x, y)] = ce^{-\frac{1}{T} \left[U(x) - \sum_{s \in S} \ln(f_{x_s}(y_s)) \right]},$$

lorsque T tend vers 0, P_T converge vers la distribution uniforme sur l'ensemble D de toutes les réalisations de X maximisant la loi a posteriori. Si la configuration optimale du champ est unique, la distribution limite correspond à une distribution de Dirac vers laquelle on espère faire converger P_T . Cette convergence est assurée, d'après les études théoriques, si la décroissance de T s'effectue suffisamment lentement, c'est à dire typiquement en $\frac{1}{\ln(n)}$, (voir [Hajek88]).

Après la $n^{\text{ième}}$ itération de l'algorithme du recuit simulé, la configuration du champs simulé sera notée:

$$x(n) = \{x_s(n), s \in S\},$$

et l'on cherche à passer à la configuration $x(n+1)$. La configuration $x(n+1)$ sera initialisée à $x(n)$. On notera, de plus, $U_t(x)$, la restriction de la fonction énergie $U(x)$ aux termes de la somme contenant t .

ALGORITHME DU RECUIT SIMULE

Cet algorithme est un échantillonneur de Gibbs où, à l'itération n , la fonction d'énergie est U_{T_n} . Les étapes de [RS₂] à [RS₅] sont présentées pour le pixel s_0 à l'itération $n+1$ et sont à répéter, à chaque itération pour tous les pixels de l'image et jusqu'à ce qu'un critère d'arrêt prédéfini soit vérifié.

[RS₁] Choix d'une suite $(T_n)_n$ décroissante vers 0 et création d'une configuration initiale $x(0)$.

[RS₂] Pour tout $\omega_i \in \Omega$, on pose $x_{s_0}(n+1) = \omega_i$ et on calcule :

$$p_i = e^{-\frac{1}{T_{n+1}} [U_{s_0}(x(n+1)) + \ln(f_{\omega_i}(y_{s_0}))]}.$$

[RS₃] Tirage de $q \in [0; 1]$ suivant la loi uniforme sur cet ensemble.

[RS₄] Détermination de l'unique j tel que $\sum_{i=1}^{j-1} p_i \leq q < \sum_{i=1}^j p_i$.

[RS₅] Alors $x_{s_0}(n+1) = \omega_j$ et on passe au pixel suivant.

La solution du MAP est directement issu de cet algorithme qui s'avère très long à converger, mais qui peut toutefois donner d'excellents résultats.

15.5 Utilisation des chaînes de Markov cachées

15.5.1 Introduction

L'inconvénient majeur du modèle des champs de Markov est que les algorithmes d'approximation qu'ils induisent, sont itératifs et utilisent des techniques de simulation de Monte-Carlo. En effet, l'algorithme de Marroquin et al. est basé sur une utilisation itérative de l'échantillonneur de Gibbs, qui est très gourmand en temps calcul.

Afin de réduire la quantité de calculs et pour obtenir des algorithmes non itératifs, on peut utiliser les chaînes de Markov cachées, en se basant sur les principes généraux développés dans [Pickard80], [Devijver85], [Devijver87] et [Devijver88]. La structure, en ligne, des chaînes n'est pas intuitivement adaptée à la construction bidimensionnelle des images. Il est, cependant, possible, grâce au choix d'un bon parcours de l'image, d'obtenir des résultats appréciables, surtout compte tenu du gain de temps. C'est l'utilisation d'un parcours de Peano qui a été choisi pour parcourir l'image (voir [Benmiloud95], [Giordana96] et [Giordana97]). Le parcours de Peano est une construction fractale récurrente permettant de décrire grâce à une chaîne, un réseau de base carrée. Son principe est développé figure FIG. III.28.

Une comparaison très complète des performances des segmentations sur des modélisations par champs et par chaînes de Markov est développée dans [Salzenstein96] une méthode automatique de choix, basée sur une estimation de la "corrélation" ρ , et de "l'homogénéité" h , de l'image, étant proposée dans [Salzenstein98].

Une chaîne de Markov cachée est un processus à temps discret doublement stochastique, composé des deux processus $X = (X_n)_{n \in \mathbb{N}}$ et $Y = (Y_n)_{n \in \mathbb{N}}$. X est une chaîne de Markov dont les réalisations sont inobservables et Y_n est réel.

DÉFINITION 15.3: Une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ à valeurs dans l'ensemble des classes $\Omega = \{\omega_1, \dots, \omega_k\}$ est une chaîne de Markov si elle vérifie pour tout $n \geq 1$

$$P[X_n = x_n / X_1 = x_1, \dots, X_{n-1} = x_{n-1}] = P[X_n = x_n / X_{n-1} = x_{n-1}]. \quad [15.70]$$

15.5.2 Loi de X

Pour obtenir la loi de X dans une chaîne de Markov nous n'avons besoin que de la loi initiale :

$$\pi_i = P(X_1 = \omega_i) \quad [15.71]$$

et de la matrice de transition, supposée constante, $A = \{a_{ij}\}_{1 \leq i, j \leq k}$ où :

$$a_{ij} = P(X_{n+1} = \omega_j / X_n = \omega_i). \quad [15.72]$$

On obtient en effet :

$$P(X_1 = \omega_{i_1}, X_2 = \omega_{i_2}, \dots, X_n = \omega_{i_n}) = \pi_{i_1} a_{i_1 i_2} \dots a_{i_{n-1} i_n}. \quad [15.73]$$

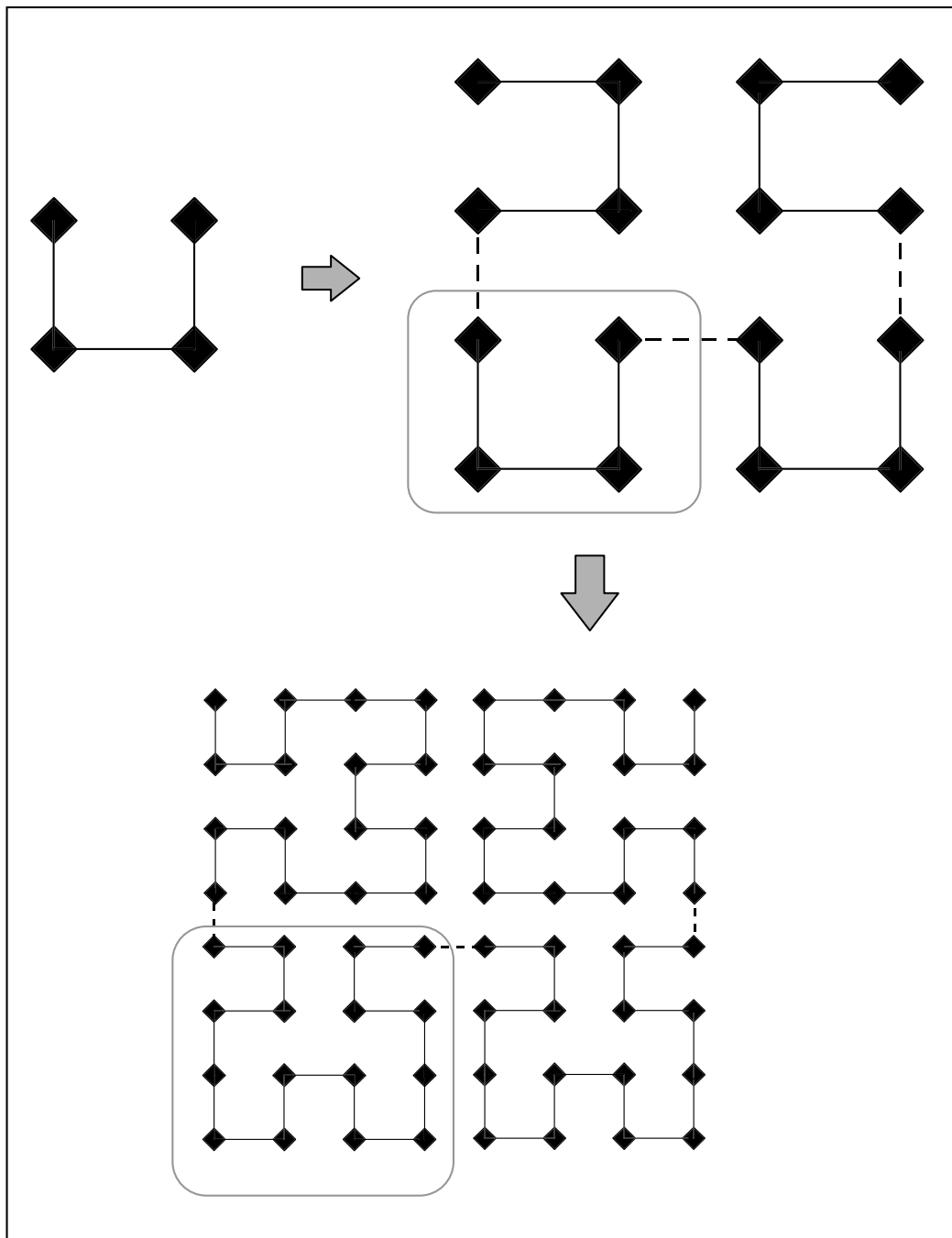


FIG. III.28. - PRINCIPE DU PARCOURS DE PEANO.

15.5.3 Loi de (X,Y).

Comme ce fut le cas pour les champs, nous supposons que les (Y_n) sont indépendants conditionnellement à X et que la loi de chaque Y_n conditionnelle à X est la loi conditionnelle à X_n .

- $(Y_i)_{1 \leq i \leq n}$ est une famille de variables indépendantes conditionnellement à toute réalisation de X .
- $\mathcal{L}(Y_i/X)$ est assimilable à $\mathcal{L}(Y_i/X_i)$, pour tout i inférieur à n . [15.74]

On note f_1, f_2, \dots, f_k les densités de Y_n respectivement conditionnellement à $X_n = \omega_1, \dots, X_n = \omega_k$.
 → les f_i définissent toutes les lois de Y conditionnelles à X .

La densité h , de la loi du couple (X, Y) en $x = (\omega_{i_1}, \dots, \omega_{i_n})$ et $y = (y_1, \dots, y_n)$ s'écrit :

$$h(x, y) = \pi_{i_1} f_{i_1}(y_1) a_{i_1 i_2} f_{i_2}(y_2) \dots a_{i_{n-1} i_n} f_{i_n}(y_n). \quad [15.75]$$

15.5.4 Méthode du MPM.

Considérons la chaîne de Markov cachée $(X_s, Y_s)_{s \in S}$, où S est un ensemble de pixels ordonnés par le choix du parcours de l'image. Quitte à opérer une réindexation, on supposera que la chaîne est indexée par les entiers $s \in \{1, \dots, |S|\}$.

Calcul du MPM

Sous les hypothèses d'indépendance énoncées ci-dessus, on souhaite calculer

$$\xi_s(i) = P[X_s = \omega_i / Y = y], \quad [15.76]$$

grâce à la formule :

$$\xi_s(i) = \frac{\alpha_s(i) \beta_s(i)}{\sum_{j=1}^k \alpha_s(j) \beta_s(j)}, \quad [15.77]$$

où

$$\begin{aligned} \alpha_s(i) &= P[X_s = \omega_i, Y_1 = y_1, \dots, Y_s = y_s] \text{ est la probabilité forward} \\ \text{et } \beta_s(i) &= P[Y_{s+1} = y_{s+1}, \dots, Y_{|S|} = y_{|S|} / X_s = \omega_i] \text{ est la probabilité backward.} \end{aligned}$$

que l'on calcule grâce à l'algorithme de Baum et Welch développé dans [Baum70].

La solution du MPM sera donnée, alors, par :

$$\left[\hat{X}_s = \omega_i \right] \iff \left[\xi_s(i) = \max_{1 \leq j \leq k} \xi_s(j) \right]. \quad [15.78]$$

Probabilités forward et backward - Algorithme de Baum-Welch.

Les probabilités forward et backward ont été introduites par Baum et al. (voir [Baum70]). Leur intérêt réside dans les relations de récurrences qui permettent leur calcul.

La probabilité forward s'obtient à la suite d'une récurrence montante induite par la relation :

$$\alpha_{s+1}(j) = \left(\sum_i \alpha_s(i) a_{ij} \right) f_j(y_{s+1}), \quad 1 \leq j \leq k, \quad 1 \leq s \leq |S|. \quad [15.79]$$

La probabilité backward s'obtient par une récurrence descendante induite par la relation :

$$\beta_s(i) = \sum_j a_{ij} f_j(y_{s+1}) \beta_{s+1}(j), \quad 1 \leq i \leq k, \quad |S| \geq s \geq 1. \quad [15.80]$$

ALGORITHME DE BAUM-WELCH

Cet algorithme est entièrement déterministe.

[BW₁] initialisation de la passe montante :

$$\alpha_1(j) = \pi_j f_j(y_1) \quad \text{pour } 1 \leq j \leq k.$$

[BW₂] passé montante : calcul des $\alpha_{s+1}(j)$ pour $1 \leq j \leq k, 1 \leq s \leq |S|$.

[BW₃] initialisation de la passe descendante :

$$\beta_{|S|}(i) = 1 \quad \text{pour } 1 \leq i \leq k.$$

[BW₄] passé descendante : calcul des $\beta_s(i)$ pour $1 \leq i \leq k, |S| \geq s \geq 1$.

[BW₅] calcul des $\xi_s(i)$, ou plutôt de

$$\xi'_s(i) = \alpha_s(i) \beta_s(i) \quad \text{pour } 1 \leq i \leq k, 1 \leq s \leq |S|.$$

[BW₆] calcul des

$$\hat{x}_s = \arg \max_{1 \leq i \leq k} \xi_s(i) = \arg \max_{1 \leq i \leq k} \xi'_s(i), \quad \text{pour } 1 \leq s \leq |S|.$$

L'estimateur du MPM est alors donné, pour tout s , par $\hat{X}_s = \hat{x}_s$.

MPM utilise donc le calcul direct des probabilités marginales a posteriori, d'où le gain de temps, par rapport aux méthodes pour les champs. Afin d'éviter les problèmes liés à la décroissance, vers 0, très rapide des quantités manipulées, on remplace les probabilités jointes par les probabilités a posteriori (voir [Devijver85]).

15.5.5 Méthode du MAP.

Nous nous plaçons toujours, dans le cas d'une chaîne de Markov cachée, indexée par un ensemble S de pixels.

Calcul du MAP.

L'estimateur par le MAP est :

$$\left[\hat{X} = \hat{x} \right] \iff \left(P[X = \hat{x} / Y = y] = \max_{x \in \Omega^{|S|}} P[X = x / Y = y] \right). \quad [15.81]$$

Soit, la suite $\delta_s(i)$ définie par

$$\delta_s(i) = \max_{x_1, \dots, x_{s-1}} P((X_1, Y_1) = (x_1, y_1), \dots, (X_{s-1}, Y_{s-1}) = (x_{s-1}, y_{s-1}), X_s = \omega_i),$$

qui est, en fait, un *MAP* pour s et X_s fixés. Cette suite peut être définie par une récurrence induite par la relation :

$$\delta_{s+1}(i) = \max_{j \in \{1, \dots, k\}} a_{ji} f_j(y_s) \delta_s(j),$$

et le *MAP* est donné finalement par

$$\max_{i \in \{1, \dots, k\}} \delta_{|S|}(i).$$

Nous notons, enfin,

$$\hat{X}_{s+1}(j) = \operatorname{argmax}_{i \in \{1, \dots, k\}} a_{ji} f_i(y_s) \delta_s(i).$$

Algorithme de Viterbi.

Cet algorithme (voir [Forney73] et [Viterbi67]) est très utilisé en reconnaissance de la parole ou en reconnaissance de caractères. L'algorithme de Viterbi est confronté à une décroissance exponentielle, (voir [Levinson83]), vers 0, des probabilités manipulées. Les problèmes découlants de cette décroissance ont été, en partie, résolus par Devijver et al dans [Devijver85], en utilisant, comme précédemment, les probabilités a posteriori, plutôt que les probabilités jointes, qui sont plus petites.

L'algorithme de Viterbi comporte une phase montante et une phase descendante.

ALGORITHME DE VITERBI

L'algorithme de Viterbi est non itératif, donc rapide, mais peut s'avérer coûteux en espace mémoire si l'image et/ou si k sont grands.

[V₁] Initialisation de la passe montante : $\delta_1(i) = \pi_i$, $1 \leq i \leq k$.

[V₂] Passe montante : pour $2 \leq s \leq |S|$ et $1 \leq i \leq k$ on calcule $\delta_s(i)$.

[V₃] Initialisation de la passe descendante : $\hat{x}_{|S|} = \operatorname{argmax}_{i \in \{1, \dots, k\}} \delta_{|S|}(i)$.

[V₄] Passe descendante : pour $|S| - 1 \geq s \geq 1$ on tire des calculs de V_2 ,

$$\hat{x}_s = \hat{X}_{s+1}(\hat{x}_{s+1}).$$

L'estimateur du *MAP* est alors :

$$\hat{X} = (\hat{x}_1, \dots, \hat{x}_{|S|}).$$

- CHAPITRE 16 -

Modélisations sur arbres.

La modélisation par arbres de Markov cachés a été récemment introduite dans le traitement statistique des images dans [Bouman94] et [Laferté96]. Alors que les calculs sont à peine plus compliqués, la structure du modèle semble, intuitivement, plus proche de celle d'une image, que celle des chaînes de Markov, même si cette question reste ouverte. Compte tenu de l'expérience que nous en avons, ce modèle semble se poser en véritable concurrent des champs de Markov.

16.1 Graphes et probabilités.

Les arbres de Markov cachés sont issus de la théorie des graphes. Commençons par donner la définition de la notion de graphe empruntée à [Whittaker90].

DÉFINITION 16.1: Un **graphe** $G = (S, L)$ est formé d'un ensemble S de sommets, encore appelés noeuds et d'un ensemble $L \subset S^2$, d'arêtes (ou arcs) reliant des sommets entre eux.

16.1.1 Graphes de dépendance

L'intérêt de l'utilisation des graphes en statistique réside dans la possibilité de représentation graphique des liens de dépendance conditionnelle entre les variables aléatoires composant un processus stochastique. Précisons cette représentation.

DÉFINITION 16.2 (CHEMIN ET SÉPARABILITÉ): Une suite de sommets $(s_i)_i$ telle que $\forall i, (s_i, s_{i+1})$ appartienne à L , est appelé **chemin**.

Un ensemble $C \subset S$ **sépare** deux ensembles $A, B \subset S$ s'il n'existe pas de chemin pour aller de A à B sans passer par C .

La notion de markoviannité sur graphe donne son sens à l'appellation de graphe de dépendance.

DÉFINITION 16.3 (MARKOVIANITÉ SUR GRAPHE.): Soit X un processus stochastique indicé par les sommets d'un graphe $G = (S, L)$.

X est dit **markovien** sur G si $\forall A, B, C \subset S$ tels que C sépare A et B , on a l'indépendance conditionnelle de X_A et X_B sachant X_C .

On dit alors que G est un **graphe de dépendance** pour X .

L'absence d'arête entre deux composantes du processus signifie une indépendance, entre ces deux variables, conditionnellement au reste des variables du processus, alors que la présence d'une arête signale la possibilité d'une non indépendance conditionnelle. En ce sens, l'absence d'arête apporte plus d'information sur le processus, que la présence d'une arête, et c'est l'indépendance conditionnelle, signalée par ces absences d'arêtes, qui permet de développer les algorithmes de calculs sur les arbres.

Le théorème de conditionnement permet de conserver la propriété de markoviannité par suppression d'une partie du graphe.

THÉORÈME 16.1 (DE CONDITIONNEMENT): Soit X un processus markovien sur $G = (S, L)$, soit $D \subset S$.

Le processus $X_{S \setminus D}$ muni de la loi conditionnelle sachant X_D est markovien sur le graphe réduit $G' = (S \setminus D, L')$ où $L' = \{(s, t) \in L \cap (S \setminus D)^2\}$.

16.1.2 Règle de retrait d'un sommet

Soit un processus X markovien sur $G = (S, L)$.

Soit alors $X' = X_{S \setminus \{s\}}$. On cherche le graphe de dépendance de X' .

Pour cela on introduit ∂_s^G , qui est l'ensemble des sommets t du graphe G reliés à s par un arc. On précise que $s \notin \partial_s^G$.

DÉFINITION 16.4 (GRAPHE COMPLET.): Un graphe est **complet** si le nombre de ses arêtes est maximal c'est à dire que tout couple de sommets forme une arête.

L'importante proposition qui suit nous permettra de comparer la modélisation par arbres avec les modélisations classiques.

PROPOSITION 16.1 (RETRAIT D'UN SOMMET.): Un graphe de dépendance pour $X_{S \setminus \{s\}}$ est donné par $G' = (S \setminus \{s\}, L')$ tel que G' contienne la restriction de G à $S \setminus \{s\}$ et que la restriction de G' à ∂_s^G soit un graphe complet.

En pratique, cela signifie que l'on joint tous les points voisins du point supprimé (le point blanc) comme on le montre FIG. III.29.

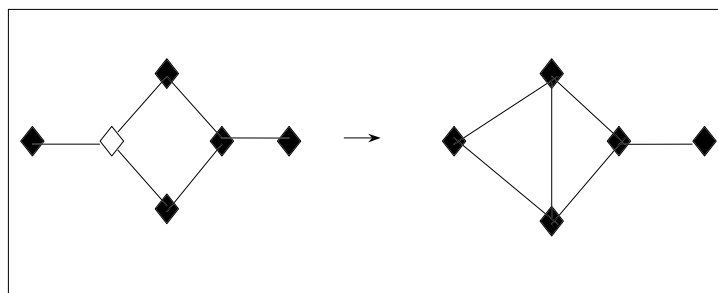


FIG. III.29. - EXEMPLE DE RETRAIT D'UN SOMMET.

16.2 Exemples de graphes de dépendances.

Les exemples qui suivent permettent d'évaluer l'aspect pratique de la représentation de certains processus par les graphes de dépendances.

EXEMPLE 16.1 (CHAÎNE DE MARKOV CACHÉE.): *Ce graphe établit visuellement les hypothèses faites en [15.74].*

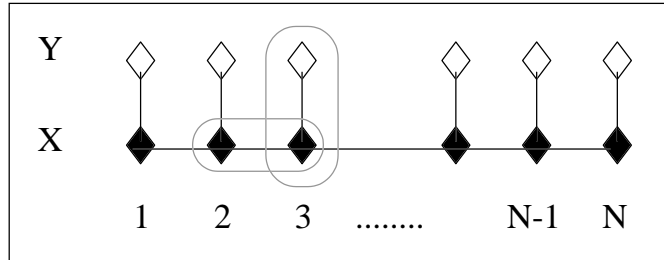


FIG. III.30. - GRAPHE DE DÉPENDANCE D'UNE CHAÎNE DE MARKOV CACHÉE.

EXEMPLE 16.2 (CHAMPS DE MARKOV DE TYPE 1): *Le graphe, FIG. III.31, met en évidence la restriction de la dépendance d'une des variables, conditionnellement au reste des variables du champs, à une dépendance par rapport à ses quatre plus proches voisins.*

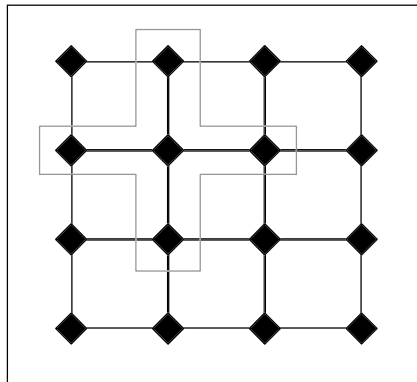


FIG. III.31. - CHAMPS DE MARKOV RELATIF AUX 4 PLUS PROCHES VOISINS.

EXEMPLE 16.3 (CHAMPS DE MARKOV DE TYPE 2): *Le graphe, FIG. III.32, met en évidence la restriction de la dépendance d'une des variables, conditionnellement au reste des variables du champs, à une dépendance par rapport à ses huit plus proches voisins.*

L'utilisation des graphes de dépendance permet une vision simplifiée des différentes modélisations. Elle met aussi en évidence la diminution, proportionnellement à l'éloignement, de l'interdépendance entre les variables, dans les champs de Markov, ce qui semble intuitivement raisonnable pour la modélisation des images. La nouvelle modélisation par arbre, compte tenu de sa structure et de sa complexité, ne permet pas une interprétation symétrique des relations de dépendance entre les variables, ce qui peut être considéré comme une faiblesse du modèle. Nous apporterons quelques précisions sur ce point, dans le chapitre suivant.

Modélisations sur arbres.

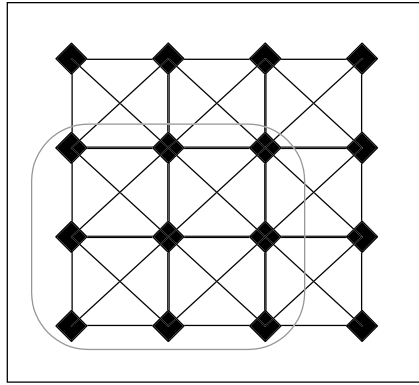


FIG. III.32. - CHAMPS DE MARKOV RELATIF AUX 8 PLUS PROCHES VOISINS.

16.3 La structure d'arbre.

Les arbres forment une sous famille particulière de la famille des graphes.

16.3.1 Définitions.

DÉFINITION 16.5: Un **graphe hiérarchique** est un graphe connexe⁹ partitionné en échelles, numérotées de 0 pour la plus grossière, à N pour la plus fine.

Un **arbre** est, alors, défini comme étant un graphe hiérarchique ne contenant pas de cycle¹⁰.

DÉFINITION 16.6: Un **arbre homogène** est un arbre dont tous les sommets, à l'exception des feuilles¹¹, sont reliés à un nombre constant de sommets.

Les quelques exemples suivants donnent une idée de la variété des graphes appartenant à la famille des graphes hiérarchiques ainsi que des spécificités des arbres.

16.3.2 Exemples.

Le graphe de FIG. III.33, appelé graphe de Kato (voir [Kato93]), est un graphe hiérarchique, mais pas un arbre car il contient plusieurs cycles. Celui de FIG. III.34 ne contient plus de cycles; c'est un arbre quelconque. L'arbre homogène est présenté sous la forme du quadarbre, FIG. III.35.

16.3.3 Interprétation de l'arbre

L'ensemble S des noeuds de l'arbre est partitionné en échelles S^n , de la plus fine 0, à la plus grossière N . On note s^- l'unique père de s et s^+ l'ensemble de ses fils. On note $s \geq$ l'ensemble des ancêtres de s et $\geq s$ l'ensemble de ses descendants.

⁹Un graphe est dit connexe, si pour tout couple de sites (s, t) il existe un chemin permettant d'aller de s à t .

¹⁰Un cycle est une suite (s_1, \dots, s_n) de sites telle que, pour tout i , (s_i, s_{i+1}) soit un arc du graphe et telle que $s_1 = s_n$.

¹¹Les feuilles sont les sommets de l'échelle 0

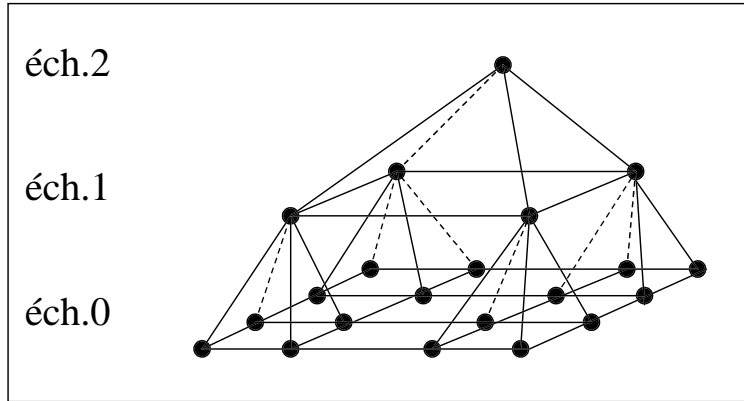


FIG. III.33. - GRAPHE DE KATO.

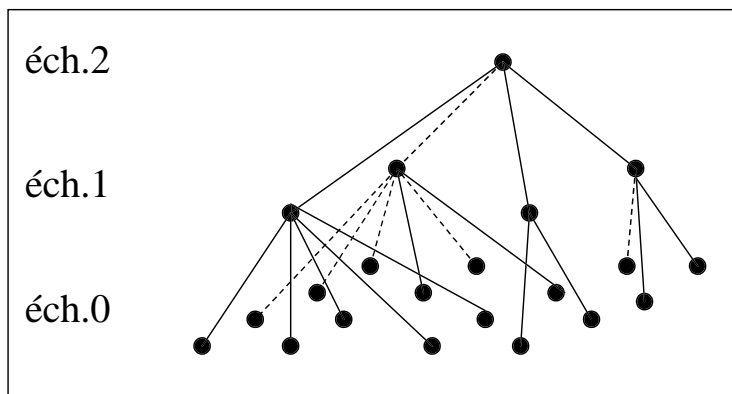


FIG. III.34. - ARBRE QUELCONQUE.

Modélisations sur arbres.

Dans un arbre, le père sépare à lui seul tous ses fils entre eux d'une part et ces derniers de leurs ancêtres d'autre part. Si l'arbre est markovien, il y a, donc, indépendance des fils conditionnellement à leur père. Cette propriété est déterminante pour les algorithmes sur les structures de type arbre.

16.3.4 Passage d'un champs aléatoire à un arbre homogène

Le lien entre la structure tridimensionnelle des arbres et celle de l'image n'est pas apparent et il est nécessaire de présenter clairement l'adaptation du squelette d'arbre à l'étude d'un problème essentiellement bidimensionnel comme celui du traitement d'image.

Principe de la modélisation

Il s'agit de construire un arbre dont les feuilles indicent les variables du champs. On "regroupe" quatre à quatre ou deux à deux les variable du champs et on construit un père à chaque groupe de feuilles. Cette construction est alors appliquée au nouveau champs formé par les pères. On répète ainsi de suite l'opération tant qu'elle est possible jusqu'à l'ancêtre unique, que l'on appelle la racine r . Pour les images 256×256 on construit donc un quadarbre à 8 niveaux d'échelles ou un diarbre à 16 niveaux.

Des arbres particuliers.

Le quadarbre.

Le quadarbre est un arbre homogène pour lequel chaque père a quatre descendants. Sa structure d'ordre quatre est bien adaptée aux images qui font couramment $2^n \times 2^n$ pixels.

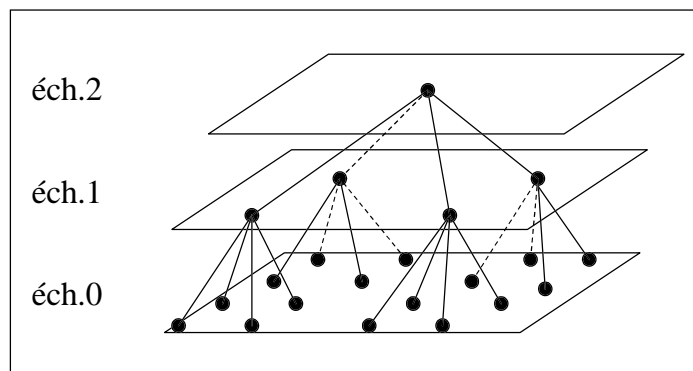


FIG. III.35. - QUADARBRE À TROIS NIVEAUX.

Le diarbre.

Le diarbre, qui n'a pas été étudié jusque là, présente la particularité de générer, une échelle sur deux, un feuillage à symétrie non carrée. Cependant, la construction d'un diarbre, à partir d'une image, se fait suivant le même schéma que pour le quadarbre. Une latitude supplémentaire s'offre, lors de la construction du diarbre, au niveau du choix du regroupement des fils. La symétrie à deux frères de ce modèle permet d'envisager beaucoup plus de constructions différentes du "branchage", que la symétrie à quatre frères qui induit naturellement un regroupement en carré. Ne serait-ce qu'en gardant cette base carrée, il y a trois façons de regrouper deux à deux quatre petits-fils.

On peut, en effet, considérer systématiquement le lien de fraternité comme un lien basé sur le voisinage droit. Les fils sont engendrés alternativement dans le sens latéral et dans la profondeur FIG. III.36.

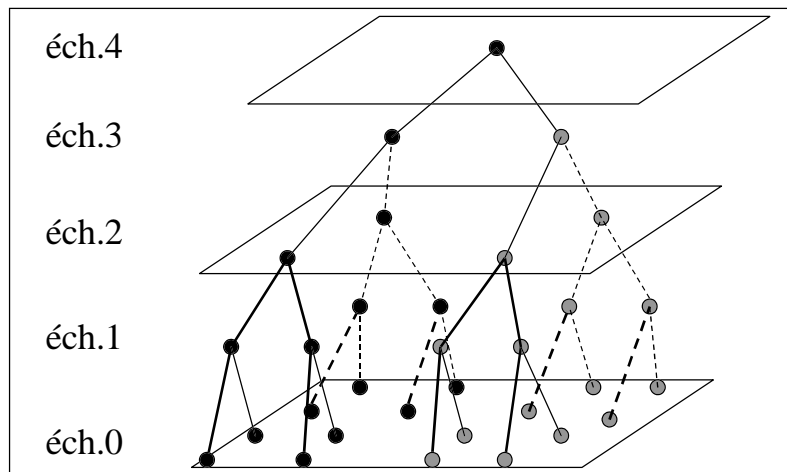


FIG. III.36. - TRANSMISSION DROITE.

Dans le second modèle FIG. III.37 la transmission se fait une fois sur deux en diagonale et dans le troisième cas, le lien de fraternité est uniquement diagonal FIG. III.38, (voir [Monfrini99]).

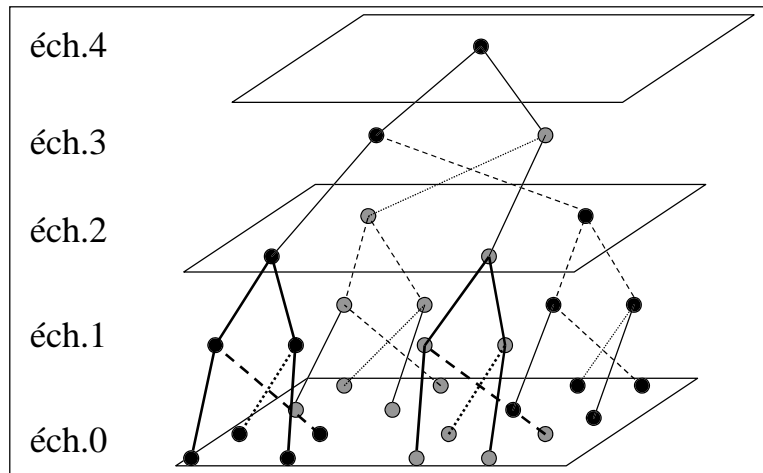


FIG. III.37. - TRANSMISSION ALTERNÉE DROITE-DIAGONALE.

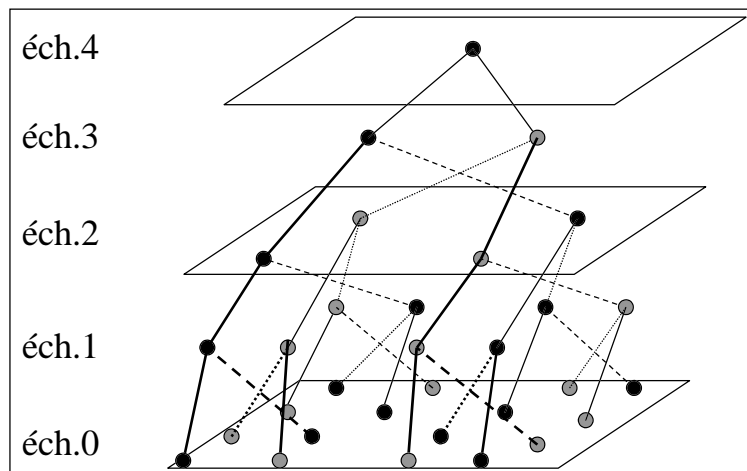


FIG. III.38. - TRANSMISSION DIAGONALE.

Comparaison de cette modélisation avec les champs de Markov.

En appliquant la règle de retrait des sommets de la PROPOSITION 16.1, à tous les noeuds de notre arbre n'appartenant pas à la résolution la plus fine, on montre que chaque variable de notre feuillage est reliée par une arête à chacune des autres variables, et que notre feuillage a donc un graphe complet.

Si l'on effectue un retrait des sommets sur un quadarbre à trois niveaux, on se familiarise avec le mécanisme qui permet d'obtenir la représentation de l'arbre ramenée à un seul niveau d'échelle. En effet si l'on retire la racine on se place dans le cas du graphe présenté FIG. III.39. Si l'on retire maintenant l'un des sommets de l'échelle 1 on relie entre eux tous ses fils et chacun de ceux-ci aux trois frères de leur père à l'échelle 1. On répète ensuite cette opération pour tous les noeuds de l'échelle 1 et on obtient un graphe, que l'on peut représenter schématiquement FIG. III.40, où tous les noeuds de l'échelle 0 sont reliés entre eux. Ce schéma tente de mettre en évidence le mécanisme

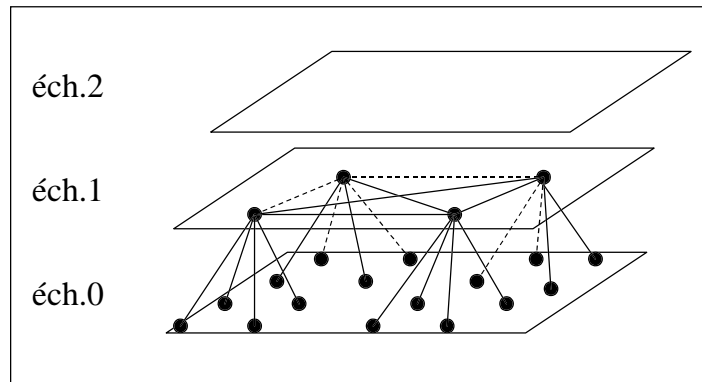


FIG. III.39. - RETRAIT DE LA RACINE.

répétitif que l'on peut appliquer à un arbre plus important. D'après ce que nous avons vu plus haut, la graphe que

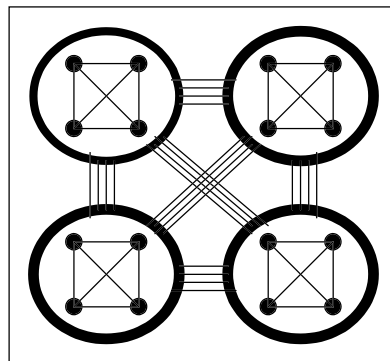


FIG. III.40. - DIAGRAMME RÉCAPITULATIF.

nous obtenons par retrait de tous les sommets de l'arbre qui ne sont pas à l'échelle 0, n'est pas markovien à l'échelle 0. La modélisation par arbre est bien différente de celle des champs et des chaînes de Markov et il n'y a d'autre moyen objectif de les comparer, que de s'intéresser aux résultats des algorithmes de segmentation qu'ils induisent.

- CHAPITRE 17 -

Application à l'imagerie.

17.1 Simulations de champs par les arbres.

L'aspect en "peau de vache" des images simulées par l'échantillonneur de Gibbs est bien connue. FIG. III.41 en rappelle l'aspect.



FIG. III.41. - IMAGE SIMULÉE PAR L'ÉCHANTILLONNEUR DE GIBBS.

Afin d'apprécier, de manière purement visuelle quelles sont les images les plus favorables à la segmentation par arbres de Markov cachés, nous proposons ici d'examiner l'aspect des images simulées par différents modèles d'arbres.

Application à l'imagerie.

17.1.1 Simulations par le quadarbre.

Le quadarbre à transitions constantes.

Le modèle le plus simple de quadarbre est celui pour lequel les probabilités de transition père-fils associées au couple de réalisations (ω_i, ω_j) sont constantes, quelle que soit l'échelle, [Laferté00]. Si l'on se place dans le cas d'un modèle à deux classes, pour lequel les probabilités de conservation de la classe sont les mêmes pour les deux classes, on obtient des images du type de celles de FIG. III.42.

Le problème, de ce type de modèle, est que l'on ne parvient pas, contrairement à ce qui se passe avec les champs de Gibbs, à obtenir des zones d'uniformité des classes. Pour remédier à ce problème nous proposons de compléxifier le modèle en faisant dépendre les probabilités de transition de l'échelle.

Le quadarbre évolutif.

L'idée est d'augmenter la probabilité de conservation de la classe au fur et à mesure de l'affinement de la résolution. Afin de ne pas multiplier les paramètres, les probabilités de transitions interéchelles dépendent d'un paramètre α constant sur l'arbre, d'un paramètre fixé ε , et de la résolution. Parmi les différentes formes de fonctions usuelles que nous avons prises en compte, nous en avons retenu deux, lesquelles répondent mieux que les autres à nos attentes, particulièrement en ce qui concerne la concentration des zones de classes uniformes. Nous avons, donc choisi de nous intéresser à des probabilités de transition père-fils de la forme (les probabilités de transition sont données à l'échelle $n > 0$) :

$$\begin{aligned} g_{i,i}^n(\alpha) &= P(X_s^n = \omega_i / X_s^{n+1} = \omega_i) = 1 - \varepsilon - \alpha \frac{\ln(N+1) - \ln(n)}{\ln(N+1)} \\ \text{et } g_{i,j}^n(\alpha) &= P(X_s^n = \omega_i / X_s^{n+1} = \omega_j) = \frac{\varepsilon + \alpha \frac{\ln(N+1) - \ln(n)}{\ln(N+1)}}{\text{Nbre de classes} - 1}, \text{ pour } i \neq j, \end{aligned} \quad [17.82]$$

ainsi que :

$$\begin{aligned} h_{i,i}^n(\alpha) &= P(X_s^n = \omega_i / X_s^{n+1} = \omega_i) = 1 - \varepsilon - \alpha \left(\frac{N+1-n}{N+1} \right)^\beta \\ \text{et } h_{i,j}^n(\alpha) &= P(X_s^n = \omega_i / X_s^{n+1} = \omega_j) = \frac{\varepsilon + \alpha \left(\frac{N+1-n}{N+1} \right)^\beta}{\text{Nbre de classes} - 1}, \text{ pour } i \neq j, \end{aligned} \quad [17.83]$$

Le paramètre β , de $h_{i,j}^n$, sera fixé à 0.5 dans tout ce qui suit, mais il pourrait être intéressant de travailler en faisant, également, varier le second paramètre, ce qui assouplirait, encore, le modèle. Le paramètre fixé, ε , permet de ne pas avoir une dernière transition égale à un, ce qui serait équivalent à la perte de l'information au niveau de la résolution la plus fine. Nous fixerons, pour la suite, $\varepsilon = 0.0001$.

Les simulations qui suivent, sont obtenues avec des probabilités de transition de forme $g_{i,j}^n$, celles obtenues avec $h_{i,j}^n$ étant équivalentes. Les simulations par ce modèle sont plus nettes que celles obtenues dans le cas simple du quadarbre constant FIG. III.42, cependant, les importantes formes rectangulaires, inhérentes au modèle et à l'origine des problèmes d'effets de blocs constatés lors des segmentations par les arbres de Markov, (voir par exemple [Laferté96]), sont encore très marquées. C'est pourquoi nous proposons les diarbres comme alternative aux quadarbres.

17.1.2 Simulations par diarbre.

Pour les diarbres nous nous placerons, encore, dans le cas où les probabilités de transition dépendent de l'échelle de manière logarithmique, comme pour le quadarbre .

Application à l'imagerie.

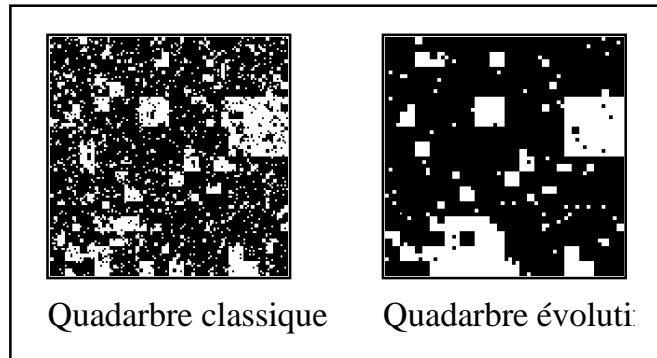


FIG. III.42. - SIMULATIONS D'IMAGES PAR LES QUADARBRES.

Simulations par le diarbre droit.

Ce type de simulation est fortement marqué par le sens de la dernière dilatation. Il permet cependant de rompre en partie les formes obtenues avec le quadarbre (image 1. FIG. III.43).

Simulations par le diarbre alterné.

L'image obtenue avec cet arbre contient des détails plus fins que la précédente, et va donc dans le sens de l'évolution que nous souhaitons apporter (image 2. FIG. III.43).

Simulations par le diarbre diagonal.

Le diarbre diagonal permet des simulations très surprenantes. En effet, de par la structure même de ce diarbre, l'image simulée fait apparaître des zones d'imbrications de croix et d'ellipses, qui s'apparentent à des classes supplémentaires. La simulation à deux classes présente trois types de zones et la simulation à trois classes en contient six. Pour n classes, il peut y avoir jusqu'à $(n + C_n^2)$ types de zones différentes (images 3. et 4. FIG. III.43).

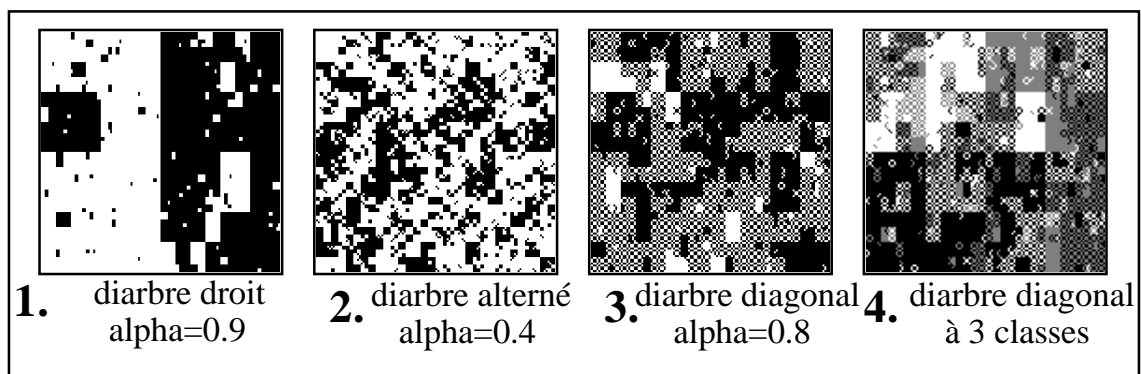


FIG. III.43. - SIMULATIONS D'IMAGES PAR LES DIARBRES.

Application à l'imagerie.

17.2 Modèle markovien caché sur arbre.

Puisque nous sommes désormais familiarisés avec le modèle, nous pouvons définir le modèle des arbres de Markov cachés. Nous noterons, comme précédemment :

$$\begin{aligned} X &= (X_s)_{s \in S} = ((X_s)_{s \in S^n})_{0 \leq n \leq N} = (X^n)_{0 \leq n \leq N} \\ \text{et } Y &= (Y_s)_{s \in S} = ((Y_s)_{s \in S^n})_{0 \leq n \leq N} = (Y^n)_{0 \leq n \leq N}, \end{aligned}$$

où S est l'ensemble des noeuds de l'arbre, S^n est l'échelle n . Nous présentons le modèle général des arbres de Markov cachés, en considérant l'existence d'une observation à chaque niveau de l'échelle.

Nous noterons x et y , les réalisations respectives de X et Y , x^n et y^n étant les réalisations respectives de X^n et Y^n . Par ailleurs, s^- est l'unique père de s et s^+ l'ensemble de ses fils, $s \geq$ est toujours l'ensemble de ses ancêtres et $\geq s$ l'ensemble de ses descendants. Enfin la racine sera notée r .

Pour assurer que (X, Y) est markovien sur l'arbre, on fait les quatre hypothèses suivantes :

►1. X est markovien en échelle:

$$P[X^n = x^n / X^k = x^k, N \geq k \geq n + 1] = P[X^n = x^n / X^{n+1} = x^{n+1}]. \quad [17.84]$$

►2. Pour compléter la première propriété, nous supposons que :

$$P[Y = y / X = x] = \prod_{n=0}^N P[Y^n = y^n / X^n = x^n]. \quad [17.85]$$

La structure du processus, en échelle, se rapproche, formellement, de celle d'une chaîne de Markov cachée.

►3. A chaque niveau d'échelle, n , $0 \leq n \leq N - 1$, les probabilités de transition inter-échelle peuvent se factoriser en

$$P[X^n = x^n / X^{n+1} = x^{n+1}] = \prod_{s \in S^n} g_n(x_s, x_{s^-}), \quad [17.86]$$

où g_n est une fonction définissant la probabilité de transition à l'échelle n , x_s et x_{s^-} étant respectivement les réalisations de X_s et X_{s^-} . Nous utiliserons, dans la suite, celles définies par [17.83] et [17.82].

►4. Enfin, à chaque niveau d'échelle, n , $0 \leq n \leq N$, :

$$P[Y^n = y^n / X^n = x^n] = \prod_{s \in S^n} f^n(y_s / x_s), \quad [17.87]$$

où f^n est la densité du bruit à l'échelle n , x_s et y_s étant respectivement les réalisations de X_s et Y_s .

De manière analogue à celle des chaînes de Markov cachées classiques, ce modèle se prête alors au développement des méthodes bayésiennes de restauration classiques *MPM* et *MAP*.

Application à l'imagerie.

17.3 Algorithmes de classification sur les arbres de Markov cachés.

Nous présentons dans cette partie des algorithmes désormais classiques en traitement d'image, ainsi que des algorithmes originaux que nous introduisons pour remédier aux travers, déjà évoqués, des méthodes existantes.

17.3.1 Algorithme déterministe du MAP sur les arbres de Markov cachés.

Cet algorithme est une généralisation de l'algorithme de Viterbi, détaillé dans la partie consacrée aux chaînes de Markov. C'est Dawid, [Dawid92], et Laferté, [Laferté96], pour ce qui est de l'image, qui ont établi ce parallèle. La méthode s'appuie, elle aussi, sur une relation de récurrence qui s'établit comme suit :

L'estimateur recherché étant :

$$\hat{x} = \arg \max_x P(X = x | Y = y),$$

il faut regarder :

$$\begin{aligned} \max_x P(X = x, Y = y) &= \max_{x_r} \left[P(X_r = x_r, Y_r = y_r) \max_{x_{>r}} P(X_{>r} = x_{>r}, Y_{>r} = y_{>r} | X_r = x_r) \right], \\ &= \max_{x_r} \left[P(X_r = x_r, Y_r = y_r) \max_{x_{>r}} \prod_{t \in r^+} P(X_{\geq t} = x_{\geq t}, Y_{\geq t} = y_{\geq t} | X_r = x_r) \right], \end{aligned}$$

et si l'on pose

$$\hat{x}_r = \arg \max_{x_r} \left[P(X_r = x_r, Y_r = y_r) \max_{x_{>r}} \prod_{t \in r^+} P(X_{\geq t} = x_{\geq t}, Y_{\geq t} = y_{\geq t} | X_r = x_r) \right],$$

on a :

$$\max_x P(X = x, Y = y) = P(X_r = \hat{x}_r, Y_r = y_r) \prod_{t \in r^+} \max_{x_{\geq t}} P(X_{\geq t} = x_{\geq t}, Y_{\geq t} = y_{\geq t} | X_r = \hat{x}_r).$$

En réitérant les calculs, on a, $\forall s \in S, s \notin S^0 \cup S^N$,

$$\underbrace{\max_{x_{\geq s}} P(X_{\geq s} = x_{\geq s}, Y_{\geq s} = y_{\geq s} | X_{s^-} = x_{s^-})}_{\Delta_s(x_{s^-})} = \max_{x_s} \left[\prod_{t \in s^+} \underbrace{\max_{x_{\geq t}} P(X_{\geq t} = x_{\geq t}, Y_{\geq t} = y_{\geq t} | X_s = x_s)}_{\Delta_t(x_s)} \right],$$

ce qui permet d'effectuer une récurrence sur les $\Delta_t(x_s)$.

Posons, par analogie avec les chaînes de Markov,

$$\hat{X}_s(x_{s^-}) = \arg \max_{x_s} P(X_s = x_s, Y_s = y_s | X_{s^-} = x_{s^-}) \prod_{t \in s^+} \Delta_t(x_s).$$

L'algorithme de Viterbi-Laferté que nous décrivons maintenant, comporte, lui aussi, deux phases principales, l'une montante l'autre descendante.

Application à l'imagerie.

ALGORITHME DE VITERBI-LAFERTE

Cet algorithme est entièrement déterministe.

[VL₁] Initialisation de la passe montante :

$$\Delta_s(x_{s-}) = \max_{x_s} f^0(y_s | x_s) g_0(x_s | x_{s-}), s \in S^0.$$

[VL₂] Passé montante : pour $1 \leq n \leq N$, $\forall s \in S^n$, on calcule, par récurrence $\Delta_s(x_{s-})$.

[VL₃] Initialisation de la passe descendante :

$$\hat{x}_r = \operatorname{argmax}_{x_r} f^N(y_r | x_r) \prod_{s \in r^+} \Delta_s(x_r).$$

[VL₄] Passé descendante : pour $N-1 \geq n \geq 0$ on tire des calculs de [VL₂],

$$\forall s \in S^n, \hat{x}_s = \hat{X}_s(\hat{x}_{s-}).$$

L'estimateur du MAP est alors :

$$\hat{X} = (\hat{x}_s)_{s \in S} ..$$

Cet algorithme donne la solution exacte du *MAP*. Il est très rapide mais très cher en espace mémoire. De plus la façon dont les erreurs sont pénalisées n'est pas très adaptée à la structure des graphes hiérarchiques. Une erreur survenant à une échelle grossière est, intuitivement, plus pénalisante qu'une erreur survenant à une échelle fine. C'est pourquoi Bouman et al. proposent dans [Bouman94] de considérer une fonction de coût qui pénalise les erreurs proportionnellement au nombre de ses descendants à l'échelle la plus fine. Le nouvel algorithme obtenu, appelé *SMAP*, est très proche du précédent.

17.3.2 Algorithme non-itératif du *MPM* sur les arbres de Markov cachés.

Cet algorithme, au sujet duquel on pourra consulter [Laferté96], [Williams99b] ainsi que [Smyth97], permet un calcul exact des probabilités marginales a posteriori. Il est, lui aussi, dérivé des résultats obtenus par Baum et Welch pour les chaînes de Markov et se déroule en deux passes, grâce aux relations de récurrence dont nous détaillons maintenant les calculs :

L'estimateur recherché étant, pour tout $s \in S$:

$$\hat{x}_s = \operatorname{argmax}_{x_s} P(X_s = x_s | Y = y),$$

il faut regarder :

$$\begin{aligned} P(X_s = x_s | Y = y) &= \sum_{x_{s-} \in \Omega} P(X_s = x_s, X_{s-} = x_{s-} | Y = y), \forall s \in S, \\ &= \sum_{x_{s-} \in \Omega} P(X_s = x_s | X_{s-} = x_{s-}, Y_{\geq s} = y_{\geq s}) P(X_{s-} = x_{s-} | Y = y), \forall s \in S, \end{aligned}$$

Application à l'imagerie.

d'après la formule de Bayes et les hypothèses de markoviannité sur arbre.

On a, alors, la relation de récurrence descendante :

$$P(X_s = x_s | Y = y) = \sum_{x_{s-} \in \Omega} \frac{P(X_s = x_s, X_{s-} = x_{s-} | Y_{\geq s} = y_{\geq s})}{\underbrace{P(X_{s-} = x_{s-} | Y_{\geq s} = y_{\geq s})}_{\phi_s(x_s, x_{s-})}} P(X_{s-} = x_{s-} | Y = y), \quad \forall s \in S.$$

Le calcul des $\phi_s(x_s, x_{s-})$, pour un s donné, ne nécessite la connaissance que des $P(X_{s-} = x_{s-} | X_s = x_s)$ et des $P(X_s = x_s | Y_{\geq s} = y_{\geq s})$, $\forall (x_{s-}, x_s) \in \Omega^2$. En effet,

$$P(X_s = x_s, X_{s-} = x_{s-} | Y_{\geq s} = y_{\geq s}) = P(X_{s-} = x_{s-} | X_s = x_s) P(X_s = x_s | Y_{\geq s} = y_{\geq s}),$$

et donc :

$$\phi_s(x_s, x_{s-}) = \frac{P(X_{s-} = x_{s-} | X_s = x_s) P(X_s = x_s | Y_{\geq s} = y_{\geq s})}{\sum_{x_s \in \Omega} P(X_{s-} = x_{s-} | X_s = x_s) P(X_s = x_s | Y_{\geq s} = y_{\geq s})}. \quad [17.88]$$

La première quantité est obtenue par :

$$P(X_{s-} = x_{s-} | X_s = x_s) = \frac{P(X_s = x_s | X_{s-} = x_{s-}) P(X_{s-} = x_{s-})}{P(X_s = x_s)},$$

où les $P(X_s = x_s)$, $\forall s \in S$, sont calculés par la récurrence :

$$P(X_s = x_s) = \sum_{x_{s-} \in \Omega} P(X_s = x_s | X_{s-} = x_{s-}) P(X_{s-} = x_{s-}),$$

cette récurrence étant initialisée, à la racine, par les paramètres du modèle :

$$P(X_r = \omega_i), \quad \forall \omega_i \in \Omega.$$

La seconde quantité est issue d'une autre relation de récurrence montante. Détaillons les calculs :

$$\begin{aligned} P(X_s = x_s | Y_{\geq s} = y_{\geq s}) &= \frac{P(X_s = x_s) P(Y_{\geq s} = y_{\geq s} | X_s = x_s)}{P(Y_{\geq s} = y_{\geq s})}, \\ &= \frac{P(X_s = x_s) P(Y_s = y_s | X_s = x_s) \prod_{t \in s^+} P(Y_{\geq t} = y_{\geq t} | X_s = x_s)}{P(Y_{\geq s} = y_{\geq s})}, \end{aligned}$$

et donc :

$$\begin{aligned} P(X_s = x_s | Y_{\geq s} = y_{\geq s}) &= \frac{P(X_s = x_s)^{-3} P(Y_s = y_s | X_s = x_s) \prod_{t \in s^+} P(Y_{\geq t} = y_{\geq t}, X_s = x_s)}{P(Y_{\geq s} = y_{\geq s})}, \\ &= \frac{P(X_s = x_s)^{-3} P(Y_s = y_s | X_s = x_s)}{P(Y_{\geq s} = y_{\geq s})} \\ &\quad \times \prod_{t \in s^+} [P(Y_{\geq t} = y_{\geq t}) P(X_s = x_s | Y_{\geq t} = y_{\geq t})], \end{aligned}$$

Application à l'imagerie.

c'est à dire :

$$P(X_s = x_s | Y_{\geq s} = y_{\geq s}) = \frac{\prod_{t \in s^+} P(Y_{\geq t} = y_{\geq t})}{\underbrace{P(Y_{\geq s} = y_{\geq s})}_{1/Z(y_{\geq s})}} P(X_s = x_s)^{-3} \\ \times P(Y_s = y_s | X_s = x_s) \prod_{t \in s^+} P(X_s = x_s | Y_{\geq t} = y_{\geq t}),$$

que l'on écrira :

$$P(X_s = x_s | Y_{\geq s} = y_{\geq s}) = \frac{P(X_s = x_s)^{-3} P(Y_s = y_s | X_s = x_s)}{Z(y_{\geq s})} \\ \times \prod_{t \in s^+} \sum_{x_t \in \Omega} P(X_s = x_s | X_t = x_t) P(X_t = x_t | Y_{\geq t} = y_{\geq t}).$$

afin de faire apparaître la récurrence montante.

D'après la forme du rapport [17.88], nous pouvons omettre le calcul des $Z(y_{\geq s})$.

ALGORITHME DE BAUM-WELCH ET LAFERTE

L'algorithme de calcul du MPM se déroule comme suit :

[BWL₁] Initialisation de la récurrence montante :

$$P(X_s = \omega_i | Y_s = y_s) = \frac{P(X_s = \omega_i) P(Y_s = y_s | X_s = \omega_i)}{\sum_{\omega_j \in \Omega} P(X_s = \omega_j) P(Y_s = y_s | X_s = \omega_j)}, s \in S^0.$$

[BWL₂] Passe montante : pour $1 \leq n \leq N-1$, $\forall s \in S^n$, on déduit de la récurrence montante les calculs de : $\phi_s(x_s, x_{s^-})$.

[BWL₃] Initialisation de la passe descendante :

$$\hat{x}_r = \operatorname{argmax}_{x_r} P(X_r = x_r | Y_{\geq r} = Y = y).$$

[BWL₄] Passe descendante : pour $N-1 \geq n \geq 0$, on tire des calculs de [BWL₂], grâce à la relation de récurrence [17.88] :

$$\forall s \in S^n, \hat{x}_s = \operatorname{argmax}_{x_s} P(X_s = x_s | Y = y).$$

L'estimateur du MPM est alors :

$$\forall s \in S, \widehat{X}_s = \hat{x}_s.$$

L'algorithme ci-dessus manipule des quantités suffisamment grandes pour ne pas être victime des problèmes d'underflow signalés dans le cas des chaînes de Markov cachées.

Dans [Chardin00], A. Chardin a développé une méthode analogue au *SMAP*, appelée *SMPM*, afin de rendre plus "équitable" la pénalisation des sites érronés. Ses travaux concernent aussi les arbres tronqués. Il s'agit de

Application à l'imagerie.

quadarbres amputés de leurs résolutions les plus grossières. Le champs obtenu à la résolution définie, alors, comme étant la plus grossière, est considéré comme un champs de Markov. Ce modèle se justifie par l'observation que les niveaux les plus fins de l'arbre suffisent à caractériser la structure d'une image. Les meilleurs résultats de segmentation sont atteints, dans le cadre d'un quadarbre comptant normalement huit niveaux, pour un nombre de niveaux inférieur ou égal à quatre. Ces travaux sont confirmés par ceux de Williams et Feng, (voir [Williams99a] et [Williams99b]), qui considèrent l'image comme une collection d'arbres disjoints.

La méthode que nous proposons est une méthode plus directe de calcul, qui se rapproche, en celà, de l'algorithme de Baum-Welch, pour le calcul du *MPM*.

17.3.3 Algorithme simplifié sur les arbres de Markov cachés.

Nous proposons un algorithme de calcul ou de simulation de la loi de X conditionnellement à $Y = y$, en deux phases. La première passe, montante, permet de calculer, de manière déterministe, les probabilités de transition a posteriori. La seconde passe, descendante, permet un calcul direct, ou une simulation d'une réalisation de X a posteriori. Cet algorithme permet, entre autre, de calculer la solution du *MPM*.

Cet algorithme se décompose comme suit :

• *une passe montante*

$n = 0$: on pose pour tout $s \in S^0$:

$$A_0(x_s^0 = \omega_i, x_{s_-}^1 = \omega_j) = P(x_s^0 = \omega_i / x_{s_-}^1 = \omega_j) f_{(x_s^0=i)}(y_s). \quad [17.89]$$

où $x_{s_-}^1$ est le père de x_s^0 et y_s est l'observation associée à x_s^0 .

Pour $1 \leq n \leq N - 1$: par récurrence, $s \in S^n$,

$$A_n(x_s^n = \omega_i, x_{s_-}^{n+1} = \omega_j) = P(x_s^n = \omega_i / x_{s_-}^{n+1} = \omega_j) \prod_{s^+ \in S_+^n} \left(\sum_{\omega \in \Omega} (A_{n-1}(x_{s^+}^{n-1} = \omega, x_s^n = \omega_i)) \right), \quad [17.90]$$

où S_+^n est l'ensemble des fils de x_s^n .

$n = N$: calcul de

$$A_N(x_r^N = \omega_i) = P(x_r^N = \omega_i) \prod_{s^+ \in S_+^N} \left(\sum_{\omega \in \Omega} (A_{N-1}(x_{s^+}^{N-1} = \omega, x_r^N = \omega_i)) \right). \quad [17.91]$$

• *une passe descendante*

$n = N$: calcul de la probabilité initiale à posteriori

$$P_N^Y(x_r^N = \omega_i) = \frac{A_N(x_r^N = \omega_i)}{\sum_{\omega \in \Omega} A_N(x_r^N = \omega)}. \quad [17.92]$$

de $n = N - 1$ à 0 : simulation de l'arbre à posteriori grâce aux probabilités de transition

Application à l'imagerie.

$$P_n^Y (x_s^n = \omega_i / x_{s-}^{n+1} = \omega_j) = \frac{A_n (x_s^n = \omega_i, x_{s-}^{n+1} = \omega_j)}{\sum_{\omega \in \Omega} A_n (x_s^n = \omega, x_{s-}^{n+1} = \omega_j)}. \quad [17.93]$$

En effectuant plusieurs simulations, on peut alors estimer les lois marginales de X^0 et effectuer une classification (qui est ici l'estimation de $X^0 = x^0$ à partir de $Y = y$) par la méthode du *MPM*.

Pour obtenir une version déterministe de l'algorithme, il suffit de calculer les probabilités a posteriori, par la récurrence descendante :

$$P_n^Y (x_s^n = \omega_i) = \sum_{\omega \in \Omega} P_n^Y (x_s^{n+1} = \omega) P_n^Y (x_s^n = \omega_i / x_{s-}^{n+1} = \omega).$$

Le principal problème qui se pose lors de la passe montante, c'est la décroissance très rapide des A_n , qui peuvent, après quelques niveaux seulement, descendre sous la précision, en 0, de la machine. Il est par conséquent nécessaire de procéder à une renormalisation lors du calcul des A_n . Au vu de la forme du calcul des P_n^Y , nous effectuerons la renormalisation :

$$A'_n (x_s^n = \omega_i, x_{s-}^{n+1} = \omega_j) = \frac{A_n (x_s^n = \omega_i, x_{s-}^{n+1} = \omega_j)}{\sum_{\omega, \omega' \in \Omega} A_n (x_s^n = \omega, x_{s-}^{n+1} = \omega')},$$

ce qui permettra de conserver les calculs décrits dans l'algorithme précédent, en remplaçant, à chaque étape, A_n par A'_n . On vérifie aisément que les probabilités de transition a posteriori s'expriment, en effet, de la même façon en fonction de A'_n et de A_n , avec l'intérêt supplémentaire que les A'_n , renormalisés à chaque étape, restent suffisamment grands pour être gérés par la machine.

Remarquons, enfin, que le calcul des probabilités a posteriori, décrit dans la passe descendante de l'algorithme, peut être mené simultanément à celui des A'_n . Cette remarque permet, lors de l'exécution du calcul, d'éviter une double manipulation inutile et coûteuse en temps calcul, de certaines quantités.

17.3.4 Segmentation supervisée sur les arbres.

Afin de tester l'efficacité de l'algorithme de classification, on peut le tester sur les images simulées par les différents types d'arbres.

Segmentations par quadarbres.

Si l'on bruite, avec un bruit gaussien de référence, des réalisations simulées par les deux types de quadarbres, les résultats obtenus donnent une bonne idée de la différence entre des deux modèles. FIG. III.44 présente les segmentations de deux images proches, la première, pour le quadarbre classique, la seconde, pour le quadarbre évolutif

Segmentations par diarbres.

Les différents bruitages, des images, FIG. III.43, qui sont proposés à la segmentation, FIG. III.45, permettent de tester la segmentation par les diarbres dans le cas de distributions normales bruitantes différenciables par leurs moyennes ou par leurs variances. Il apparaît clairement que la méthode est fiable, indépendamment des paramètres du bruit ou de ceux du modèle.

Application à l'imagerie.

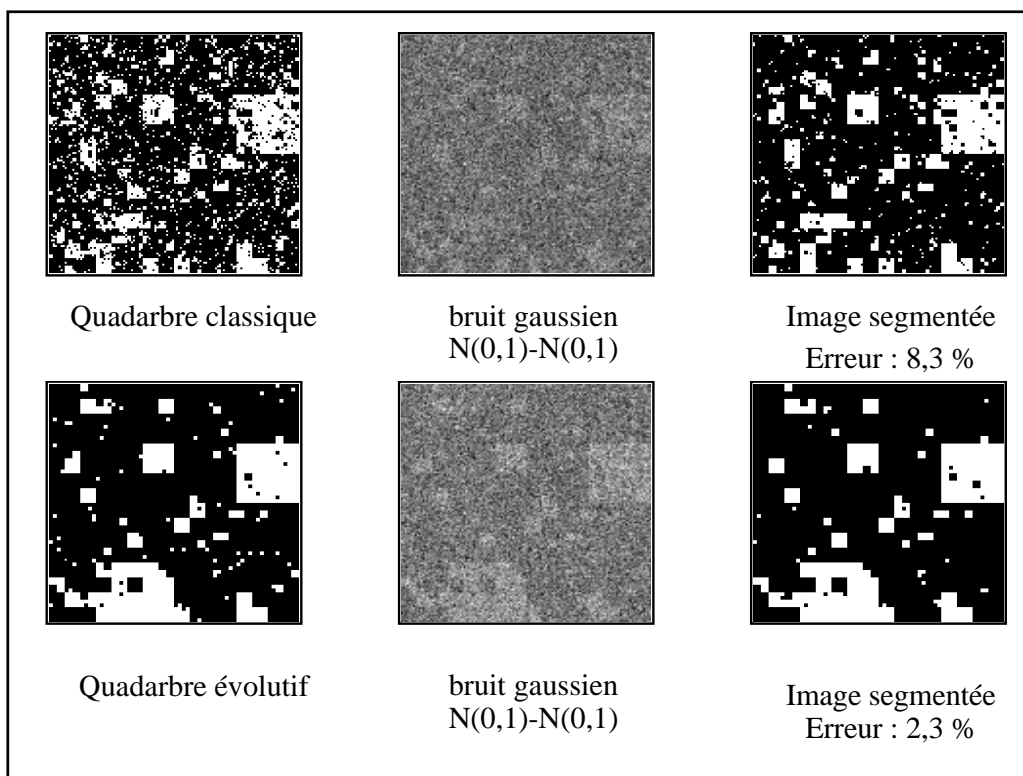


FIG. III.44. - SEGMENTATIONS D'IMAGES SIMULÉES PAR QUADARBRES.

La segmentation par le diarbre droit illustre le cas où les deux classes, de l'image, sont très efficacement différenciées, mais où les "couleurs" les représentant sont inversées. Ce phénomène est fonction du bruitage, puisqu'on peut remarquer que la classe "blanche" de l'image segmentée correspond à la classe "la plus claire" de l'image bruitée. La segmentation d'images de synthèse issues du modèle à 3, 5 ou 7 classes, montre que la qualité de la segmentation est tout aussi probante, lorsque le nombre de classes augmente. Dans les exemples de la figure FIG. III.46, les classes sont bruitées avec des distributions normales de moyennes échelonnées de 1 en 1, et de même variance 1. Les segmentations des images à 3 et 5 classes sont effectuées pour des diarbres droits, alors que les segmentations à 7 classes sont présentées, dans l'ordre, pour les diarbres droits, alternés et diagonaux.

En conclusion, il apparaît que les différents exemples traités par la méthode *MPM* de classification pour les différents diarbres de Markov cachés, assurent d'une certaine fiabilité de la méthode. Pour être étendue à d'autres types d'images, la méthode de classification nécessite une phase préalable d'estimation des paramètres du bruit, et des paramètres du modèle. Cet aspect est traité dans le chapitre suivant.

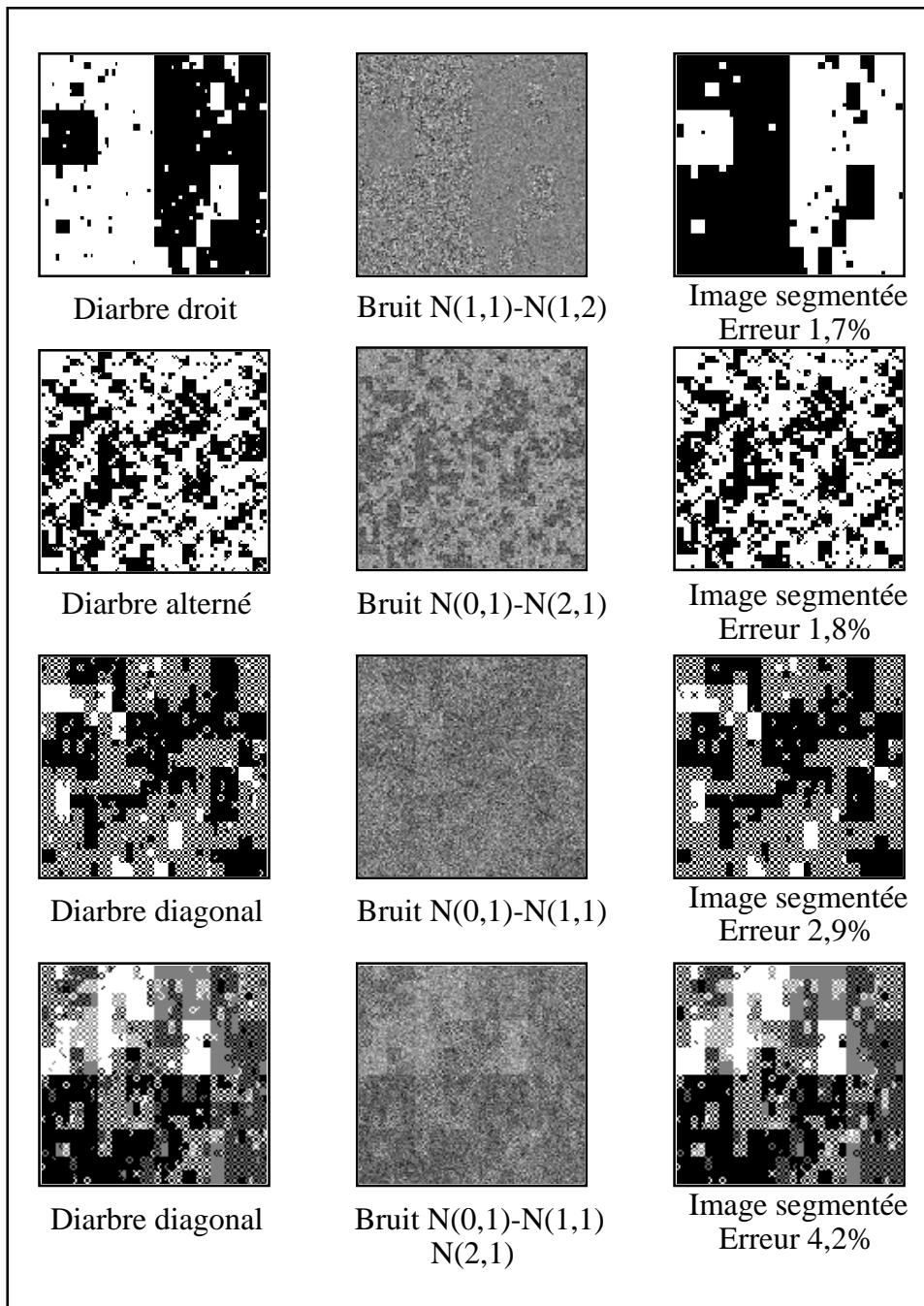
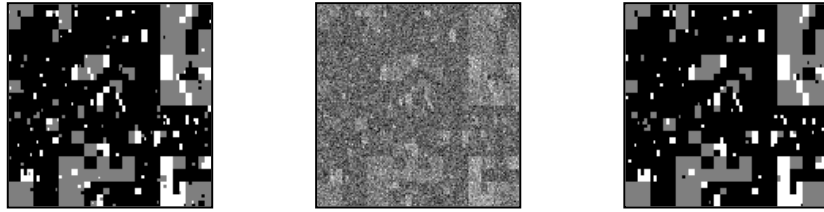


FIG. III.45. - SEGMENTATIONS D'IMAGES SIMULÉES PAR DIARBRES.

Application à l'imagerie.

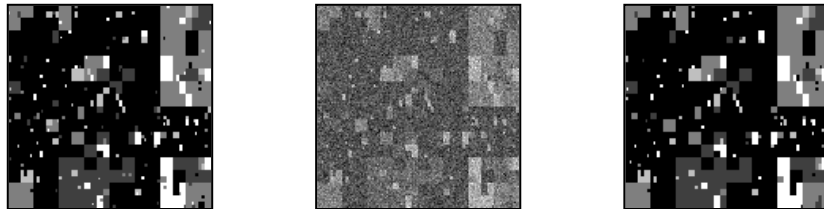
Segmentation à 3 classes : bruitage gaussien $N(0,1) - N(1,1) - N(2,1)$:

Diarbre droit :
Erreur : 2.6 %



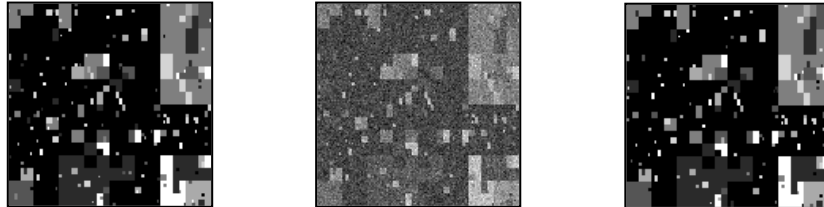
Segmentation à 5 classes : bruitage gaussien $N(0,1) - N(1,1) - N(2,1) - N(3,1) - N(4,1)$:

Diarbre droit :
Erreur : 2.3 %

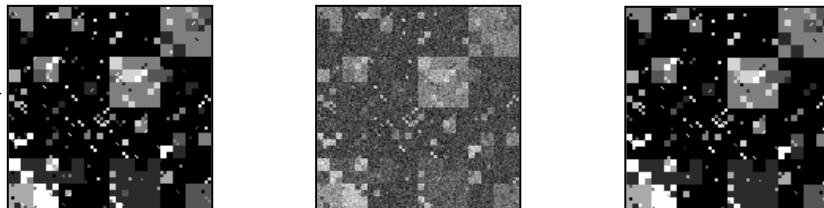


Segmentation à 7 classes : bruitage gaussien $N(0,1) - N(1,1) - N(2,1) - N(3,1) - N(4,1) - N(5,1) - N(6,1)$:

Diarbre droit :
Erreur : 1.7 %



Diarbre alterné :
Erreur : 1.7 %



Diarbre diagonal
Erreur : 1.8 %



FIG. III.46. - SEGMENTATIONS D'IMAGES SIMULÉES PAR LES DIARBRES.

- CHAPITRE 18 -

Estimation des paramètres dans les modèles cachés.

Nous cherchons, ici, à estimer les paramètres $\theta = (\theta_1, \theta_2)$ où θ_1 est l'ensemble des paramètres concernant la loi de X et θ_2 l'ensemble de ceux concernant les lois de Y conditionnellement à X , que l'on suppose généralement gaussiennes. Cette estimation se fait à partir de la seule observation des réalisations de Y .

18.1 Cas des champs de Markov.

La loi de X est une distribution de Gibbs par rapport au type de voisinage V [15.63]. Les $m = |\Omega|^{|V|}$ configurations qui existent dans le cadre de ce type de voisinage pour l'ensemble des classes Ω correspondent aux m probabilités conditionnelles, $p_{i\nu}$, qu'il faut estimer pour décrire θ_1 ,

$$p_{i\nu} = P[X_s = \omega_i / V_s \text{ de type } \nu], \forall s \in S. \quad [18.94]$$

Si l'on suppose Y gaussien conditionnellement à X , les paramètres θ_2 du bruit sont les $k = |\Omega|$ couples moyennes-variances définissant ces distributions.

18.1.1 L'algorithme EM Gibbsien :

Cet algorithme, développé par Chalmond (voir [Chalmond89]) utilise le principe de *EM*, (voir [Dempster77], [Redner84] et [Peng95]), avec une réestimation effectuée grâce à l'échantillonneur de Gibbs.

Il s'agit de maximiser, par rapport à θ , la loi du couple (X, Y) , notée $L_\theta(X, Y)$, conditionnellement à $Y = y$.

Cependant, le problème déjà évoqué de l'ignorance de la constante de normalisation dans la distribution Markovienne du champs ne permet pas le calcul direct par cette méthode. L'idée de Besag, (voir [Besag74]), d'avoir recours à la pseudo-vraisemblance permet alors d'éviter cet écueil. Il s'agit de remplacer la vraisemblance par le produit des distributions conditionnelles qui ne contiennent plus de constante de renormalisation. Pour un type, noté V ,

de voisinage, la pseudo-vraisemblance l_θ est définie par :

$$l_\theta(x, y) = \prod_{s \in S} P(X_s = x_s / X_{V_s} = x_{V_s}, \theta) P(Y_s = y_s / X_s = x_s),$$

et nous recherchons :

$$\hat{\theta}_{EM} = \arg \max_{\theta} E[l_\theta(x, y) / Y = y, \theta]. \quad [18.95]$$

La maximisation étant recherchée itérativement, on définit, à partir d'une valeur initiale $\theta^{(0)}$, une suite $(\theta^{(q)})_{q \in \mathbb{N}}$ strictement croissante qui converge vers l'argument du maximum de vraisemblance. La suite est définie par la récurrence :

$$\theta^{(q+1)} = \arg \max_{\theta} E[l_\theta(x, y) / Y = y, \theta^{(q)}]. \quad [18.96]$$

Le calcul des espérances, comme nous l'avons signalé précédemment, ne pouvant être effectué directement, on les approche stochastiquement, grâce à la simulation, par exemple par l'échantillonneur de Gibbs, d'un grand nombre de champs tirés suivant la loi à posteriori en fonction de $\theta^{(q)}$. Grâce à ces tirages, on réestime les $P_{\theta^{(q)}}[X_t = \omega_i / Y = y]$, pour se ramener aux formules de réestimation, pour les paramètres du bruit :

$$\begin{aligned} \mu_i^{q+1} &= \frac{\sum_{t \in S} P_{\theta^{(q)}}(X_t = \omega_i / Y = y) y_t}{\sum_{t \in S} P_{\theta^{(q)}}(X_t = \omega_i / Y = y)}, \\ (\sigma^2)_i^{q+1} &= \frac{\sum_{t \in S} P_{\theta^{(q)}}(X_t = \omega_i / Y = y) (y_t - \mu_i^{q+1})^2}{\sum_{t \in S} P_{\theta^{(q)}}(X_t = \omega_i / Y = y)} \end{aligned} \quad [18.97]$$

et pour tout $s \in S$, $\omega_i \in \Omega$,

$$P_{\theta^{(q+1)}}(X_s = \omega_i / X_{V_s}, \text{ avec } V_s \text{ de type } \nu) = \frac{E[N_{i\nu}^q / Y = y, \theta^{(q)}]}{\sum_{i=1}^k E[N_{i\nu}^q / Y = y, \theta^{(q)}]}, \quad [18.98]$$

où $N_{i\nu}^q$ est le nombre d'occurrences de la configuration $X_s = \omega_i$ pour V_s de type ν , dans la réalisation de X à l'itération q . Pour simplifier l'énoncé de l'algorithme, posons :

$$\begin{aligned} p_{i\nu}^q &= P_{\theta^{(q)}}(X_s = \omega_i / X_{V_s}, \text{ avec } V_s \text{ de type } \nu) \\ \text{et } \gamma_t^q(i) &= P_{\theta^{(q)}}(X_t = \omega_i / Y = y). \end{aligned}$$

L'algorithme *EM* Gibbsien se déroule comme suit :

ALGORITHME EM GIBBSIEN

Les étapes [EMG₂] à [EMG₄] sont présentées à l'itération q et sont à répéter jusqu'à convergence (jusqu'à ce qu'un critère d'arrêt soit vérifié).

[EMG₁] Initialisation de $\theta^{(0)}$.

[EMG₂] Tirage d'un grand nombre de champs par échantillonneur de Gibbs suivant la loi :

$$P(X = x/Y = y, \theta^{(q)}).$$

[EMG₃] Approximation, par la loi des grands nombres, de $\gamma_t^q(i)$ et N_{iv}^q pour tout $\omega_i \in \Omega$, $t \in S$ et tout type de voisinage v .

[EMG₄] Calcul de p_{iv}^{q+1} , μ_i^{q+1} , $(\sigma^2)_i^{q+1}$ par les formules précédentes.

Le choix de l'initialisation des paramètres du modèle influence la rapidité de convergence de l'algorithme. Il y a aussi des risques de convergence vers un maximum local éloigné du maximum absolu.

18.1.2 L'algorithme du gradient stochastique GS .

GS recherche directement le maximum de la log vraisemblance $\log[L_\theta(X, Y)]$. On cherche le paramètre θ maximisant $\log[L_\theta]$, pour l'énergie $\Delta(\theta, x, y)$ et $\Delta'(\theta, x, y)$, son gradient par rapport à θ .

Younes montre dans [Younes88] et [Younes89], que

$$\frac{\partial}{\partial \theta} \log[L_\theta(x, y)] = E_\theta[\Delta'] - E_\theta[\Delta'/Y = y], \quad [18.99]$$

grâce à quoi l'on se ramène à résoudre une équation de type $E_\theta[Z_\theta] = 0$, soluble par une méthode de gradient stochastique. Si (A_q, B_q) suit la même loi que (X, Y) , et C_q est une variable suivant la loi conditionnelle de X par rapport à $Y = y$, ces variables peuvent être simulées, avec l'échantillonneur de Gibbs, en fonction de la valeur courante de θ_q . La réestimation se fait par:

$$\theta_{q+1} = \theta_q + \frac{d}{q+1} [\Delta'(\theta, A_q, B_q) - \Delta'(\theta, C_q, y)], \quad [18.100]$$

où d est une constante, qui peut, par exemple, être choisie inversement proportionnelle au nombre de pixels de l'image.

18.1.3 L'algorithme Iterative Conditional Estimation (ICE)

C'est une procédure générale d'estimation des paramètres développée par Pieczynski (voir [Pieczynski90], [Pieczynski92] et [Pieczynski94]) qui ne vise pas à rechercher le maximum de vraisemblance. L'idée est d'approcher, par une fonction des seules données observées, un estimateur $\hat{\theta}$ de θ , défini à partir des données complètes (x, y) , minimal au sens de l'erreur quadratique. C'est l'espérance conditionnelle :

$$\mathbb{E}[\hat{\theta}(X, Y)/Y = y],$$

qui réalise la meilleure approximation. Le principe d'*ICE* consiste donc, par un procédé itératif, à calculer, si cela est possible, cette espérance conditionnelle ou à l'approcher si cela s'avère nécessaire, par une démarche stochastique de tirage d'échantillon suivant la loi a posteriori pour le paramètre courant. A chaque itération on réestime, donc, le paramètre courant par :

$$\theta^{(q+1)} = \mathbb{E}_{\theta^{(q)}} \left[\widehat{\theta}(X, Y) / Y = y \right].$$

Il est possible d'estimer séparément ou en parallèle les deux composantes $\widehat{\theta}_1$ et $\widehat{\theta}_2$ de $\widehat{\theta}$:

$$\begin{aligned} \theta_1^{(q+1)} &= \mathbb{E}_{\theta^{(q)}} \left[\widehat{\theta}_1(X, Y) / Y = y \right] \\ \text{et } \theta_2^{(q+1)} &= \mathbb{E}_{\theta^{(q)}} \left[\widehat{\theta}_2(X, Y) / Y = y \right]. \end{aligned}$$

D'une façon générale, les estimateurs choisis pour les paramètres du bruit seront les moments empiriques.

Notons que lorsque l'estimateur $\widehat{\theta}$ que l'on choisit est l'estimateur du maximum de vraisemblance, *ICE* et *EM* se rejoignent si l'on peut faire entrer l'opérateur $\arg \max_{\theta}$ dans l'espérance conditionnelle (voir [Delmas95] et [Delmas97]).

Pour conclure cette présentation d'introduction à *ICE*, nous dirons qu'*ICE*, qui ne cherche pas, nécessairement, l'estimateur du maximum de vraisemblance, est applicable dans un plus grand nombre de cas que *EM*.

L'algorithme *ICE* sous sa forme générale se déroule comme suit :

ALGORITHME ICE DE PIECZYNSKI

Les étapes [ICEP₂] à [ICEP₄] sont présentées à l'itération q et sont à répéter jusqu'à convergence (jusqu'à ce qu'un critère d'arrêt soit vérifié).

[ICEP₁] Initialisation de $\theta^{(0)}$ et choix de $\widehat{\theta}_1$ et $\widehat{\theta}_2$.

[ICEP₂] Simulation de N réalisations du champ suivant la loi à posteriori, grâce à l'échantillonneur de Gibbs.

[ICEP₃] Estimation de $\widehat{\theta}_{1l}^{(q)}$ et $\widehat{\theta}_{2l}^{(q)}$, $1 \leq l \leq N$, pour chaque réalisation précédente.

[ICEP₄] Réestimation des paramètres, par les moyennes :

$$\begin{aligned} \theta_1^{(q+1)} &= \frac{1}{N} \sum_{l=1}^N \widehat{\theta}_{1l}^{(q)} \\ \text{et } \theta_2^{(q+1)} &= \frac{1}{N} \sum_{l=1}^N \widehat{\theta}_{2l}^{(q)}. \end{aligned}$$

Comme nous l'avons déjà mentionné, la réestimation des paramètres des lois de Y_t conditionnellement à X_t se fait, en général, par les moyennes et variances empiriques:

$$\widehat{\mu}_i^{(q+1)} = \frac{\sum_{j=1}^n Y_j 1_{(X_j=\omega_i)}}{\sum_{j=1}^n 1_{(X_j=\omega_i)}}, \quad [18.101]$$

Estimation des paramètres dans les modèles cachés.

$$\left(\widehat{\sigma_i^2}\right)^{(q+1)} = \frac{\sum_{j=1}^n \left(Y_j - \widehat{\mu}_i^{(q+1)}\right)^2 \mathbf{1}_{(X_j=\omega_i)}}{\sum_{j=1}^n \mathbf{1}_{(X_j=\omega_i)}}$$

et, pour les paramètres du modèle :

$$\text{pour tout } s \in S, \omega_i \in \Omega, P_{\theta^{(q+1)}}(X_s = \omega_i / X_{V_s}, \text{ avec } V_s \text{ de type } \nu) = \frac{E \left[N_{i\nu}^q / Y = y, \theta^{(q)} \right]}{\sum_{i=1}^k E \left[N_{i\nu}^q / Y = y, \theta^{(q)} \right]}, \quad [18.102]$$

où $N_{i\nu}^q$ est le nombre d'occurrences de la configuration $X_s = \omega_i$ pour V_s de type ν , dans la réalisation de X à l'itération q .

18.2 Cas des chaînes de Markov.

Nous exposons dans ce qui suit les principales méthodes d'estimation des paramètres dans le cadre des chaînes de Markov, cadre qui est proche de la modélisation sur arbre qui fait l'objet de notre étude. Ce paragraphe est fortement inspiré de l'article de Benmiloud et Pieczynski [Benmiloud95] très complet sur l'estimation des paramètres dans les chaînes de Markov cachées avec application à la segmentation d'image.

Dans la modélisation par les chaînes de Markov cachées, les paramètres à estimer sont:

- les probabilités initiales π_i et la matrice de transition a_{ij} .
- Pour le bruit gaussien, associé à la classe ω_i , la moyenne μ_i et la variance σ_i .

18.2.1 L'algorithme Expectation Maximisation (EM classique).

Cet algorithme a été développé initialement par Baum dans [Baum72], avant même que la procédure générale de EM ne soit définie dans [Dempster77].

Comme les algorithmes précédemment exposés dans le cadre des chaînes de Markov cachées, cet algorithme est entièrement déterministe. L'EM sur les chaînes de Markov cachées se déroule itérativement et permet de définir une suite croissante $(\theta^{(q)})_q$ croissante et convergente sous de bonnes conditions initiales. Le problème de l'initialisation est discuté dans [Benmiloud95].

A chaque itération on utilise les formules de réestimation des paramètres a priori suivantes :

$$\begin{aligned} \pi_i^{(q+1)} &= \frac{1}{|S|} \sum_{t \in S} \xi_t^{(q)}(i) \\ \text{et } a_{ij}^{(q+1)} &= \frac{\sum_{t \in S'} \Psi_t^{(q)}(i, j)}{\sum_{t \in S'} \xi_t^{(q)}(i)}, \end{aligned} \quad [18.103]$$

où S' est l'ensemble S privé du dernier élément de la chaîne de Markov et où

$$\Psi_t(i, j) = P(X_t = \omega_i, X_{t+1} = \omega_j / Y = y)$$

se calcule par,

$$\Psi_t(i, j) = \frac{\alpha_t(i) a_{ij} f_j(y_{t+1}) \beta_{t+1}(j)}{\sum_{l=1}^k f_t(y_{t+1}) \sum_{j=1}^k \alpha_t(j) a_{jl}}, \quad [18.104]$$

ξ_t ayant été défini en [15.77],

Pour les paramètres du bruit on aura :

$$\begin{aligned} \mu_i^{(q+1)} &= \frac{\sum_{t \in S} \xi_t^{(q)}(i) y_t}{\sum_{t \in S} \xi_t^{(q)}(i)} \\ \text{et } (\sigma^2)_i^{(q+1)} &= \frac{\sum_{t \in S} \xi_t^{(q)}(i) (y_t - \mu_i^{(q+1)})^2}{\sum_{t \in S} \xi_t^{(q)}(i)}. \end{aligned} \quad [18.105]$$

Compte tenu de ce que nous avons déjà vu sur les chaînes de Markov cachées, l'algorithme *EM* de Baum-Welch se déroule de la façon suivante :

ALGORITHME EM DE BAUM-WELCH

Les étapes [EMB_W₂] à [EMB_W₄] sont présentées à l'itération q et sont à répéter jusqu'à convergence (jusqu'à ce qu'un critère d'arrêt soit vérifié).

[EMB_W₁] Initialisation des paramètres $\theta^{(0)}$.

[EMB_W₂] Calcul des probabilités forward et backward pour les paramètres courant.

[EMB_W₃] Calcul de

$$\xi_t^{(q)}(i) \text{ pour tout } t \in S' \text{ et } \omega_i \in \Omega$$

$$\text{et } \Psi_t^{(q)}(i, j) \text{ pour tout } t \in S' \text{ et } (\omega_i, \omega_j) \in \Omega^2.$$

[EMB_W₄] Réestimation des paramètres :

$$\pi_i^{(q+1)}, a_{ij}^{(q+1)}, \mu_i^{(q+1)}, (\sigma^2)_i^{(q+1)} \text{ pour tout } (\omega_i, \omega_j) \in \Omega^2,$$

grâce aux formules exposées précédemment.

Le choix du critère d'arrêt de cet algorithme est aussi discuté dans [Benmiloud95].

18.2.2 Algorithmes *ICE* et *MICE* de Benmiloud-Pieczynski.

Ces algorithmes développés dans le cadre des chaînes de Markov stationnaires permettent de travailler sur les probabilités jointes de la loi a priori, ce qui permet de calculer analytiquement les

$$c_{ij} = P(X_t = \omega_i, X_{t+1} = \omega_j).$$

Estimation des paramètres dans les modèles cachés.

En effet, si l'on choisit les fréquences empiriques comme estimateurs de ces paramètres,

$$\widehat{c}_{ij}(X) = \frac{1}{|S'|} \sum_{t \in S'} 1_{[X_t = \omega_i, X_{t+1} = \omega_j]},$$

la formule de réestimation, à la $(q+1)^{i\grave{e}me}$ itération sera :

$$c_{ij}^{(q+1)} = \frac{1}{|S'|} \sum_{t \in S'} P \left[X_t = \omega_i, X_{t+1} = \omega_j / Y = y, \theta^{(q)} \right],$$

ce qui nous ramène, si l'on reprend la paramétrisation utilisée pour *EM*, à une réestimation pour les composantes de la matrice de transition qui est :

$$a_{ij}^{(q+1)} = \frac{\sum_{t \in S'} \Psi_t^{(q)}(i, j)}{\sum_{t \in S'} \xi_t^{(q)}(i)}.$$

La réestimation des $a_{ij}^{(q+1)}$, que l'on sait calculer analytiquement, se fait donc, pour *ICE* et *MICE*, de la même façon que pour *EM*. Les paramètres du bruit étant estimés par les moments empiriques, à partir des échantillons simulés selon la loi a posteriori en fonction du paramètre courant, cette méthode est à la fois déterministe et stochastique. Ceci la différencie de *EM* et *SEM* et constitue un juste milieu entre les avantages et les inconvénients de chacune de ces deux autres méthodes.

C'est précisément au niveau de la simulation de la réalisation de la chaîne selon la loi a posteriori en fonction du paramètre courant, que se situe la différence entre *ICE* et *MICE*. L'algorithme *MICE* diffère de la procédure générale en ce qu'il propose, plutôt qu'une simulation suivant la loi a posteriori, un tirage, à chaque instant t de l'itération q , suivant la loi marginale a posteriori, $\xi_t^{(q)}(i)$.

ALGORITHMES ICE ET MICE

Les étapes [ICEM₂] à [ICEM₄] sont présentées à l'itération q et sont à répéter jusqu'à convergence (jusqu'à ce qu'un critère d'arrêt soit vérifié).

[ICEM₁] Initialisation des paramètres $\theta^{(0)}$.

[ICEM₂] Calcul des probabilités forward et backward pour les paramètres courants.

[ICEM₃] Calcul de

$$\xi_t^{(q)}(i) \text{ pour tout } t \in S' \text{ et } \omega_i \in \Omega$$

et

$$\Psi_t^{(q)}(i, j) \text{ pour tout } t \in S' \text{ et } (\omega_i, \omega_j) \in \Omega^2.$$

[ICEM₄] Réestimation des paramètres :

$$\pi_i^{(q+1)}, a_{ij}^{(q+1)} \text{ pour tout } (\omega_i, \omega_j) \in \Omega^2.$$

[ICEM₅] Simulation d'une réalisation de X , en fonction de l'algorithme *ICE* ou *MICE* choisi, selon le paramètre courant $\theta^{(q)}$.

Estimation des paramètres dans les modèles cachés.

[ICEM₆] Estimation par les moments empiriques des paramètres du bruit :

$$\mu_i^{(q+1)} = \frac{\sum_{t \in S} y_t 1_{[X_t = \omega_i]}}{\sum_{t \in S} 1_{[X_t = \omega_i]}} \text{ pour tout } \omega_i \in \Omega$$

$$\text{et } (\sigma^2)_i^{(q+1)} = \frac{\sum_{t \in S} \left(y_t - \mu_i^{(q+1)} \right)^2 1_{[X_t = \omega_i]}}{\sum_{t \in S} 1_{[X_t = \omega_i]}} \text{ pour tout } \omega_i \in \Omega.$$

18.2.3 L'algorithme SEM

La version stochastique de l'algorithme EM (voir [Broniatowski83], [Celeux86], [Celeux92] et [Celeux95]), qui peut permettre, comme nous l'avons déjà signalé, de ne pas rester coincé au voisinage d'un maximum local s'avère très utile dans le cadre des chaînes de Markov cachées. Tous les paramètres sont estimés par une procédure stochastique, aussi bien ceux liés au modèles que ceux caractérisant le bruit.

La procédure est la suivante :

ALGORITHME SEM

Les étapes [SEM₂] à [SEM₄] sont présentées à l'itération q et sont à répéter jusqu'à convergence (jusqu'à ce qu'un critère d'arrêt soit vérifié).

[SEM₁] Initialisation des paramètres $\theta^{(0)}$.

[SEM₂] Simulation d'une réalisation X selon le paramètre courant $\theta^{(q)}$.

[SEM₃] Estimation des paramètres du modèle :

$$c_{ij}^{(q+1)} = \frac{1}{|S'|} \sum_{t \in S'} 1_{[X_t = \omega_i, X_{t+1} = \omega_j]} \text{ pour tout } (\omega_i, \omega_j) \in \Omega^2.$$

[SEM₄] Estimation par les moments empiriques des paramètres du bruit :

$$\mu_i^{(q+1)}, (\sigma^2)_i^{(q+1)} \text{ pour tout } \omega_i \in \Omega.$$

18.3 Cas des arbres de Markov cachés.

Les paramètres à estimer, dans le modèle qui nous intéresse sont :

→ Les probabilités de la racine sur chaque classe,

$$\pi_i = P(x_r^N = \omega_i),$$

Estimation des paramètres dans les modèles cachés.

ainsi que les probabilités de transition père-fils à l'échelle n ,

$$g_{i,j}^n(\alpha) = P(x_s^n = \omega_i, x_{s-1}^{n+1} = \omega_j), \text{ (resp. } h_{i,j}^n(\alpha) \text{ suivant le modèle),}$$

entièrement définies, dans notre modèle, par la connaissance du paramètre α introduit au chapitre précédent.

→ Les paramètres du bruit, typiquement, les moyennes et variances des distributions normales :

pour tout $\omega_i \in \Omega$, μ_i et σ_i définissant f_i .

18.3.1 Algorithmes *EM* sur le diarbre.

L'algorithme *EM* que nous développons ici, est inspiré de celui que nous venons de décrire sur les chaînes de Markov. On pourra consulter à ce sujet [Bouman94], [Laferté96], [Smyth97] et [Williams99b].

EM donne, pour les paramètres liés au modèle, les formules de réestimation :

$$\pi_i^{(q+1)} = P(X_r = \omega_i | Y = y, \theta^{(q)}) \quad [18.106]$$

et, pour tout n , $1 \leq n \leq N$,

$$g_{i,j}^n(\alpha^{(q+1)}) \text{ (resp. } h_{i,j}^n(\alpha^{(q+1)})) = \frac{\sum_{s \in S^n} P(X_s^n = \omega_i | X_{s-1}^{n+1} = \omega_j, Y = y, \theta^{(q)})}{|S^n|}, \quad [18.107]$$

ce qui permet de tirer, selon le modèle,

$$\alpha^{(q+1)} = \frac{1}{|C| \sum_{c \in C} |S^c|} \left[\sum_{c \in C} |S^c| \left[(g_{i,i}^c)^{-1} \circ g_{i,i}^c(\alpha^{(q+1)}) \right] \right], \quad C \subset \mathbb{N} \cup [1; N], \quad [18.108]$$

$$\text{ou } \alpha^{(q+1)} = \frac{1}{|C| \sum_{c \in C} |S^c|} \left[\sum_{c \in C} |S^c| \left[(h_{i,i}^c)^{-1} \circ h_{i,i}^c(\alpha^{(q+1)}) \right] \right], \quad C \subset \mathbb{N} \cup [1; N]. \quad [18.109]$$

C étant un ensemble à déterminer pour optimiser l'estimation. Pour les paramètres du bruit, la mise à jour des paramètres se fera par :

$$\mu_i^{(q+1)} = \frac{\sum_{s \in S^0} y_s P(X_s^0 = \omega_i | Y = y, \theta^{(q)})}{\sum_{s \in S^0} P(X_s^0 = \omega_i | Y = y, \theta^{(q)})} \quad [18.110]$$

$$\text{et } (\sigma_i^2)^{(q+1)} = \frac{\sum_{s \in S^0} (y_s - \mu_i^{(q+1)})^2 P(X_s^0 = \omega_i | Y = y, \theta^{(q)})}{\sum_{s \in S^0} P(X_s^0 = \omega_i | Y = y, \theta^{(q)})}. \quad [18.111]$$

Cet algorithme, entièrement déterministe, se déroule comme suit :

Estimation des paramètres dans les modèles cachés.

ALGORITHME EM SUR LE DIARBRE

Les étapes [EMD₂] à [EMD₄] sont présentées à l'itération q et sont à répéter jusqu'à convergence (jusqu'à ce qu'un critère d'arrêt soit vérifié).

[EMD₁] Initialisation des paramètres $\theta^{(0)}$.

[EMD₂] Calcul d'une réalisation X selon le paramètre courant q .

[EMD₃] Estimation des paramètres du modèle :

$$\pi_i^{(q+1)} = P\left(X_r = \omega_i \mid Y = y, \theta^{(q)}\right) \text{ pour tout } \omega_i \in \Omega$$

$$\text{et } \alpha^{(q+1)} = \frac{\sum_{c \in C} |S^c| g_c^{-1} \left(\frac{\sum_{s \in S^c} P(X_s^c = \omega_i \mid X_{s-1}^c = \omega_j, Y = y, \theta^{(q)})}{|S^n|} \right)}{|C| \sum_{c \in C} |S^c|}, \text{ les } g_c \text{ dépendant du modèle.}$$

[EMD₄] Estimation des paramètres du bruit :

$$\mu_i^{(q+1)} = \frac{\sum_{s \in S^0} y_s P\left(X_s^0 = \omega_i \mid Y = y, \theta^{(q)}\right)}{\sum_{s \in S^0} P\left(X_s^0 = \omega_i \mid Y = y, \theta^{(q)}\right)}$$

$$\text{et } (\sigma_i^2)^{(q+1)} = \frac{\sum_{s \in S^0} \left(y_s - \mu_i^{(q+1)}\right)^2 P\left(X_s^0 = \omega_i \mid Y = y, \theta^{(q)}\right)}{\sum_{s \in S^0} P\left(X_s^0 = \omega_i \mid Y = y, \theta^{(q)}\right)}.$$

18.3.2 Algorithmes ICE et MICE.

Nous mettons en place, sur ce même modèle deux autres procédures d'estimation des paramètres, *ICE* et *MICE*. Les paramètres du modèle sont calculés de façon déterministe, comme pour *EM*, grâce à [18.106]-[18.109], alors que les paramètres du bruit, à l'itération $(q+1)$, sont réestimés au moyen des fréquences empiriques,

$$\mu_i^{(q+1)} = \frac{\sum_{s \in S^0} y_s \mathbf{1}_{[(x_s^0)^q = \omega_i]}}{\sum_{s \in S^0} \mathbf{1}_{[(x_s^0)^q = \omega_i]}} \quad [18.112]$$

$$\text{et } (\sigma_i^2)^{(q+1)} = \frac{\sum_{s \in S^0} \left(y_s - \mu_i^{(q+1)}\right)^2 \mathbf{1}_{[(x_s^0)^q = \omega_i]}}{\sum_{s \in S^0} \mathbf{1}_{[(x_s^0)^q = \omega_i]}}, \quad [18.113]$$

calculées pour une réalisation du champ obtenue, par simulation selon la loi de X a posteriori, avec les paramètres estimés à l'itération q . Pour *MICE*, nous simulons uniquement une réalisation de X^0 , selon les lois marginales a posteriori. En cela, la version sur arbre, de cet algorithme, diffère légèrement de celle sur les chaînes.

Ces algorithmes se résument à :

ALGORITHMES ICE ET MICE SUR LE DIARBRE

Estimation des paramètres dans les modèles cachés.

Les étapes [DICE₁₂] à [DICE₁₄] sont présentées à l'itération q et sont à répéter jusqu'à convergence (jusqu'à ce qu'un critère d'arrêt soit vérifié).

[DICE₁₁] Initialisation des paramètres $\theta^{(0)}$.

[DICE₁₂] Calcul selon le paramètre courant $\theta^{(q)}$, d'une réalisation de X par la loi a posteriori, pour *ICE* et d'une réalisation de X^0 , selon les lois marginales a posteriori, pour *MICE*.

[DICE₁₃] Estimation des paramètres du modèle :

$$\pi_i^{(q+1)} = P\left(X_r = \omega_i \mid Y = y, \theta^{(q)}\right) \text{ pour tout } \omega_i \in \Omega$$

$$\text{et } \alpha^{(q+1)} = \frac{\sum_{c \in C} |S^c| g_c^{-1} \left(\frac{\sum_{s \in S^c} P(X_s^c = \omega_i \mid X_{s-1}^{c+1} = \omega_j, Y = y, \theta^{(q)})}{|S^n|} \right)}{|C| \sum_{c \in C} |S^c|}, \text{ les } g_c \text{ dépendant du modèle.}$$

[DICE₁₄] Estimation des paramètres du bruit:

$$\mu_i^{(q+1)} = \frac{\sum_{s \in S^0} y_s \mathbf{1}_{[(x_s^0)^q = \omega_i]}}{\sum_{s \in S^0} \mathbf{1}_{[(x_s^0)^q = \omega_i]}}$$

$$\text{et } (\sigma_i^2)^{(q+1)} = \frac{\sum_{s \in S^0} \left(y_s - \mu_i^{(q+1)} \right)^2 \mathbf{1}_{[(x_s^0)^q = \omega_i]}}{\sum_{s \in S^0} \mathbf{1}_{[(x_s^0)^q = \omega_i]}}.$$

18.3.3 Algorithme *SEM* sur le diarbre.

On ne calcule pas les probabilités a posteriori en chacun des noeuds, mais on simule une réalisation de l'arbre à partir de la probabilité de la racine et les probabilités interéchelles calculées lors de la passe montante. On considère, par conséquent, les estimateurs empiriques, basés sur les fréquences d'apparition des différents couples père-fils, dans la réalisation simulée de l'arbre, pour les paramètres du modèle, et les moyennes et variances empiriques, [18.112] et [18.113], pour les paramètres du bruit.

ALGORITHME SEM SUR LE DIARBRE

Les étapes [DSEM₂] à [DSEM₄] sont présentées à l'itération q et sont à répéter jusqu'à convergence (jusqu'à ce qu'un critère d'arrêt soit vérifié).

[DSEM₁] Initialisation des paramètres $\theta^{(0)}$.

[DSEM₂] Simulation d'une réalisation X a posteriori, suivant le paramètre courant $\theta^{(q)}$.

[DSEM₃] Estimation des paramètres du modèle :

$$\pi_i^{(q+1)} = P\left(X_r = \omega_i \mid Y = y, \theta^{(q)}\right) \text{ pour tout } \omega_i \in \Omega$$

Estimation des paramètres dans les modèles cachés.

$$\text{et } \alpha^{(q+1)} = \frac{\sum_{c \in C} |S^c| g_c^{-1} \left(\frac{\sum_{s \in S^c} 1_{[X_s^c = \omega_i]} \cdot 1_{[X_{s-}^{c+1} = \omega_j]}}{\sum_{s \in S^c} 1_{[X_{s-}^{c+1} = \omega_j]}} \right)}{|C| \sum_{c \in C} |S^c|}, \text{ les } g_c \text{ dépendant du modèle.}$$

[DSEM₄] Estimation des paramètres du bruit:

$$\mu_i^{(q+1)} = \frac{\sum_{s \in S^0} y_s 1_{[(x_s^0)^q = \omega_i]}}{\sum_{s \in S^0} 1_{[(x_s^0)^q = \omega_i]}} \text{ et } (\sigma_i^2)^{(q+1)} = \frac{\sum_{s \in S^0} (y_s - \mu_i^{(q+1)})^2 1_{[(x_s^0)^q = \omega_i]}}{\sum_{s \in S^0} 1_{[(x_s^0)^q = \omega_i]}}.$$

Estimation des paramètres dans les modèles cachés.

- CHAPITRE 19 -

Segmentation non supervisée par arbres de Markov cachés.

19.1 Choix de l'algorithme d'estimation des paramètres.

Nous sommes désormais en présence de quatre algorithmes d'estimation des paramètres, adaptés à deux modèles de diarbres, différenciables par la fonction définissant leurs probabilités de transition interéchelle. Nous comparons l'efficacité de ces huit algorithmes pour la segmentation de cinq images, choisies en fonction du degré de difficulté croissant qu'elles présentent. Nous utilisons le critère général d'homogénéité de Salzenstein (voir [Salzenstein96] et [Salzenstein98]), pour donner une mesure de la difficulté, a priori, pour les cinq images de classes FIG. III.47. Ce

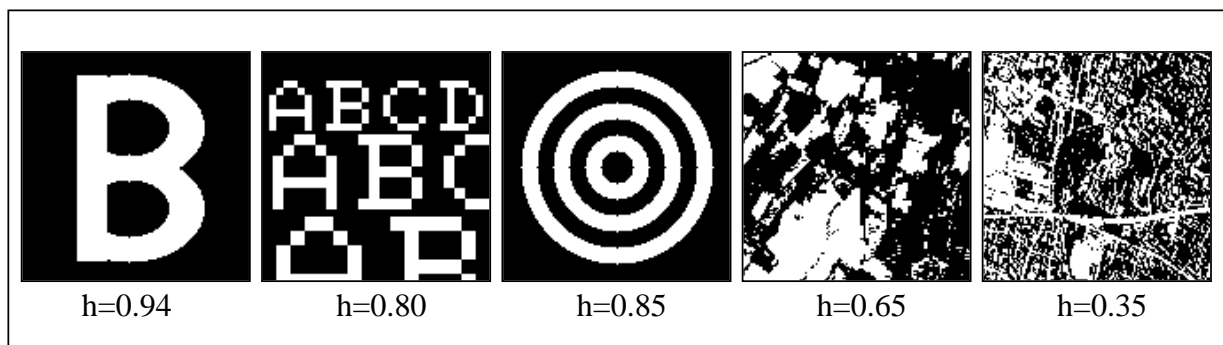


FIG. III.47. - CARACTÉRISATION DE L'HOMOGENÉITÉ DES IMAGES DE CLASSES.

sont des images à segmenter suivant deux classes, que l'on a bruitées avec deux distributions normales, $N(0, 1)$ et $N(1, 1)$. Notre objectif est de déterminer l'algorithme le mieux adapté à chacun des modèles d'arbres, et de préciser les domaines d'application des deux modèles. Pour cela, nous faisons tourner cent fois, les huit algorithmes sur chacune des cinq images, bruitées indépendamment, pour chacune des cent segmentations, avec les distributions normales ci-dessus.

Le tableau FIG. III.48 récapitule, pour chaque image et chaque algorithme, l'erreur moyenne sur les cent segmentations, ainsi que l'erreur minimale et l'erreur maximale commises au cours des cent expériences. Nous y donnons également, l'erreur globale commise par chaque algorithme, au cours des cinq cent segmentations.






Image	SEM type 1	ICE type1	MICE type1	EM type 1	SEM type 2	ICE type2	MICE type2	EM type 2	
	3.41%	3.51%	3.43%	3.51%	3.42%	3.40%	3.42%	3.43%	Erreur mini.
	4.29%	4.63%	4.66%	4.48%	4.72%	4.50%	4.46%	4.64%	Erreur maxi.
	3.91%	4.01%	3.98%	3.97%	4.01%	3.97%	3.95%	3.97%	Erreur moy.
	4.12%	4.23%	4.28%	4.03%	4.09%	4.02%	4.02%	4.12%	Erreur mini.
	6.41%	7.50%	6.45%	6.61%	6.52%	6.68%	7.84%	6.73%	Erreur maxi.
	5.04%	5.48%	5.32%	5.23%	5.16%	5.32%	5.22%	5.34%	Erreur moy.
	7.66%	8.14%	8.23%	7.84%	7.71%	7.60%	7.87%	7.67%	Erreur mini.
	9.88%	11.16%	11.08%	10.08%	9.81%	10.16%	10.35%	10.17%	Erreur maxi.
	8.86%	9.44%	9.41%	8.95%	8.75%	8.88%	8.88%	8.94%	Erreur moy.
	13.50%	14.87%	14.41%	13.83%	13.26%	13.78%	13.79%	13.89%	Erreur mini.
	15.39%	16.91%	16.82%	15.64%	15.05%	15.77%	15.80%	15.71%	Erreur maxi.
	14.51%	15.79%	15.62%	14.88%	14.15%	14.79%	14.68%	14.96%	Erreur moy.
	29.75%	32.23%	31.99%	30.37%	27.03%	29.87%	29.31%	31.15%	Erreur mini.
	31.91%	51.46%	47.42%	33.10%	32.24%	35.05%	34.01%	48.82%	Erreur maxi.
	30.66%	34.06%	33.88%	31.47%	28.46%	31.61%	31.45%	32.98%	Erreur moy.
	12.60%	13.76%	13.64%	12.90%	12.11%	12.91%	12.83%	13.24%	Erreur globale

FIG. III.48. - COMPARAISON DES MÉTHODES D'ESTIMATION.

Les algorithmes basés sur le *SEM* et les diarbres stochastiques donnent de meilleurs résultats que leurs concurrents, avec une différence significative pour le *SEM* de type 2, sur la dernière image.

Le tableau FIG. III.49, dans lequel la proportion de la classe 1 est notée p , donne les estimations des paramètres du bruit, pour le bruitage, de chaque image, ayant entraîné la plus grande erreur, en moyenne sur les huit algorithmes. Ces résultats sont significatifs de la qualité des algorithmes, en ce qu'ils illustrent les limites de ces algorithmes dans l'estimation des paramètres du bruit, sur les images très détaillées. En effet, la qualité de la classification baisse significativement, avec la baisse de la qualité de l'estimation de ces paramètres.

Image	SEM type 1	ICE type1	MICE type 1	EM type 1	SEM type 2	ICE type2	MICE type 2	EM type 2	
	-0.01	0.02	0.02	0.01	0.02	0.00	-0.01	0.01	Moyenne 1
	1.03	1.05	1.03	1.04	1.04	1.04	1.00	1.04	Variance 1
	0.69	0.70	0.71	0.70	0.71	0.70	0.71	0.70	p
	0.96	0.96	0.97	0.96	0.96	0.96	1.04	0.96	Moyenne 2
	1.04	1.06	1.03	1.06	1.06	1.06	0.98	1.06	Variance 2
	0.01	0.06	0.07	0.04	0.01	0.05	0.05	0.04	Moyenne 1
	1.03	1.06	1.06	1.04	1.03	1.05	1.05	1.05	Variance 1
	0.73	0.77	0.77	0.75	0.73	0.76	0.76	0.76	p
	1.02	1.02	1.04	1.02	1.03	1.02	1.03	1.03	Moyenne 2
	0.95	1.00	1.01	0.98	0.94	1.00	0.98	0.98	Variance 2
	0.02	0.05	0.04	0.03	0.03	0.03	0.03	0.03	Moyenne 1
	1.03	1.04	1.04	1.03	1.03	1.03	1.05	1.03	Variance 1
	0.65	0.67	0.68	0.66	0.67	0.67	0.69	0.67	p
	1.04	0.89	0.95	0.92	0.93	0.95	0.96	0.93	Moyenne 2
	1.10	1.10	1.04	1.06	1.05	1.03	1.04	1.05	Variance 2
	0.08	0.15	0.08	0.09	0.05	0.10	0.06	0.11	Moyenne 1
	1.12	1.15	1.13	1.12	1.10	1.13	1.16	1.13	Variance 1
	0.63	0.70	0.64	0.65	0.63	0.66	0.63	0.68	p
	0.91	0.95	0.97	0.93	0.96	0.96	0.99	0.96	Moyenne 2
	1.07	1.08	1.06	1.06	1.02	1.05	1.04	1.05	Variance 2
	0.24	0.37	0.00	0.26	0.03	0.31	-0.13	0.23	Moyenne 1
	1.20	1.25	1.11	1.21	1.09	1.22	1.06	1.21	Variance 1
	0.65	0.87	0.22	0.65	0.50	0.79	0.37	0.51	p
	0.78	0.85	0.53	0.75	0.83	0.90	0.68	0.64	Moyenne 2
	1.20	1.17	1.25	1.20	1.11	1.15	1.25	1.24	Variance 2

FIG. III.49. - RÉSULTATS DE L'ESTIMATION DES PARAMÈTRES.

Les conclusions que nous tirions de FIG. III.48, sont confirmées par une meilleure estimation des paramètres pour les *SEM*, la différence la plus marquée apparaissant pour le *SEM* de type 2, sur la dernière image.

Les compte rendus visuels des segmentations, présentés FIG. III.50, FIG. III.51, FIG. III.52, FIG. III.53 et FIG. III.54, sont représentatifs de l'erreur moyenne commise par les algorithmes de segmentation. Ils permettent une compréhension intuitive de la difficulté croissante de la segmentations des images présentées. L'image du "B" et celle de l'alphabet sont plus favorables au modèle que les deux cartes, alors que la cible pose le problème de la restitution, lors de la classification par arbres, des formes arrondies.

Conformément à ce que suggèrent les différents résultats que nous avons présentés, nous poursuivrons notre exposé en utilisant l'algorithme *SEM* sur le diarbre stochastique, pour l'estimation des paramètres.

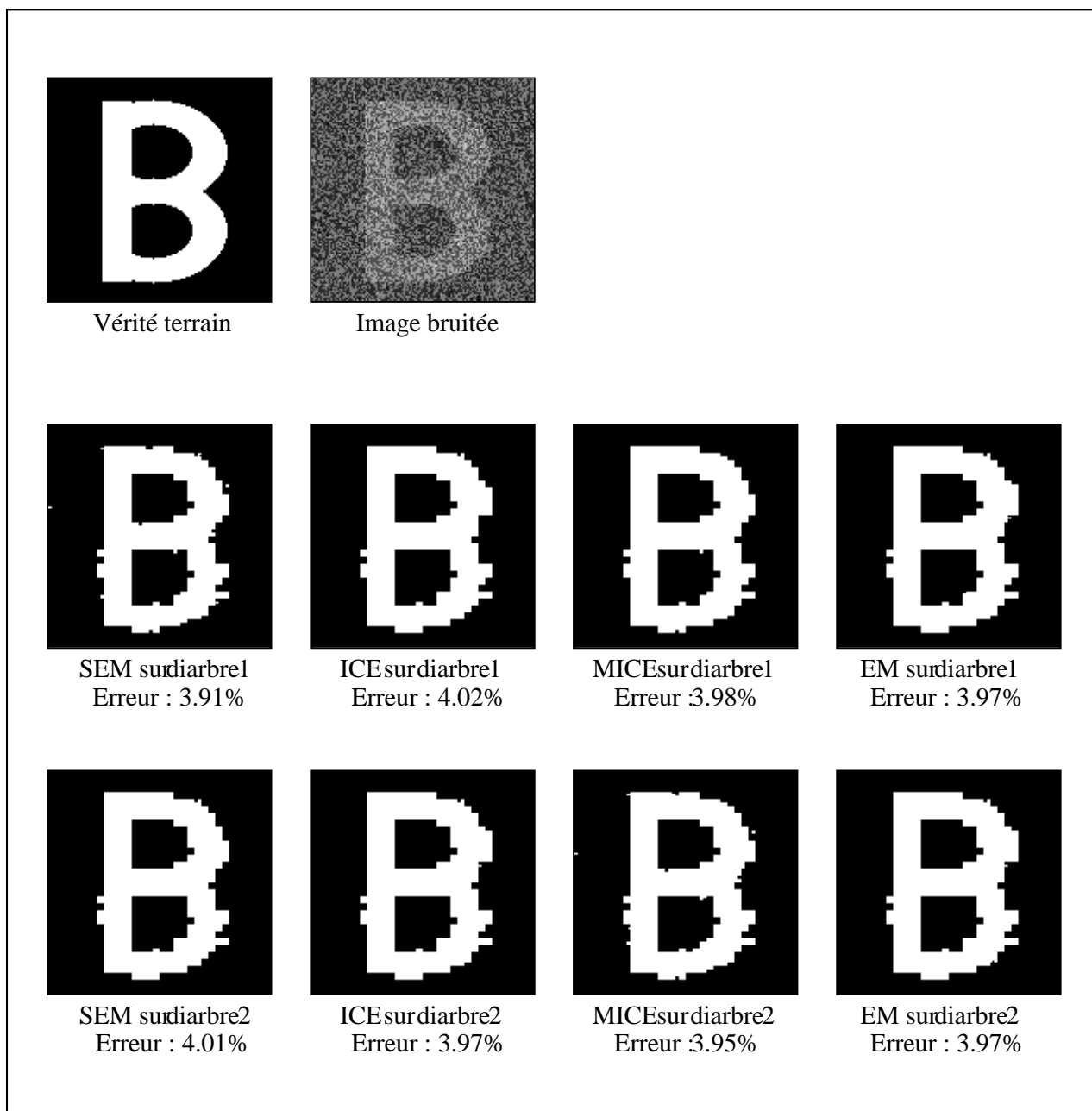


FIG. III.50. - SEGMENTATIONS COMPARÉES POUR UNE IMAGE D'HOMOGÉNÉITÉ $h = 0.94$.

Segmentation non supervisée par arbres de Markov cachés.

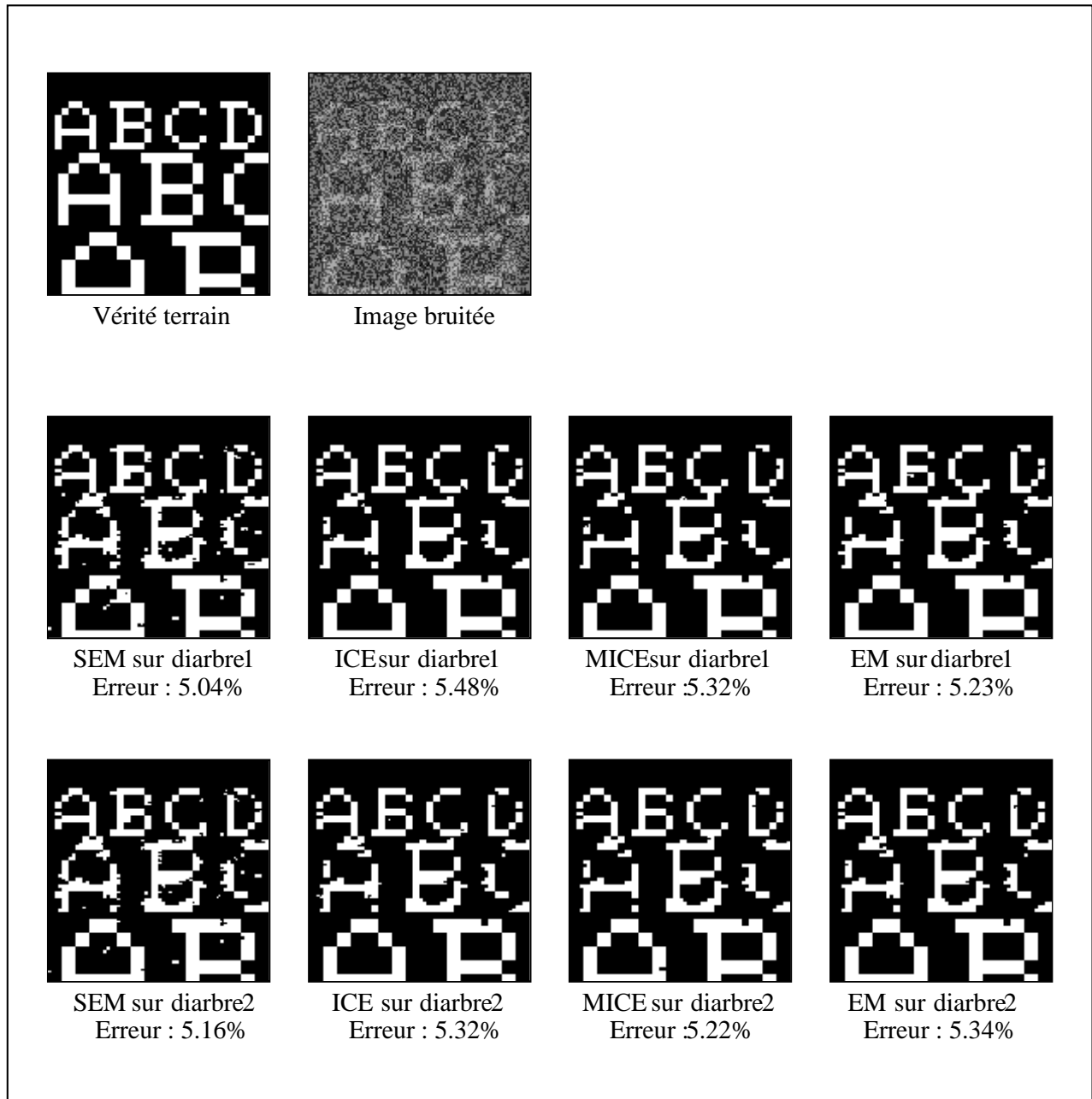


FIG. III.51. - SEGMENTATIONS COMPARÉES POUR UNE IMAGE D'HOMOGENÉITÉ $h = 0.80$.

Segmentation non supervisée par arbres de Markov cachés.

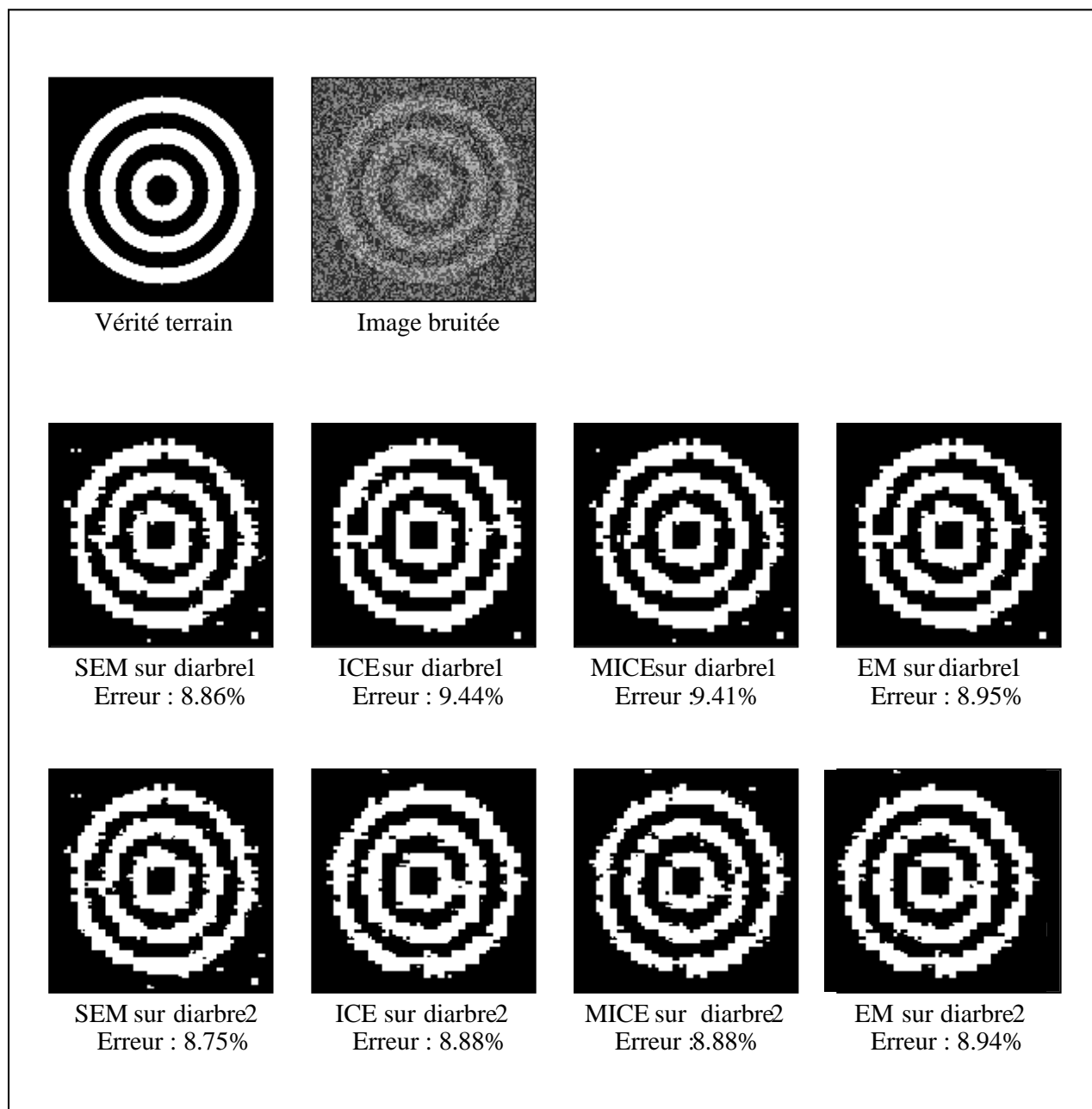


FIG. III.52. - SEGMENTATIONS COMPARÉES POUR UNE IMAGE D'HOMOGÉNÉITÉ $h = 0.85$.

Segmentation non supervisée par arbres de Markov cachés.

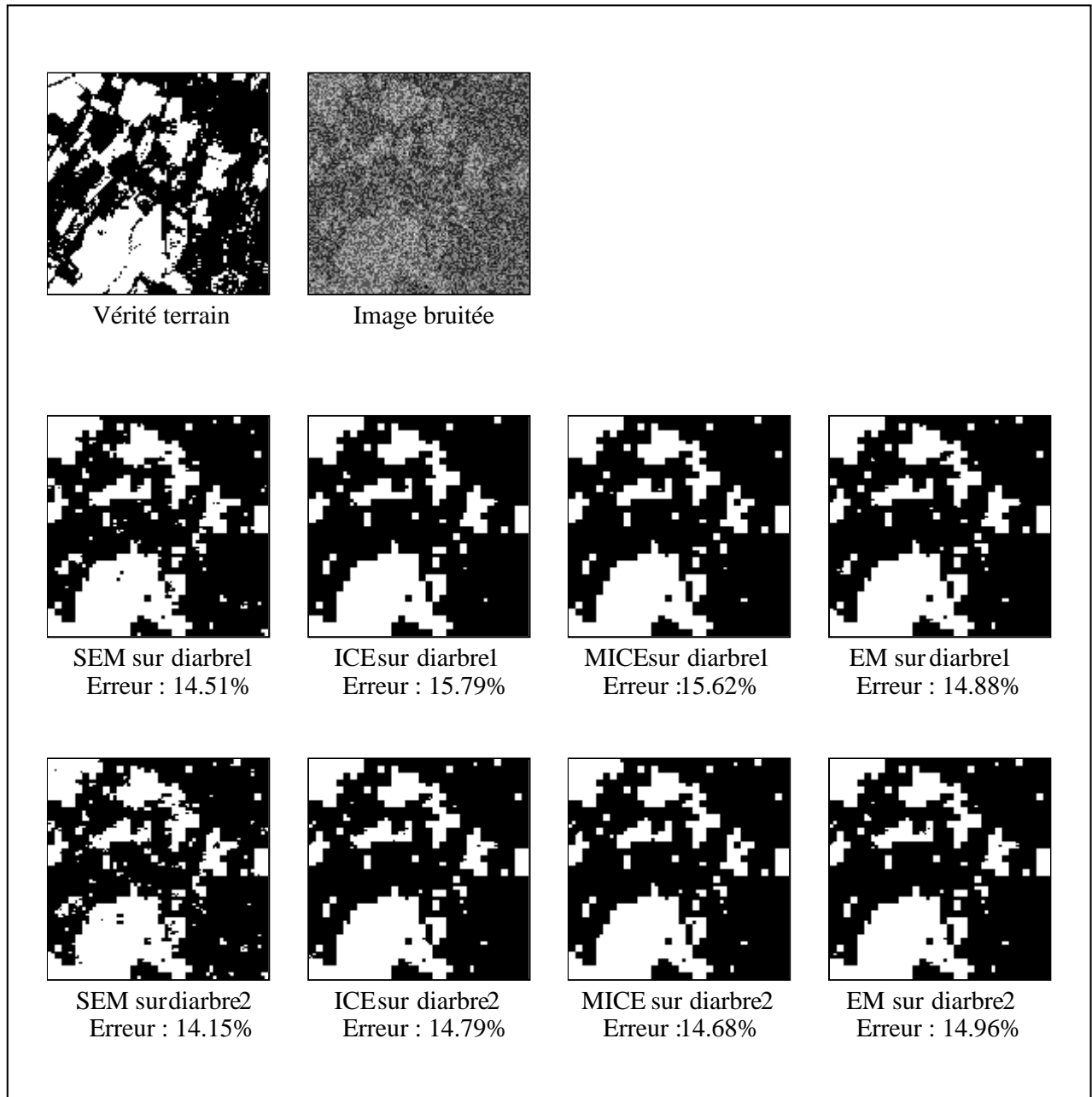


FIG. III.53. - SEGMENTATIONS COMPARÉES POUR UNE IMAGE D'HOMOGENÉITÉ $h = 0.65$.

Segmentation non supervisée par arbres de Markov cachés.

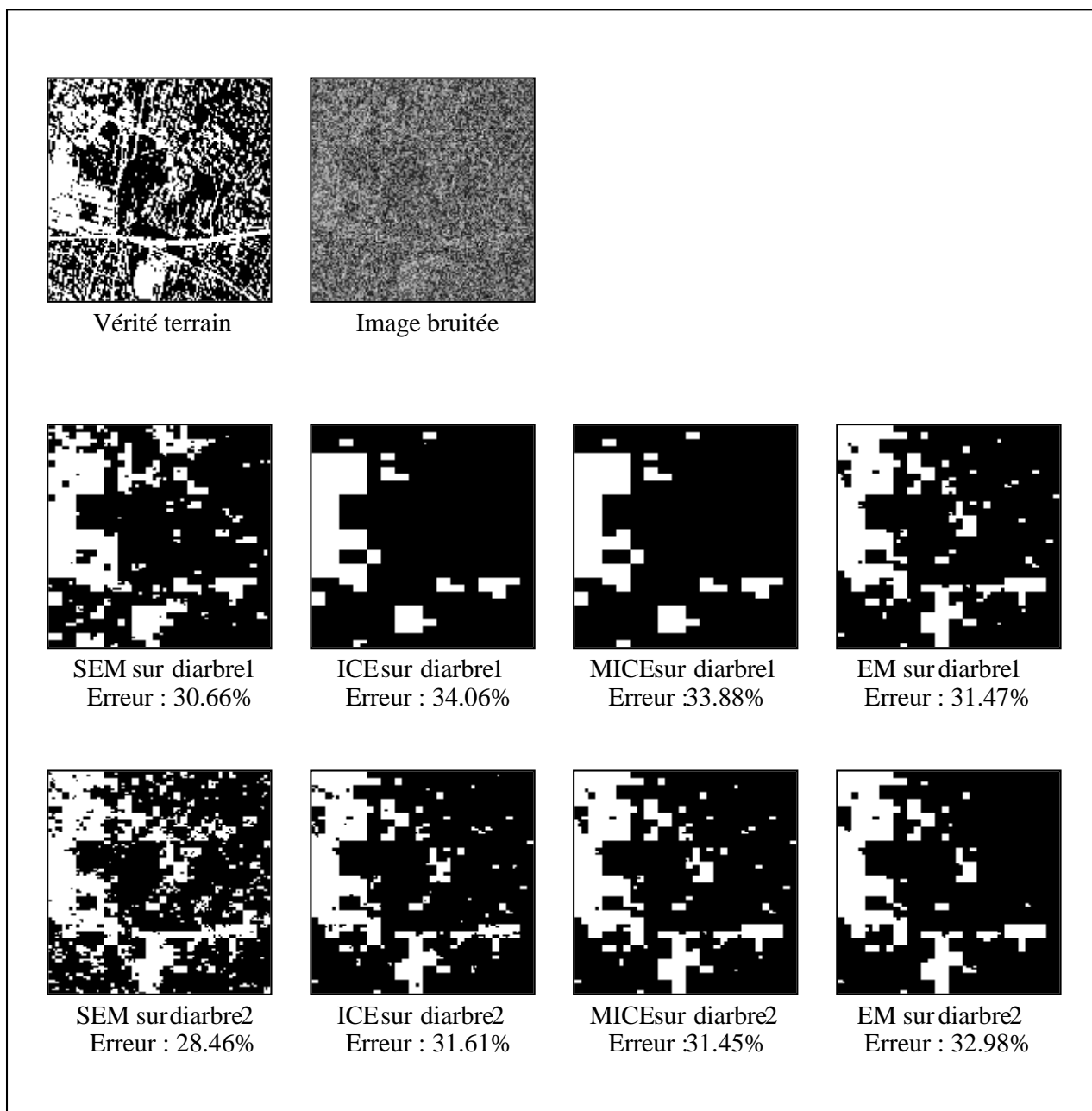


FIG. III.54. - SEGMENTATIONS COMPARÉES POUR UNE IMAGE D'HOMOGENÉITÉ $h = 0.35$.

Segmentation non supervisée par arbres de Markov cachés.

19.2 Comparaison des performances des modèles sur arbres de Markov et champs de Markov.

19.2.1 Images simulées par les différents modèles.

Dans cette partie nous exposons les résultats de segmentations des différents modèles markoviens sur des images quelconques. L'estimation des paramètres est faite par le gradient stochastique pour les champs de Markov et par *SEM* pour les arbres.

Nous commencerons par comparer les modèles markoviens, sur des images simulées par ces différents modèles. Le premier cas correspond à une vérité terrain simulée par un quadarbre et segmentée (après estimation des paramètres) par les champs de Markov cachés, les quadarbres et les diarbres de Markov cachés. Notons, que la modélisation par champs de Markov est effectuée dans le cadre le plus simple de paramétrisation de ce modèle.

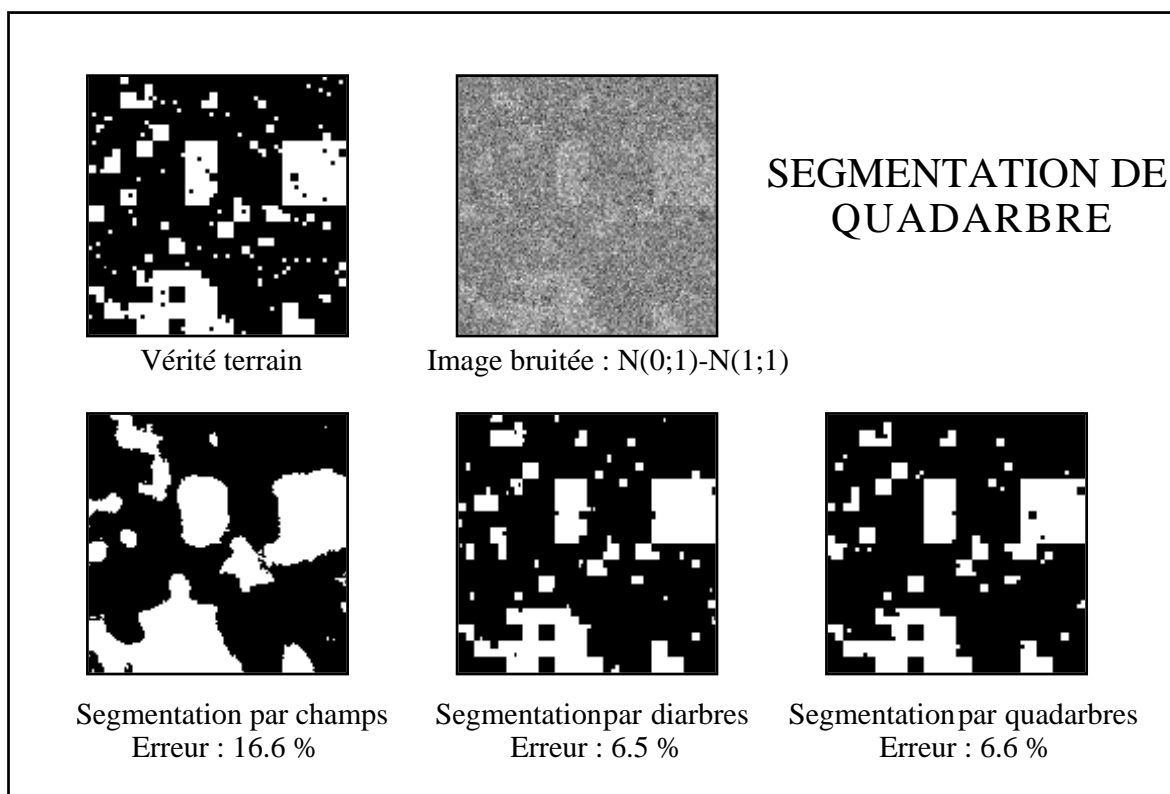


FIG. III.55. - SEGMENTATION DE QUADARBRES.

Segmentation non supervisée par arbres de Markov cachés.

La segmentation suivante est celle d'une image simulée d'après un diarbre droit. La différence dans la qualité de la segmentation entre les champs et les arbres peut être expliquée, d'une part, par la forme de la vérité terrain pour laquelle les champs ne sont pas particulièrement adaptés, ainsi que par la version non optimisée, de la modélisation par champs, que nous utilisons. La dernière étape présente la segmentation d'une image simulée par le modèle de

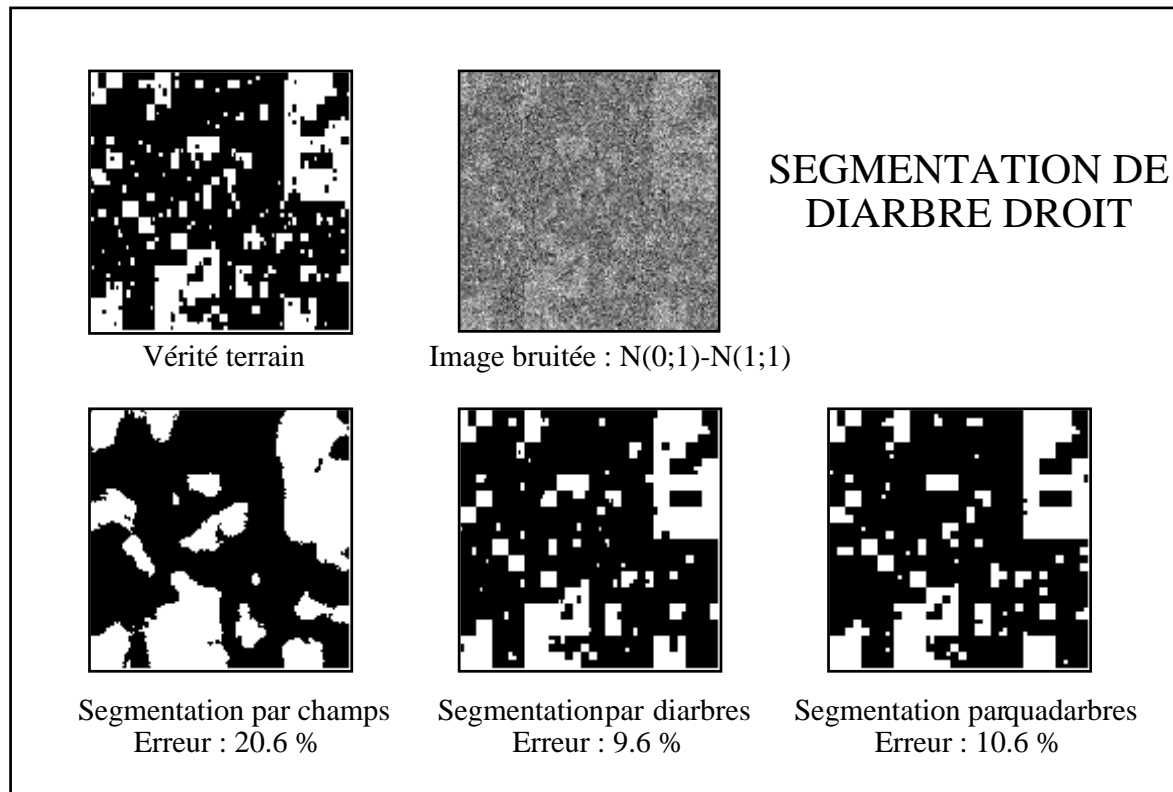


FIG. III.56. - SEGMENTATION DE DIARBRE DROIT.

champs de Markov que nous utilisons.

Les résultats équivalents des trois méthodes, dans un cas avantageant manifestement les champs, permet de justifier la compétitivité des méthodes par arbres de Markov cachés comme concurrentes des méthodes par champs de Markov.

Segmentation non supervisée par arbres de Markov cachés.

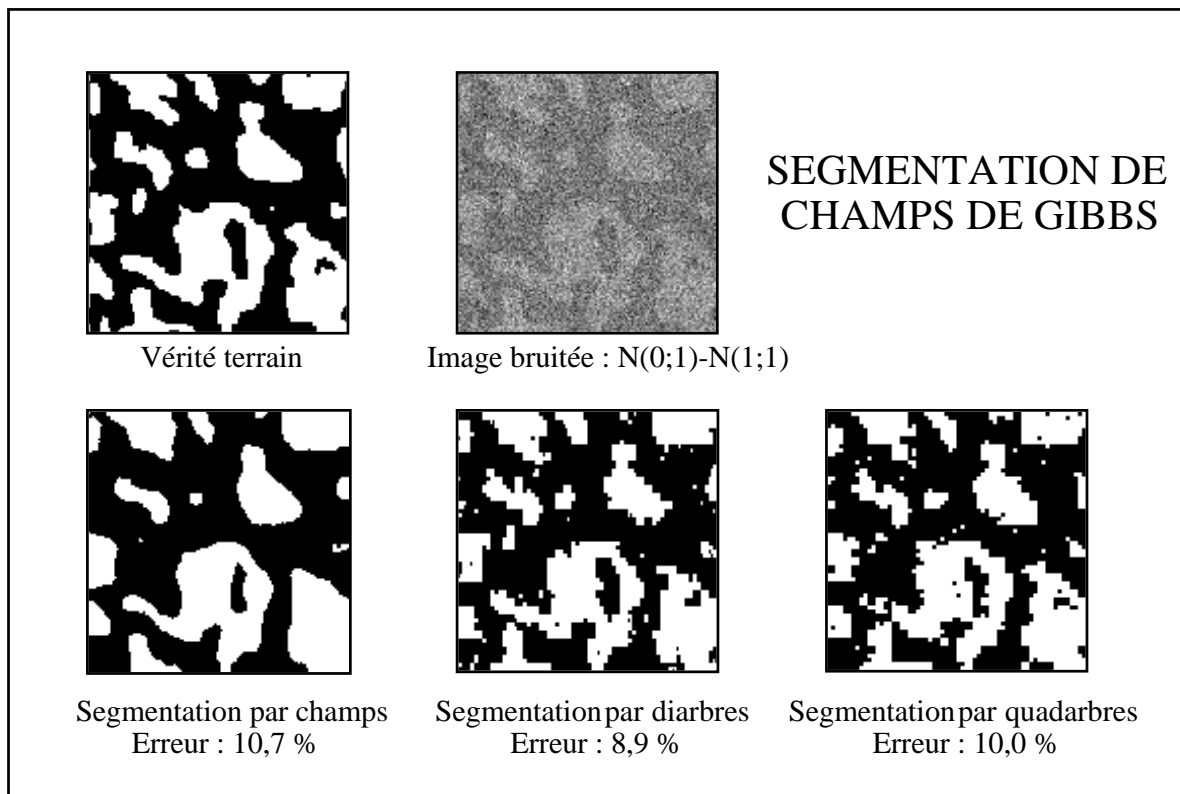


FIG. III.57. - SEGMENTATION DE CHAMPS DE GIBBS.

Segmentation non supervisée par arbres de Markov cachés.

19.2.2 Images de synthèse quelconques.

Nous passons désormais à des images de synthèse, ne privilégiant pas, a priori l'un ou l'autre des types de modèles markoviens. Les bruits différents apportent des éléments supplémentaires pour la comparaison des deux modèles. Nous introduisons, aussi, des bruitages exponentiels, mais nous ne travaillons pas sur des mélanges généralisés au sens propre du terme, puisque nous supposons connue la famille de chacune des distributions bruitantes.

Pour les figures (FIG. III.58, FIG. III.59, FIG. III.60, FIG. III.61) sur la première ligne, après l'image bruitée, nous présentons, dans l'ordre, les segmentations par champs (Ch), par quadarbre classique (Qc) et évolutif (Qé), et sur la seconde ligne celles par les diarbres droits (DD), alternés (Da) et diagonaux (Dd). La méthode d'estimation des paramètres sera, d'après les constatations faites lors des segmentations comparées, *SEM* de type 2 pour les deux premières images, et *SEM* de type 1 pour les deux dernières. Signalons, cependant, que les deux méthodes donnent des résultats équivalents sur ces exemples. La carte, FIG. III.58, est bruitée plutôt légèrement, mais la

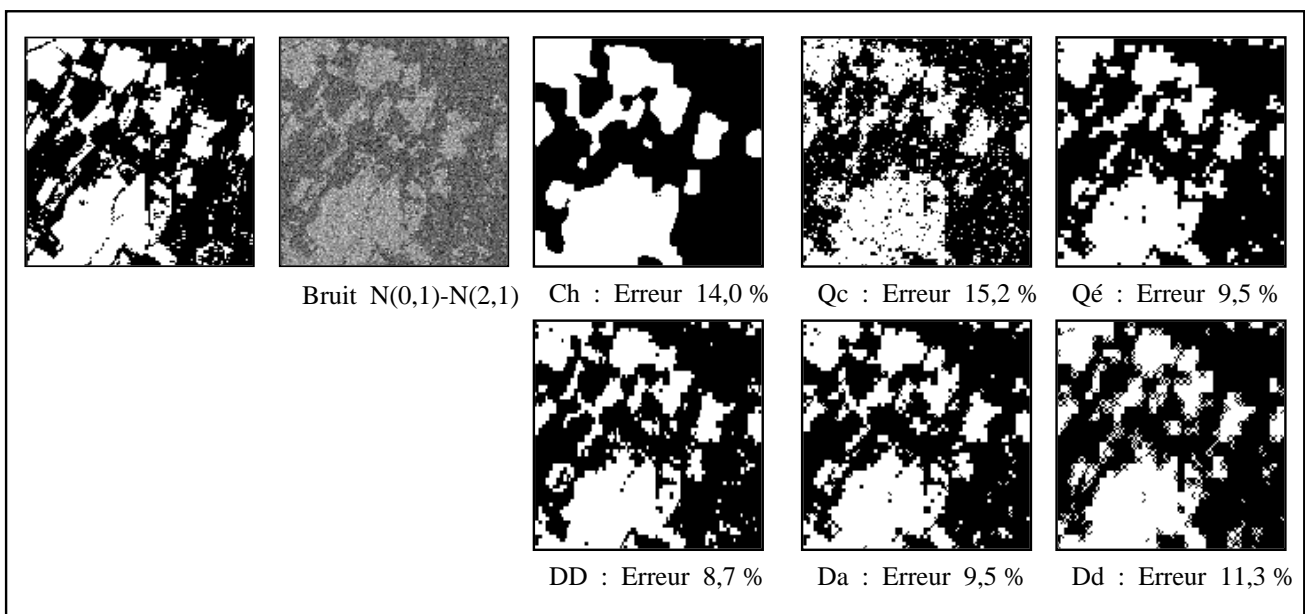


FIG. III.58. - SEGMENTATION POUR UN BRUITAGE GAUSSIEN LÉGER.

difficulté vient, une fois encore, de la finesse des détails de ce type d'images. Le quadarbre évolutif et le diarbre alterné donnent des taux d'erreur sous les 10 %, sensiblement moins bons que celui obtenu pour le diarbre droit. Bruitée avec des lois exponentielles, la cible, FIG. III.59, est difficile à segmenter par les algorithmes sur arbres qui donnent de meilleurs résultats sur des images contenant des zones rectangulaires. La différence, pour le bruitage de FIG. III.60, se situe au niveau des variances. Les résultats sont équivalents à ceux obtenus pour la carte.

Les lois qui ont servis au bruitage du "B", FIG. III.61, ne sont pas différenciables par leurs deux premiers moments. Les résultats de la segmentation sont très bons pour tous les types de modèles, ce qui est dû, en partie, comme nous le constatons plus haut, à l'image utilisée.

On peut conclure des différentes segmentations présentées dans cette partie que les arbres de Markov donnent des résultats légèrement meilleurs que les champs de Markov les plus simples. Parmi les différents types d'arbres, le diarbre droit semble un peu meilleur que les diarbres alternés et les quadarbres évolutifs, dans la mesure où il donne, avec un écart plus ou moins grand par rapports aux autres types, toujours un meilleur résultat.

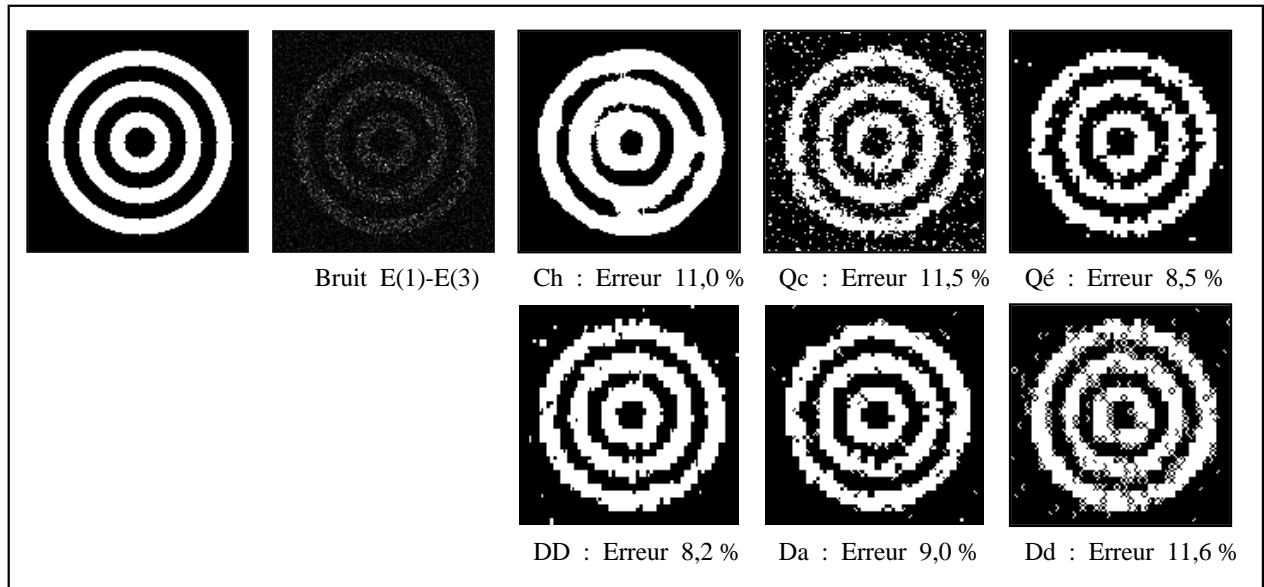


FIG. III.59. - SEGMENTATION POUR UN BRUITAGE EXPONENTIEL.

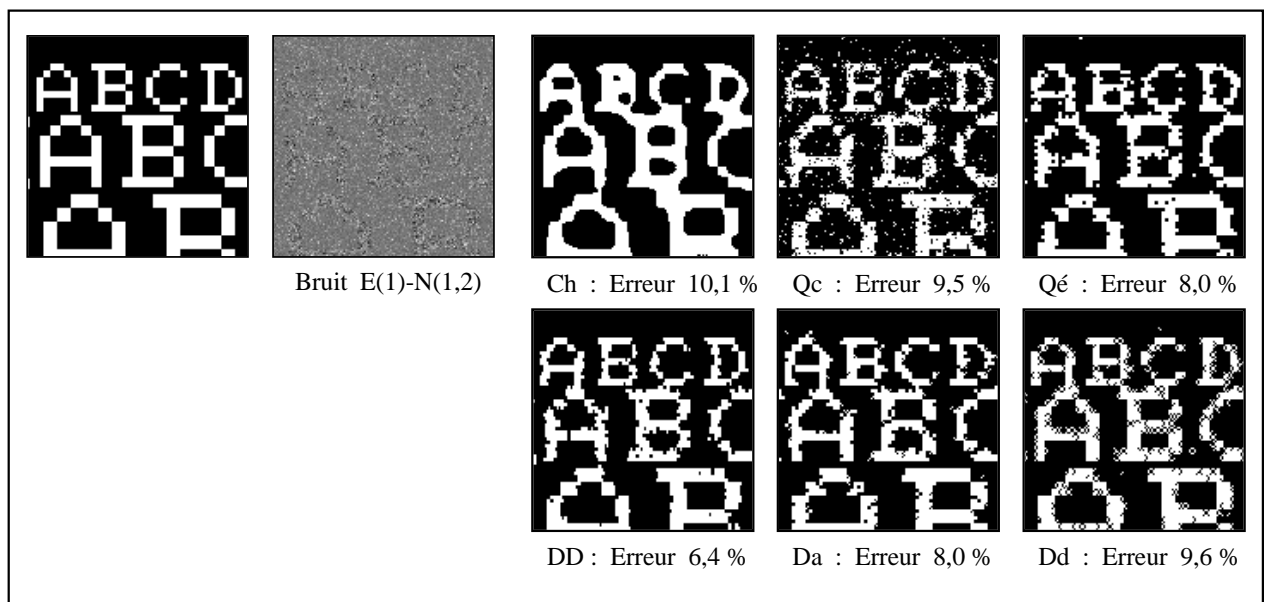


FIG. III.60. - SEGMENTATION POUR UN BRUITAGE OÙ LES MOYENNES SONT ÉGALES.

Segmentation non supervisée par arbres de Markov cachés.

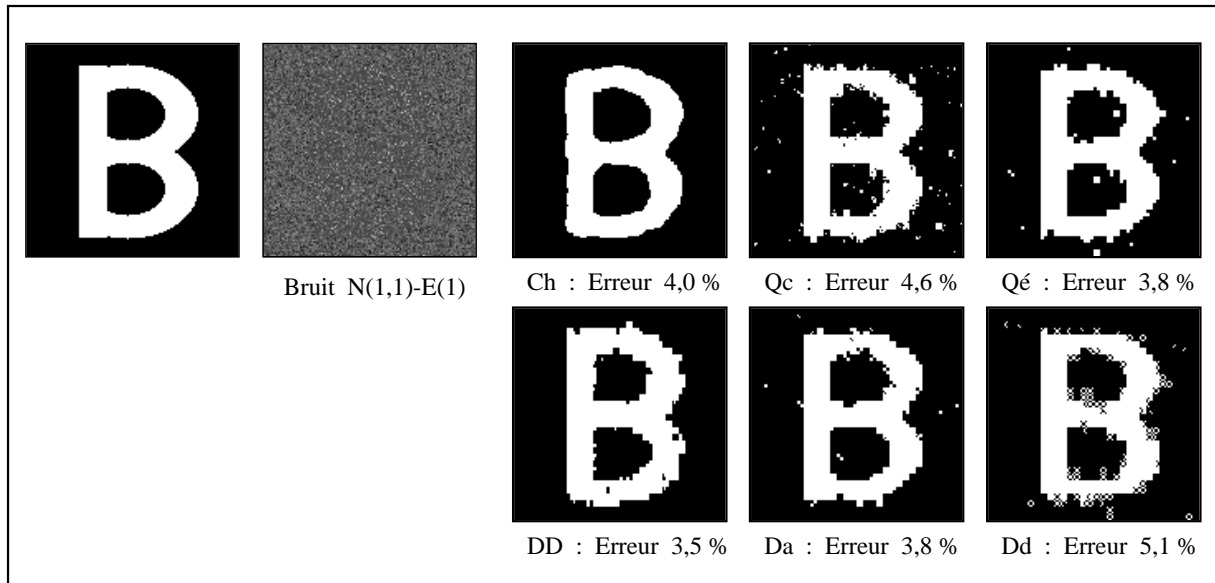


FIG. III.61. - SEGMENTATION POUR UN BRUITAGE OÙ LES MOYENNES ET VARIANCES SONT ÉGALES.

19.3 Limites du modèle par arbres.

Comme le modèle par champs de Markov que nous utilisons n'est pas optimisé, et que nous chercherons pas à l'améliorer ici, nous ne testerons pas les limites du modèle par champs de Markov cachés. Nous n'étudierons, donc, dans cette partie que les différents modèles d'arbres, à savoir les diarbres droits (DD) et alternés (Da) ainsi que le quadarbre évolutif (Qé).

Des bruitages très forts sur une vérité terrain favorable, permettent de tester la valeur des algorithmes d'estimation des paramètres du bruit, sur les arbres de Markov cachés. Comme précédemment, et en vertu des résultats des segmentations comparées, nous utiliserons une estimation des paramètres par le *SEM* de type 1. Les différents bruitages appliqués à la vérité terrain sont précisés dans chacun des cas présentés FIG. III.62.

Il ressort de l'étude de ces cas extrêmes que l'estimation des paramètres reste bonne, même pour de tels bruitages, mais la segmentation devient plus difficile. Il est intéressant de constater que l'on voit réapparaître le problème, que nous avons, en partie réussi à contrôler jusque là, des effets de bloc. D'une façon un peu surprenante, c'est probablement en conséquence de cet effet indésirable, que les méthodes sur le quadarbre donnent, pour cette image en particulier, d'aussi bons résultats.

A travers ces exemples, nous mettons en évidence la stabilité des algorithmes d'estimation et de classification sur les arbres. Alors que leur structure semble, intuitivement, moins adapté à la modélisation des images que les champs de Markov, les arbres de Markov permettent la mise au point d'algorithmes de segmentation statistique d'images très résistants. Nous avons, en effet, pu constater, les performances constantes de ces algorithmes, aussi bien pour des images peu homogènes, que pour d'autres, bruitées exagérément. Il apparaît, en conclusion, à cette étude sur la segmentation non supervisée d'images de synthèse, destinée à étalonner les algorithmes de segmentation d'images sur les arbres de Markov, que ces méthodes sont très stables, même lorsqu'on tend vers les cas limites, et aussi fiables, dans les cas moins complexes, que les méthodes de segmentations sur les modèles de Markov classiques.

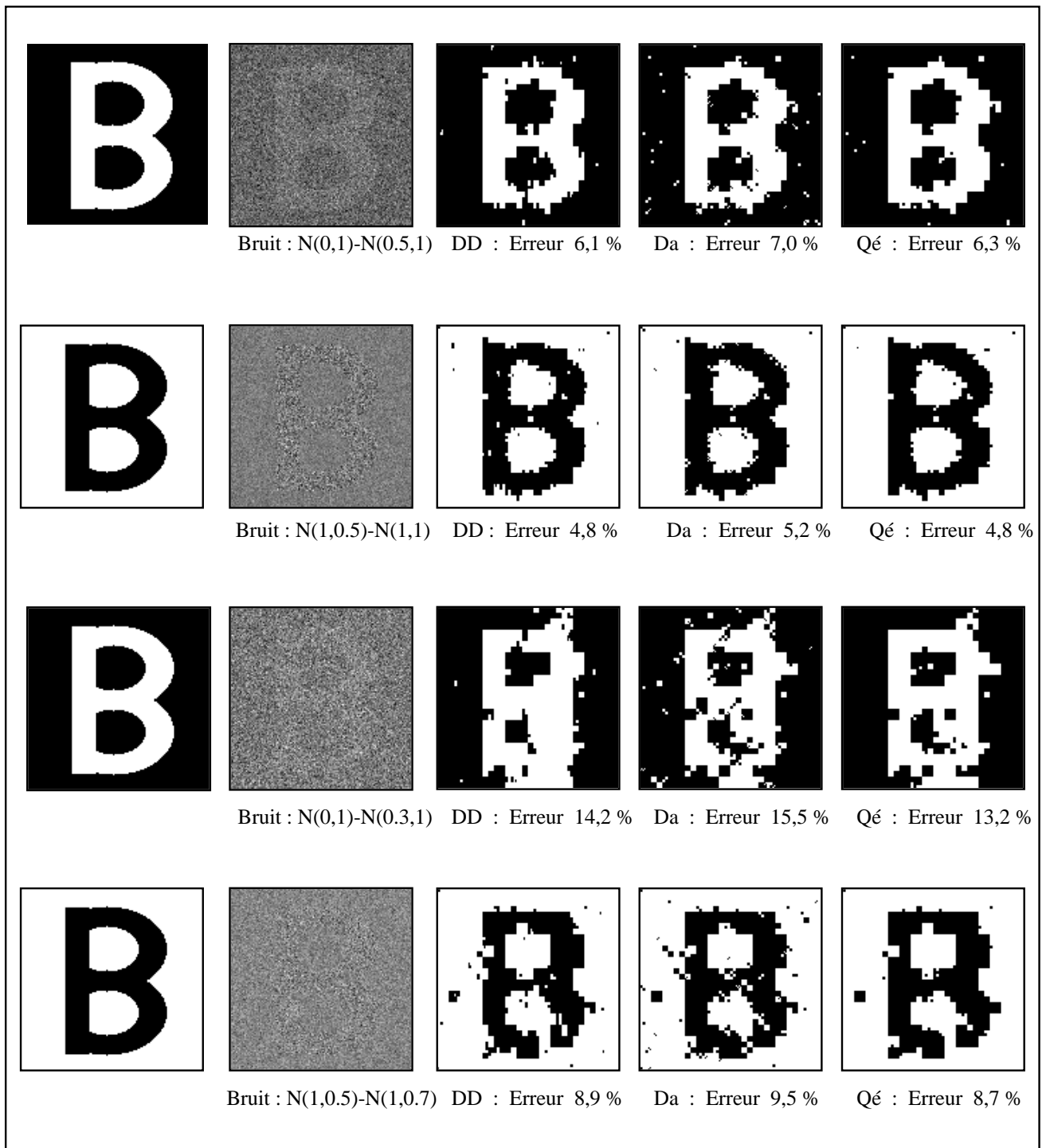


FIG. III.62. - SEGMENTATIONS DANS DES CAS DE BRUITAGES EXTRÊMES.

Segmentation non supervisée par arbres de Markov cachés.

19.4 Images réelles.

Ce paragraphe, est destiné à illustrer les possibilités des algorithmes de segmentation statistique d'images sur les diarbres droits. Nous traitons ici plusieurs cas d'images réelles, dont nous menons les segmentations pour un grand nombre de classes. Les distributions du bruit seront supposées être des distributions gaussiennes. Nous entrons, en particulier, dans un domaine d'estimation des paramètres de mélanges, que nous n'avons pas su traiter dans la partie théorique de ce travail. La bonne qualité des estimations, qui se traduit par celle des classifications, montre l'efficacité de ces méthodes itératives, pour l'estimation des paramètres de certains types de mélanges finis.

L'image spot FIG. III.63, a été segmentée, selon cinq classes, aux deux formats 128×128 et 64×64 , ce dernier format dégradant sensiblement la qualité des résultats. Compte tenu de l'importance du bruit sur l'image originale, les différentes segmentations semblent très satisfaisantes.

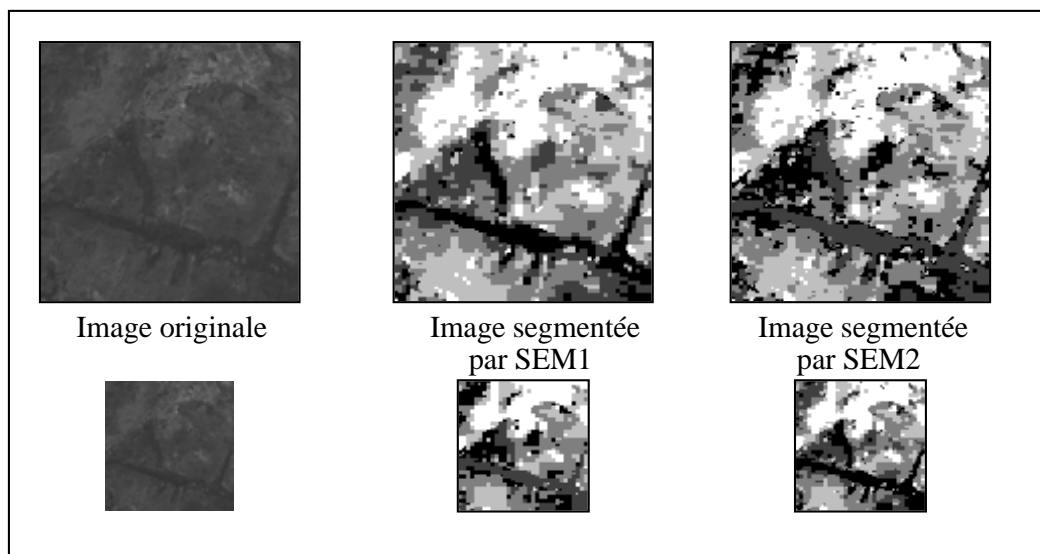


FIG. III.63. - SEGMENTATION D'UNE IMAGE SPOT AUX FORMATS 128×128 ET 64×64 .

Le second cas que nous traitons est celui d'une photographie aérienne que nous segmentons successivement en deux, trois, quatre puis cinq classes. Les résultats des segmentations présentées FIG. III.64, le sont pour le diarbre droit, diarbre alterné ainsi que pour le quadarbre évolutif. Les résultats pour les trois méthodes sont équivalents, et semblent être de bonne qualité.

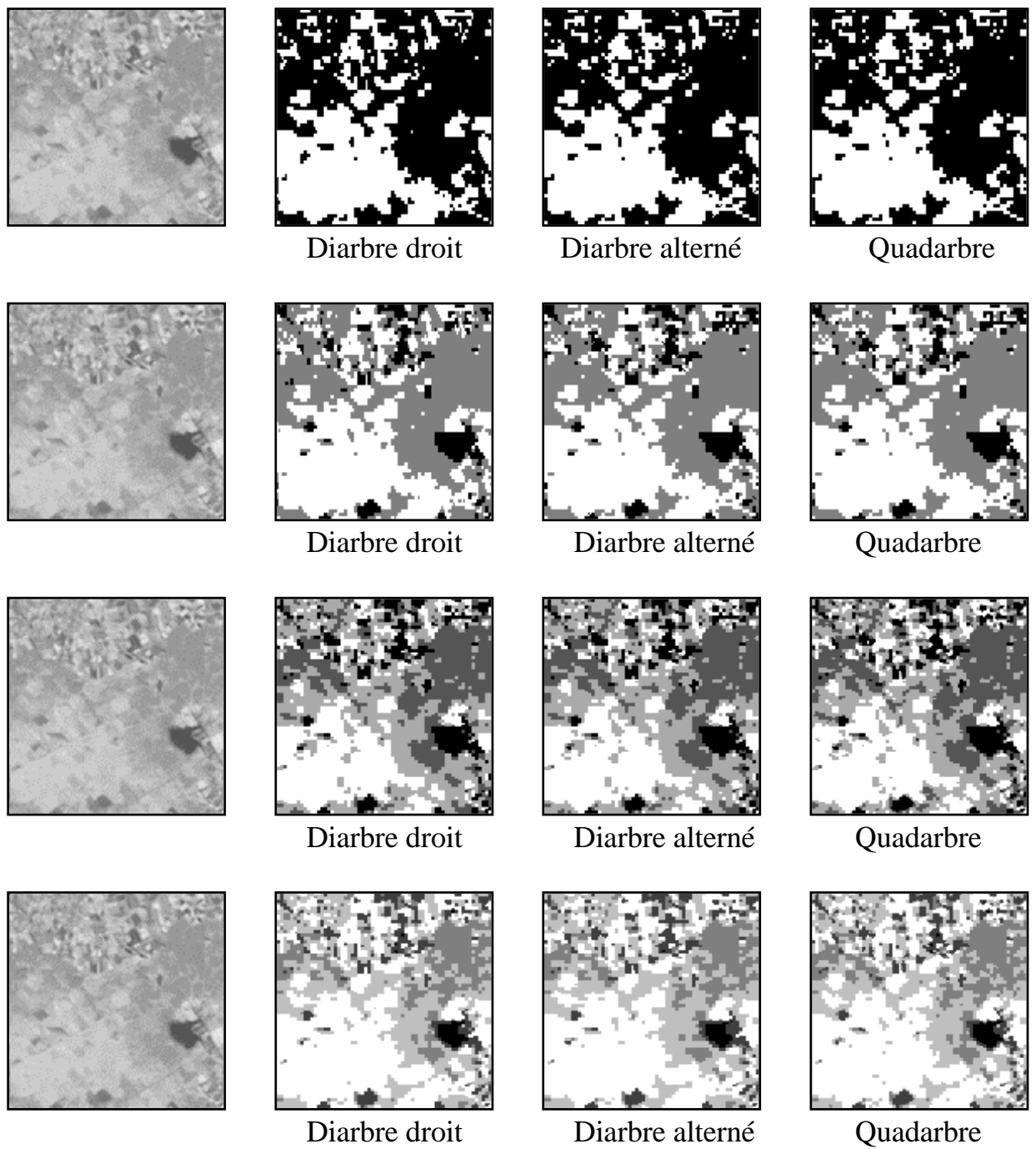


FIG. III.64. - SEGMENTATIONS, PAR ARBRES, D'UNE PHOTOGRAPHIE AÉRIENNE.

Segmentation non supervisée par arbres de Markov cachés.

La segmentation de l'image FIG. III.65 et FIG. III.66, image radar de la forêt amazonienne, est présentée pour les deux formats d'images, 256×256 et 128×128 et effectuée suivant cinq classes. Nous disposons, pour cette image de la vérité terrain, qui nous, permet, par conséquent, de nous faire une idée plus précise de la qualité de la segmentation. Les deux résolutions auxquelles sont effectuées ces segmentations, donnent une idée de la faible perte d'information pour une telle diminution de la résolution. Notons que l'estimation par *SEM2*, à la résolution 128×128 , induit, pour la classification, un résultat plus proche de la vérité terrain que celui issu de *SEM1*.

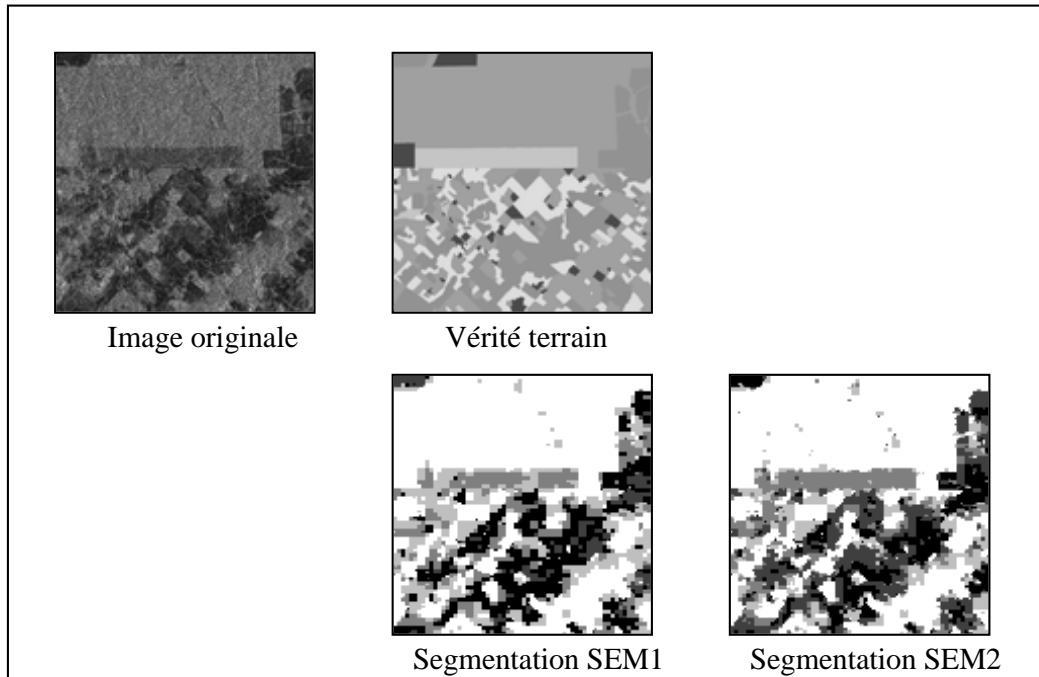


FIG. III.65. - SEGMENTATION D'UNE IMAGE RADAR AU FORMAT 128×128 .

Les résultats obtenus sur ces quelques images réelles attestent, une fois encore, de la grande souplesse de ce modèle, tout à fait adapté aux contraintes de la segmentation d'images réelles.

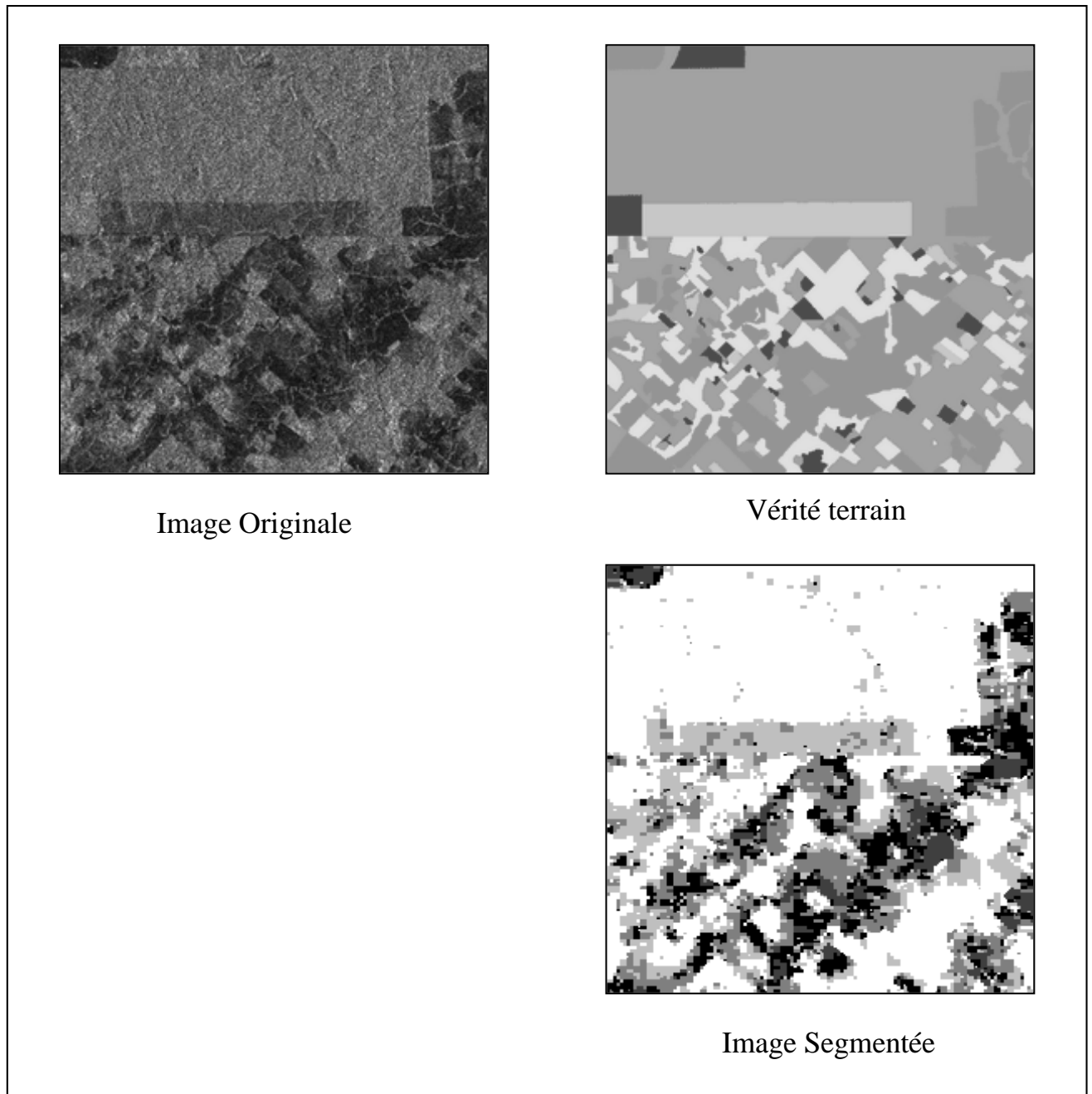


FIG. III.66. - SEGMENTATION D'UNE IMAGE RADAR AU FORMAT 256×256 .

Segmentation non supervisée par arbres de Markov cachés.

- CHAPITRE 20 -

Le modèle des mélanges généralisés.

20.1 Présentation de la méthode.

Justifiée par les estimations des paramètres sur des images réelles issues d'IRM, de satellites ou de radars, [Maffet91], la modélisation du bruit par des mélanges généralisés a été introduite récemment. Le principe de cette modélisation est très proche du cas classique des bruits gaussiens. Il s'agit de considérer, pour chaque variable, des distributions modélisant le bruit, qui ne sont pas uniquement caractérisées par leurs paramètres, mais par un ensemble (type de la distribution)-(paramètres associés). Sans que l'on puisse le justifier théoriquement, la procédure consistant à optimiser cet ensemble à chaque itération d'un algorithme classique d'estimation, est convergente, et permet une estimation assez fiable, pour les mélanges généralisés (voir [Bendjebour00], [Delignon97] et [Fouque99]).

La rapidité des calculs dans le cadre des arbres de Markov cachés permet une autre approche des modèles généralisés. Il devient, en effet, possible de faire une estimation complète des paramètres pour chaque cas de mélange de lois autorisé par les familles de distributions considérées. Cette méthode permet de se concentrer sur l'estimation des paramètres puisque les distributions sont fixées. Il faut, cependant, une fois les différentes estimations effectuées, choisir parmi l'ensemble de solutions, celle qui décrit le mieux l'échantillon étudié. On peut, par exemple, choisir la solution qui minimisera la distance entre la distribution théorique calculée avec les paramètres estimés, et la distribution de l'échantillon, approchée par un histogramme, ou par un estimateur à noyau de la distributions de mélange. On pourra consulter à ce propos la comparaison des différentes méthodes présentée dans [Bendjebour00], ainsi que [Giordana96], [Giordana97], [Delignon97] et [Fouque99]. Ces méthodes de choix sont celles utilisées habituellement dans les algorithmes d'estimation des paramètres des mélanges généralisés. Nous proposons une innovation en la matière, puisque, nous choisissons de comparer les moments empiriques calculés à partir de l'image bruitée, avec ceux issus "théoriquement" des distributions estimées dans la première partie de l'algorithme. Le meilleur compromis semble, après avoir testé différentes combinaisons, de faire cette comparaison sur la somme des quatre premiers moments.

20.2 Etalonnage de la méthode proposée.

Si nous reprenons les exemples que nous avons traités plus haut, en ne préjugant pas, cette fois, des distributions bruitantes utilisées, nous pouvons éprouver cette nouvelle méthode d'estimation pour les mélanges généralisés. Compte tenu de la méthode d'estimation des mélanges généralisés que nous proposons, effectuant des estimations de paramètres complètes pour les différents types de mélanges, le problème du choix final du type estimé est crucial. Il est intéressant de constater que la méthode de test effectue le bon choix du type de mélange dans les quatre cas présentés

La différence entre les deux algorithmes *SEM1* et *SEM2* n'est pas significative, et la solution la plus raisonnable, pour faire un choix entre les images classifiées à partir des paramètres issus de ces deux algorithmes d'estimation, est de prendre celle pour laquelle l'ensemble des paramètres estimés minimise le test d'adéquation avec les moments empiriques. Les différents résultats de segmentation effectués dans un cadre de mélanges généralisés sont référencés figures FIG. III.68 à FIG. III.71. Les segmentations présentées sur la première ligne sont obtenues par les algorithmes développés sur le diarbre 1, celles de la seconde ligne viennent du diarbre 2. Le graphique du mélange estimé qui accompagne les images segmentées est celui du mélange, estimé par l'algorithme *SEM1* ou , minimisant le test d'adéquation avec les moments empiriques.

Dans le tableau de la figure FIG. III.67, nous présentons les résultats des estimations des paramètres, pour les quatre images tests, par les méthodes généralisée issues de *SEM* sur les arbres de type 1 et 2, ainsi que celles obtenues par la méthode des moments généralisée énoncée dans la deuxième partie. Le fait que nous travaillions à une résolution d'image de 128×128 n'est pas fait pour favoriser la méthode des moments, qui échoue à choisir le bon type de mélange dans le cas du mélange exponentielle-exponentielle. Ce résultat était prévisible, compte tenu des remarques que nous avons faites, dans le cadre de ce type de mélange, dans la deuxième partie. Les algorithmes itératifs d'estimation des paramètres donnent d'excellentes estimations dans les quatre cas présentés ici, et comme nous l'avons déjà signalé, semblent fiables au niveau du choix du type du mélange. A moins que nous ne disposions d'une image de grande résolution, typiquement supérieure à 512×512 , la méthode des moments ne peut être raisonnablement utilisée.

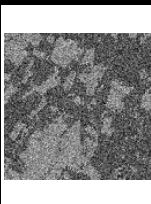
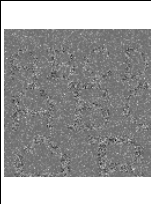
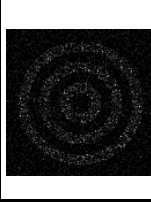
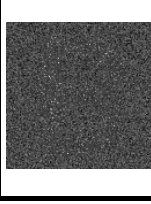
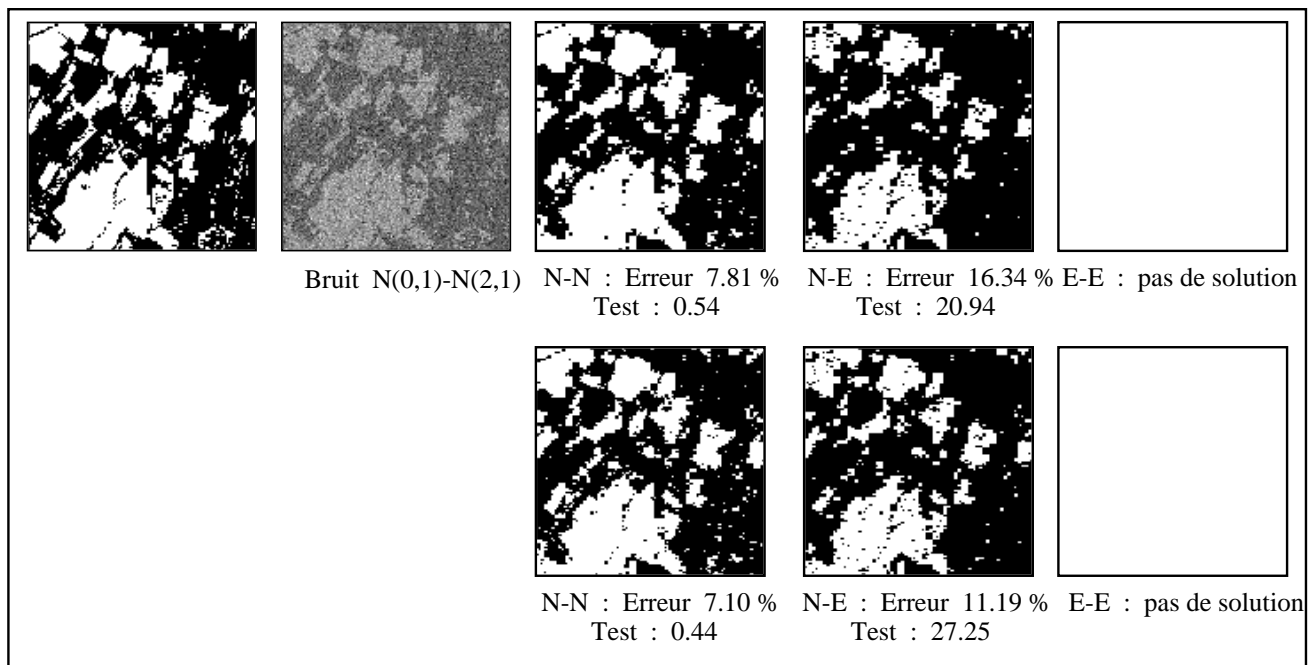
Image	Type et paramètres théoriques	SEM type 1	SEM type 2	Méthode des moments	
	N-N	N-N	N-N	N-N	Type
	0.0	0.05	0.01	-0.04	Moyenne 1
	1.0	1.13	1.05	0.97	Variance 1
	0.60	0.61	0.61	0.60	p
	2.0	1.22	1.97	2.00	Moyenne 2
	1.0	1.04	1.14	1.04	Variance 2
	N-E	N-E	N-E	N-E	Type
	1.0	0.99	1.00	1.01	Moyenne 1
	2.0	1.99	2.02	1.93	Variance 1
	0.28	0.27	0.28	0.30	p
	1.0	0.99	0.98	0.98	Moyenne 2
	/	/	/	/	Variance 2
	E-E	E-E	E-E	N-E	Type
	1.0	1.01	1.03	0.91	Moyenne 1
	/	/	/	0.58	Variance 1
	0.68	0.64	0.67	0.62	p
	3.0	2.83	2.84	2.86	Moyenne 2
	/	/	/	/	Variance 2
	N-E	N-E	N-E	N-E	Type
	1.0	1.00	1.00	0.99	Moyenne 1
	1.0	1.03	1.03	1.00	Variance 1
	0.70	0.70	0.70	0.74	p
	1.0	1.01	1.02	1.06	Moyenne 2
	/	/	/	/	Variance 2

FIG. III.67. - ESTIMATION DES PARAMÈTRES DANS LE CADRE DES MÉLANGES GÉNÉRALISÉS.

Le modèle des mélanges généralisés.



Graphique du mélange estimé par SEM2 type N-N.

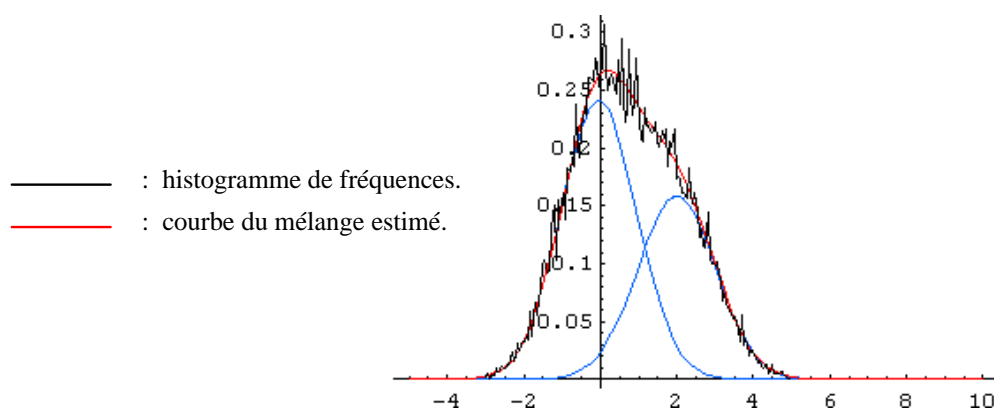
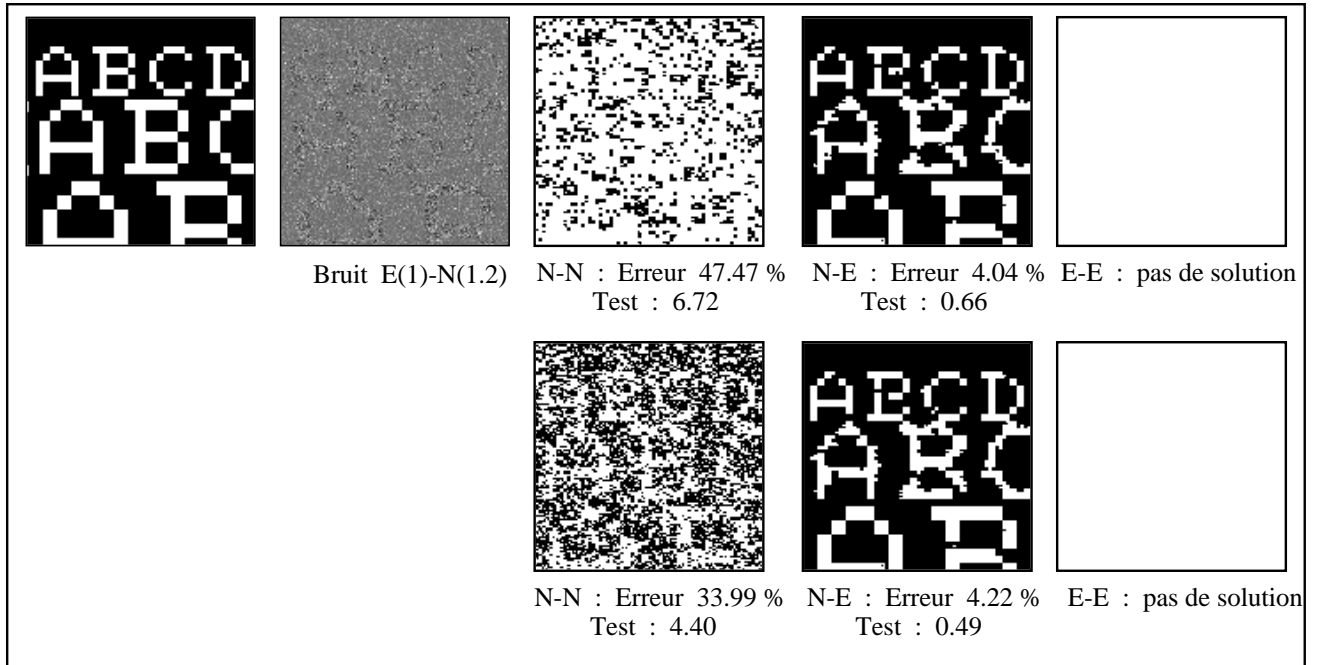


FIG. III.68. - SEGMENTATION D'IMAGE SUR LE MODÈLE DES MÉLANGES GÉNÉRALISÉS.

Le modèle des mélanges généralisés.



Graphique du mélange estimé par SEM2 type N-E.

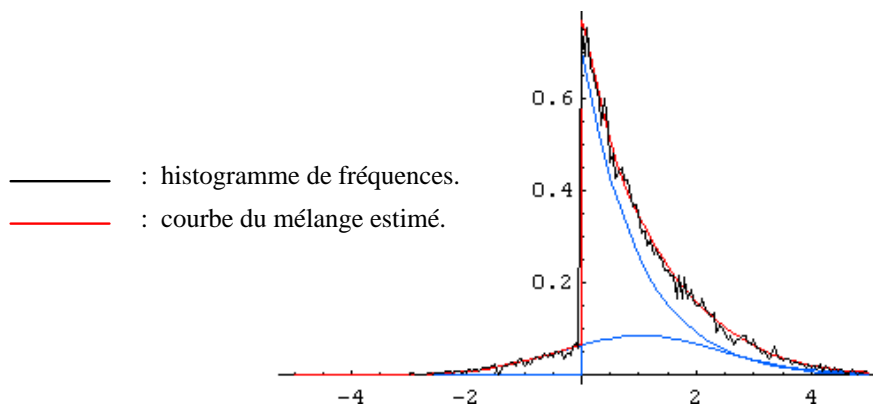
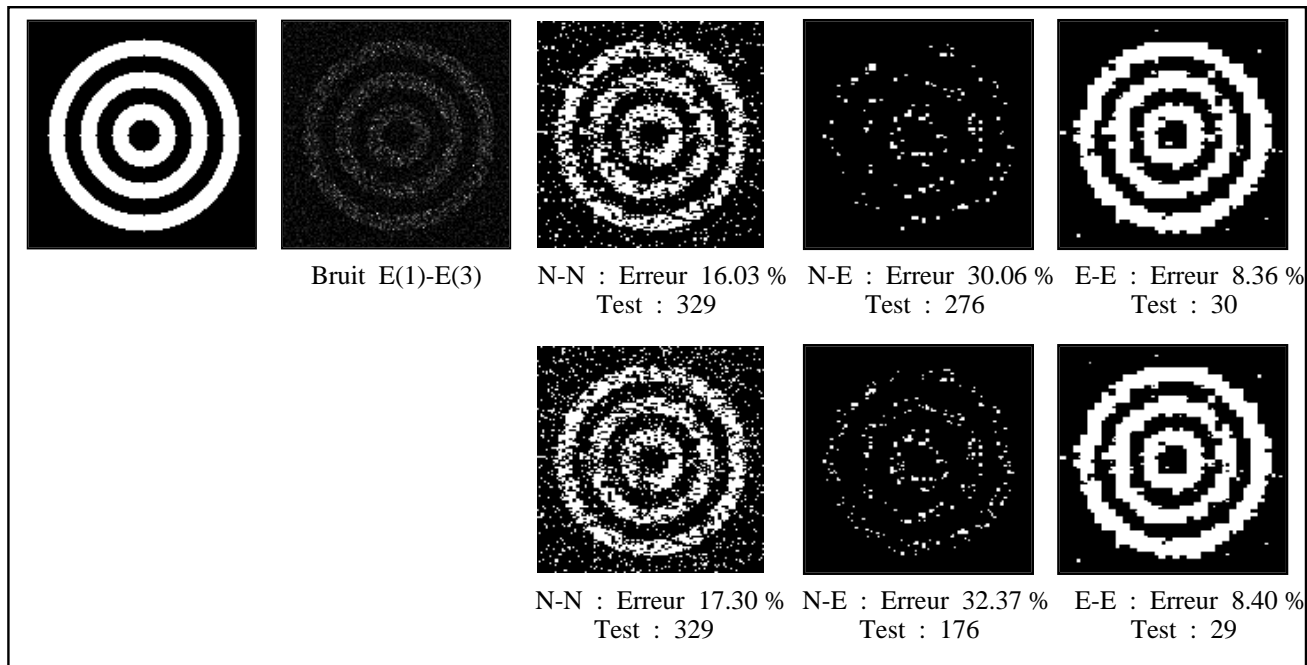


FIG. III.69. - SEGMENTATION D'IMAGE SUR LE MODÈLE DES MÉLANGES GÉNÉRALISÉS.

Le modèle des mélanges généralisés.



Graphique du mélange estimé par SEM2 type E-E.

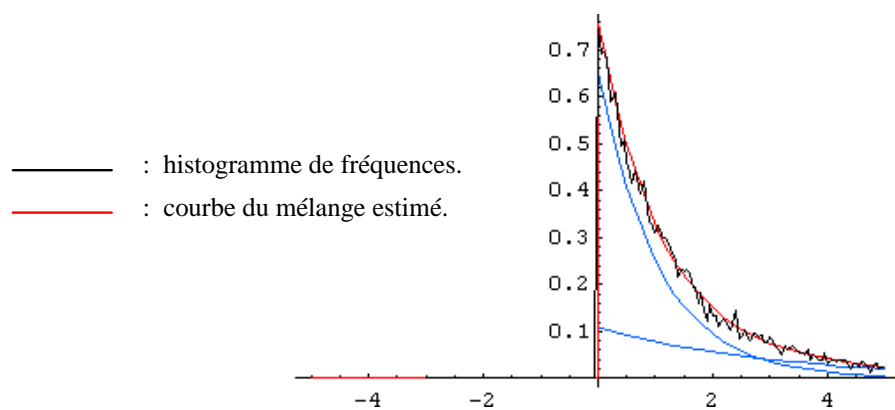
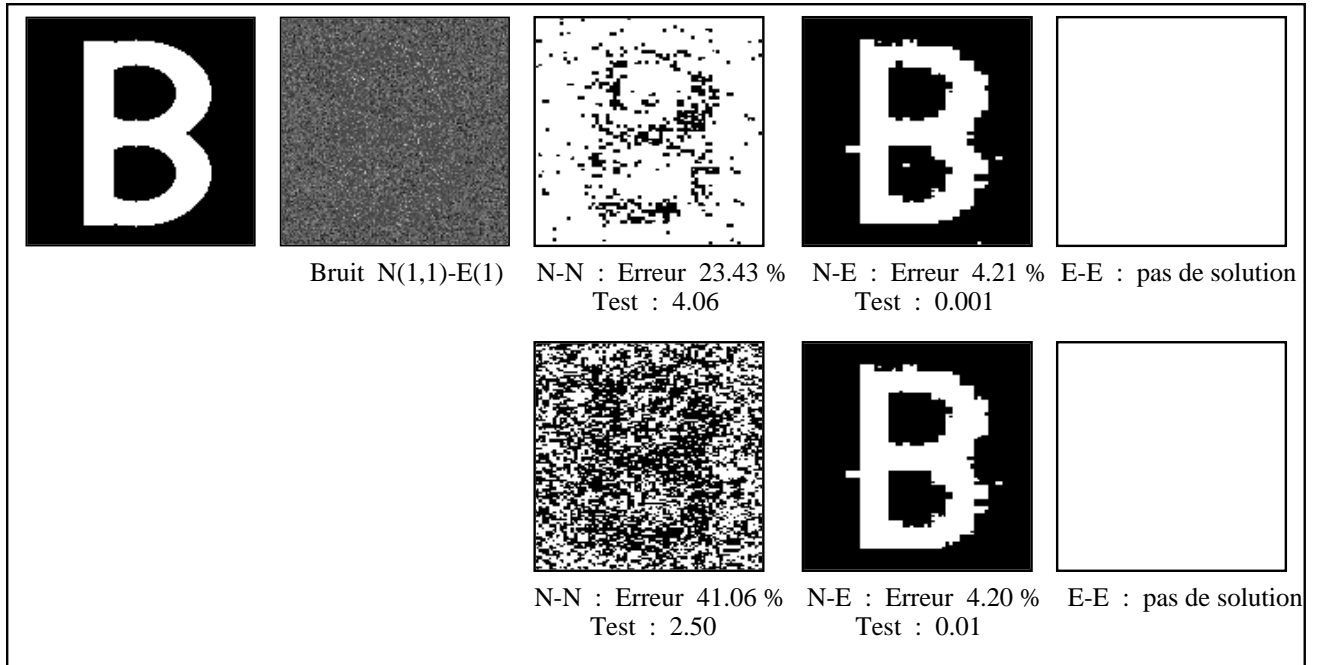


FIG. III.70. - SEGMENTATION D'IMAGE SUR LE MODÈLE DES MÉLANGES GÉNÉRALISÉS.

Le modèle des mélanges généralisés.



Graphique du mélange estimé par SEM1 type N-E.

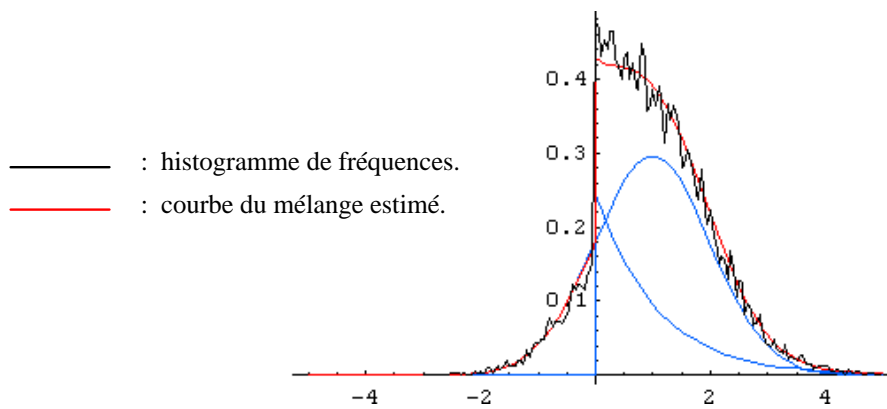


FIG. III.71. - SEGMENTATION D'IMAGE SUR LE MODÈLE DES MÉLANGES GÉNÉRALISÉS.

Le modèle des mélanges généralisés.

20.3 Lecture de cartons d'orgue de barbarie.

20.3.1 Description d'un carton.

Sur une demande de l'association Argentique¹², et afin d'illustrer la rapidité et la précision des algorithmes de segmentation d'images développés sur les arbres de Markov cachés, nous nous sommes intéressés à la mise au point d'un protocole de lecture d'un carton d'orgue de barbarie à 27 touches. Ce type de carton mesure 154mm de large. La note la plus haute, qui est un *Ré7* est située à 5mm du bord gauche du carton. Chaque note est constituée par un trou rectangulaire, de 3,5mm de largeur, et dont la longueur est fonction de la durée de la note dans la partition originale, et de la vitesse choisie pour le défilement standard du carton. Deux notes voisines, dans la succession non chromatique des notes d'un tel orgue, sont séparées, sur le carton, par un écart de 2mm, ce qui positionne la note la plus basse, un *Do4*, à 9mm du bord droit du carton.

20.3.2 Protocole d'acquisition de l'image.

Le carton est entraîné à une vitesse de défilement constante et filmé par une caméra grand angle, à une résolution de 320×320 pixels. On reconstitue, ensuite, à partir des portions non déformées des images de la séquence filmée, l'image que nous devons segmenter.

Dans le cadre de notre travail, la vitesse de défilement du carton est limitée par la résolution à laquelle nous travaillons, à cause de la nécessaire précision sur l'interprétation de la durée des notes. En tenant compte de la précision de lecture de l'orgue mécanique, nous pouvons, raisonnablement, travailler avec une marge d'erreur, longitudinale de 2mm. En particulier, l'espace minimum entre deux répétitions d'une même note sera de 2mm et la longueur minimale d'une note sera de 4mm, la taille du poinçon permettant la perforation des cartons étant $3.5 \times 3.5mm^2$. La vitesse de défilement du carton peut donc être fixée à $0.04ms^{-1}$, ce qui est un peu en dessous de la vitesse moyenne à laquelle le tourneur interprète son morceau. Nous ne pouvons cependant pas nous permettre de descendre sous la précision de segmentation longitudinale $1\text{ pixel} \rightarrow 1mm$.

La largeur de l'image permet une plus grande précision latérale globale, de $1\text{ pixel} \rightarrow 0.5mm$, qui compense la faible marge d'erreur longitudinale. Si l'on considère que la largeur théorique d'un trou correspond, alors, à 7 pixels, centrés sur une ligne dont la position est connue, parce que fixée par l'axe théorique de la note qu'ils représentent, il est raisonnable de considérer comme déterminante de l'une des deux classes, trou ou carton, la classe dominante, après segmentation, sur les 14 pixels formant l'unité de travail que nous nous sommes fixée.

Les deux images auxquelles nous nous intéresserons dans cette dernière partie du travail sont présentées figures FIG. III.72 et FIG. III.73. La première recense toutes les difficultés musicales caractérisant les cartons d'orgue de barbarie, alors que la seconde soulève le problème des cartons anciens, tachés ou partiellement dégradés. L'étude de ces deux images nous permettra de mettre au point une méthode de lecture de cartons, selon les critères que nous nous sommes fixés.

¹²Argentique est une association à but non lucratif, qui aide de jeunes artistes en participant à la production de films courts-métrages et de documentaires, ainsi qu'à l'organisation de spectacles de rue.

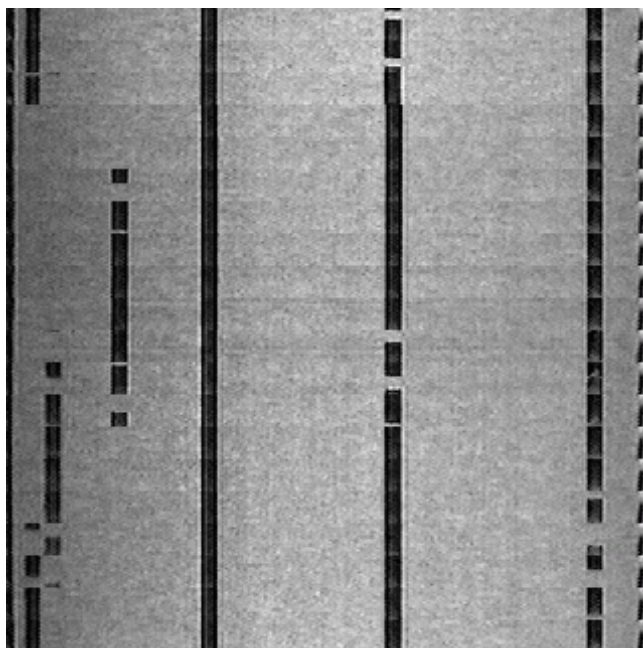


FIG. III.72. - IMAGE DE CARTON 1.

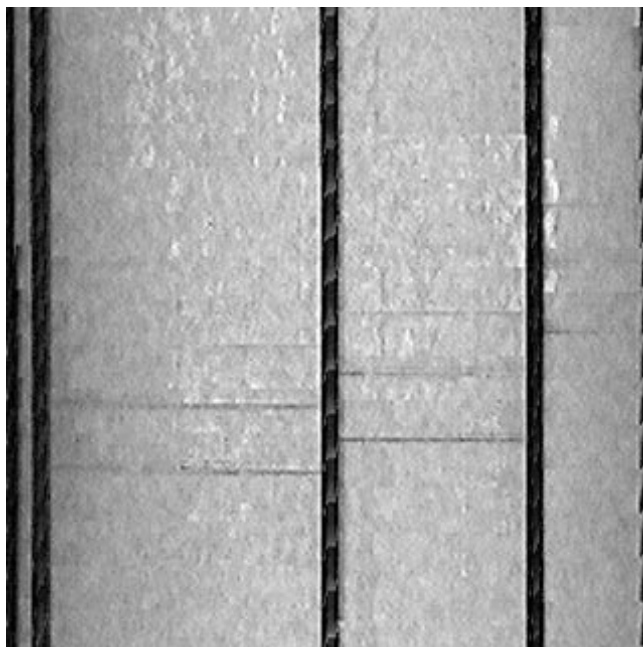


FIG. III.73. - IMAGE DE CARTON 2 (TACHÉE).

Le modèle des mélanges généralisés.

20.3.3 Segmentation de l'image du carton 1.

Les trois images FIG.III.74, FIG.III.75 et FIG.III.76, déduites de l'image carton 1 et obtenues sans avoir recours à un algorithme de segmentation, montrent l'impossibilité d'exploiter directement l'information contenue dans le fichier de capture. L'image classant les pixels en deux catégories est clairement inutilisable, à cause de la partie noire de la gauche de l'image. La répartition des classes dans les images FIG.III.75 et FIG.III.76, ne permet pas, non plus, une lecture raisonnable. A moins de perfectionner le dispositif d'acquisition des images, le recours à une méthode statistique de segmentation d'images apparaît comme une solution secourable.

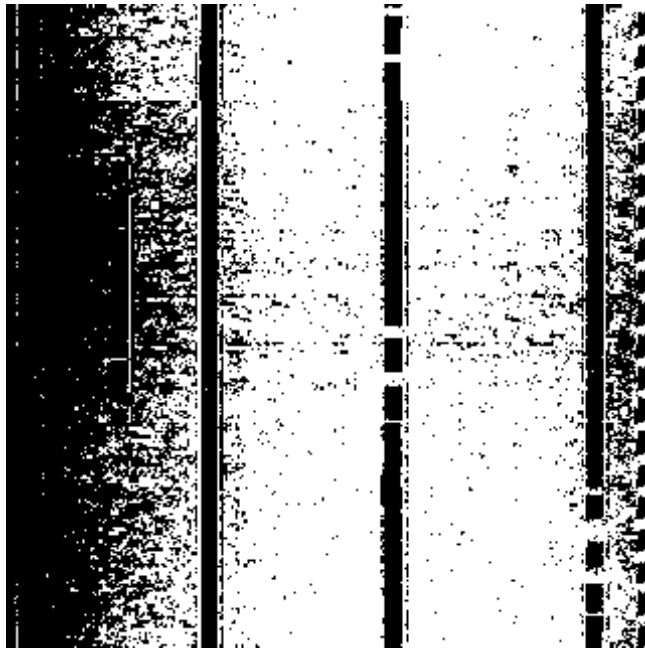


FIG. III.74. - EXPLOITATION DIRECTE DE L'INFORMATION EN DEUX CLASSES.

Pour des raisons de rapidité de la segmentation, et dans un souci de qualité des résultats, il est nécessaire de morceler l'image originale à la résolution 320×320 , en 25 sous-images à la résolution 64×64 . Cette solution permet de réaliser la segmentation de l'image de carton en un peu moins de 8 secondes, ce qui équivaut à une interprétation en temps réel de l'image acquise par la caméra. La rapidité des algorithmes de segmentation développés sur les arbres de Markov, permet, sans perte de temps, d'ajouter une dimension statistique à l'interprétation des données issues de la capture.

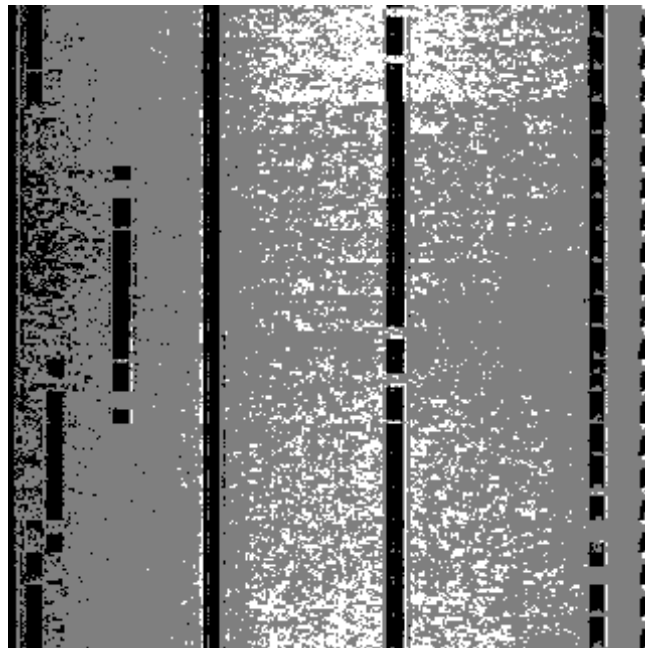


FIG. III.75. - EXPLOITATION DIRECTE DE L'INFORMATION EN TROIS CLASSES.

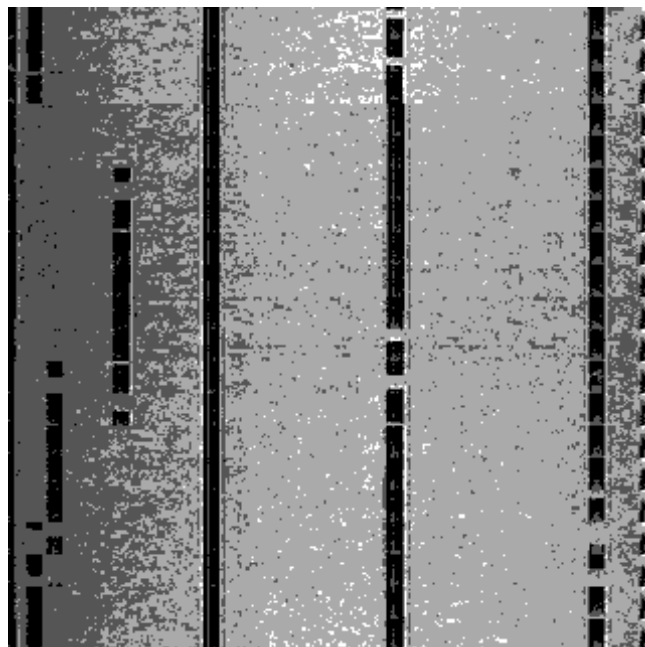


FIG. III.76. - EXPLOITATION DIRECTE DE L'INFORMATION EN QUATRE CLASSES.

Le modèle des mélanges généralisés.

Les tentatives de segmentation de l'image en plus de deux classes dans le cas gaussien, afin d'optimiser la qualité de l'interprétation que l'on veut en faire, s'avèrent peu fructueuses. En effet, une segmentation en trois classes, FIG. III.77, pose, comme nous l'avons vu sur l'image FIG. III.75, le problème de l'affectation des pixels de la classe grise intermédiaire. La segmentation en quatre classes, FIG. III.78, plus intéressante, provoque des échecs, à cause du trop grand nombre de classes, lors de l'estimation des paramètres de certaines zones de l'image.

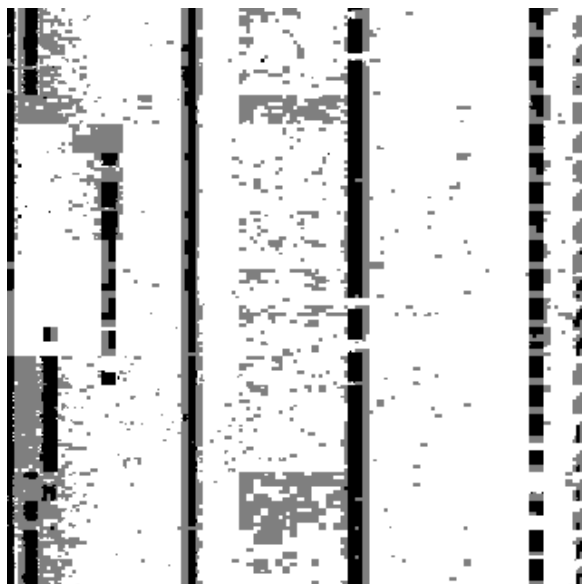


FIG. III.77. - SEGMENTATION DE L'IMAGE CARTON 1 EN 3 CLASSES.

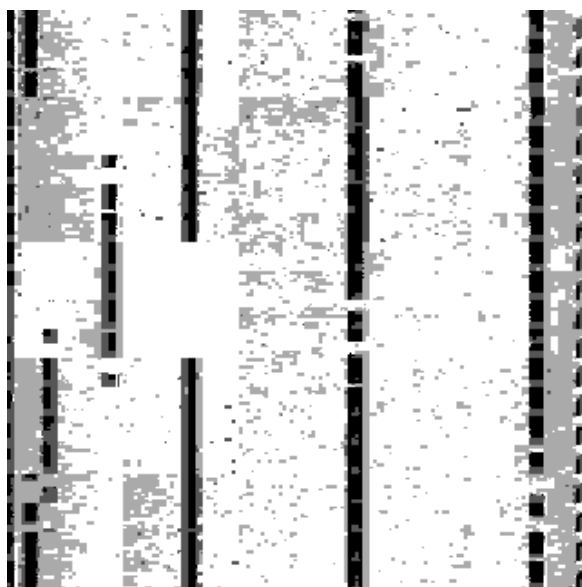


FIG. III.78. - SEGMENTATION DE L'IMAGE CARTON 1 EN 4 CLASSES.

Le modèle des mélanges généralisés.

La segmentation sera donc menée selon deux classes, ce qui permettra une construction naturelle du fichier "lu", comme nous l'avons décrite plus haut. La première segmentation selon deux classes, FIG. III.79, qui améliore incontestablement les résultats obtenus par la classification sans calculs, est effectuée dans le cadre d'un mélange de type normale-normale. L'image "interprétée", FIG. III.80, à partir de cette segmentation est très satisfaisante, aux approximations du protocole près.

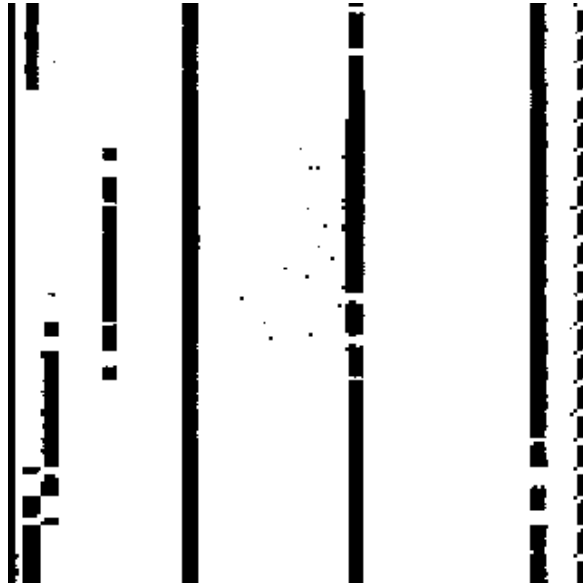


FIG. III.79. - IMAGE CARTON 1 SEGMENTÉE SUR LE MODÈLE DE TYPE N-N.

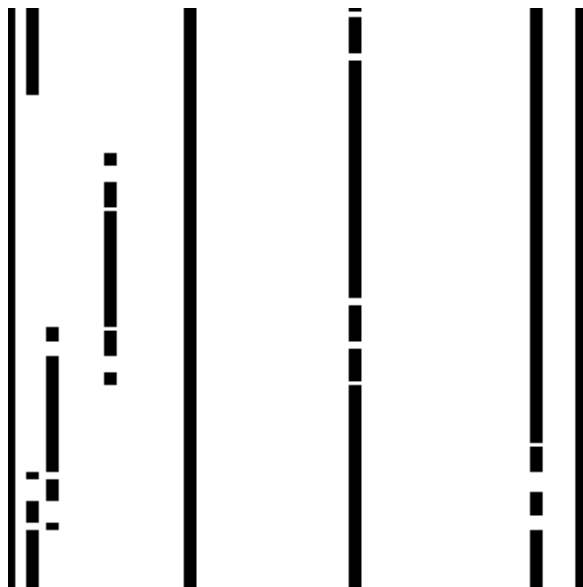


FIG. III.80. - IMAGE CARTON 1 "INTERPRÉTÉE".

Le modèle des mélanges généralisés.

Signalons aussi, qu'une segmentation, de chacune des 25 sous-images, par la méthode de mélanges généralisés, sélectionne, les 25 fois, le type de mélange normale-normale. Les segmentations dans les cas normale-exponentielle, et exponentielle-exponentielle sont proposées pour information. La première, FIG. III.81, est très proche de celle obtenue dans le cas normale-normale, à cela près que les classes "trou" sont, en général, plus larges. La connaissance théorique des zones de l'image susceptibles de porter la classe "trou", permet cependant une très bonne interprétation de cette image. La seconde, FIG. III.82, au contraire, réduit la largeur de ces classes, qui apparaissent très nettement. Les travers de cette segmentation apparaissent, notamment dans la partie gauche de l'image, ou la restriction extrême des zones de classes "trou", fausse l'interprétation que nous pourrions en faire.

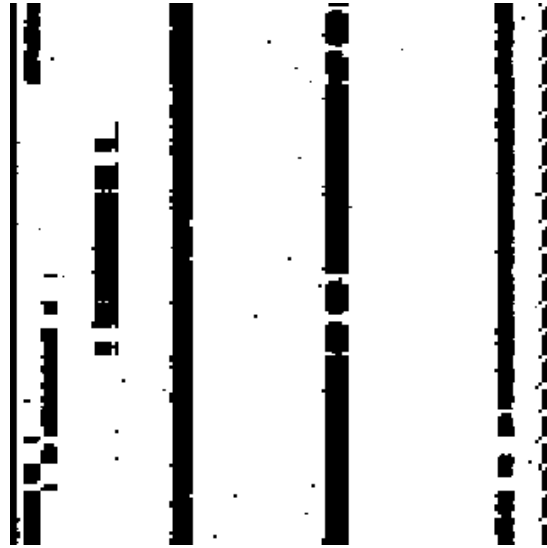


FIG. III.81. - IMAGE CARTON 1 SEGMENTÉE SUR LE MODÈLE DE TYPE N-E.

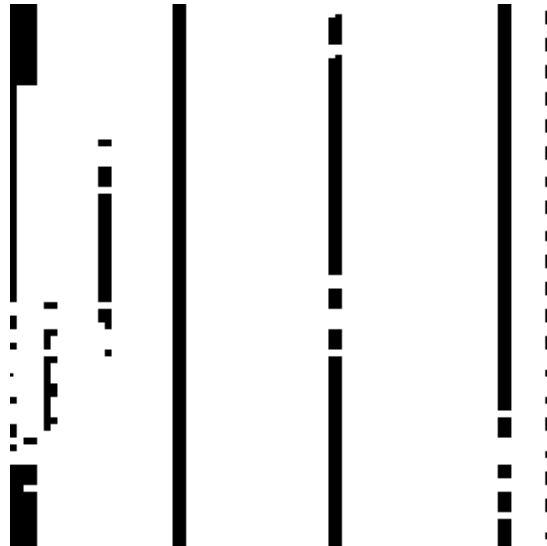


FIG. III.82. - IMAGE CARTON 1 SEGMENTÉE SUR LE MODÈLE DE TYPE E-E.

Le modèle des mélanges généralisés.

20.3.4 Segmentation de l'image du carton sale.

Le carton que nous essayons de lire ici, bien qu'il soit très simple d'un point de vu musical, soulève deux types de difficultés, qui ne touchaient pas le premier carton. Il a d'une part, été taché par le vernis utilisé pour durcir et éviter l'usure des bordures des trous, et il a, d'autre part, été plié à plusieurs endroits. Ces problèmes se manifestent, au niveau de l'image à segmenter, par des taches anormalement claires pour le vernis, et par des traits sombres que l'on peut confondre avec les trous, pour les pliures.

La complexification du travail devient claire lorsque l'on tente de segmenter l'image carton sale par la méthode des mélanges généralisés, qui, à nouveau ne retient que les estimations de type normale-normale. L'image segmentée, FIG. III.83, n'est pas exploitable pour une interprétation musicale correcte du carton.

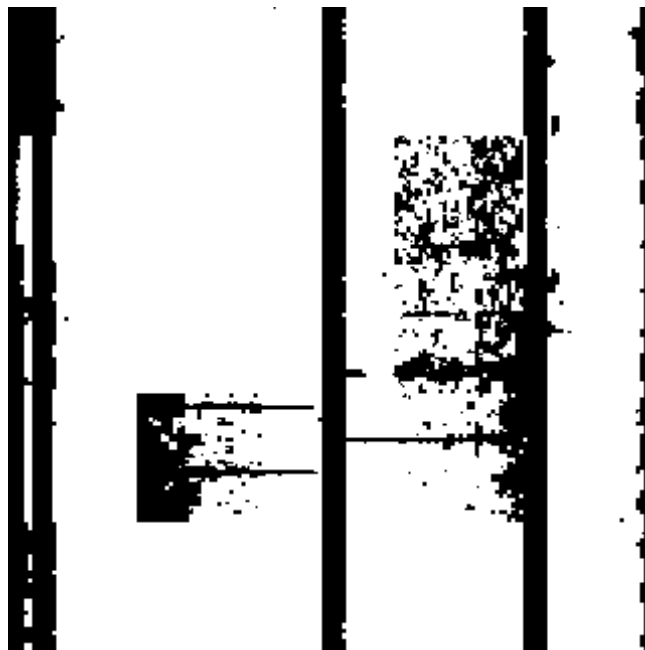


FIG. III.83. - IMAGE CARTON SALE SEGMENTÉE PAR LA MÉTHODE DES MÉLANGES GÉNÉRALISÉS.

Il apparait, en fait, que si la segmentation sur le modèle exponentielle-exponentielle, FIG. III.84, n'est pas plus pertinente, il n'en est pas de même de celle effectuée sur le modèle normale-exponentielle présentée FIG. III.85. Cette dernière segmentation identifie les trois notes, comme le montre son interprétation, FIG. III.86, et n'entraîne pas l'apparition artificielle de fausses notes.



FIG. III.84. - IMAGE SEGMENTÉE SUR LE MODÈLE DE TYPE E-E.

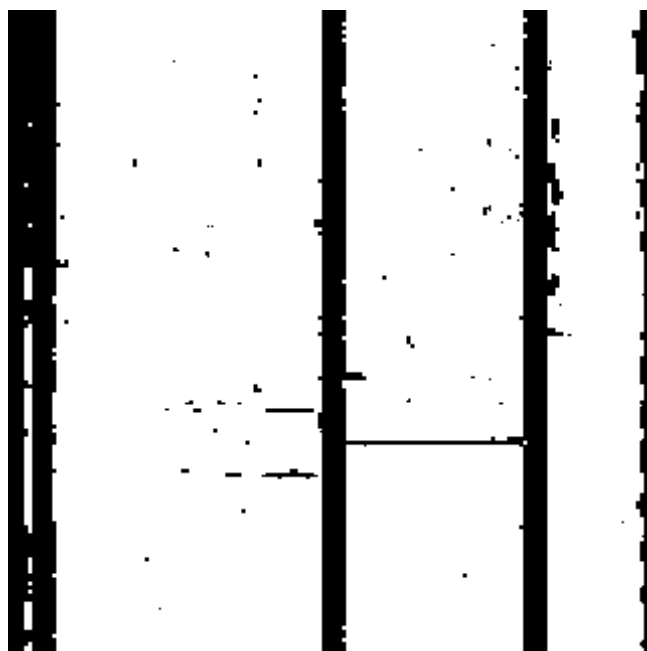


FIG. III.85. - IMAGE SEGMENTÉE SUR LE MODÈLE DE TYPE N-E.

Le modèle des mélanges généralisés.

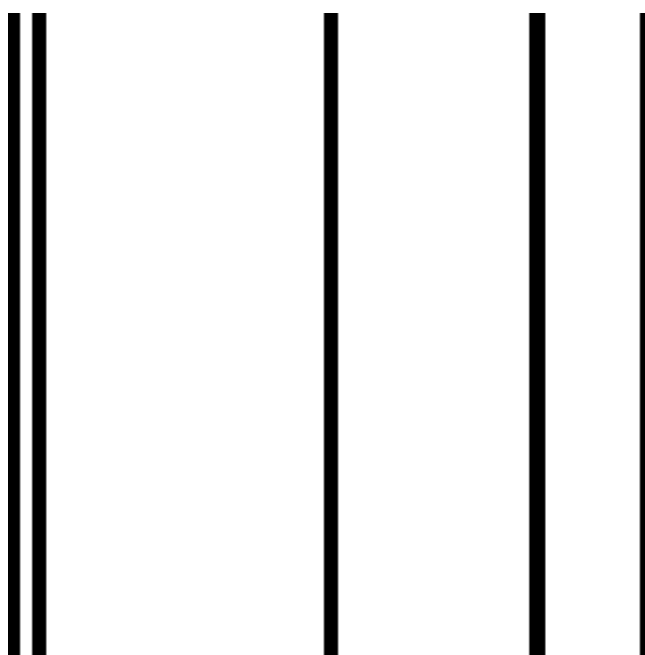


FIG. III.86. - IMAGE CARTON SALE "INTERPRÉTÉE".

Le modèle des mélanges généralisés.

20.3.5 Une méthode de lecture.

Le problème qui se pose dans ce cas très particulier de segmentation d'images réelles, typiquement adapté au cas des mélanges généralisés, c'est que le critère de choix général que nous avons décrit dans le chapitre précédent, induise un choix de segmentation que nous savons inapproprié. Il convient, donc, de définir un critère de choix pour la méthode des mélanges généralisés, spécifiquement adapté au problème qui nous intéresse. L'interprétation du carton demande, en effet, une grande précision dans la segmentation, que l'on peut atteindre grâce à une bonne connaissance des contraintes a priori.

La nouvelle méthode d'estimation des paramètres de mélanges généralisés que nous proposons, spécialement définie pour la segmentation d'images de cartons d'orgue de barbarie, multiplie par deux la durée totale de la segmentation, menée initialement sur une seule famille de distribution. Le nouveau critère consiste à choisir, pour chacune des 25 sous-images, celle des deux segmentations issues des estimations pour des mélanges de type N-N et N-E, qui aura le moins souvent attribué la classe "trou", à des variables que nous savons nécessairement classées "carton".

Cette procédure, appliquée à l'image carton sale, permet effectivement de corriger, par le choix de la segmentation issue du type N-E, les quatre sous-images dégradant précédemment le résultat de la segmentation N-N. L'image ainsi obtenue, FIG. III.87, optimise la segmentation et par conséquent, la qualité de la lecture du carton. Notons, pour conclure, que cette méthode, appliquée à l'image carton 1, permet d'aboutir, conformément à ce qu'on pouvait prévoir, à la même segmentation que celle obtenue dans le paragraphe précédent.

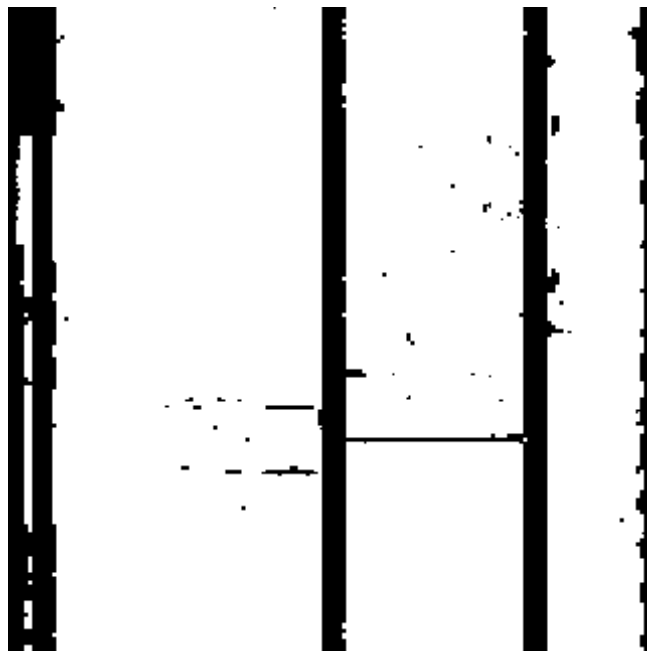


FIG. III.87. - IMAGE CARTON SALE POUR UNE SEGMENTATION OPTIMISÉE.

Cette méthode de lecture de cartons d'orgue de barbarie, testée sur plusieurs cartons complets, donne de très bons résultats. Nous ne pouvons, bien sûr, éviter les erreurs sur la durée de certaines notes, ou l'apparition de quelques fausses notes. Les retouches musicales à apporter sur le fichier interprété sont cependant raisonnablement peu fréquentes. Les erreurs sont, généralement, concentrées sur deux ou trois mesures, pour lesquelles les contraintes de virtuosité du morceau, ou la qualité locale de l'acquisition de l'image segmentée, perturbent la classification.

Les principales limites de cette technique de lecture de cartons d'orgue de barbarie, sont celles fixées par le protocole d'acquisition de l'image. Un résolution de capture supérieure, permettrait évidemment de gagner en

fiabilité. Une exploitation, en parallèle, des différents canaux RVB d'une capture couleur, optimisée par la possibilité de colorer le support, qui caractérise la couleur de la classe "trou", est actuellement en cours de développement.

Appendice A: Théorème fondamental de Robbins.

Soit le vecteur de Ω^m , $y = (y_1, \dots, y_m)$ et $G(y)$ une fonction de répartition dont la probabilité sur \mathcal{B}_m associée à G , est ν . On se donne une fonction $F(x, y)$ telle que, pour ν -presque tout y , $F(x, y)$ est une fonction de répartition, de x , et pour tout x une fonction borélienne de y . La probabilité sur \mathcal{B}_n associée à la fonction de répartition $F(x, y)$, de x , sera notée μ_y .

THÉORÈME : *La fonction*

$$H(x) = \int_{\Omega^m} F(x, y) dG(y) \quad [20.114]$$

est une fonction de répartition de \mathbb{R}^n . Si μ est la probabilité sur \mathcal{B}_n associée à H , pour tout $S \in \mathcal{B}_n$, $\mu_y(S)$ est une fonction borélienne de y et

$$\mu(S) = \int_{\Omega^m} \mu_y(S) dG(y). \quad [20.115]$$

Démonstration : On notera C , la classe de tous les boréliens S de \mathbb{R}^n tels que $\mu_y(S)$ est une fonction borélienne de y . Montrons que C est normale :

- Soit S_1, S_2, \dots une suite d'ensembles disjoints de C . Si $S = \bigcup_{n=1}^{+\infty} S_n$, on a :

$$\mu_y(S) = \mu_y\left(\bigcup_{n=1}^{+\infty} S_n\right) = \sum_{n=1}^{+\infty} \mu_y(S_n),$$

qui est une série convergente de fonctions boréliennes et donc une fonction borélienne elle-même.

- Soit $S_1 \supset S_2 \supset \dots$ une suite décroissante d'ensembles de C . Si $S = \bigcap_{n=1}^{+\infty} S_n$, on a :

$$\mu_y(S) = \mu_y\left(\bigcap_{n=1}^{+\infty} S_n\right) = \lim_{n \rightarrow \infty} \mu_y(S_n),$$

qui est la limite d'une suite de fonctions boréliennes et donc une fonction borélienne elle-même.

C est donc une classe normale.

Comme $\mu_y(S_x) = F(x, y)$ est une fonction borélienne de y pour tout x , C contient tous les S_x , on a donc $C = \mathcal{B}_n$.

Or on montre facilement (voir l'article de Robbins [Robbins48]) que si μ est une mesure sur \mathcal{B}_n , que μ_y est une probabilité sur \mathcal{B}_n et que $\nu(\mathbb{R}^m) = 1$, alors $\mu(\mathbb{R}^n) = 1$. On en déduit donc que la fonction μ définie en [1.4] est une probabilité sur \mathcal{B}_n associée à la fonction de répartition $H(x)$ définie en [1.3]. ♦

: Identifiabilité de la classe de mesures produits.

Appendice B: Identifiabilité de la classe de mesures produits.

Si l'on note, en se référant aux notations de la définition des classes de mélanges, :

$$F_{n,mn}^* = \left\{ F^*(x, \tilde{\theta}) : F^*(x, \tilde{\theta}) = \prod_{i=1}^n F(x, \theta^i), \text{ où } F(x, \theta^i) \in F_{1,m}, 1 \leq i \leq n \right\},$$

pour $\tilde{\theta} = (\theta^1, \dots, \theta^n) \in \Omega^{nm} = \underbrace{\Omega^m \times \Omega^m \times \dots \times \Omega^m}_{n \text{ fois}}$,

on définit une classe $F_{n,m-n}^*$ particulière, sur Ω^{nm} , dont la distribution générique est la mesure produit $F^*(x, \tilde{\theta})$. En fait, $F_{n,mn}^*$ est la famille des distributions jointes des échantillons de n variables indépendantes ayant chacune une distribution dans $F_{1,m}$. Nous avons alors le résultat suivant :

THÉORÈME : *Si la classe de tous les mélanges de $F_{1,m}$ est identifiable, alors pour tout $n > 1$, la classe des mélanges de $F_{n,mn}^*$ est aussi identifiable. Réciproquement, si pour au moins un $n > 1$, la classe de tous les mélanges de $F_{n,mn}^*$ est identifiable, il en est de même pour $F_{1,m}$.*

Démonstration : La réciproque est évidente puisqu'il suffit, pour $F \in F_{1,m}$, de multiplier par $\prod_{i=1}^{n-1} F(x_i, \theta_0)$, $\theta_0 \in \Omega^m$, les deux membres de

$$\int F(x, \theta) dG(\theta) = \int F(x, \theta) d\hat{G}(\theta), \quad \forall x,$$

pour avoir $I_{\theta_0} \dots I_{\theta_0} G = I_{\theta_0} \dots I_{\theta_0} \hat{G}$, I_{θ_0} étant la fonction de répartition de la probabilité ne chargeant que le point θ_0 , et donc $G(\theta_0) = \hat{G}(\theta_0)$, pour tout $\theta_0 \in \Omega^m$.

Montrons que si $F_{n,mn}^*$ est identifiable, alors il en est de même de $F_{n+1,m(n+1)}^*$, ce qui établira la première partie du théorème 2.5. Supposons, pour $F^* \in F_{n,mn}^*$ et $F \in F_{1,m}$, que l'on a, pour tout $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$:

$$\int_{\Omega^{(n+1)m}} F^*(x, \tilde{\theta}) F(y, \theta) dG(\tilde{\theta}, \theta) = \int_{\Omega^{(n+1)m}} F^*(x, \tilde{\theta}) F(y, \theta) d\hat{G}(\tilde{\theta}, \theta). \quad [20.116]$$

On note $G_2(\theta)$ et $\hat{G}_2(\theta)$ les distributions marginales, suivant θ , de G et \hat{G} respectivement. $G(\tilde{\theta} | \theta)$ et $\hat{G}(\tilde{\theta} | \theta)$ sont les probabilités telles que pour tout θ , $G(\tilde{\theta} | \theta)$ et $\hat{G}(\tilde{\theta} | \theta)$ sont des fonctions de répartition de la variable $\tilde{\theta}$, et pour tout $\tilde{\theta}$, $G(\tilde{\theta} | \theta)$ et $\hat{G}(\tilde{\theta} | \theta)$ sont presque partout des fonctions mesurables de θ . Alors [20.116] devient

$$\int_{\Omega^m} F(y, \theta) H(x, \theta) dG_2(\theta) = \int_{\Omega^m} F(y, \theta) \hat{H}(x, \theta) d\hat{G}_2(\theta), \quad [20.117]$$

où

$$H(x, \theta) = \int_{\Omega^{nm}} F^*(x, \tilde{\theta}) d_{\tilde{\theta}} G(\tilde{\theta} | \theta), \quad [20.118]$$

: Identifiabilité de la classe de mesures produits.

$$\widehat{H}(x, \theta) = \int_{\Omega^{nm}} F^*(x, \tilde{\theta}) d_{\tilde{\theta}} \widehat{G}(\tilde{\theta} | \theta) .$$

On peut réécrire [20.117]

$$\int_{\Omega^m} F(y, \theta) dJ_x(\theta) = \int_{\Omega^m} F(y, \theta) d\widehat{J}_x(\theta) , \tag{20.119}$$

où

$$\begin{aligned} J_x(\theta) &= \int_{-\infty}^{\theta} H(x, z) dG_2(z) \leq G_2(\theta) , \\ \widehat{J}_x(\theta) &= \int_{-\infty}^{\theta} \widehat{H}(x, z) d\widehat{G}_2(z) \leq \widehat{G}_2(\theta) . \end{aligned} \tag{20.120}$$

Par la convergence dominée de [20.119], on a pour tout x , $J_x(\Omega^m) = \widehat{J}_x(\Omega^m)$, finies d’après [20.120]. D’après la réciproque déjà montrée, du théorème 2.5, la classe des mélanges de la famille $F_{n, mn}^*$ étant identifiable, $F_{1, m}$ engendre une classe de mélanges identifiables. on a donc, d’après [20.119], on a, pour tout x , $J_x = \widehat{J}_x$, ce qui donne pour [20.120], pour tout $\theta \in \mathbb{R}^n$ et tout x ,

$$\int_{-\infty}^{\theta} H(x, z) dG_2(z) = \int_{-\infty}^{\theta} \widehat{H}(x, z) d\widehat{G}_2(z) . \tag{20.121}$$

D’un autre côté, par passage à la limite infinie, en x , dans [20.118] puis dans [20.117], on a, pour tout y ,

$$\int_{\Omega^m} F(y, \theta) dG_2(\theta) = \int_{\Omega^m} F(y, \theta) d\widehat{G}_2(\theta) ,$$

ce qui implique comme ci-dessus que

$$G_2 = \widehat{G}_2 . \tag{20.122}$$

D’après [20.122] et [20.121], on a, pour presque tout θ , $H(x, \theta) = \widehat{H}(x, \theta)$. Grâce à [20.118] et à l’identifiabilité de $F_{n, mn}^*$, on a, alors, pour presque tout θ ,

$$G(\cdot | \theta) = \widehat{G}(\cdot | \theta) ,$$

ce qui, combiné à [20.122], donne finalement $G(\cdot, \cdot) = \widehat{G}(\cdot, \cdot)$, $F_{n+1, m(n+1)}^*$ est bien identifiable. ♦

Appendice C: Théorème d'identifiabilité de Teicher.

THÉORÈME : Soit $F = \{F_i(x) = F(x, \theta^i), x \in \mathbb{R}^n, \theta^i \in \Omega^m\}$ une famille de fonction de répartition pour lesquelles il existe une transformation Φ , de domaine de définition S_Φ , telle que $M : F \mapsto \Phi$, soit linéaire et bijective. Supposons de plus qu'il existe, sur F , un ordre total, noté \trianglelefteq , tel que $F_1 \triangleleft F_2$ implique :

- $S_{\Phi_1} \subseteq S_{\Phi_2},$
- $\exists t_1 \in \overline{S_{\Phi_1}}^{13}, (t_1 \text{ indépendant de } S_{\Phi_2}) \text{ tel que } \lim_{t \rightarrow t_1} \frac{\Phi_2(t)}{\Phi_1(t)} = 0.$

Alors la classe \mathcal{H}' de tous les mélanges finis de F est identifiable. Démonstration : Supposons qu'il existe deux sous ensembles finis de F ,

$$F_1 = \{F_{1,i}, 1 \leq i \leq k\} \text{ et } F_2 = \{F_{2,j}, 1 \leq j \leq k'\},$$

ainsi que deux vecteurs $c_1 = (c_{1,i})_{1 \leq i \leq k}$ et $c_2 = (c_{2,j})_{1 \leq j \leq k'}$ tels que :

$$\forall x \in \mathbb{R}^n, \sum_{i=1}^k c_{1,i} F_{1,i}(x) = \sum_{j=1}^{k'} c_{2,j} F_{2,j}(x), \text{ avec} \quad [20.123]$$

$$\sum_{i=1}^k c_{1,i} = \sum_{j=1}^{k'} c_{2,j} = 1 \text{ et } \forall i, 1 \leq i \leq k, c_{1,i} > 0 \text{ et } \forall j, 1 \leq j \leq k', c_{2,j} > 0.$$

On peut, quitte à regrouper, dans chaque membre, les éléments identiques de F et à réindexer la somme, supposer que les $F_{1,i}$ et $F_{2,j}$ sont ordonnés dans un ordre strictement croissant, c'est à dire que $F_{1,i} \triangleleft F_{1,j}$ et $F_{2,i} \triangleleft F_{2,j}$ si $i < j$. On peut aussi, toujours sans perte de généralité, supposer que $F_{1,1} \trianglelefteq F_{2,1}$. Dans ce cas, si $F_{1,1} \triangleleft F_{2,1}$, on a $S_{\Phi_{1,1}} \subseteq S_{\Phi_{2,1}}$ et donc par transformation par M de [20.123] on a :

$$c_{1,1} + \sum_{i=2}^k c_{1,i} \left(\frac{\Phi_{1,i}(t)}{\Phi_{1,1}(t)} \right) = \sum_{j=1}^{k'} c_{2,j} \left(\frac{\Phi_{2,j}(t)}{\Phi_{1,1}(t)} \right), \forall t \in T_{1,1} = \{t \in S_{\Phi_{1,1}}; \Phi_{1,1}(t) \neq 0\}, \quad [20.124]$$

et par passage à la limite, lorsque $t \rightarrow t_1$, dans l'égalité [20.124], on a $c_{1,1} = 0$ ce qui contredit les hypothèses de [20.123]. On a, par conséquent, $F_{1,1} = F_{2,1}$ (au sens de la relation d'ordre \trianglelefteq), ce qui entraîne, par bijectivité de M , $\Phi_{1,1} = \Phi_{2,1}$ et par la linéarité de la transformation par M de [20.123] on a :

$$(c_{1,1} - c_{2,1}) + \sum_{i=2}^k c_{1,i} \left(\frac{\Phi_{1,i}(t)}{\Phi_{1,1}(t)} \right) = \sum_{j=2}^{k'} c_{2,j} \left(\frac{\Phi_{2,j}(t)}{\Phi_{1,1}(t)} \right), \forall t \in T_{1,1} = \{t \in S_{\Phi_{1,1}}; \Phi_{1,1}(t) \neq 0\}. \quad [20.125]$$

Par passage à la limite, lorsque $t \rightarrow t_1$, dans l'égalité [20.125], on a $c_{1,1} = c_{2,1}$ et donc [20.123] devient

$$\forall x \in \mathbb{R}^n, \sum_{i=2}^k c_{1,i} F_{1,i}(x) = \sum_{j=2}^{k'} c_{2,j} F_{2,j}(x).$$

En répétant ce raisonnement $\min(k, k')$ fois, on se ramène, en supposant par exemple que $k > k'$, à

$$\forall x \in \mathbf{R}^n, \sum_{i=k'+1}^k c_{1,i} F_{1,i}(x) = 0.$$

On montre alors, comme précédemment, que $c_{1,i} = 0$, $k' < i \leq k$, ce qui est impossible d'après [20.123]. On a donc $k = k'$, ce qui achève de montrer l'identifiabilité de \mathcal{H}' . ♦

Appendice D: CNS d'identifiabilité de Yakowitz et Spragins.

Soit la famille $F = \{F_i = F(x, \theta^i), \theta^i \in \Omega^m, x \in \mathbb{R}^n\}$, d'une distribution paramétrique donnée.

THÉORÈME : *La classe de tous les mélanges finis de F est identifiable si et seulement si F est une famille linéairement indépendante sur \mathbb{R}^{14} .*

Démonstration : Si $\sum_{i=1}^k a_i F_i = 0$, $a_i \in \mathbb{R}$ et les F_i différents deux à deux, on peut supposer, quitte à réindexer les a_i , qu'il existe M , $0 \leq M \leq k$, tel que $a_i < 0 \iff i \leq M$. alors :

$$\sum_{i=1}^M |a_i| F_i = \sum_{i=M+1}^k |a_i| F_i. \quad [20.126]$$

Les F_i étant des fonctions de répartition, $F_i(x) \xrightarrow{x \rightarrow (+\infty; \dots; +\infty)} 1$. En passant à la limite dans [20.126], on a donc

$\sum_{i=1}^M |a_i| = \sum_{i=M+1}^k |a_i| = c > 0$. Il suffit alors de poser $a'_i = \frac{|a_i|}{c}$ et l'on a :

$$\sum_{i=1}^M a'_i F_i = \sum_{i=M+1}^k a'_i F_i,$$

c'est à dire deux représentations différentes du même mélange (les F_i étant différentes deux à deux). Il est donc nécessaire que F soit linéairement indépendante. Si c'est le cas, F est une base de l'espace vectoriel $\langle F \rangle$, sur \mathbf{R} , engendré par F . Comme $\mathcal{H} \subset \langle F \rangle$, l'unicité de la décomposition d'un élément de $\langle F \rangle$ sur sa base F , assure l'identifiabilité de \mathcal{H} . La condition est donc suffisante. ♦

¹⁴ F est une famille linéairement indépendante sur R , si $\forall k \in N; \forall x \in R^n; (a_1, \dots, a_k) \in R^k; \forall i \neq j, F_j \neq F_i, \sum_{i=1}^k a_i F_i(x) = 0 \Rightarrow (a_1, \dots, a_k) = (0, \dots, 0)$.

Appendice E: Les moments.

E.1. Moments et moments centrés.

L'espérance mathématique d'une variable aléatoire, X , définie sur l'espace (Ω, \mathcal{F}, P) , est définie, si elle existe, par :

$$M = \mathbb{E}[X] = \int_{\Omega} X(\omega) P(d\omega).$$

Si, l'on note μ la mesure de probabilité associée à la variable aléatoire, et que l'on considère la fonction borélienne g définie sur \mathbb{R} , on aura :

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) \mu(dx).$$

Si, de plus, la variable a une densité, f , l'espérance de $g(x)$ est donnée par :

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f(x) dx.$$

Si, enfin, F est la fonction de répartition de X , on a :

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) dF(x).$$

Le moment d'ordre $k \in \mathbb{N}^*$, si il existe, d'une variable aléatoire X est défini par :

$$M_k = \mathbb{E}[X^k] = \int_{\mathbb{R}} x^k \mu(dx) = \int_{\mathbb{R}} x^k dF(x).$$

et le moment centré d'ordre $k > 1$, est, alors, défini par :

$$M_k^c = \mathbb{E}[(X - \mathbb{E}[X])^k] = \int_{\mathbb{R}} (x - M_1)^k \mu(dx) = \int_{\mathbb{R}} (x - M_1)^k dF(x).$$

Signalons que, par définition, le premier moment centré est nul, et que les moments centrés d'ordre impair des distributions symétriques sont tous nuls.

E.2. Conditions suffisantes à la caractérisation d'une distribution par ses moments.

Sous certaines conditions, une distribution est définie de manière univoque par ses moments. Le théorème suivant, dont on trouve la preuve dans [Billingsley68], est la solution la plus connue au "Problème des moments de Stieljes".

THÉORÈME : *Soit une mesure de probabilité, μ , définie sur le même espace que précédemment et ayant des moments, M_k , finis à tous les ordres.*

: *Les moments.*

Si la série entière S :

$$S(x) = \sum_{k \geq 1} \frac{M_k x^k}{k!}$$

a un rayon de convergence positif, alors μ est la seule mesure de probabilité ayant les moments $M_k, k \geq 1$.

Si une mesure de probabilité, μ , associée à une distribution F , vérifie les conclusions de ce théorème, la distribution est déterminée par ses moments.

Différentes études permettent de compléter ce théorème. Tout d'abord, l'article de Dharmadhikari [Dharmadhikari65] montre que cette condition n'est pas nécessaire. L'article de Cramer et Wold, [Cramer36], donne une condition suffisante à la résolution du problème des moments de Stieljes, dans le cas des variables aléatoires de dimension 2. Enfin, dans son livre, Rao, [Rao73], donne une série de conditions suffisantes en fonction du rang de la variable aléatoire étudiée.

E.3. Conditions nécessaires et suffisantes à l'existence d'une distribution correspondant à une suite de moments.

Dans les deux articles [Hamburger20] et [Hamburger21], Hamburger développe une CNS au problème des moments de Stieljes, dans le cas particulier des mesures positives croissantes ayant un nombre infini de points de croissance. Ceci s'applique donc, aux fonctions de répartitions des variables aléatoires chargeant un nombre non fini de points. Avant de développer ce théorème, commençons par définir, pour une suite de réels $(s_k)_{k \in \mathbb{N}}$, la suite des déterminants, $k \in \mathbb{N}$:

$$D_k [(s_k)_{k \in \mathbb{N}}] = D_k = \begin{vmatrix} 1 & s_1 & \dots & s_k \\ s_1 & s_2 & \dots & s_{k+1} \\ \dots & \dots & \dots & \dots \\ s_k & s_{k+1} & \dots & s_{2k} \end{vmatrix}.$$

DÉFINITION : une suite de réels $(s_k)_{k \in \mathbb{N}}$ est définie positive si, pour tout $k \in \mathbb{N}$, le déterminant D_k est positif.

Forts de ces précisions, nous pouvons énoncer le théorème suivant :

THÉORÈME : Une condition nécessaire et suffisante pour qu'il existe une fonction de répartition, F , définie sur \mathbb{R} , ayant un nombre infini de points de croissance et telle que :

$$M_k = \int_{\mathbb{R}} x^k dF(x), k \in \mathbb{N},$$

et que la suite $(M_k)_{k \in \mathbb{N}}$ soit définie positive.

Le théorème de Riesz, qui suit, généralise le théorème de Hamburger. Il utilise la notion de suite non négative dont la définition est donnée ci-dessous. Il permet de statuer sur tous les types de distributions et de raisonner sur une suite finie de moments.

: Les moments.

DÉFINITION : une suite $(s_k)_{1 \leq k \leq n}$, où $n \leq +\infty$, est non négative relativement à l'intervalle $]a, b[$, où $a \geq -\infty$ et $b \leq +\infty$ si pour tout polynôme R , de degré $\leq n$,

$$R(x) = a_0 + a_1x + \dots + a_nx^n, x \in]a, b[,$$

tel que :

$$\forall x \in]a, b[, R(x) \geq 0,$$

on a :

$$a_0 + a_1s_1 + \dots + a_ns_n \geq 0.$$

Le théorème de Riesz, pour les distributions s'énonce alors comme suit :

THÉORÈME : Une CNS pour qu'il existe une fonction de répartition, F , pour laquelle :

$$M_k = \int_{\mathbb{R}} x^k dF(x), 1 \leq k \leq n \leq +\infty$$

et que la suite $(M_k)_{1 \leq k \leq n}$ soit non négative, n étant, bien sûr, supposé être pair.

Voir aussi [Akhiezer65], [Akhiezer88] et [Landau87].

E.4. Cumulants.

Les cumulants d'une distribution sont dérivés des moments et nous permettent de rendre certaines manipulations algébriques plus claires.

DÉFINITION : Si une distribution F possède des moments finis à tous les ordres, on définit la suite $(k_r)_{r \in \mathbb{N}}$ des cumulants par :

$$\exp\left(\sum_{r=1}^{\infty} \frac{k_r t^r}{r!}\right) = \sum_{r=1}^{\infty} \frac{M_r t^r}{r!},$$

k_r est le cumulant d'ordre r de la distribution F .

Peu utilisables sous cette forme, les cumulants vérifient la proposition suivante :

PROPOSITION : pour tout $(r, j) \in \mathbb{N}^* \times \mathbb{N}^*$, on a :

$$\frac{\partial M_r}{\partial k_r} = \binom{r}{j} M_{r-j}.$$

: Les moments.

Grâce à cette propriété, on a au rang $j = 1$:

$$\frac{\partial M_r}{\partial k_1} = rM_{r-1},$$

ce qui permet de tirer, par récurrence :

$$\begin{aligned} M_1 &= k_1, \\ M_2 &= k_2 + k_1^2, \\ M_3 &= k_3 + 3k_1k_2 + k_1^3, \\ M_4 &= k_4 + 4k_1k_3 + 3k_2^2 + 6k_1^2k_2 + k_1^4, \\ &\text{et ainsi de suite} \end{aligned}$$

Pour les moments centrés on obtient des relations similaires :

$$\begin{aligned} 0 &= k_1, \\ M_2^c &= k_2, \\ M_3^c &= k_3, \\ M_4^c &= k_4 + 3k_2^2, \\ M_5^c &= k_5 + 10k_3k_2, \\ M_6^c &= k_6 + 15k_4k_2 + 10k_3^2 + 15k_3k_2^2, \\ &\text{et ainsi de suite} \end{aligned}$$

ce qui permet d'exprimer les cumulants en fonction des moments centrés :

$$\begin{aligned} k_1 &= 0, \\ k_2 &= M_2^c, \\ k_3 &= M_3^c, \\ k_4 &= M_4^c - 3(M_2^c)^2, \\ k_5 &= M_5^c - 10M_3^cM_2^c, \\ k_6 &= M_6^c - 15M_4^cM_2^c - 10(M_3^c)^2 + 30(M_2^c)^3, \\ &\text{et ainsi de suite} \end{aligned}$$

On pourra trouver des développements supplémentaires sur les cumulants dans le livre de Kendall et Stuart [Kendall87], ou dans [Haberman96].

: *Les moments empiriques.*

Appendice F: Les moments empiriques.

F.1. Définition des moments empiriques.

Soit l'échantillon X_1, \dots, X_n de la distribution F . Si l'on suppose que pour tout i , $1 \leq i \leq n$, $X_i \in \mathbb{R}$, considérons la droite réelle munie de la tribu, \mathcal{B} , de ses boréliens et définissons sur $(\mathbb{R}, \mathcal{B})$ la mesure de probabilité P_n :

$$\forall B \in \mathcal{B}, P_n(B) = \frac{1}{n} \sum_{i=1}^n I(X_i, B) \text{ avec } I(X, B) = \begin{cases} 1, & \text{si } X \in B \\ 0, & \text{si } X \notin B \end{cases} .$$

En fait, P_n est une probabilité concentrée en les X_i et telle que la probabilité de chacun d'eux est égale à $\frac{1}{n}$. P_n est appelée mesure empirique associée à l'échantillon X_1, \dots, X_n . Les moments empiriques (resp. les moments centrés empiriques) d'un échantillon issu de la distribution F , sont les moments par rapport à la mesure de probabilité empirique associée à cet échantillon. Le premier moment empirique, \mathfrak{M} , est la moyenne de l'échantillon. Le moment empirique d'ordre k et le moment centré empirique d'ordre k sont donnés, respectivement, par :

$$\begin{aligned} \mathfrak{M}_k &= \frac{1}{n} \sum_{i=1}^n X_i^k, \\ \text{et } \mathfrak{M}_k^c &= \frac{1}{n} \sum_{i=1}^n (X_i - \mathfrak{M}_1)^k. \end{aligned}$$

F.2. Convergence des moments empiriques vers les moments théoriques.

On se donne un échantillon, X_1, \dots, X_n , issu d'une distribution F_0 appartenant à une famille \mathcal{F} de distributions. Supposons que G est une fonctionnelle de la forme :

$$G(F) = h \left[\int g(x) dF(x) \right],$$

où g est une fonction borélienne et h une fonction continue au point $a = \int g(x) dF_0(x)$. Nous pouvons, alors, définir la fonction \overline{g}_n :

$$\overline{g}_n = \frac{1}{n} \sum_{i=1}^n g(X_i).$$

On nommera, par ailleurs, statistique de type I , une statistique S_n^* de type :

$$S_n^* = G(F_n) = h(\overline{g}_n),$$

où F_n est la fonction de répartition associée à P_n .

THÉOREME : *Sous les hypothèses précédentes :*

$$G(F_n) \xrightarrow[p.s.]{} G(F_0), \text{ quand } n \rightarrow \infty,$$

ou encore :

$$S_n^* = h(\bar{g}_n) \xrightarrow[p.s.]{} h(a), \text{ quand } n \rightarrow \infty.$$

En particulier, $\forall k \in \mathbb{N}$,

$$\mathfrak{M}_k = \frac{1}{n} \sum_{i=1}^n X_i^k \xrightarrow[p.s.]{} \mathbb{E}[X_1^k], \text{ quand } n \rightarrow \infty$$

et

$$\mathfrak{M}_k^c = \frac{1}{n} \sum_{i=1}^n (X_i - \mathfrak{M}_1)^k \xrightarrow[p.s.]{} \mathbb{E}[(X_1 - M_1)^k], \text{ quand } n \rightarrow \infty,$$

c'est à dire que les moments empiriques centrés et non centrés convergent presque sûrement vers les moments théoriques, respectivement centrés et non centrés.

Si l'on suppose, de plus, que h est dérivable au point a , et que

$$\int g^2(x) dF_0(x) < \infty,$$

on peut établir le théorème suivant :

THÉOREME : *Sous les hypothèses précédemment énoncées,*

$$\sqrt{n}(S_n^* - h(a)) \Longrightarrow h'(a) \Phi_{0,\sigma^2},$$

où Φ_{0,σ^2} est une distribution normale centrée et de variance :

$$\sigma^2 = \int (g(x) - a)^2 dF_0(x).$$

En particulier, pour tout $k \in \mathbb{N}$,

$$\sqrt{n}(\mathfrak{M}_k - M_k) \Longrightarrow \Phi_{0,\sigma^2}, \text{ où } \sigma^2 = M_{2k} - M_k^2$$

et

$$\sqrt{n}(\mathfrak{M}_k^c - M_k^c) \Longrightarrow \Phi_{0,\sigma_c^2}, \text{ où } \sigma_c^2 = M_{2k}^c - (M_k^c)^2.$$

Tous ces développements peuvent être retrouvés dans [Borovkov87].

: *La méthode des moments.*

Appendice G: La méthode des moments.

G.1. Estimation par méthode de substitution. Convergence.

On se donne un échantillon, X_1, \dots, X_n , issu d'une distribution P appartenant à une famille \mathcal{P} de distributions. On suppose que le paramètre θ , inconnu et caractérisant P se représente par une fonctionnelle G de la distribution P , c'est à dire :

$$\theta = G(P).$$

On nommera, alors, estimateur de substitution de θ , l'estimateur θ_n^* tel que :

$$\theta_n^* = G(P_n).$$

Rappelons, par ailleurs, qu'un estimateur θ_n^* est dit convergent vers θ s'il converge en probabilité, et fortement convergent s'il converge presque sûrement.

Pour les estimateurs de substitution, on peut établir le théorème suivant :

THÉORÈME : *Soit l'échantillon $(X_n)_{n \geq 1}$ issu de la distribution F_0 . Supposons que $\theta = G(P)$, où G est une fonctionnelle de la forme :*

$$G(P) = h \left[\int g(x) dF(x) \right],$$

où g est une fonction borélienne et h une fonction continue au point $a = \int g(x) dF_0(x)$.

Alors, $\theta_n^* = G(P_n)$ est un estimateur fortement convergent.

G.2. Normalité asymptotique des estimateurs de substitution.

Rappelons qu'un estimateur $\theta_n^* = (\theta_{n,1}^*, \dots, \theta_{n,k}^*)$ de $\theta = (\theta_1, \dots, \theta_k)$ est dit asymptotiquement normal de matrice $\sigma^2 \geq 0$, si la distribution limite de $\sqrt{n}(\theta_n^* - \theta)$ est une distribution normale (multivariée) centrée de matrice σ^2 . Pour les estimateurs de substitution, on peut établir le théorème suivant :

THÉORÈME : *Soit l'échantillon $(X_n)_{n \geq 1}$ issu de la distribution F_0 . θ étant défini par une fonctionnelle du type du théorème précédent, où $g = (g_1, \dots, g_s)$ et $h(t) = (h_1(t), \dots, h_k(t))$, (où $t = (t_1, \dots, t_s)$) avec les hypothèses supplémentaires que h admet en $a = (a_1, \dots, a_s)$ (où $a_j = \int g_j(x) dF_0(x)$) les dérivées partielles :*

$$h_{i,j} = \frac{\partial h_i}{\partial t_j}(a), \quad 1 \leq i \leq k, \quad 1 \leq j \leq s,$$

et que

$$\int g_i(x) g_j(x) dF_0(x) < \infty, \quad 1 \leq i \leq s, \quad i \leq j \leq s.$$

Alors, θ_n^* est un estimateur asymptotiquement normal de matrice :

$$\sigma^2 = Hd^2H^T,$$

où $d^2 = [d_{i,j}]$ est une matrice carrée $s \times s$, symétrique, avec

$$d_{i,j} = \mathbb{E}[(g_i(X_1) - a_i)(g_j(X_1) - a_j)],$$

et $H = [h_{i,j}]_{\substack{1 \leq i \leq k \\ 1 \leq j \leq s}}$.

Remarquons que σ^2 est une matrice $k \times k$, alors que d^2 est une matrice $s \times s$. Voir [Borovkov87] ou [Haberman96]

G.3. Estimation par la méthode des moments, convergence et normalité asymptotique.

Soit l'échantillon $(X_n)_{n \geq 1}$ issu de la distribution F_θ . Si l'on se place toujours dans le cas où θ est un paramètre à k dimensions. Considérons une fonction $g = (g_1, \dots, g_k)$, et notons

$$m_j(\theta) = \int g_j(x) P_\theta(dx) = \mathbb{E}_\theta[g_j(X_1)], \quad \text{pour } 1 \leq j \leq k.$$

Choisissons, de plus, la fonction g de telle sorte que l'équation :

$$m(\theta) = t, \quad \text{avec } t = (t_1, \dots, t_k),$$

admette une unique solution continue $\theta = m^{-1}(t)$, dans l'ensemble Θ des valeurs de θ . Nous pouvons, alors, définir la fonction \bar{g}_n :

$$\bar{g}_n = \left(\frac{1}{n} \sum_{i=1}^n g_1(X_i), \dots, \frac{1}{n} \sum_{i=1}^n g_k(X_i) \right).$$

D'une façon générale l'estimateur $\theta_n^* = m^{-1}(\bar{g}_n)$ est appelé estimateur de θ par la méthode des moments. Le cas le plus souvent utilisé est celui où $g_k(x) = x^k$ et les équations sont les équations ordinaires des moments. Les équations des moments centrés, correspondant au cas où $g_k(x) = (x - M)^k$, entrent aussi dans le cadre de cette méthode. Nous noterons respectivement $\bar{\theta}_n$ et $\bar{\theta}_n^c$ ces deux derniers types d'estimateurs.

Il apparaît clairement que les estimateurs des moments sont des **estimateurs de substitution** et qu'il sont, à ce titre, **fortement convergents**. La normalité asymptotique doit être étudiée au cas par cas en vérifiant les hypothèses du théorème précédent ou, dans le cas le plus simple, celles du théorème suivant :

: La méthode des moments.

THÉORÈME : Soit l'échantillon $(X_n)_{n \geq 1}$ issu de la distribution F_θ telle que

$$M_i(\theta) = \mathbb{E}_\theta [X_1^i] < \infty, 1 \leq i \leq 2k.$$

Si l'on suppose de plus que m admet en $a = [M_1(\theta), \dots, M_s(\theta)]$ les dérivées partielles :

$$h_{i,j} = \frac{\partial m_i}{\partial \theta_j}(a), 1 \leq i, j \leq k.$$

et que $H = [h_{i,j}]_{1 \leq i, j \leq k}$ est de rang k .

Alors, $\overline{\theta}_n$ est un estimateur asymptotiquement normal de matrice :

$$\sigma^2 = H^{-1} \Sigma [H^T]^{-1},$$

où $\Sigma = [\sigma_{i,j}]_{1 \leq i, j \leq k}$, avec

$$\sigma_{i,j} = M_{i+j} - M_i M_j.$$

: Les moments de mélanges.

Appendice H: Les moments de mélanges.

H.1. Equations des moments des mélanges de deux distributions.

Si l'on considère, pour la clarté du propos, deux distributions F_1 et F_2 , mélangées en proportions respectives, p et $1 - p$. Si X_1 suit une distribution F_1 et X_2 une distribution F_2 , on aura, pour tout $n \geq 1$:

$$p\mathbb{E}[X_1^n] + (1 - p)\mathbb{E}[X_2^n] = M_n^{ix},$$

où M_n^{ix} est le $n^{\text{ième}}$ moment de la distribution de mélange. Cette équation est la $n^{\text{ième}}$ équation de moments d'un mélange de deux distributions. La $n^{\text{ième}}$ équation des moments centrés d'un mélange de deux distributions est, sous une forme brute

$$p\mathbb{E}\left[(X_1 - M_1^{ix})^n\right] + (1 - p)\mathbb{E}\left[(X_2 - M_1^{ix})^n\right] = M_n^{ixC},$$

où M_n^{ixC} est le $n^{\text{ième}}$ moment centré de la distribution de mélange.

Si l'on pose : $x_1 = \mathbb{E}[X_1] - M_1^{ix}$ et $x_2 = \mathbb{E}[X_2] - M_1^{ix}$ ces équations deviennent :

$$p\mathbb{E}[(x_1 + (X_1 - \mathbb{E}(X_1)))^n] + (1 - p)\mathbb{E}[(x_2 + (X_2 - \mathbb{E}(X_2)))^n] = M_n^{ixC}. \quad [20.127]$$

La première équation est, alors :

$$px_1 + (1 - p)x_2 = 0, \quad [20.128]$$

et pour $n > 1$:

$$px_1^n + (1 - p)x_2^n = M_n^{ixC} - \sum_{i=2}^n \binom{i}{n} (p\mathbb{E}(X_1^i) x_1^{n-i} + (1 - p)\mathbb{E}(X_2^i) x_2^{n-i}). \quad [20.129]$$

Ce sont ces équations que nous utiliserons dans la suite.

H.2. Cas particulier du mélange de deux distributions gaussiennes.

Dans le cas du mélange de distributions gaussiennes $N(\mu_1, \sigma_1^2)$ et $N(\mu_2, \sigma_2^2)$, en proportions respectives p et $(1 - p)$, le nombre des paramètres à estimer étant 5, nous tenterons d'appliquer la méthode des moments par le

système :

$$\begin{cases} 0 = px_1 + (1-p)x_2 \\ M_2^{ixC} = px_1^2 + (1-p)x_2^2 + (p\sigma_1 + (1-p)\sigma_2) \\ M_3^{ixC} = px_1^3 + (1-p)x_2^3 + 3(p\sigma_1x_1 + (1-p)\sigma_2x_2) \\ M_4^{ixC} = px_1^4 + (1-p)x_2^4 + 6(p\sigma_1x_1^2 + (1-p)\sigma_2x_2^2) + 3(p\sigma_1^2 + (1-p)\sigma_2^2) \\ M_5^{ixC} = px_1^5 + (1-p)x_2^5 + 10(p\sigma_1x_1^3 + (1-p)\sigma_2x_2^3) + 15(p\sigma_1^2x_1 + (1-p)\sigma_2^2x_2) \end{cases} . \quad [20.130]$$

Nous verrons cependant qu'il nous faudra avoir recours à une sixième équation de moments, qui est :

$$M_6^{ixC} = p(15\sigma_1^3 + 45\sigma_1^2x_1 + 15\sigma_1x_1^4 + x_1^6) + (1-p)(15\sigma_2x_2^4 + 45\sigma_2^2x_2^2 + 15\sigma_2^3 + x_2^6).$$

H.3. Convergence des estimateurs des moments des mélanges.

D'après ce que nous avons vu dans l'Appendice G, les estimateurs des moments des mélanges sont fortement convergents, la normalité asymptotique étant à vérifier grâce au théorème ci-dessus.

Appendice I: Le Théorème des fonctions implicites.

I.1. Matrice jacobienne.

Soit une fonction $f = (f_1, \dots, f_k)$, où f :

$$\begin{aligned} E &\subset \mathbb{R}^q \longrightarrow F \subset \mathbb{R}^k \\ t &= (t_1, \dots, t_q) \rightarrow (f_1(t), \dots, f_k(t)). \end{aligned}$$

Si f est une fonction différentiable, c'est à dire si chacune des composantes de f est dérivable par rapport à chacune de ses variables, la matrice jacobienne de f , est la matrice :

$$Jac(f) = \left[\frac{\partial f_i}{\partial t_j} \right]_{\substack{1 \leq i \leq k \\ 1 \leq j \leq q}}.$$

Si les dimensions de E et F sont les mêmes, le jacobien de f est le déterminant de $Jac(f)$.

I.2. Théorème d'inversion locale des fonctions.

Le théorème des fonctions implicites donne une condition à l'inversion d'un système, qu'il soit linéaire ou non.

THÉORÈME : Soient E, F, G trois espaces de Banach. Soit, encore, une fonction f , continuellement différentiable d'un sous ensemble ouvert A de $E \times F$ dans G . Soit, enfin, un élément $(x_0, y_0) \in A$ tel que $f(x_0, y_0) = 0$. Si le jacobien de $f(x_0, \cdot)$ ne s'annule pas en y_0 , alors il existe un voisinage ouvert U_0 de x_0 inclus dans E , tel que pour tout sous ensemble U , ouvert et convexe, de U_0 il existe une unique fonction continue u de U dans G , telle que :

$$u(x_0) = y_0 \text{ et } f(x_0, u(x_0)) = 0.$$

On pourra consulter [Dieudonné70] pour davantage de précisions

Appendice J: La méthode de Newton de résolution des systèmes algébriques non linéaires.

Alors que l'on dispose de méthodes générales de résolution des systèmes algébriques linéaires, l'inversion des systèmes algébriques non-linéaires ne peut, le plus souvent, pas être réalisée littéralement. La seule possibilité est, alors, de se tourner vers une méthode numérique et itérative de résolution. Parmi ces méthodes, les plus efficaces sont celles dérivant de la méthode de Newton, dont nous présentons ici le principe (voir par exemple [Bates88] et [Kress98]). Considérons un ensemble $D \subset \mathbb{R}$, et une fonction $f : D \rightarrow \mathbb{R}$, continuellement différentiable et dont la matrice jacobienne, f' est non-singulière pour tout $x \in D$. Nous cherchons à résoudre l'équation :

$$f(x) = 0 \text{ où } 0 \text{ est le vecteur nul.}$$

D'après la formule de Taylor généralisée, nous savons qu'au voisinage d'un point x_0 proche d'un zéro de la fonction f , on a :

$$f(x) = f(x_0) + f'^T(x_0)(x - x_0) + (x - x_0)\varepsilon(x), \text{ avec } (x - x_0)\varepsilon(x) \xrightarrow{x \rightarrow x_0} 0.$$

Si un point x_1 est suffisamment proche de x_0 on peut négliger le dernier terme de l'égalité ci-dessus et considérer que :

$$f(x_1) = 0 \text{ se ramène à } x_1 = x_0 - [f'^T(x_0)]^{-1} f(x_0).$$

La méthode itérative de résolution de Newton consiste à construire la suite x_n définie par :

$$x_{n+1} = x_n - [f'(x_n)]^{-1} f(x_n) \text{ à partir d'un } x_0 \in D,$$

qui, sous certaines conditions convergera vers un zéro de notre système. Le théorème suivant, précise des conditions suffisantes de convergence de la méthode de Newton.

THÉORÈME : *Sous les hypothèses énoncées ci-dessus, D étant un ouvert, si f est deux fois continuellement différentiable et que r est un zéro de f tel que la matrice jacobienne de f n'est pas singulière en r , alors, la méthode de Newton est localement convergente. (i.e. il existe un voisinage B de r , tel que pour tout $x_0 \in B$, la méthode de Newton converge.)*

La méthode de Newton-Raphson utilise le développement de Taylor au rang supérieur. Le principe reste donc le même, puisqu'au voisinage d'un point x_0 proche d'un zéro de la fonction f , on a :

$$f(x) = f(x_0) + f'^T(x_0)(x - x_0) + (x - x_0)^T \frac{f^{(2)}(x_0)}{2} (x - x_0) + (x - x_0)^2 \varepsilon'(x),$$

avec $(x - x_0)^2 \varepsilon'(x) \xrightarrow{x \rightarrow x_0} 0.$

où $f^{(2)}$ est le hessien de f . Avec les mêmes approximations que précédemment et en adaptant le théorème précédent, on obtient une méthode qui converge très rapidement, si l'initialisation est raisonnable. En effet, le principal obstacle à l'utilisation de la méthode de Newton-Raphson est la nécessité d'initialiser l'algorithme avec une valeur initiale susceptible de justifier les approximations, c'est à dire, proche d'une solution. Cette initialisation est d'autant plus complexe que la dimension de la fonction à inverser est importante.

: La méthode de Newton de résolution des systèmes algébriques non linéaires.

BIBLIOGRAPHIE.

- [Abramowitz65] Abramowitz M., Stegun J.A. (1965). *Handbook of Mathematical Functions*, Dover, New York.
- [Andrews74] Andrews D.F., Mallows C.L. (1974). *Scale mixture of normal distributions*, J Roy. Statist. Soc., ser. B, **36**, 99-102.
- [Akhiezer65] Akhiezer N.I., (1965). *The classical Moment Problem*, Edimburgh : Oliver and Boyd.
- [Akhiezer88] Akhiezer N.I., (1988). *Lectures on integral transforms*, Translations of Mathematical monographs, American Mathematical Society.
- [Bach87] Bach A., Plachky D., Thomson W. (1987). *A characterization of identifiability of mixtures of distributions*, in Mathematical Statistics and Probability Theory, vol. A, pp. 15-21, M.L. Puri, P. Révész ed., Reidel, Dordrecht.
- [Bagirov88] Bagirov E.B., (1988). *Some remark on mixtures of normal distributions*, Theor. Probab. Appl., **33**, 709-710.
- [Barndorff65] Barndorff-Nielsen O. (1965). *Identifiability of finite mixtures of exponential families*, .J. math. Anal. appl., **21**, 115-121.
- [Barndorff82] Barndorff-Nielsen O., Kent J., Sorensen. M. (1982). *normal variance-mean mixture and z-distributions*, Int. Statist. Rev., **50**,145-159.
- [Bates88] Bates D.M., Watts D.G., (1988). *Nonlinear regression analysis and its applications*, Wiley Series in Proba. and Math. Stat.
- [Baum70] Baum L.E., Petrie T., Soules G., Weiss N., (1970). *A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains*, Ann. Math. Stat., **41**, 164-171.
- [Baum72] Baum L.E., (1972). *An inequality and associated maximization technique in statistical estimation for probabilistic function of Markov processes*, Inequalities, **3**, 433-454.
- [Bendjebour00] Bendjebour A., (2000). *Segmentation d'images multisenseur par fusion de Dempster-Shafer dans un contexte markovien*, Thèse de doctorat de l'Université Paris VI.
- [Bendjebour01] Bendjebour A., Delignon Y., Fouque L. Samson V., Pieczynski W., (2001), *Multisensor Images Segmentation Using Dempster-Shafer Fusion in Markov Fields Context*. IEEE Transaction on Geoscience an Remote Sensing, Vol39, **8**, 1789-1798.
- [Benmiloud95] Benmiloud B., Pieczynski W., (1995). *Estimation des paramètres dans les chaînes de Markov*

cachées et segmentation d'images, Traitement du signal, Vol 12, No 5.

[Benveniste87] Benveniste A., Métivier M., Priouret B., (1987). *Algorithmes adaptatifs et approximations stochastiques*, Techniques stochastiques, Masson, Paris.

[Besag74] Besag J., (1974). *Spatial interaction and the statistical analysis of lattice systems*, J. Roy. Stat. Soc., Ser. B, **36**, 192-236.

[Besag86] Besag J., (1986). *On the statistical analysis of dirty pictures*, J. Roy. Stat. Soc., Ser. B, **48**, 259-302.

[Bhattacharya67] Bhattacharia C. G., (1967). *A simple method of resolution of a distribution into Gaussian components*. Biometrics, **23**, 115-135.

[Bickel93] Bickel P.J., Klaassen C.A., Ritov Y., Wellner J.A., (1993). *Efficiency and Adaptive Estimation for Semiparametric Models*, John Hopkins University Press, Baltimore.

[Billingsley68] Billingsley P. (1968). *Convergence of probability measures*. Wiley, New-York.

[Blum77] Blum J.R., Sursala V. (1977). *Estimating the parameters of a mixing distribution. function*, Ann. Probab., **5**, 200-209.

[Bondesson92] Bondesson L. (1992). *Generalized gamma Convolutions and Related Classes of Distribution*, Lecture Notes in Statist., **76**, Springer, New York.

[Borges99] Borges C.F., (1999). *On the estimation of Markov random field parameters*, IEEE Transactions PAMI, Vol 21, **3**, 216-224.

[Borovkov87] Borovkov A.A, (1987). *Statistique Mathématique*, Ed. Mir Moscou.

[Bouman94] Bouman C., Shapiro, (1994). *A multiscale random field model for bayesian image segmentation*, IEEE Transactions On image Processing, Vol 3, **2**, 162-177.

[Bowman73] Bowman K. O., Shenton L. R., (1973). *Space of solution for a normal mixture*. Biometrika, **60**, 629-636.

[Braathen93] Braathen B., Masson P., Pieczynski W., (1993). *Global and local methods of unsupervised Bayesian segmentation of images*, Machine Graphics and Vision, Vol 2, **1**, 39-52.

[Broniatowski83] Broniatowski M., Celeux G., Diebolt J, (1983). *Reconnaissance de mélanges de densités par un algorithme d'apprentissage probabiliste*, In data Analysis and Informatics 3, E. Diday ed., North Holland, Amsterdam, 359-373.

[Bruni85] Bruni C., Koch G., (1985). *Identifiability of continuous mixtures of Unknown Gaussian distributions*, Ann. Probab., **13**, 1341-1357.

[Bryant83] Bryant J. L., Paulson A. S., (1983). *Estimation of mixing proportions via distance between characteristic functions*. Commun. Statist. A, **12**, 1009-1029.

[Caillol93] Caillol H., Hillion A., Pieczynski W., (1993). *Fuzzy random fields and unsupervised image segmentation*, IEEE Transactions on GRS, Vol 31, **4**, 801-810.

[Caillol97] Caillol H., Hillion A., Pieczynski W., (1997). *Estimation of Fuzzy Gaussian Mixture and Unsupervised Statistical Image Segmentation*, IEEE Transaction on Image Processing, Vol 6, **3**, 425-440.

[Casella99] Casella G., Robert C. P., (1999). *Monte Carlo Statistical Methods*. Springer Texts in statistics, Springer Verlag, New-York.

[Celeux86] Celeux G., Diebolt J., (1986). *L'algorithme SEM: un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densités*, Vol 43, **2**, 35-51.

[Celeux92] Celeux G., Diebolt J., (1992). *A stochastic approximation type EM algorithm for the mixture problem*, Stochastics and stochastics reports, **41**, 119-134.

- [Celeux95] Celeux G., Chauveau D., Diebolt J., (1995). On stochastic Versions of the EM Algorithm, Rapport de recherche **2514**, INRIA.
- [Chalmond89] Chalmond B., (1989). *An iterative Gibbsian technique for reconstruction of m-ary images*, Pattern Recognition, Vol 22, **6**, 747-761.
- [Chandra77] Chandra S., (1977). *On the mixture of probability distributions*, Scand J. Statist., **4**, 105-112.
- [Chardin00] Chardin A., (2000). *Modèles énergétiques hiérarchiques pour la résolution des problèmes inverses en analyse d'images - Application à la télédétection*. Thèse de l'Université de Rennes I.
- [Charlier24] Charlier C. V. L., Wicksell S. D., (1924). *On the dissection of frequency functions*. Arkiv. Mat. Astron. och Fysik., Bd. **18**, No 6.
- [Chellapa93] Chellapa R., (1993), Markov Random Fields, Academic Press.
- [Chen95] Chen J., (1995). *Optimal rate of convergence for finite mixture models*, Ann. Statist., **23**, 221-223.
- [Choi68] Choi K., Bulgren W.B., (1968). *An estimation procedure for mixture of distributions*, J. Roy, Statist. Soc., ser. **B**, 444-460.
- [Cohen67] Cohen A.C., (1967). *Estimation in Mixtures of Two Normals Distributions*. Technometrics Vol. 9 No 1. pp 15-28.
- [Conway73] Conway J.B. (1973). *Functions of One Complex Variable*, Springer New York.
- [Cox74] Cox D. R., Hinkley D. V., (1974). *Theoretical statistics*. London, Chapman and Hall.
- [Cramer36] Cramer, Wold, (1936). Some theorems on distribution functions, J. London Math. Soc., **11**, 290-294.
- [Dawid92] Dawid A., (1992). *Application of a general propagation algorithm for probabilistic expert systems*, Statistics and Computing, **2**, 25-36.
- [Deely68] Deely J.J., Kuse R.L., (1968). *Construction of sequences estimating the mixture distribution*, Ann. Math. statist., **39**, 286-288.
- [Delignon97] Delignon Y., Marzouki A., Pieczynski W., (1997). *Estimation of generalized Mixture and its application in Image Segmentation*, IEEE Transaction on Image Processing, Vol 6, **10**, 1364-1374.
- [Delmas95] Delmas J.P., (1995). *Relations entre les algorithmes d'estimation itérative EM et ICE avec des exemples d'application*, Quinzième Colloque GRETSI, Juan-Les-Pins, 185-188.
- [Delmas97] Delmas J.P., (1997). *An equivalence of the EM and ICE algorithm for Exponential family*. IEEE Transactions on Signal Processing, Vol 45, **10**, 2613-2615.
- [Delyon99] Delyon B., Lavielle M., Moulines E., (1999). *Stochastic computation - convergence of a stochastic approximation version of the EM algorithm*, Annals of Statistics, Vol 27, **1**, 94-128.
- [Dempster77] Dempster M.M., Laird N.M., Rubin D.B., (1977). *Maximum Likelihood from incomplete data via th EM algorithm*, J. Roy. Stat. Soc., Ser. B, **39**, 1-38.
- [Derin87] Derin H., Elliott H., (1987). *Modeling and Segmentation of noisy and textured images using Gibbs Random Fields*, IEEE Transactions on PAMI, Vol 9, **1**, 39-55.
- [Descombes97] Descombes X., Lorette A., Zerubia J., (1997). *Quelques améliorations à la segmentation d'images bayésienne - première partie : modélisation.*, Traitement du Signal, Vol 14, **4**, 373-382.
- [Descombes99] Descombes X., Morris R., Zerubia J., Berthod M., (1999). *Estimation of Markov random field prior parameters using Markov chain Monte-Carlo maximum Likelihood*, IEEE Transactions on Image Processing, Vol 8, **7**, 954-963.

- [Devijver85] Devijver P.A., (1985). *Baum's forward-backward algorithm revisited*, Pattern Recognition Letters **3**, 369-373.
- [Devijver87] Devijver P.A., Dekesel M., (1987). *Algorithmes d'apprentissage d'un modèle Markovien d'images*, Actes du 6ème Congrès, RFIA, Antibes, 183-207.
- [Devijver88] Devijver P.A., Dekesel M. (1988). *Champs aléatoires de Pickard et modélisation d'images digitales*, Traitement du signal, Vol 5, **5**.
- [Devijver90] Devijver P.A., (1990). Image segmentation using causal markov random fields models, Int. Conf. Pattern Rec., Atlantic City, 153-158,
- [Devroye90] Devroye L., (1990). *Consistent deconvolution in density estimation*, Canad. J. Statist., **17**, 235-239.
- [Dharmadhikari65] Dharmadhikari S., (1965). An example in the problem of moments, Ann. Math. Monthly, **72**, 302-303.
- [Dharmadhikari88] Dharmadhikari S., Joag-Dev K., (1988). *Unimodality, Convexity and Applications*, Academic Press, New-York.
- [Diebolt90] Diebolt J., Robert C., (1990). *Estimation des paramètres d'un mélange par échantillonnage bayésien*, Compte rendu de l'Académie des Sciences, Paris, Vol 311, **10**, 653-658.
- [Dieudonné70] Dieudonné J., (1970). *Eléments d'analyse, Tome 1 - Fondements de l'analyse moderne*, Cahiers scientifiques, Gauthier-Villars, Paris.
- [Doukhan94] Doukhan P., (1994). *Mixing : Properties and Examples*. Lecture Notes in Statistic, Springer, New York.
- [Eagleson75] Eagleson G.K., (1975). *Martingale convergence to mixture of infinitely divisible laws*, Ann. probab, **3**, 557-562.
- [Edelman88] Edelman D., (1988). *Estimation of the mixing distribution for a normal mean with applications to the compound decision problem*, Ann. Statist., **16**, 1602-1622.
- [Efron78] Efron B., Olshen R., (1978). *How broad is the class of normal scale mixtures?*, Ann. Statist., **6**, 1159-1164.
- [Eggermont95] Eggermont P., LaRiccia V.N., (1995). *Maximum smoothed likelihood density estimation for inverse problems*, Ann. Statist, **23**, 199-220.
- [Everitt81] Everitt B.S., Hand D.J., (1981). *Finite Mixture Distributions*, Chapman and Hall, London.
- [Feller71] Feller W., (1971). *Theory of Probability and its Applications*, vol. 2, Wiley, New York.
- [Fiorin01] Fiorin S., (2001). *Inconsistency for roots of likelihood equations which are relative maxima of the likelihood function*. Rapport technique ISUP-LSTA, 2001-6.
- [Forney73] Forney G.D., (1973). *The Viterbi Algorithm*, Proc. IEEE, **61**, 268-278.
- [Fouque99] Fouque L., Tebbache A.N., Pieczynski W., (1999). *Sur les segmentations statistiques non supervisées dans le contexte des mélanges généralisés*, Dix-septième Colloque GRETSI, Vannes, **1**, 639-642.
- [Fowlkes79] Fowlkes E. B., (1979). *Some methods for studying the mixture of two normal (Lognormal) distributions*. J. Amer. Statist. Assoc., **74**, 561-575.
- [Fraser81] Fraser M.D., Hsu Y.S., Walker, J.J., (1981). *Identifiability of finite mixture of von Mises distributions*, Ann. statist., **9**, 1130-1131.
- [Fryer70] Fryer J. G., Robertson C. A., (1970). *The bias and accuracy of moment estimators*. Biometrika, **57**, 57-65.

- [Fryer72] Fryer J. G., Robertson C. A., (1972). *A comparison of some methods for estimating mixed normal distributions*. *Biometrika*, **59**,639-648.
- [Furman94a] Furman D., Lindsay B. G., (1994). *Testing for the number of components in a mixture of normal distributions using moments estimators*. *Comput. Statist. Data Anal.*, **17**, 473-492.
- [Furman94b] Furman D., Lindsay B. G., (1994). *Measuring the relative effectiveness of moment estimators as starting values in maximizing mixture likelihoods*. *Comput. Statist. Data Anal.*, **17**, 493-507.
- [Geman84] Geman S., Geman D, (1984). *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*, *IEEE Transactions on PAMI*, Vol 6, 721-741.
- [Geman86] Geman S., Graffigne C., (1986). *Markov random field image and their applications to computer vision*, *Proceedings of ICM*, **2**, 1496-1517.
- [Giordana96] Giordana N., (1996). *Segmentation non supervisée d'images multispectrales par chaines de Markov cachées*, Thèse de doctorat de l'Université de Technologie de Compiègne.
- [Giordana97] Giordana N., Pieczynski W., (1997). *Estimation of Generalized multisensor Hidden Markov Chains and Unsupervised Image Segmentation*, *IEEE Transactions PAMI*, Vol 19, **5**, 465-476.
- [Groeneboom85] Groeneboom P., (1985). *Estimating a monotone density*, in *Proc. Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer*, L.M. Lecam, R.A. Olshen eds., **2**, 539-555, Wadsworth, Belmont, CA.
- [Haberman96] Haberman S.J., (1996). *Advanced Statistics, Vol1*, Springer, New-York.
- [Hajek88] Hajek B., (1988). *Cooling schedules for optimal annealing*, *Math. Oper. Res.*, **13**, 197-243.
- [Hamburger20] Hamburger, (1920). *Über eine Erweiterung des Stieljesschen Momenten Problem*, **81**, 231-319.
- [Hamburger21] Hamburger, (1921). *Über eine Erweiterung des Stieljesschen Momenten Problem*, **82**, 120-187.
- [Haralick86] Haralick R., Hyonam J., *A context classifier*, *IEEE Transaction on Geoscience and Remote Sensing*, **24**, 997-1007.
- [Hartigan85] Hartigan J.A., Hartigan P.M., (1985). *The dip test of unimodality*, *Ann. Statist.*, **13**, 70-80.
- [Heitz93] Heitz F., Bouthemy P., (1993). *Multimodal estimation of discontinuous optical flow using Markov Random Fields*, *IEEE Transactions PAMI*, Vol 15, **12**, 1217-1232.
- [Hillion92] Hillion A., (1992). *Les approches statistiques pour la reconnaissance des images de télédétection*, *Atti della XXXVI Riunione Scientifica, SIS*, Vol 1, 287-297.
- [Jensen96] Jensen F., (1986). *An introduction to Bayesian Networks*, University College London Press, London.
- [Jewel82] Jewel, N. (1982). *Mixtures of exponential distributions*, *Ann. Statist*, **10**, 479-484.
- [Johnson70] Johnson N.L., Kotz S., (1970). *Distributions in Statistics : Continuous Univariate Distributions*, Wiley, New York.
- [Kato93] Kato Z., Berthod M., Zerubia J., (1993). *Multiscale Markov random field models for parallel image classification*, *Proc. Int. Conf. Computer Vision*, Berlin, 253-257.
- [Keilson72] Keilson J., Steutel F.W., (1972). *Families of infinite divisible distributions closed under mixing and convolutions*, *Ann. Math. Statist.*, **43**, 242-250.
- [Kelker71] Kelker D., (1971). *Infinite divisibility and variance mixture of the normal distribution*, *Ann. Math. Statist.*, **42**, 802-808.
- [Kendall87] Kendall M.G., Stuart A., Ord J.K., (1987). *Kendall's advanced Theory of Statistics, fifth edition of Vol1 - Distribution theory*, Charlie Griffin & Company, Limited, London.

- [Kent81] Kent J.T., (1981). *Convolution mixture of infinitely divisible distributions*, Math. Proc. Camb. Phil. Soc., **90**, 141-153.
- [Kent83] Kent J.T., (1983). *Identifiability of finite mixture for directional data*, Ann. Statist., **11**, 984-988.
- [Kress98] Kress R., (1998). *Numerical analysis*, Graduate texts in Mathematics, 181, Springer Verlag.
- [Laferté96] Laferté J.M., (1996). *Contribution à l'analyse d'images par modèles markoviens sur des graphes hiérarchiques - Application à la fusion de données multirésolution*, Thèse de doctorat de l'Université de Rennes I.
- [Laferté00] Laferté J.M., Perez P., Heitz F., (2000). *Discrete Markov image modeling and inference on the quad-tree*, IEEE Transactions on Image Processing, **9**, 390-404.
- [Laird78] Laird N., (1978). *Nonparametric maximum-likelihood estimation of a mixing distribution*, J. Amer. Statist. Assoc., **73**, 805-811.
- [Lambert84] Lambert D., Tierney L., (1984). *Asymptotic properties of maximum likelihood estimates in the mixed Poisson model*, Ann. Statist., **12**, 1388-1399.
- [Landau87] Landau H.J., (1987). *Moments in Mathematics, Proceedings of Symposium in applied Mathematics*, American Mathematical Society.
- [Lemdani95] Lemdani M. (1995). *Tests dans le Cas d'un Melange de Lois dans des Modeles Parametriques et non Paramétriques*, Ph.D. thesis, Ecole Polytechnique.
- [Levinson83] Levinson S.E., Rabiner L.R., Sondhi M.M., (1983). *An introduction to the application of the theory of probabilistic functions of a Markov process in automatic speech recognition*, BSTJ, **62**, 1035-1074.
- [Liang92] Liang Z., Jaszczak R.J., Coleman R.E., (1992). *Parameter Estimation of Finite Mixture Using the EM Algorithm and Information Criteria with Application to medical Image Processing*, IEEE Transactions on Nuclear Science, Vol 39, **4**, 1126-1133.
- [Lindsay83a] Lindsay B.G., (1983). *The geometry of mixture likelihood : a general theory*, Ann.Statist., **11**, 86-94.
- [Lindsay83b] Lindsay B.G., (1983). *Efficiency of the conditional score in a mixture setting*, Ann. statist., **11**, 486-497.
- [Lindsay95] Lindsay B.G., (1995). *Mixture Models : Theory, geometry and Applications*, NSFCBMS Regional Conference Series in Probability and Statistics, **vol.5**, IMS, Hayward, CA.
- [Luxmann87] Luxmann-Ellingaus U., (1987). *On the identifiability of mixtures of infinitely divisible power series distributions*, Statist. Probab. Lett., **5**, 375-378.
- [Maffet91] Maffet A.L., Wackerman C.C., (1991). *The modified beta density function as a model for synthetic aperture radar clutter statistics*, IEEE Transactions on Geoscience and Remote sensing, Vol 29, **2**, 277-283.
- [Marhic91] Marhic N., Masson P., Pieczynski W., (1991). *Mélange de lois et segmentation non supervisée des données Spot*, Statistique et Analyse des données, Vol 16, **2**.
- [Marroquin87] Marroquin J., Mitter S., Poggio T., (1987). *Probabilistic solution of ill-posed problems in computational vision*, Journal of the American Statistical Association, **82**, 76-89.
- [Masson93] Masson P., Pieczynski W., (1993). *SEM Algorithm and unsupervised segmentation of satellite images*, IEEE Transactions on Geoscience and Remote Sensing, Vol 31, **3**, 618-833.
- [McDonald71] Mc Donald P.D.M., (1971). *Comment on "An estimation procedure for mixture of distributions" by Choi and Bulgren*, J. Roy. Statist. Soc., **ser. B**, **33**, 326-329.
- [McLachlan88] McLachlan G.J., Basford K.E., (1988). *Mixture Models : inference and Applications to Clus-*

tering, Marcel Dekker, New York.

- [McLachlan97] McLachlan G.J., Krishnan T., (1997). *The EM Algorithm and Extensions*. Wiley series in probability and statistics, John Wiley and sons, Inc.
- [McLachlan00] McLachlan G., Peel D., (2000). *Finite Mixture Models*. John Wiley & Sons, New-York.
- [Mammen92] Mammen E., Maron S., Fisher N.I., (1992). *Some asymptotic for multimodality tests based on kernel density estimator*, Probab. Theor. Rel. Fields, **91**, 115-132.
- [Martin72] Martin R.D., Schwartz S.C., (1972). *On mixture, quasi-mixture and nearly normal random processes*, Ann. Math. Statist., **40**, 948-967.
- [Medgyessy77] Medgyessy P., (1977). *Decomposition of Superposition of Density Functions and Discrete Distributions*. Adam Hilger Bristol.
- [Metropolis53] Metropolis N., Rosenbluth A.M., Teller A.E., (1953). *Equation of state calculations by fast computing machines*, J. Chem. Phys., **21**, 1087-1092.
- [Mignotte99] Mignotte M., Collet C., Perez P, Bouthemy P., (2000). *Sonar image segmentation using an unsupervised hierarchical MRF model*, IEEE Transactions on Image Processing, **9**, 1216-1231.
- [Monfrini99] Monfrini E., Ledru T., Vaie E., Pieczynski W., (1999). *Segmentation non supervisée d'images par arbres de Markov cachés*, Dix-septième Colloque GRETSI, Vannes, **1**, 43-46.
- [Newcomb1886] Newcomb S., (1886). *A generalized theory of the combination of observations so as to obtain the best result*. Amer. J. Math., **8**,343-366.
- [Patil66] Patil, G.P., Bildikar, S. (1966). *Identifiability of countable mixtures of discrete probability distribution using methods of infinite matrices*, Proc. Camb. Phil. Soc., **62**, 485-494.
- [Pearson1894] Pearson K., (1894). *Contribution to the mathematical theory of evolution*, Phil.Trans.Roy. Soc., **A. 185**, 71-110.
- [Pearson1895] Pearson K., (1895). *Contribution to the mathematical theory of evolution*, Phil.Trans.Roy. Soc., **A. 186**, 343-414.
- [Peng95] Peng A., Pieczynski W., (1995). *Adaptive Mixture Estimation and Unsupervised Local Bayesian Image Segmentation*, Graphical Models and Image Processing, Vol 57, **5**, 389-399.
- [Pfanzagl88] Pfanzagl J., (1988). *Consistency of maximum likelihood estimators for certain nonparametric families, in particular mixtures*, J. Statist. Plann. Inf, **19**, 137-158.
- [Perez96] Perez P., Heitz F., (1996). *Restriction of a Markov random field on a graph and multiresolution statistical image modeling*, IEEE Transactions on Information Theory, Vol 42, **1**, 180-190.
- [Pickard80] Pickard D., (1980). *Unilateral Markov Fields*, Adv. Appl. Prob., **12**, 655-671.
- [Pieczynski90] Pieczynski W., (1990). *Mixture of distributions, Markov random fields and unsupervised bayesian segmentation of images*, Rapport technique, **122**, LSTA, Université Paris VI.
- [Pieczynski92] Pieczynski W., (1992). *Statistical image segmentation*, Machine Graphics and vision, Vol 1, **1**, 261-268.
- [Pieczynski94] Pieczynski W., (1994). *Champs de Markov cachés et estimation conditionnelle itérative*, Traitement du signal, Vol 11, **2**, 141-153.
- [Pieczynski99] Pieczynski W., (1999). *Hidden Evidential Markov Trees and Image Segmentation*, Proceeding's of ICIP'99, Kobe, Japon.
- [Pieczynski00] Pieczynski W., Tebbache A.N., (2000). *Pairwise Markov random fields and segmentation of textured images*, Mach. Graph. Vis., Vol9, **3**, 705-718.

- [Pollard84] Pollard D. (1984). *Convergence of Stochastic Processes*, Springer, New York.
- [Prakasa69] Prakasa Rao B.L.S, (1969). *Estimating a unimodal density*, Sankhya, **Ser A**, **31**, 23-36.
- [Prakasa92] Prakasa Rao B.L.S, (1992). *Identifiability in Stochastic Models : Characterization of Probability Distributions*, Academic Press, New York.
- [Qian89] Qian W., Titterington D.M., (1989). *On the use of Gibbs Markov chain models in the analysis of images based on second order pairwise interactive distributions*, J. Appl. Stat., Vol 16, **2**, 267-282.
- [Quandt78] Quandt R. E., Ramsey J. B., (1978). *Estimating mixtures of normal distributions and switching regressions*. J. Amer. Statist. Assoc, **73**, 730-738.
- [Rao48] Rao C. R., (1948). *The utilization of multiple measurements in problems of biological classification*. J. Roy. Statist. Soc. B, **10**, 159-203.
- [Rao73] Rao C.R., (1973). *Linear Statistical Inference and its Applications*, Wiley Series in Proba. and Math. Stat.
- [Redner84] Redner R., Walker H.F., (1984). *Mixture densities, maximum likelihood and the E.M. algorithm*, SIAM Review, **26**, 195-239
- [Rennie72] Rennie R.R., (1972). *On the independence of the identifiability of finite multivariate mixture and the identifiability of the marginal mixtures*, Sankya, **Ser. A**, **34**, 449-452.
- [Rider61] Rider P. R., (1961). *The method of moments applied to a mixture of two exponential distributions*. Ann. Math. Stat., **32**, 143-147.
- [Rider62] Rider P. R., (1962). *Estimating the parameters of mixed Poisson, binomial and Weibull distributions by the method of moments*. Bull. ISI, **39**, Part 2, 225-232.
- [Robbins48] Robbins H., (1948). *Mixture of distributions*, Ann. Math. statist., **19**, 360-369.
- [Robbins49] Robbins H. E. , Pitman E. J. G., (1949). *Application of the method of mixtures to quadratic forms in normal variates*, Ann. of Math. Stat., Vol. 20 pp.552-560.b., **38**, 211-216.
- [Robert92] Robert C.P., (1992). *L'analyse statistique bayésienne*, Economica.
- [Robert96] Robert C.P., (1996). *Méthodes de Monte-Carlo par chaînes de Markov*, Economica.
- [Rootzen77] Rootzen H., (1977). *A note on convergence to mixture of normal distribution*, Zeit. Wahrsch. Theor. verw. Ge
- [Salzenstein96] Salzenstein F., (1996). *Modèles Markoviens flous et segmentation statistique non supervisée d'images*, Thèse de doctorat de l'université de Rennes I.
- [Salzenstein98] Salzenstein F, Pieczynski W., (1998). *Sur le choix de méthodes de segmentation statistique d'images*, Traitement du signal, Vol 15, **2**, 119-128.
- [Shorack86] Shorack G.R., Wellner J.A., (1986). *Empirical processes with Applications to Statistics*, Wiley, New York.
- [Silverman81] Silverman B.W., (1981). *Using kernel density estimates to investigate multimodality*, J. Roy. Statist. Soc., **Ser. B**, **43**, 97-99.
- [Simar76] Simar L., (1976). *Maximum likelihood estimation of a compound Poisson process*, Ann. Statist., **4**, 1200-1209.
- [Smyth97] Smyth P., Heckerman D., Jordan M., (1997). *Probabilistic independence networks for hidden Markov probability models*, Neural Computation, Vol9, **2**, 227-269.
- [Swain81] Swain P., Vardeman V., Tilton J., (1981). *Contextual classification of multispectral image data*,

- Pattern Recognition, **13**, 428-441.
- [Tallis69] Tallis G.M., (1969). *The identifiability of mixtures of distributions*, J. Appl. Probab., **6**, 389-398.
- [Tallis82] Tallis G.M., Chesson P., (1982). *Identifiability of mixture*, J. Austr. Math. Soc., **32**, 339-348.
- [Tan72] Tan W. Y., Chang W. C, (1972). *Some comparisons of the method of moments and the method of maximum likelihood in estimating parameters of a mixture of two normal densities*. J. Amer. Stat. Assoc, **67**, 702-708.
- [Tanner62] Tanner W. F., (1962). *Components of the hypsometric curve of the earth*. J. Geophys. Res., **67**, 2841-2843.
- [Teicher54] Teicher H.1954. *On the convolution of distributions*, Ann. of Math. Stat., Vol. 25 pp. 775-778.
- [Teicher60] Teicher H., (1960). *On the mixture of distributions*, Ann. Math. statist., **31**, 55-77.
- [Teicher61] Teicher H., (1961). *Identifiability of mixtures*, Ann. Math. statist., **32**, 244-248.
- [Teicher63] Teicher H., (1963). *Identifiability of finite mixtures*, Ann. Math. Statist., **34**, 1265-1269.
- [Teicher67] Teicher H., (1967). *Identifiability of mixtures of product measures*, Ann. Math. Statist., **38**, 1300-1302.
- [Thierney84] Thierney L., Lambert D., (1984). *Asymptotic efficiency of estimators of functional of mixed distributions*, Ann. Statist., **12**, 1380-1387.
- [Titterington85] Titterington D.M., Smith A.F.M., Makov U.E., (1985). *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York.
- [Tricomi57] Tricomi F. G., (1957). *Integral Equations*, Interscience Publishers Inc., New-York.
- [VanDeGeer93] Van de Geer S., (1993). *Rates of convergence for the maximum likelihood estimator in mixture models*, Technical Report, **TW 93-09**, University of Leiden.
- [VanDeGeer94] Van de Geer S., (1994). *Asymptotic normality in mixture model*, Technical report, **TW 94-03**, University of Leiden..
- [Viterbi67] Viterbi A.J., (1967). *Error bounds for convolutional codes and asymptotically optimum decoding algorithm*, IEEE Transactions on Information Theory, Vol 13, 260-269.
- [Whittaker90] Whittaker J., (1990). *Graphical Models in Applied Multivariate Statistics*, Wiley.
- [Williams99a] Williams C.K.L., Adams A.J., (1999). *DTs: dynamic trees*, Advances in Neural Information Processing System.
- [Williams99b] Williams C.K.L., Feng X., (1999). *Tree-Structured belief networks as models of images*, Proceedings of ICANN'99.
- [Wolfowitz57] Wolfowitz J., (1957). *The minimum distance method*. Ann. Math. Stat., **28**, 75-88.
- [Yakowitz68] Yakowitz S.L., Spragins J.D., (1968). *On the identifiability of finite mixture*, Ann. Math. Statist., **40**, 1728-1735.
- [Yao89] Yao J.F., (1989). *Segmentation bayésienne d'images : comparaison de méthodes contextuelles et globales*, Cahier du centre d'Etudes et de Recherches, Op. 30, **4**, 269-290.
- [Younes88] Younes L., (1988). *Estimation and annealing for Gibbsian fields*, Annales de l'institut Henri Poincaré, **24**, 269-294.
- [Younes89] Younes L., (1989), *Parametric inference for imperfectly observed Gibbsian fields*, Probability Theory and related Fields, **82**, 625-645.

- [Zhang90] Zhang C-H., (1990). *Fourrier methods for estimating mixing densities and distributions*, Ann. Statist., **18**, 806-831.
- [Zhang95] Zhang C-H., (1995). *On estimating mixing densities in discrete exponential family models*, Ann. Statist., **23**, 929-945.
- [Zhang93] Zhang J., (1993). *The mean fields theory in EM procedures for blind Markov random field image restoration*, IEEE Transactions on Image Processing, Vol 2, **1**, 27-40.

